Reviewer 1 (Stephen Curry)

This manuscript reports the results of an interesting discrete choice experiment designed to probe the values and interests that inform researchers' decisions on where to publish their work.

Although I am not an expert in the design of discrete choice experiments, the methodology is well explained and the design of the study comes across as well considered, having been developed in a staged way to identify the most appropriate pairings of journal attributes to include.

The principal findings to my mind, well described in the abstract, include the observations that (1) researchers' strongest preference was for journal impact factor and (2) that they were prepared to remove results from their papers if that would allow publication in a higher impact factor journal. The first of these is hardly surprising – and is consistent with a wide array of literature (and ongoing activism, e.g. through DORA, CoARA). The second is much more striking – and concerning for the research community (and its funders). This is the first time I have seen evidence for such a trade-off.

Overall, the manuscript is very clearly written. I have no major issues with the methods or results. However, I think but some minor revisions would enhance the clarity and utility of the paper.

First, although it is made clear in Table 1 that the researchers included in the study are all from the medical and clinical sciences, this is not apparent from the title or the abstract. I think both should be modified to reflect the nature of the sample. In my experience researchers in these fields are among those who feel most intensely the pressure to publish in high IF journals. The authors may want also to reflect in a revised manuscript how well their findings may transfer to other disciplines.

Response: We agree and have now written "Health and medical researchers" in the title, abstract, and opening of the discussion. We have added a paragraph to our limitations section to state that our results may not be generalisable to other scientific fields.

Second, in several places I felt the discussion of the results could be enriched by reference to papers in the recent literature that are missing from the bibliography. These include (1) Muller and De Rijcke's 2017 paper on Thinking with Indicators, which discusses how the pressure of metrics impacts the conduct of research (https://doi.org/10.1093/reseval/rvx023); (2) Bjorn Brembs' analysis of the reliability of research published in prestige science journals (https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2018.00376/full; and (3) McKiernan's et al.'s

examination of the use of the Journal Impact Factor in academic review, promotion, and tenure evaluations (https://pubmed.ncbi.nlm.nih.gov/31364991/).

Response: Thanks for these suggestions. We were unaware of Müller and de Rijcke's paper, which we enjoyed reading and is highly relevant. We now cite this as further evidence of the focus on "where" rather than "what" to publish. We were aware of Brembs paper and agree that it is relevant here as our results may partly explain their finding that papers in high impact factor journals are not reliable (now cited in Discussion). We were unaware of McKiernan et al's paper and this provides useful background on why researchers are forced to consider impact factors (now cited in Discussion).

Third, although the text and figures are nicely laid out, I would recommend using a smaller or different font for the figure legends to more easily distinguish them from body text.

Response: We agree and have changed all table and figure legends as suggested.

Reviewer 2 (Tony Ross-Hellauer)

Peer Review of Preprint: "Researchers Are Willing to Trade Their Results for Journal Prestige: Results from a Discrete Choice Experiment", https://doi.org/10.31219/osf.io/uwt3b

Tony Ross-Hellauer, tross@know-center.at, 2nd Nov 2024

In "*Researchers Are Willing to Trade Their Results for Journal Prestige: Results from a Discrete Choice Experiment*", the authors investigate researchers' publication preferences using a discrete choice experiment in a cross-sectional survey of international health and medical researchers. The study investigates publishing decisions in relation to negotiation of trade-offs amongst various factors like journal impact factor, review helpfulness, formatting requirements, and usefulness for promotion in their decisions on where to publish. The research is timely; as the authors point out, reform of research assessment is currently a very active topic. The design and methods of the study are suitable and robust. The use of focus groups and interviews in developing the attributes for study shows care in the design. The survey instrument itself is generally very well-designed, with important tests of survey fatigue, understanding (*dominant choice task*) and respondent choice consistency (repeat choice task) included. Respondent performance was good or excellent across all these checks. Analysis methods (pMMNL and latent class analysis) are well-suited to the task. Pre-registration and sharing of data and code show commitment to transparency. Limitations are generally well-described.

In the below, I give suggestions for clarification/improvement. Except for some clarifications on limitations and one narrower point (reporting of qualitative data

analysis methods), my suggestions are only that – the preprint could otherwise stand, as is, as a very robust and interesting piece of scientific work.

Response: Thanks for these observations. We spent time carefully planning the discrete choice experiment and our team includes authors with great experience in discrete choice design.

1. Respondents come from a broad range of countries (63), with 47 of those countries represented by fewer than 10 respondents. Institutional cultures of evaluation can differ greatly across nations. And we can expect variability in exposure to the messages of DORA (seen, for example, in level of permeation of DORA as measured by signatories in each country, https://sfdora.org/signers/). In addition, some contexts may mandate or incentivise publication in some venues using measures including IF, but also requiring journals to be in certain databases like WoS or Scopus, or having preferred journal lists). I would suggest the authors should include in the Sampling section a rationale for taking this international approach, including any potentially confounding factors it may introduce, and then adding the latter also in the limitations.

Response: We did consider this in our design and discussed targeting specific countries. However, health and medical research is an international endeavour and there are pressures that apply to all researchers. The Müller and de Rijcke paper (recommended by reviewer 1) mentions the "global pressures" and the common metrics used across countries. A recent international survey found that the "pressure to publish" was the factor most often cited as "always" contributing to irreproducibility (DOI: 10.1371/journal.pbio.3002870). However, we acknowledge that practices and the extents of the pressures will vary between countries, and we now include a paragraph on this in the limitations in terms of the generalisability of our results. We have also added a justification sentence to the Sampling section.

Another reason for including all researchers is that it's not possible to completely reliably identify authors' countries using *PubMed,* as addresses can be missing or incomplete. Some authors may also have moved country.

2. Reporting of qualitative results: In the introduction and methods, the role of the focus groups and interviews seems to have been just to inform the design of the experiment. But then, results from that qualitative work then appear as direct quotes within the discussion to contextualise or explain results. In this sense though, the qualitative results are being used as new data. Given this, I feel that the methods section should include description of the methods and tools used for qualitative data analysis (currently it does not). But in addition, to my understanding (and this may be a question of disciplinary norms – I'm not a health/medicine researcher), generally new data should not be introduced in the discussion section of a research paper. Rather the discussion is meant to interpret, analyse, and provide context for the results that have

already been presented. I personally hence feel that the paper would benefit from the qualitative results being reported separately within the results section.

Response: We agree and have now expanded the first paragraph in our methods section and added another sentence on our distilling approach which includes using the qualitative results to help interpret the quantitative results.

In relation to presenting new results in the discussion, we think that the quotes provide context for the discussion. If the quotes were in the results section, they would either need to be repeated in the discussion or the reader would have to find them in the results, impeding readability. We note that the paper by McKiernan et al, recommended by Reviewer 1, also presents quotes for the first time in the discussion.

3. Impact factors – Discussion section: While there is interesting new information on the relative trade-offs amongst other factors, the most emphasised finding, that impact factors still play a prominent role in publication venue decisions, is hardly surprising. More could perhaps be done to compare how the levels of importance reported here differ with previous results from other disciplines or over time (I know a like-for-like comparison is difficult but other studies have investigated these themes, e.g., https://doi.org/10.1177/01655515209585). In addition, beyond the question of whether impact factors are important, a more interesting question in my view is why they still persist. What are they used for and why are they still such important "driver[s] of researchers' behaviour"? This was not the authors' question, and they do provide some contextualisation by quoting their participants, but still I think they could do more to contextualise what is known from the literature on that to draw out the implications here. The attribute label in the methods for IF is "ranking", but ranking according of what and for what? Not just average per-article citations in a journal over a given time frame. Rather, impact factors are used as a proxy indicators of less-tangible desirable qualities – certainly prestige (as the title of this article suggests), but also quality, trust (as reported by one quoted focus group member "I would never select a journal without an impact factor as I always publish in journals that I know and can trust that are not predatory", p.6), journal visibility, importance to the field, or improved chances of downstream citations or uptake in news media/policy/industry etc. Picking apart the interactions of these various factors in researchers' choices to make use of IFs (which is not in all cases bogus or unjustified) could add valuable context. I'd especially recommend engaging at least briefly with more work from Science and Technology Studies - especially Müller and de Rijcke's excellent Thinking with Indicators study (doi: 10.1093/reseval/rvx023), but also those authors other work, as well as work from Ulrike Felt, Alex Rushforth (esp https://doi.org/10.1007/s11024-015-9274-5), Björn Hammerfelt and others.

Response: We agree that the dominance of the impact factor is not a surprise, although we do present some new information on how it scales relative to other decisions such as the speed of peer review.

The continued importance of impact factors is a fascinating problem, and our results suggest it becomes even more important over career time; Figure 3 which shows a stronger preference for impact factors for researchers with more papers.

We have added text to the end of the first paragraph of the discussion that compare our results to those from Rushforth and de Rijcke, as we similarly found that there were researchers who thought of impact factors as measuring quality, but others who were more indifferent to how impact factors influenced what journal to submit to. Further research could explore why some researchers are indifferent to impact factors whilst others are enslaved to them.

We have also cited the Rowley et al paper, as they similarly found that authors value helpful peer review.

4. Disciplinary coverage: (1) A lot of the STS work I talk about above emphasises epistemic diversity and the ways cultures of indicator use differ across disciplinary traditions. For this reason, I think it should be pointed out in the limitations that this is research in Health/Med only, with questions on generalisability to other fields. (2) Also, although the abstract and body of the article do make clear the disciplinary focus, the title does not. Hence, I believe the title should be slightly amended (e.g., "Health and Medical Researchers Are Willing to Trade …")

Response: Thank you for your comment. This point was also raised by Reviewer #1 and we agree that it is important to emphasise the area of research. We have amended the title so that it now includes "Health and medical", and we have included a paragraph in our limitations section regarding generalisability.

Editorial assessment

In this article the authors use a discrete choice experiment to study how health and medical researchers decide where to publish their research, showing the importance of impact factors in these decisions. The article has been reviewed by two reviewers. The reviewers consider the work to be robust, interesting, and clearly written. The reviewers have some suggestions for improvements. One suggestion is to emphasize more strongly that the study focuses on the health and medical sciences and to reflect on the extent to which the results may generalize to other fields. Another suggestion is to strengthen the embedding of the article in the literature. Reviewer 2 also suggests to extend the discussion of the sample selection and to address in more detail the question of why impact factors still persist.

Response: We are grateful for these two useful peer reviews.

We have clarified that our study targeted health and medical researchers and might not be generalisable to other fields. We have read the helpful papers suggested by the reviewers, and expanded our reflections in the discussion to include these papers and some of their ideas.

We have rearranged the paper so that the methods section is before the results section. We've also added our ORCIDs to the first page.

Competing interest: Ludo Waltman is Editor-in-Chief of MetaROR working with Adrian Barnett, a co-author of the article and a member of the editorial team of MetaROR.