Meta-Psychology, 2025, vol 9, MP.2025.4738
https://doi.org/10.15626/MP.2025.4738
Article type: Commentary

Open data: Yes
Open materials: Not Applicable
Open and reproducible analysis: Yes
Open reviews and editorial process: Yes
Preregistration: No

Edited by: Felix D. Schönbrodt
Reviewed by: Invited commentaries
Analysis reproduced by: Lucija Batinović
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/T5RPZ

Replication value as a function of citation impact and sample size: response to commentaries

Peder Mortvedt Isager[1], Anna E. van 't Veer[2], and Daniël Lakens[3]
[1]Department of Psychology, Oslo New University College
[2]Methodology and Statistics unit, Institute of Psychology, Leiden University
[3]Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology

The primary goal of our target article (Isager et al., 2025) is to give the research community an example of what a well-justified replication value metric could look like, and to encourage discussion of how replication value could be quantified in practice. Furthermore, in the target article we discuss practical hurdles to quantification and possible practical applications for RV_{Cn} and other metrics. As that article proposes a method for how to do research—in this case a method to select which claims in the literature need replication most—it is important to receive criticism, feedback, and viewpoints from a diverse range of authors interested in this topic. We are delighted to read the many thoughtful yet critical commentaries, several of which proposing adjustments or alternatives to the equations we have proposed in the target article. This is very encouraging to see, as our aim with initiating this call in *Meta-Psychology* was to create an open dialogue in the scientific record. RV_{Cn} is an efficient but limited metric. Its limitations should be laid bare, and we fully expect that improved metrics and selection procedures can be created in the future. We hope our target article and these commentaries together will inspire readers to continue the discussion of how to efficiently and transparently select studies for replication. In this rejoinder we will summarize what we see as the major themes touched on in the commentaries, and we will reply to some of the specific proposals and criticisms brought up by different commentary authors.

Keywords: RVcn, replication value, replication

Practical use cases for RV_{Cn} and other metrics

RV_{Cn} and the four-step selection process we propose (see Figure 1) solve a quite specific problem; how to select studies for replication when our goal is to maximize the expected utility of the replication study, and when we have many candidate studies to choose from. Feldman (2025) rightly points out that replication can serve other goals that RV_{Cn} does not help us achieve, such as generalizing findings to new contexts, validating measurement instruments, or bringing renewed attention to neglected research. It is therefore important to consider the various practical contexts RV_{Cn} could be applied to, and to think carefully about whether it is actually useful for the goal we are trying to achieve. We are glad to see several commentary authors propose and critically discuss practical uses of RV_{Cn}.

An excellent example is Beerdsen (2025) who considers how RV_{Cn} might be used to replicate research used in legal proceedings. The context described by Beerdsen seems to fit well with the general utility of RV_{Cn}. A wide variety of claims are used to support arguments in court, some of which may be more uncertain than others, and some of which may be more important for the outcome of the case. However, as Beerdsen points out, 'value' would need to be operationalized differently in this context. Academic citations are relevant for scientific impact, but may have little bearing on legal impact. We expect that adjustments to the RV_{Cn} equation will have to be made for it to work in more applied contexts (legal, medical, financial settings, etc.).

Takashima and Yamada (2025) propose a context to which RV_{Cn} is more directly applicable; helping academic mentors recommend useful replication projects to their students. This is exactly the kind of context we had in mind when designing RV_{Cn}. The person selecting (the mentor or supervisor) wants to prioritize important and uncertain studies for replication, there are many studies to choose from, there is limited time to get familiar with each study, and a "good enough" decision needs to be made rather quickly. We hope the selection procedure we have proposed (or an improved version of it) can help facilitate more student replication projects

in the future.

We can think of a few more contexts in which RV_{Cn} and similar metrics may be useful, in addition to those proposed by the commentary authors:

- For scientific communities, to coordinate and discuss which replications are most needed within a research field, and to issue calls for these replications.
- For scientific journals, to promote continual quality control of the science they publish (e.g., through Registered Verification Reports; Isager et al., 2024).
- For large scale collaborations like the Psychological Science Accelerator (Moshontz et al., 2018), to select which replication studies to spend their many-lab resources on.
- For initiatives that facilitate replications, like CREP (McLaughlin et al., 2013), to select which studies to recommend replication of.
- For funders, to select which replication projects to fund or earmark funding for (e.g. NWO, 2019).

Responsible use of RVcn

Several of the commentaries rightfully indicate that metrics for replication study selection, as any metric, should be used responsibly. In the case of RV_{Cn}, this means interpreting numerical results together with (subjective, qualitative) contextual information, as the metric itself will not capture all aspects relevant to the most optimal study selection. In a first step, as our target article depicted and as is visualised here in Figure 1, study selection will be guided by the replication goals of the author(s). This could mean that the candidate set of articles are chosen to be in line with the authors interest, expertise or possibilities (e.g., a set of studies is selected that would be feasible for a student to replicate; Takashima and Yamada, 2025). In Step 3 and 4, where an in-depth inspection of a subset of candidates iteratively leads to replication target selection, there is again room to tailor the decision to the most fitting study based on subjective qualitative information such as feasibility to conduct this replication within the author team (Pittelkow et al., 2025). Factors such as those proposed by Bekkers (2025) – societal impact, generality of claim, effect size, test informativeness, etc. – can also be considered at this point, and should override the RC_{Cn} rank order whenever appropriate. This procedure, with qualitative and quantitative steps interwoven (Hessels et al., 2022), facilitates responsible use of RV_{Cn}.

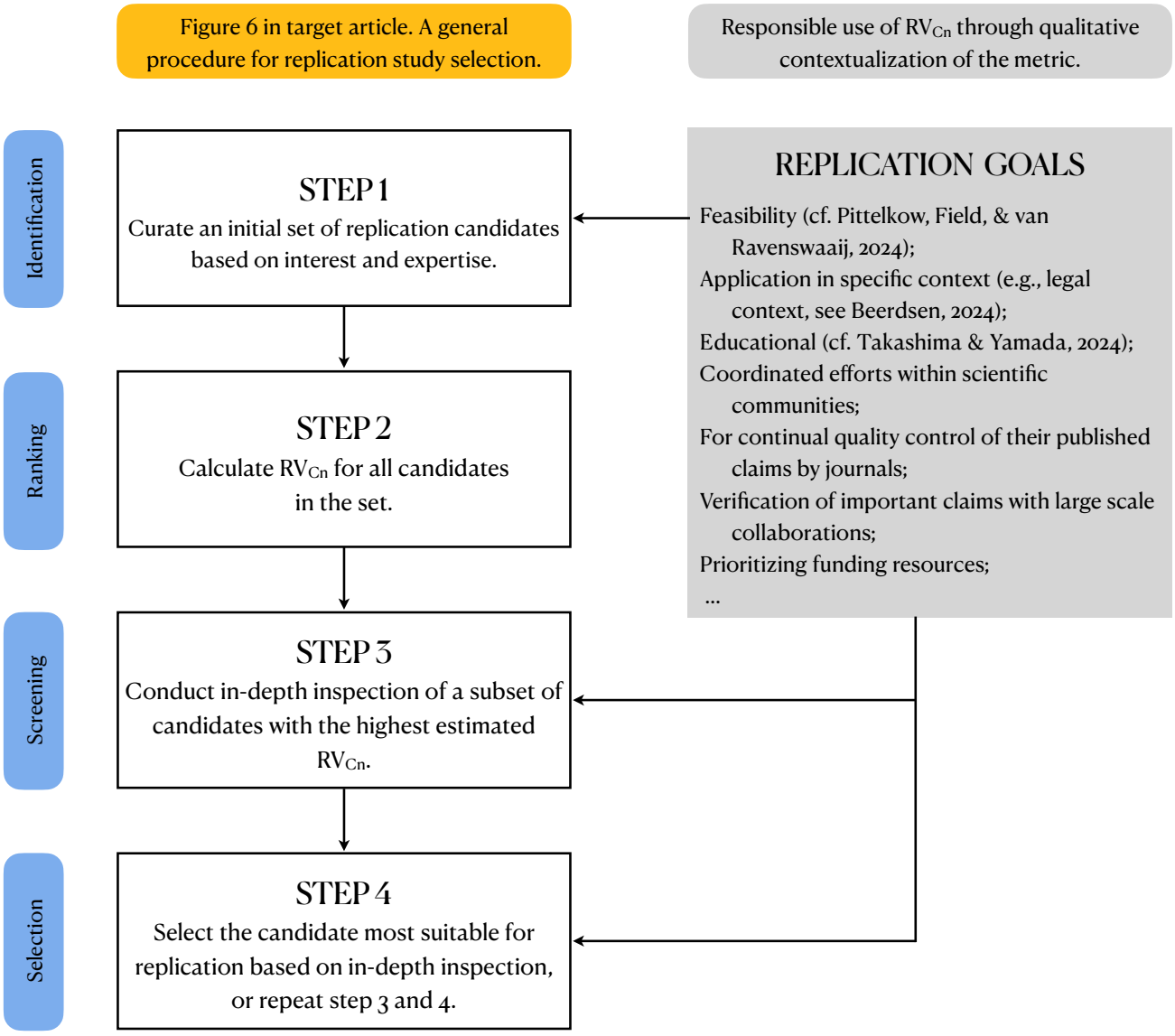
Pittelkow and colleagues proposed quantitative metrics of replication value in the past (Field et al., 2019; Pittelkow et al., 2021), but in their commentary indicate they have "shifted towards more qualitative and open approaches for describing replication value, exemplified by the checklist for transparent replication target selection" (Pittelkow et al., 2025). Checklists are a useful tool for researchers to transparently log why they decided to replicate a study. However, formulas such as RV_{Cn} serve an additional goal. Such formulas not only help to make the selection rationale transparent after the fact, they also help make the actual selection process more efficient, i.e., they are *part of* the selection rationale.

When using a replication value metric in isolation, for instance, a journal can—in a relatively fast and efficient way—state that it will publish any replication study for findings published in the same journal that have a RV_{Cn} higher than some value (van 't Veer et al., 2025). RV_{Cn} does not 'fail' to capture the complexity of replication target selection, but it intentionally simplifies this complexity by reducing the features that form the basis of a decision. There are situations where this is more desirable, and situations where this is less desirable. The goal of this simplification and quantification is to develop a consensus about which replication studies are valuable. This can be very useful in a time where there is great uncertainty about whether top journals will publish a replication study, or whether funders will award a grant to a proposal to replicate studies. If stakeholders agree on a specific replication value formula, uncertainty about the expected utility of performing the replication study is reduced, which can be useful to motivate researchers to perform valuable replication studies, journals to publish them, and funding agencies to award grants for this work. Of course, researchers remain free to replicate any other study they feel like, and hope that their subjective arguments for why a replication study is important will convince reviewers and editors. But it is more difficult to see how mere subjective criteria offer a systematic and efficient way to verify the most uncertain yet impactful claims.

As Rainey (2025) mentions, there is a risk researchers will mindlessly use a replication value formula, without subsequently evaluating the target candidates in depth. A related problem might be that researchers believe only replication studies with a high RV_{Cn} deserve to be published, as Feldman (2025) notes, without considering that there are additional criteria that can justify performing a replication study. We fully agree with Rainey that disagreements about which studies deserve to be replicated should be embraced, as it means scientists are finally explicitly discussing which

Figure 1

A general procedure for replication study selection as proposed in Isager et al. (2025).



research is valuable to perform. We hope that each field will reach consensus on a number of reasonable replication value formulas, which should ensure a certain amount of diversity, while still guiding researchers to subsets of studies that are most valuable to replicate given limited resources.

The responsible use of a replication value formula (including RV_{Cn}) requires explicitly stating its limitations. As Takashima and Yamada (2025) note, sample size is only an indicator of uncertainty, and in some cases computing standard errors is necessary. Citations in the scientific literature are a crude measure of impact, as there are many reasons why some papers receive more citations than others (such as prestige bias, as noted by Fillon and Chandrashekar, 2025). Future research should work towards identifying and, where possible, resolving these issues when replication value formulas are proposed. In the end, any proposed procedure for identifying high-utility replication candidates will have limitations. The trick is to decide which lim-

itations are acceptable, and to balance the limitations of any given selection method against the limitations of having no method for selecting studies at all.

Proposed modifications and alternative metrics

RV_{Cn} is a measurement instrument. It aims to measure the expected utility of replicating a given claim. Like any measurement instrument, it is subject to issues with reliability and validity, some of which may be resolved through adjustments to the formula. In addition, 'value' is a social construct whose definition may change across contexts. There is no *one true* definition of value. Therefore, there is no *one true* measure of replication value. Any measure of replication value must be adapted to fit the users' definition of value.

Several such issues are identified in the target article, and some adjustments to improve measurement validity are proposed in appendixes. We are very excited to see commentary authors propose further adjustments and adaptations. All these proposals move the field forward, and their diversity allows different stakeholders to pick a replication value formula that they consider appropriate for a specific time and place. We will first discuss alternative ways to quantify or weigh uncertainty that were raised in the commentaries, and then discuss different ways to quantify impact.

Alternative quantifications of uncertainty

Bakker et al. (2025) question whether our proposed metric appropriately weighs impact and uncertainty. As the sample size grows, its influence on RV_{Cn} becomes smaller. Bakker and colleagues point out that in fields with large sample sizes, one may want to preserve the impact of sample size on relative replication value by for example taking the log of the sample size. We think it is very sensible to create variations of a replication value for fields with predominantly large sample sizes, as long as relatively smaller sample sizes in this field are still considered a reason to prioritize replication studies.

Pittelkow et al. (2025) show how Bayes factors yield a different relative ordering of studies than RV_{Cn}. Where RV_{Cn} operationalizes uncertainty in terms of estimation (and hence only focuses on sample size) a Bayes factor operationalizes uncertainty in terms of informativeness with respect to distinguishing between a specific null and alternative hypothesis. Here, a Bayes factor of 1 implies an uninformative test result, and values further away from 1 provide stronger support for one hypothesis over another. Pittelkow et al. (2021) propose the use of a 'default' Bayes factor, because these are widely used in psychology. This is true, but the use of a default prior is also widely criticized as it does not actually reflect an alternative hypothesis of interest in most cases. A replication value metric based solely on a Bayes factor also ignores the impact of an original study (which RV_{Cn} incorporates in terms of citations). Because the two metrics attempt to measure different things, estimates produced by the approach suggested by Pittelkow et al. (2021) should not be expected to be strongly related to RV_{Cn} estimates.

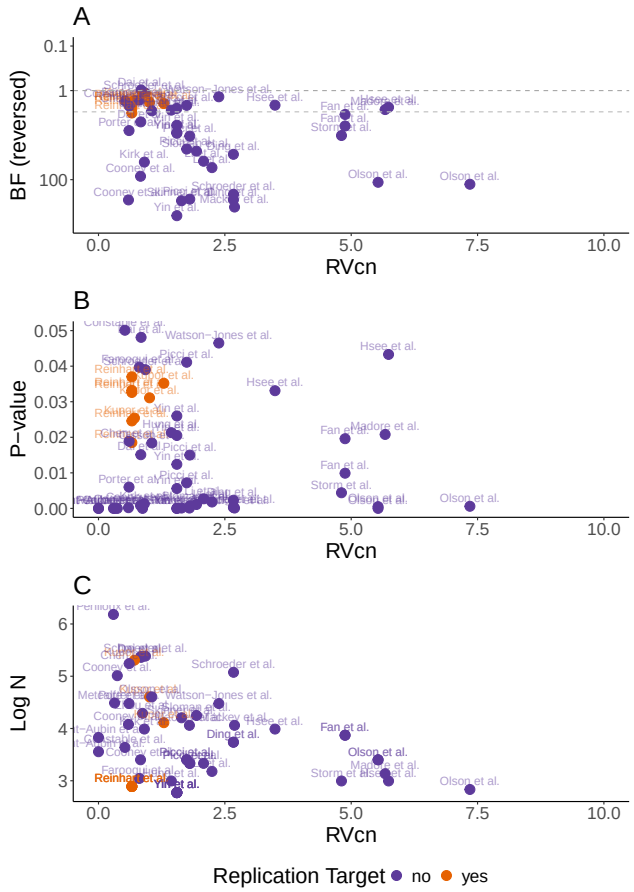
Due to the direct relation between Bayes factors and *p*-values (Francis, 2017) it is perhaps more intuitive for readers to see the relation between RV_{Cn} and *p*-values alongside the relation between RV_{Cn} and Bayes factors (see figure 2). This comparison makes it clear that the Bayes factor approach will lead to the recommendation to replicate studies with a *p*-value relatively close to 0.05, regardless of how often the study has been cited. *P*-values (and correspondingly, Bayes factors) are in part a function of the sample size (the larger the sample size, the lower the *p*-value and the further a Bayes factor will be from 1) but the *p*-value and Bayes factor also depend on the *t*-value. As a consequence, there is no one-to-one relation between the sample size and the Bayes factor (or *p*-value), and as a consequence, the RV_{Cn} can differ from Bayes factors. The authors empirically show the two measures are indeed unrelated. This nicely illustrates the importance of carefully considering how we want to operationalize the value of replication studies.

It is useful to examine a few specific examples where the RV_{Cn} differs from a Bayes factor. For example, a study by Olson et al. (2015) is valuable to replicate based on RV_{Cn} because the small sample size (16 and 17 observations per condition) combined with the large citation count (303 times) warrant a replication. The Bayes factor (126.1) and *p*-value ($p < .001$), however, give little reason to doubt the null hypothesis can be rejected, given the observed effect size of $d = 1.06$. This highlights an important distinction: when there is little uncertainty about whether the effect is nonzero (indicated by the high Bayes factor and low *p*-value) the size of the effect can still be highly uncertain (reflected in the low sample size and subsequently wide confidence intervals). Which of these uncertainties should matter more for study selection is a topic for future debate. An opposite example is provided by a study by Dai et al. (2015) which has a relatively low RV_{Cn}, as the sample size was relatively large ($N = 213$), and the citation count not exceptionally large (123 citations). However, the Bayes factor is 0.97, which is very close to 1, and the *p*-value is 0.0481, which is very close to 0.05. There can be good reasons to value replicating just-significant effects in the scientific literature, although given limited resources, it might still be desirable to take the impact of studies into account.

Finally, Takashima and Yamada (2025) raise the im-

Figure 2

Scatterplots plotting replication value (RV_{Cn}) against Bayes Factors (BF), p-values and log sample size. Adopted from Pittelkow et al. (2025)



portant point that in some fields small sample sizes do not have high uncertainty, such as in vision science. We fully agree. When RV_{Cn} is computed across disciplines, the final ranking will not adequately reflect the uncertainty in findings, because too many aspects of study design that influence uncertainty are allowed to vary. The candidate set curated in step 1 must be constrained to studies for which sample size plays a comparable role in determining overall uncertainty. We noted how sample size fails to adequately capture uncertainty "in repeated measures designs where the number of repeated samples per participant can matter more than the number of participants per se", and whenever there is a substantial number of such designs, researchers will need to rely on a measure of uncertainty that more adequately tracks the standard error. The authors also note how uncertainty needs to take previous replication studies into account. We agree, and we actually already proposed how to achieve this in supplementary material SM1 of the target article (https://osf.io/rdhw3).

Alternative quantifications of impact

Beerdsen (2025) points out that in the legal field citation counts in scientific papers do not fully capture impact, as references to studies in non-scientific documents, such as those found in legal databases. The proposal to weight references in legal documents equally to citations is perhaps even too modest, and even scientists might agree that studies used repeatedly in legal documents quickly become valuable enough to replicate, regardless of scientific impact. The proposal by Beerdsen provides a wonderful illustration of an easily quantifiable measure of a specific type of societal impact, which we agree should be taken into account when relevant and possible.

Both Bekkers (2025) and Fillon and Chandrashekar (2025) are concerned with biases in citations and sample sizes. For example, they point out that citations can be bought, and that citations are impacted by the prestige of authors. These biases reduce the connection between citation count and actual scientific impact, which is essential for the validity of RV_{Cn}. It is possible that there are fields where citations do not adequately track impact, and then another measure of impact should be used. It is also possible to exclude outliers due to questionable citation practices in the in-depth manual stage (step 3) of selecting a replication study (see figure 1). The authors also suggest that replication value should be related to "how a finding or an effect of interest maps onto the causality scheme of the theory". We agree that effects central to a theory are more important than peripheral effects. If researchers find a way to determine the centrality of effects to a theory, and this information is easily available, it can be incorporated into a replication value formula.

Rainey (2025) similarly notes that citations should not be the only measure of impact, because claims are important for other reasons, such as social or clinical impact. We fully agree, and hope researchers take Beerdsen (2025) as an example of how to discuss and concretize other sources of impact.

Alternative quantifications of value

Some proposals move beyond adjustments or modifications of impact and uncertainty, and propose an altogether different quantification of value. Feldman (2025) articulates the most extreme viewpoint possible: Every study that has been published is valuable to replicate, because if a study is valuable enough to appear in the scientific literature, it is valuable enough

to replicate. Although we sympathise with this viewpoint, we do not think it works in practice. First, some journals publish articles only based on methodological quality, and explicitly informs researchers to not judge whether a submission reaches a (vaguely specified) level of importance. For such papers, the consideration whether the study was valuable enough to perform is decoupled from study selection, and a replication study does not automatically inherit an earlier value judgment. It is noteworthy that for example PlosOne, which only evaluates novel research based on whether it is technically rigorous and meets the scientific and ethical standards, requires authors to add a justification for why a replication study is worthwhile to perform, and states that "Submissions that replicate or are derivative of existing work will likely be rejected if authors do not provide adequate justification" (https://journals.plos.org/plosone/s/reviewer-guidelines).

Feldman argues that even though "It would be best to prioritize which replications to run to maximize utility and impact", it will take years to reach a ratio of novel studies versus replication studies where we need to start worrying about which replication studies should be performed. We agree with the sad reality that it will likely take a long time before replication studies are commonplace, and we greatly appreciate the contribution Feldman has made to perform and publish high-quality replication studies. We nevertheless disagree that scientists only need to worry about the value of replication studies after a certain ratio of novel to replication studies is reached. We believe scientists should want to reach a healthy balance of novel and replication studies by performing the most important replication studies first. Indeed, given that so few researchers choose to perform and publish replication studies, it would be a shame if these individuals choose to perform replications that are deemed to have relatively little value by scientific peers, or that repeat an already relatively certain claim.

We fully agree with Feldman that our question 'which replication studies are valuable enough to do' is a subset of a more important question: Which research is worth doing? This is in line with Rainey, who notes that it is valuable to perform a replication study whenever one wants to build on previous work, and therefore "Arguments for the value of building upon a prior claim are arguments for the importance of the prior claim itself." Feldman argues for broader indications of value, such as replication studies that identify errors and inadequate methods (in line with Rainey, who notes studies with large samples sizes are worth replicating if the design or measures of the study were inadequate), that contribute to generalizability (e.g., replicating studies originally performed decades ago), that make theoretical assumptions more explicit, contribute to the evaluation of measurement, or that reignite interest in forgotten studies. All these aspects of replication studies are valuable (although some would require conceptual, and not direct replications), but they are all also difficult to identify in advance. Therefore, in absence of work that shows how we can identify replication studies that score high on these attributes, they will be difficult to use systematically in the selection of studies for replication.

Riesthuis et al. (2024) raise the important issue of being able to interpret the results of a replication study, which requires that it is clear in advance when a replication study failed to confirm a previously observed effect. This requires researchers to specify which effects would not be considered support for a prediction, for example by specifying a smallest effect size of interest. This value judgment is not unique to replication studies, and also applies to original studies. One can certainly question how useful it is to perform a replication study if it can not demonstrate the absence of a meaningful effect. We do believe that this criteria is more suitable for the 4th step in our proposal, where replication candidates are selected based on a more in depth evaluation.

Criticism of the empirical data we present

Pittelkow et al. (2025) claim that the validation dataset is subject to sampling bias because meta-analyses include effects that were secondary to the original publication, and citations could be unrelated to these secondary effects. We completely agree that the preliminary validation study we performed is not a severe test, and perhaps even the least severe test that RV_{Cn} should have passed. As we wrote in Isager et al. (2023): "Any operationalization of replication value will require validation. At the very least, we should make sure that our assessment strategy will often indicate a high replication value for claims that we are intuitively confident would be worth replicating, and a low replication value for claims we are intuitively confident would not be worth replicating. More severe validation studies would certainly be desirable, though we are not at present sure what such studies would look like." We still do not know what severe validation studies would look like.

Pittelkow and colleagues explore the consequences of using predicted citation counts (building on Figure 2 of our target manuscript). They find an even more pronounced difference in the replication value of replicated and non-replicated studies (based on Figure 7 in our target manuscript), and—not surprisingly—different absolute values for the RV_{Cn}. If citation counts can be ac-

curately predicted, that would be extremely useful beyond quantifying the value of replication studies. However, it is not yet clear whether the gamma parameterization proposed by Pittelkow et al. actually predicts future citation count better than the raw average of past yearly citations. This makes differences hard to interpret. Whatever the case, RV_{Cn} should always adopt the best prediction algorithm available.

Kamermans et al. (2025) provide a very interesting preliminary empirical overview of the motivations that researchers provide when they publish replication studies. Although uncertainty (i.e., replication crisis, methodological concerns, and conflicting results) and impact (i.e., influential/fundamental) are mentioned, currently reward structures push research more to novel contributions, so it is not surprising that conceptual replications are more popular, and generalizability and extensions are a more important motivation to replicate a study. This ties in with a concern by Pittelkow and colleagues who ask if RV_{Cn} is prescriptive or descriptive. We believe the answer may be 'both'. RV_{Cn} is designed to help identify which studies *should* be selected to maximize the expected utility of replication. This is clearly prescriptive. However, RV_{Cn} is only an estimate of expected utility gain; it is a measurement instrument. As such, it is clearly descriptive. Our proposal is based on discussions between peers about how to quantify the value of replication studies, combined with descriptive work which indicate researchers often report being guided in their selection for a replication study based on impact, uncertainty, and feasibility (Isager et al., 2023). In our preliminary validation study we examined if our operationalization of impact and uncertainty was descriptively related to decisions researchers made in practice. If this validation would have revealed that researchers do not select replication studies based on uncertainty and/or impact, this would have been surprising, and a reason to organize broader consensus meetings. In the end, every version of a replication value formula is a tool for rational decision making for those researchers who agree that the formula correctly specifies the expected utility of replications.

Author Contact

Peder Mortvedt Isager: https://orcid.org/0000-0002-6922-3590 Anna E. van 't Veer: https://orcid.org/0000-0002-2733-1841 Daniël Lakens: https://orcid.org/0000-0002-0247-239X

Correspondence concerning this article should be addressed to Peder Mortvedt Isager, Ullevålsveien 76, 0454 Oslo, Norway. Email: pederisager@gmail.com

Conflict of Interest and Funding

The authors declare no conflicts of interest.

Author Contributions

The authors made the following contributions. Peder Mortvedt Isager: Investigation, Visualization, Writing - Original Draft, Writing - Review & Editing; Anna E. van 't Veer: Investigation, Visualization, Writing - Original Draft, Writing - Review & Editing; Daniël Lakens: Investigation, Supervision, Visualization, Writing - Original Draft, Writing - Review & Editing.

Open Science Practices

This article earned the Open Data and Open Code badge for making the data, and code openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews (published commentaries), is published in the online supplement.

References

Bakker, B. N., Bomm, L., & Peterson, D. (2025). Commentary on Isager et al. (2021) Reflections on the Replication Value (RV) and a Proposal for Revision. *Meta-Psychology*, *9*. https://doi.org/10.15626/MP.2024.4324

Beerdsen, E. (2025). Replication Value in the Courtroom; a Commentary on Isager, van 't Veer & Lakens. *Meta-Psychology*, *9*. https://doi.org/10.15626/MP.2024.4325

Bekkers, R. (2025). Replication Value Increases With Transparency, Test Severity, and Societal Impact. *Meta-Psychology*, *9*. https://doi.org/10.15626/MP.2024.4194

Dai, H., Milkman, K. L., & Riis, J. (2015). Put Your Imperfections Behind You: Temporal Landmarks Spur Goal Initiation When They Signal New Beginnings. *Psychological Science*, *26*(12), 1927–1936. https://doi.org/10.1177/0956797615605818

Feldman, G. (2025). The value of replications goes beyond replicability and is associated with the value of the research it replicates: Commentary on Isager et al. (2024). *Meta-Psychology*, *9*. https://doi.org/10.15626/MP.2024.4326

Field, S. M., Hoekstra, R., Bringmann, L., & Van Ravenzwaaij, D. (2019). When and Why to Replicate: As Easy as 1, 2, 3? *Collabra: Psychology*, *5*(1), 46. https://doi.org/10.1525/collabra.218

Fillon, A. A., & Chandrashekar, S. P. (2025). The Replication Dilemma: Potential Challenges in Measuring Replication Value - A Commentary on Isager, van't Veer, & Lakens (2024). *Meta-Psychology*, *9*. https://doi.org/10.15626/MP.2024.4312

Francis, G. (2017). Equivalent statistics and data interpretation. *Behavior Research Methods*, *49*(4), 1524–1538. https://doi.org/10.3758/s13428-016-0812-3

Hessels, L., van Drooge, L., Holtrop, T., & Costas, R. (2022). Responsible metrics for societal value of scientific research. Retrieved April 2, 2025, from https://www.leidenmadtrics.nl/articles/responsible-metrics-for-societal-value-of-scientific-research

Isager, P. M., van Aert, R. C. M., Bahník, Š., Brandt, M. J., DeSoto, K. A., Giner-Sorolla, R., Krueger, J. I., Perugini, M., Ropovik, I., van 't Veer, A. E., Vranka, M., & Lakens, D. (2023). Deciding what to replicate: A decision model for replication study selection under resource and knowledge constraints. *Psychological Methods*, *28*(2), 438–451. https://doi.org/10.1037/met0000438

Isager, P. M., van 't Veer, A. E., Freeman, Z., Martinovici, A., Breemer, L., Van Ravenzwaaij, D., Liem, C. C. S., Hoekstra, R., & Rasti, S. (2024). Hackathon proceedings for: Perspectives on Scientific Error–Coordinating Quality Control in Practice. https://doi.org/10.31222/osf.io/94a6f

Isager, P. M., van 't Veer, A. E., & Lakens, D. (2025). Replication value as a function of citation impact and sample size. *Meta-Psychology*, *9*. https://doi.org/10.15626/MP.2022.3300

Kamermans, K. L., Dudda, L., Daikoku, T., & Verheyen, S. (2025). The is-ought problem in deciding what to replicate: Which motives guide current replication practices? [Preprint]. https://doi.org/10.31234/osf.io/6xdy2

McLaughlin, H., Peng, C., France, H., McFall, J., Baughman, K., Hildebrandt, L., Wamba, T., Pazda, A., Levitan, C., Peck, T., Lazarevic, L., VanBenschoten, A., Wiggins, B. J., Christopherson, C. D., Grahe, J., Adetula, A., Chartier, C. R., IJzerman, H., Brandt, M., … LePine, S. (2013). Collaborative Replications and Education Project (CREP). https://doi.org/10.17605/OSF.IO/WFC6U

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Protzko, J., Flake, J. K., Forero, D. A., Janssen, S. M., Keene, J., Aczel, B., Ansari, D., Antfolk, J., Baskin, E., Batres, C., … Chartier, C. R. (2018). Psychological Science Accelerator: Advancing Psychology through a Distributed Collaborative Network. *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.1177/2515245918797607

NWO. (2019). Replication Studies. Retrieved August 28, 2019, from https://web.archive.org/web/20190601002622/https://www.nwo.nl/en/funding/our-funding-instruments/sgw/replication-studies/replication-studies.html

Olson, K. R., Key, A. C., & Eaton, N. R. (2015). Gender Cognition in Transgender Children. *Psychological Science*, *26*(4), 467–474. https://doi.org/10.1177/0956797614568156

Pittelkow, M.-M., Field, S. M., & Van Ravenzwaaij, D. (2025). Thinking Beyond RVCN: Addressing the Complexity of Replication Target Selection. *Meta-Psychology*, *9*. https://doi.org/10.15626/MP.2024.4323

Pittelkow, M.-M., Hoekstra, R., Karsten, J., & van Ravenzwaaij, D. (2021). Replication Target Selection in Clinical Psychology: A Bayesian and Qualitative Reevaluation. *Clinical Psychology: Science and Practice*, *28*(2), 210–221. https://doi.org/10.1037/cps0000013

Rainey, C. (2025). Use and Misuse of a Fast Approximation: Not a Criticism, but a Caution. *Meta-Psychology*, *9*. https://doi.org/10.15626/MP.2024.4216

Riesthuis, P., Mesquida, C., & Cribbie, R. (2024). Statistical (non)Significance ≠ (un)Successful Replication: The Importance of the Smallest Effect Size of Interest. https://doi.org/10.31234/osf.io/s3zfy

Takashima, K., & Yamada, Y. (2025). Valuing replication value. *Meta-Psychology*, *9*. https://doi.org/10.15626/MP.2024.4210

van 't Veer, A. E., Freeman, Z., Hoekstra, R., Isager, P. M., Martinovici, A., Van Ravenzwaaij, D., & Rasti, S. (2025). Registered Verification Reports—A model for continual quality control by journals. https://doi.org/10.17605/OSF.IO/CEH4R