RESEARCH PAPER

Health and medical researchers are willing to trade their results for journal impact factors: results from a discrete choice experiment

Natalia Gonzalez Bohorquez[1], Sucharitha Weerasuriya[1], David Brain[1], Sameera Senanayake[2], Sanjeewa Kularatna[2] and Adrian Barnett[1]

[1]School of Public Health and Social Work, Queensland University of Technology, Brisbane
[2]Health Services and Systems Research, Duke-NUS Medical School, Singapore

Submission date: 24 March 2025; Acceptance date: 24 May 2025; Publication date: 5 December 2025

ABSTRACT

The fixation of the research community on the prestige of journals harms research quality as some researchers focus on where to publish instead of what. We examined researchers' publication preferences using a discrete choice experiment in a cross-sectional survey of international health and medical researchers. We asked researchers to consider two hypothetical journals and decide which one they would prefer. The hypothetical journals varied in their impact factor, formatting requirements, speed of peer review, helpfulness of peer review, editor's request to cut results and whether the paper would be useful for their next promotion. These attributes were designed using focus groups and interviews with researchers, with the aim of creating a tension between personal and societal benefit. Our survey found that the researchers' strongest preference was for the highest impact factor and the second strongest for a moderate impact factor. The least important attribute was a preference to make changes in format and wording instead of cutting results. Some respondents were willing to cut results in exchange for a higher impact factor. Despite international efforts to reduce the importance of the impact factor, it remains a driver of researchers' behaviour. The most prestigious journals may have the most partial evidence, as researchers are willing to trade their results for prestige.

DOI: 10.13169/Prometheus.40.4.0002

ORCID: https://orcid.org/0000-0001-6349-3170
ORCID: https://orcid.org/0000-0002-7622-3329
ORCID: https://orcid.org/0000-0002-6612-348X
ORCID: https://orcid.org/0000-0002-5606-2046
ORCID: https://orcid.org/0000-0001-5650-154X
ORCID: https://orcid.org/0000-0001-6339-0374
CONTACT: a.barnett@qut.edu.au
ACCEPTING EDITOR: Stuart Macdonald

Introduction

Peer-reviewed publications are academic currency (Génova *et al.,* 2015). Having sufficient publications in the bank is important for recruitment, promotion and funding (Schimanski and Alperin, 2018; Rice *et al.,* 2020). Publications are also a vital record of evidence that can improve policy and practice, and direct future research (Dawes *et al.,* 2005). Ideally, publications could be both useful as academic currency and sources of evidence for scientific progress. However, the value of publications as a currency may be trumping their main purpose to provide reliable evidence (Smaldino and McElreath, 2016). The intrinsic motivation of a 'taste for science' (described by Merton and Storer, 1973) may have been superseded by the extrinsic motivation of a 'taste for publications' (Binswanger, 2013). In the 'publish or perish' world, researchers may 'prefer popularity to intrinsic value' and hence focus on where to publish instead of what to publish (Génova *et al.,* 2015; Rushforth and de Rijcke, 2015; Müller and de Rijcke, 2017).

Most researchers regularly make considered decisions about which journal to submit to and how to navigate peer review (Rowley *et al.,* 2022). Factors include the journal's prestige (often defined using the impact factor), the target audience, the article processing charges, the required formatting, and the journal's rejection rate and turnaround times. The perfect home for a paper is rare (Maggio *et al.,* 2024), and researchers often need to make compromises to be successful (Anderson *et al.,* 2007). We aimed to study some of the important compromises that researchers make and thus examine how researchers publish their research. We were especially interested in the trade-offs that researchers make in personal benefits and wider benefits for society. We aimed to test trade-offs between earning academic currency and creating an accurate record of the evidence.

Methods

Designing the discrete choice experiment

We used a discrete choice experiment to examine researchers' publication preferences as this is well-suited to testing the multiple trade-offs that researchers make when publishing papers. We used a mixed methods approach (Bohorquez *et al.,* 2024b; Galárraga *et al.,* 2020) in which the rigorous development of the qualitative component informed the attribute list for the discrete choice experiment survey and the interpretation of the quantitative results. After gathering the quantitative results, we referred back to the results of the qualitative component, specifically verbatim quotes related to attributes, to identify possible explanations for the quantitative results. This triangulation is used in our discussion section.

We used multiple stages to design and implement the discrete choice experiment (see Figure 1 and Appendix 1 for details). With the aim of considering a wide range of attributes, we started with a review of the literature that examined one or more potential attributes. The review collected 77 potential attributes about publications, with most concerning the journal (e.g., impact factor), the impact (e.g., social media discussion) and the characteristics of the article (e.g., article with novel findings).

We used focus groups and in-depth interviews with health researchers from Australian academic institutions to explore the most important attributes, collect new attributes and test potential trade-offs. We recruited participants from our networks and maximized the variation in career stage, gender and research field. We conducted focus groups in clinical sciences (eight participants), public health and health services research (eight participants), and used interviews for the two participants in fundamental science as there were too few people for a focus group. The focus group was piloted with nine participants from health services research. The sample sizes for the focus groups and interviews were arbitrarily determined by the number of interested participants.

We used a semi-structured interview guide with an adapted nominal group technique without consensus (Bohorquez *et al*., 2024b). Participants were asked to imagine they had written a

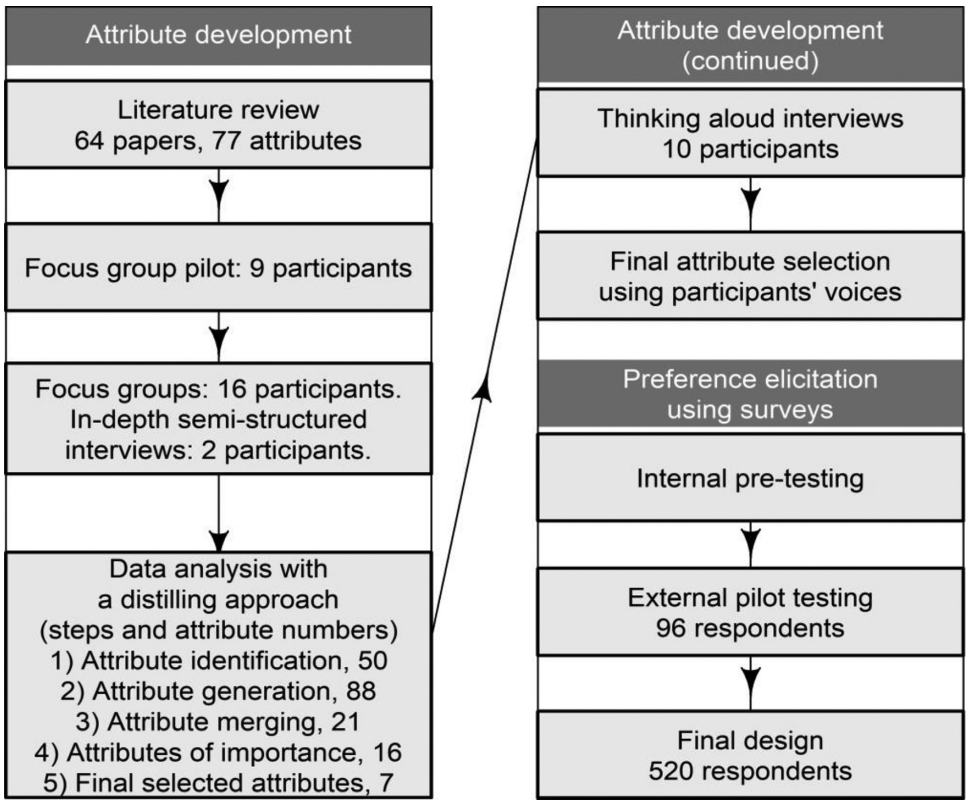

Figure 1. The stages of designing and deploying the discrete choice experiment to elicit researchers' publication preferences

paper and were now thinking of submitting it to a journal. They were asked about the attributes they consider most important when submitting to a journal. Each participant talked through up to ten attributes with the group and explained their choices. The attributes mentioned were then added to an online survey and participants voted for the most important attributes, explaining the rationale for their choices. We analysed and selected the attributes using the five steps of attribute development with a distilling approach (Bohorquez *et al*., 2024a). Qualitative data analysis using a distilling approach involves identifying attributes of importance by coding verbatim transcriptions and indexing quotes from participants in the resulting codebook or attribute matrix (Bohorquez *et al*., 2024a).

An initial design of eight attributes was tested using a thinking-aloud exercise with ten researchers (Leighton, 2017). Researchers were shown a choice task and were asked to discuss their thoughts aloud on whether:

- they had any comments on the content or wording
- there were any levels they struggled to understand or that seemed unrealistic
- the gaps between any levels were too jarring or obvious
- there was anything missing.

This exercise identified that an attribute on journal prestige was sometimes contradictory to an attribute on journal ranking, and hence the prestige attribute was removed.

Attribute and level selection

The final attributes and levels are in Table 1. In this section, we explain the choices behind the attributes and levels, and explain the perceived importance of some attributes and why some attributes were

Table 1. The six attributes and their levels for the discrete choice experiment

Label	Attribute	Level
Ranking	In your field the journal's impact factor is	- the highest - moderate - not available
Formatting	To fit the journal's style requirements your paper will need	- major formatting - minor formatting
Speed	Colleagues say that the journal's decisions are usually	- fast - slow
Reviews	Colleagues say that the journal's reviews will often	- help you improve the paper - be contradictory and unhelpful
Editor	After peer review the editor has indicated the paper will be accepted if you	- make changes in format and wording - cut a table and analysis to reduce the word count to 3,000
Promotion	Considering your next application for promotion or a fellowship, this paper will be	- useful - not useful

Note: The first column is a short label used to refer to the attributes.

excluded.The impact factor of the journal was the most common attribute in the literature review and was also frequently mentioned in focus groups and interviews. Participants suggested that its importance relied on self-serving purposes, such as job promotions, grants and funding, but it was also perceived as a reflection of the excellence of the researcher and a way to quantify the worth of their work. Related to the impact factor was the idea of predatory journals, which raised strong feelings of aversion due to reputational damage; e.g., 'I avoid them like the plague'. For the levels, we decided against numeric impact factors because these numbers vary by field (Althouse *et al*., 2008). Therefore we used a relative field ranking of the highest, middle and a journal without an impact factor, which could represent a new journal or a potentially predatory journal.

Formatting was often considered as a 'painful' process. Concerns were mentioned about the time needed to fit a journal's style requirements, and respondents wanted to avoid onerous systems. We used two simple levels of minor and major formatting. Peer review was widely discussed with researchers interested in the speed and quality of reviews. We framed both of these attributes by what their colleagues had told them, as colleagues were an important source of information about prospective journals. We used the relative labels of 'slow' and 'fast' rather than numeric review times (e.g., 30 days) because average times vary by field (Publons, 2018).

The focus group discussions uncovered a new issue as some researchers raised experiences of being asked by a journal editor to cut results from their paper at the peer-review stage. There were multiple potential reasons including the need to reduce word count, maintain a 'clean story', make the story 'digestible', remove results that contradicted previous findings, or remove findings that were not of interest to journals or colleagues. We included this as an attribute as it suited the tension we were aiming to test, being a trade-off between the loss of evidence from presenting an abridged version of the work against the potential benefit of earning a publication. A difference between this attribute and the others is that it occurs post-submission.

The final attribute was a direct appeal to personal benefit, as it concerned whether the paper was useful or not for their next promotion or fellowship application. An example of a good quality paper that researchers might not use in a fellowship application is a 'negative' study, where, for example, a new intervention or treatment did not work (often judged by the arbitrary statistical significance threshold of $p < 0.05$). 'Negative' studies can be less cited and receive less publicity than 'positive' studies (Koren, 1991; Greenberg, 2009), highlighting their reduced value as academic currency.

Article processing charges (APCs) were often discussed, but we excluded them as an attribute because they could often not be traded; for example, when researchers had no budget to pay APCs. Using 'charges' could have introduced a hypothetical bias, as researchers mostly do not personally pay the APCs and therefore the choices would not be as meaningful (Hensher, 2010). Citations were a common attribute in the literature review, but focus group discussions revealed that they were considered beyond the control of the researchers and somewhat subject to chance. Hence, it would not be plausible to use varying citation numbers as attribute levels. Supporting this decision, a prospective study of journal editors finds that citation counts are difficult to predict (Schroter *et al*., 2022).

Scenario

The scenario in Box 1 was shown at the beginning of the survey and was repeated in every choice task. The scenario framed the choice tasks and included some attributes of journal choice relevant to decision-making where (1) they could not be measured independently as they overlapped with other attributes, or (2) their importance was relative across participants or deterministic. For example, the scope and readership of the journal were often mentioned in focus groups as one of the most important attributes. However, as researchers were strongly unwilling to submit to journals outside their scope, we added it to the scenario. The two scenario endings were created because in the focus groups some researchers mentioned how they could change behaviour after experiencing rejections. To test this potential difference in the survey, the researchers were randomly assigned to view one or the other scenario ending in a 1:1 ratio. Some focus group participants mentioned that their previous experiences with a journal, good or bad, would strongly influence their choices. To avoid this concern, the scenario stated that the researchers did not have any experience with the journal. Similarly, we stated that they did not know the editorial staff, as this also influences researchers' journal choices.

Box 1. Scenario for the discrete choice experiment

> Imagine you have written a paper and are now trying to get it published in a journal. Your paper contains original research and is around 4,000 words long with tables and figures. Your paper is relevant to your field and you believe it is good quality. You will only consider journals that fit the scope of your paper and are read by your target audience. You have no previous experience of the journals (good or bad). You do not have any personal or professional relationships with the journal editors or publishers. You are the first author and will make all decisions on behalf of your co-authors.
> Scenario 1 ending: Your paper has not yet been submitted to any journal.
> Scenario 2 ending: Your paper has been submitted and desk-rejected (rejected without peer reviews) by two journals.

Dominant task

An example of a discrete choice task is shown in Figure 2. The choice tasks were unlabelled as the hypothetical journals were A and B. This example is the dominant choice task where journal A is clearly the most desirable. It was used to examine whether the respondents understood the task. It was shown as the first task to warm up respondents and was not used in data analysis. In Figure 2, journal A has the levels we assumed most respondents would prefer.

Survey of discrete choice tasks

The online survey started with a link to the participant information sheet (Appendix 2) and asked the researchers to indicate their consent. Fourteen respondents did not consent. Those who

Attribute	Journal A	Journal B
In your field the journal's impact factor is	the highest	not available
To fit the journal's style requirements your paper will need	minor formatting	major formatting
Colleagues say that the journal's decisions are usually	fast	slow
Colleagues say that the journal's reviews will often	help you improve the paper	be contradictory and unhelpful
After peer review the editor has indicated the paper will be accepted if you	make changes in format and wording	cut a table and an analysis to reduce the word count to 3,000
Considering your next application for promotion or a fellowship, this paper will be	useful	not useful

Figure 2. Example discrete choice task showing the attributes and levels

consented were shown the scenario (Box 1) and dominant task (Figure 2). The respondents next answered eight choice tasks. The final task was a repeat task of one of the eight. This was used to assess the stability of the participants' responses based on the percentage of respondents who gave the same answer as the original task (Özdemir *et al*., 2009). Differing answers could be a result of learning effects or fatigue. The repeat task was not used in the analysis.

The final section of the survey asked respondents if they found the choice tasks easy or difficult. We also gathered the following information from the respondents: their broad research area, gender, years of experience in research, number of published papers, country, and their perceived publication pressure. Lastly, respondents could add optional comments. Respondents could skip any question. The complete survey is available in Appendix 2.

The NGene software (version 13.0) was used to select 24 pairwise choice tasks based on the D-error to give an efficient fractional design. This D-efficient design was developed using the Modified Federov algorithm to estimate a multinomial logit model. For the final D-efficient design, the weights were selected from the pilot test. The 24 choice tasks were divided into three blocks of eight. Using a fractional design maximized the statistical efficiency of the design, while giving a manageable number of choice tasks of ten (eight plus the dominant task and retest).

Statistical methods

We used the panel mixed multinomial logit (pMMNL) model and the panel Latent Class Model (pLCM) for the main analysis. We also used the pMMNL model to examine whether preferences differed systematically on the basis of the characteristics of the respondents. The results are presented as mean utilities with 95% confidence intervals and the estimated importance of the attribute (Gonzalez, 2019). Subgroup analyses were performed using the following characteristics: years of experience, gender, number of publications, having a publication target and the hypothetical paper's prior rejection. The pLCM was used to capture non-systematic heterogeneity in preferences among respondents, assuming that differences in preferences manifest as discrete groups or latent classes (Greene and Hensher, 2003). The ideal number of classes was determined using the Akaike Information Criterion (AIC). A pLCM was used to assess the non-attendance of the task, incorporating a 'garbage class' to identify the respondents who provided non-informative responses (Gonzalez *et al*., 2024). This approach enabled an evaluation of preference heterogeneity that distinguished between attentive and non-attentive participants. Our data collection and analyses were preregistered in a study protocol (Barnett *et al*., 2024). The only change from our planned design was

that we did not use the pre-notification email for most invites, as it did not appear to increase the response rate. Our R code and data are available on GitHub (Barnett, 2024).

Sample size

Sample size formulae are not available for discrete choice experiments and estimates are often made using rules of thumb or simulations (Lancsar and Louviere, 2008; Reed Johnson *et al*., 2013). We faced uncertainty in selecting plausible model parameters with one to two parameters per discrete choice attribute and no similar previous studies. Hence, our final sample size was based on a pilot. Pilot testing has been recommended for informing sample size calculations for complex interventions (Lancaster *et al*., 2010). We analysed the pilot data of 51 respondents to inform the final design. The required sample size based on minimizing the D-error was 309. Both the pilot and the final design had 24 choice tasks in three blocks of eight. The attributes and levels were the same in the pilot and final design, hence we combined respondents from the pilot and final surveys in our analyses.

Sampling frame

Our target population was current health and medical researchers. We included authors from any country as research is a global endeavour, and hence we preferred not to focus on particular countries. We approached this population by creating a sampling frame of researchers extracted from papers on the PubMed database, which is a widely used search engine that contains the MEDLINE database of published papers in life sciences and biomedical topics (Sayers *et al*., 2021). To capture current researchers, we restricted the search to the year 2022 onwards. We used the 'publication type' search field to exclude non-research papers such as obituaries. We extracted only researchers who had an email available. The search was conducted on 11 April 2024. The search returned more than 140,000 papers, which we randomly reordered and iteratively extracted no more than one unique email per paper until we had a sample of 9,000 researchers. Randomly selected researchers from the sampling frame were sent an initial email with reminders one and two weeks later.

Results

Sample description

The surveys were collected between 26 March 2024 and 30 May 2024 (66 days) (see Appendix 3). The median time to complete the survey was seven minutes. We received 616 responses from 7,376 invitations, giving a response rate of 8.5% (this excludes 170 emails that were no longer active). A classification tree found that the response rate varied by email domain, with a higher predicted response rate of 21% for – amongst others – Australia, Switzerland and the UK, and a lower response rate of 3% for – amongst others – China, Germany and Japan (see Appendix 4). The questions were generally well completed, but there was some survey fatigue, with under 1% missing the first choice task and 15% missing the tenth and last choice task (Appendix 5).

Of the respondents 13% found answering the hypothetical choices difficult or very difficult. More than 99% of the respondents selected the dominant choice task, indicating an excellent understanding of the attributes and levels. The repeat choice task had the same answer as the original for 79% of the respondents, indicating good internal consistency. Summary statistics for the sample are in Table 2. Respondents had been working in research for a median of ten years and had a median of 43 peer-reviewed papers; 47% were female. The most popular broad research area was Clinical Sciences (57%). A personal target for their annual number of publications was noted by 40% of respondents.

Table 2. Summary statistics on the respondents' characteristics

Continuous characteristics		**Median (Q1 to Q3)**
		Years of experience 10 (6, 20)
		Number of peer-reviewed papers 43 (14, 100)
Categorical characteristics		
Characteristic	**Level**	***n* (%)**
Gender	Male	264 (51)
	Female	242 (47)
	Prefer not to say	5 (1)
	Non-binary/third gender	4 (1)
	I use a different term	2 (0.4)
Broad research Area	Clinical Science	297 (57)
	Public Health	72 (14)
	Other	56 (11)
	Health Services Research	49 (9)
	Basic/Fundamental Science	38 (7)
	Medical Education	11 (2)
Understanding of the hypothetical choices	Very difficult	7 (1)
	Difficult	61 (12)
	Moderate	165 (31)
	Easy	198 (38)
	Very easy	95 (18)
Target number of publications	Mandatory target from department/group	26 (6)
	Suggested target from department/group	60 (14)
	Personal target	173 (40)
	No target	200 (47)

Note: Whether researchers had a target number of publications was asked only in the final sample; respondents could tick multiple answers for this question. Q1 = first quartile, Q3 = third quartile. The sample included responses from 63 countries, with the three most common being USA (15%), UK (11%) and Australia (10%) (table of all countries in Appendix 6).

Researchers' preferences

The utilities for each attribute are in Figure 3 and Table 3. The Figure also shows the utilities stratified by the respondents' characteristics and the scenario wording concerning prior rejections.

The strongest preference was for the highest impact factor and the second strongest for the moderate impact factor. The least important attribute was a preference for making changes in format and wording compared with cutting a table and analysis. After the impact factor, the next strongest preference was for a helpful review. The utilities were similar for a fast review and minor formatting. Researchers had a clear preference for papers that were useful for their promotion. More experienced researchers had a stronger preference for the highest impact factor and minor formatting. Researchers who had more peer-reviewed papers had a much stronger preference for the highest and moderate impact factors. Female researchers had slightly stronger preferences for helpful reviews and papers that were useful for their promotion. There was little difference in researchers' preferences by whether the paper had been previously rejected.

The results of the latent class are in Figure 4. The optimal number of groups according to the AIC was four. The largest group had the strongest preferences for impact factor, a relatively small preference for fast results and a slight preference for cutting results over minor formatting. The second largest group had the strongest preference for a helpful review, with a much reduced – although still positive – preference for journal impact factor. The third group was not concerned

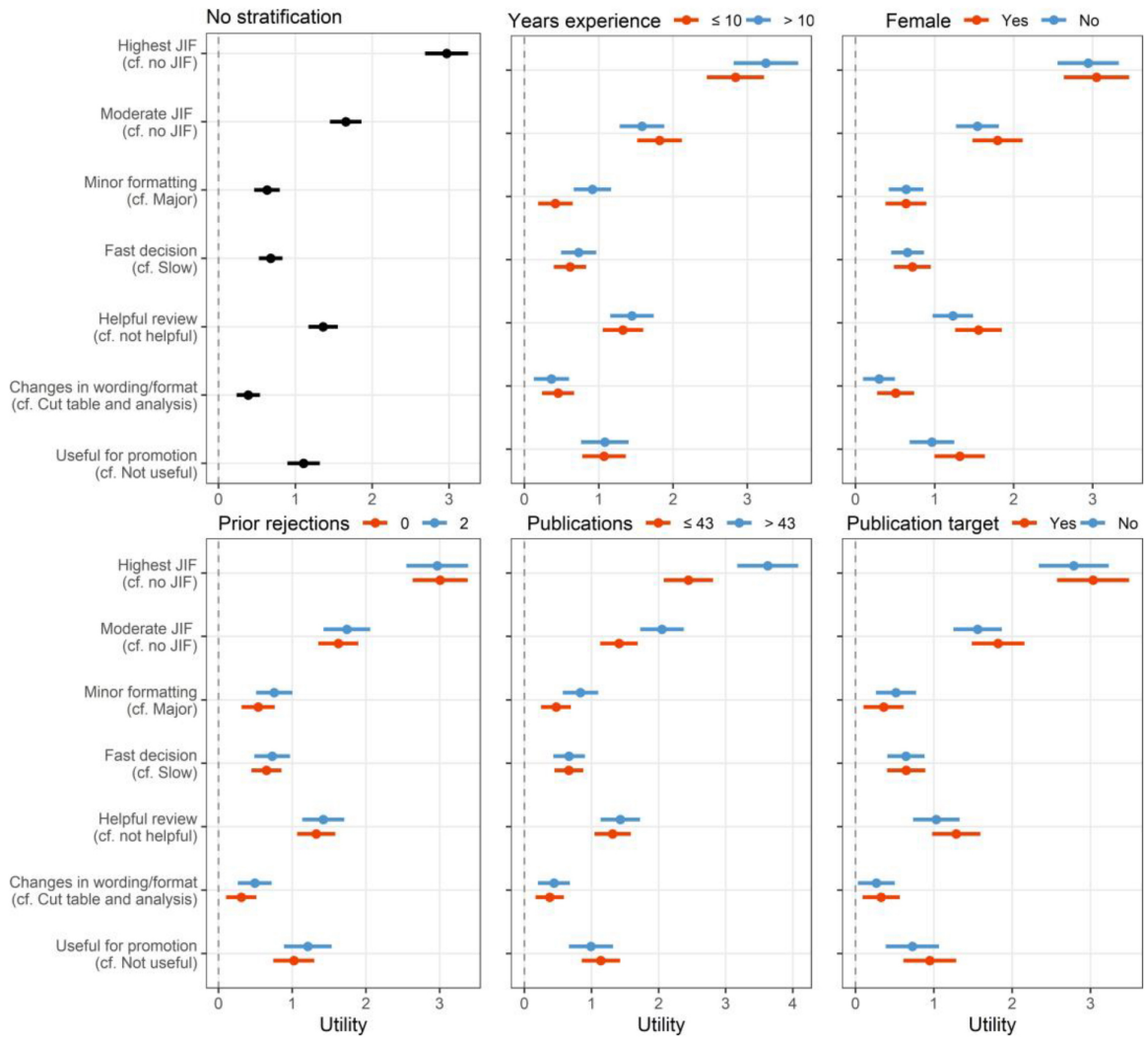

Figure 3. Utility estimates and 95% confidence intervals for the six attributes

Note: The dotted vertical line at zero is for no difference in utility. Forty-three publications was the sample median. JIF = journal impact factor. The upper limit on the utility axis scale varies by panel.

Table 3. Utilities for the journal preferences and attribute importance

Attribute and level (Reference level)	Mean	95% CI	Attribute importance (%)
Highest JIF (No JIF)	2.97	2.69 to 3.25	34
Moderate JIF (No JIF)	1.66	1.45 to 1.86	19
Minor formatting (Major)	0.63	0.46 to 0.80	7.2
Fast decision (Slow)	0.68	0.52 to 0.83	7.7
Helpful review (Unhelpful)	1.36	1.17 to 1.55	16
Changes in wording/format (Cut table and analysis)	0.39	0.23 to 0.54	4.4
Useful for promotion (Not useful)	1.11	0.90 to 1.32	13

Note: See Table 1 for the full wording of the attributes and levels. JIF = journal impact factor.

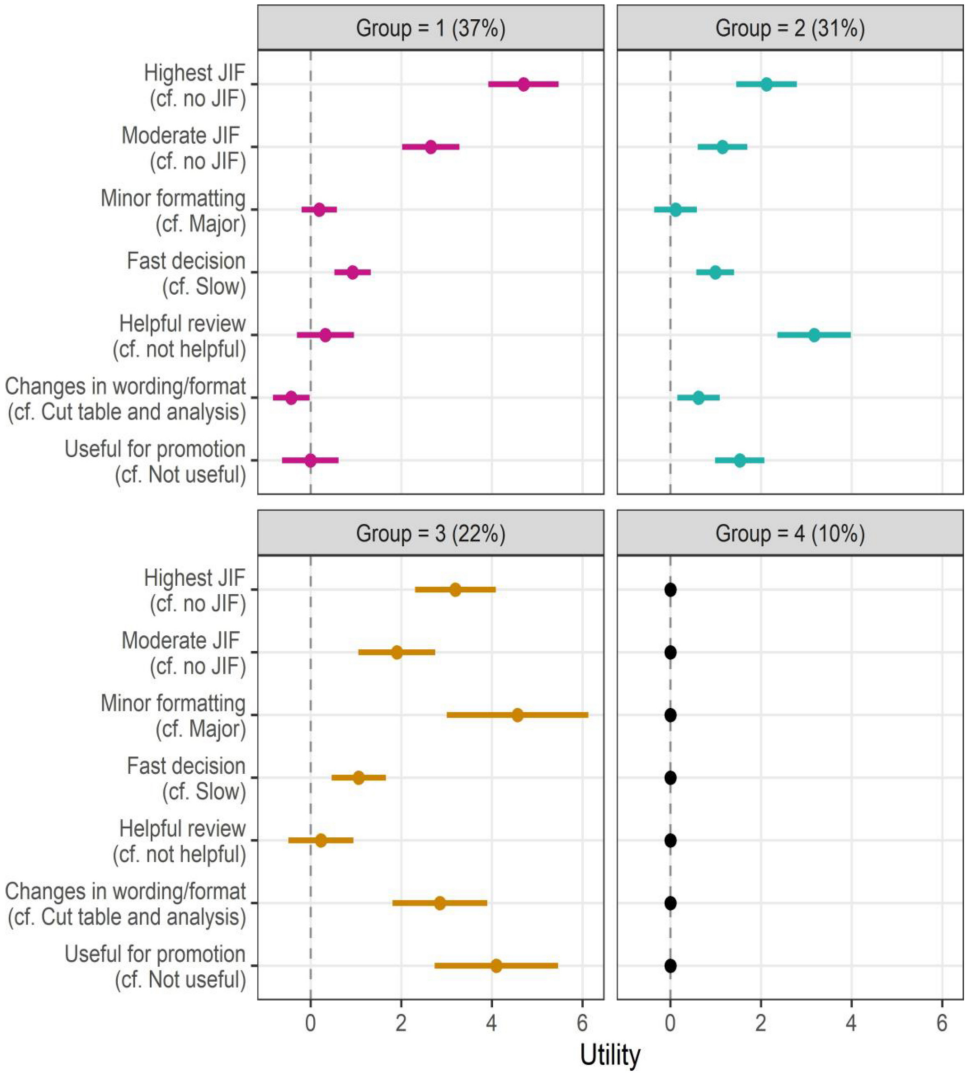

Figure 4. Utility estimates and 95% confidence intervals for the six attributes using a latent class model

Note: The percentages in the panel headers are the group sizes. The dotted vertical line at zero is for no difference in utility. JIF = journal impact factor.

about a helpful review, but strongly preferred minor over major formatting and a paper that was useful for their promotion. Non-informative responses were provided by 10% of respondents.

Interactions

The five planned interactions are plotted in Figure 5 with the estimates in Appendix 7. When the journal did not have an impact factor, there was a stronger preference for faster review. The journal rank had a similar interaction with both the editor's requests and the style requirements, since there was no difference in utility when the journal had no impact factor. This could indicate an indifference by researchers about their papers in journals without an impact factor. There was an interaction between the editor's requests and a helpful review: if the review was not helpful, then there was a stronger preference for formatting and wording changes over cutting results. However, for helpful reviews, researchers showed little difference between the editor's requests, which could be because they interpreted all requests as helpful. There was a small interaction between a helpful review and speed, as researchers were more willing to wait for a helpful review.

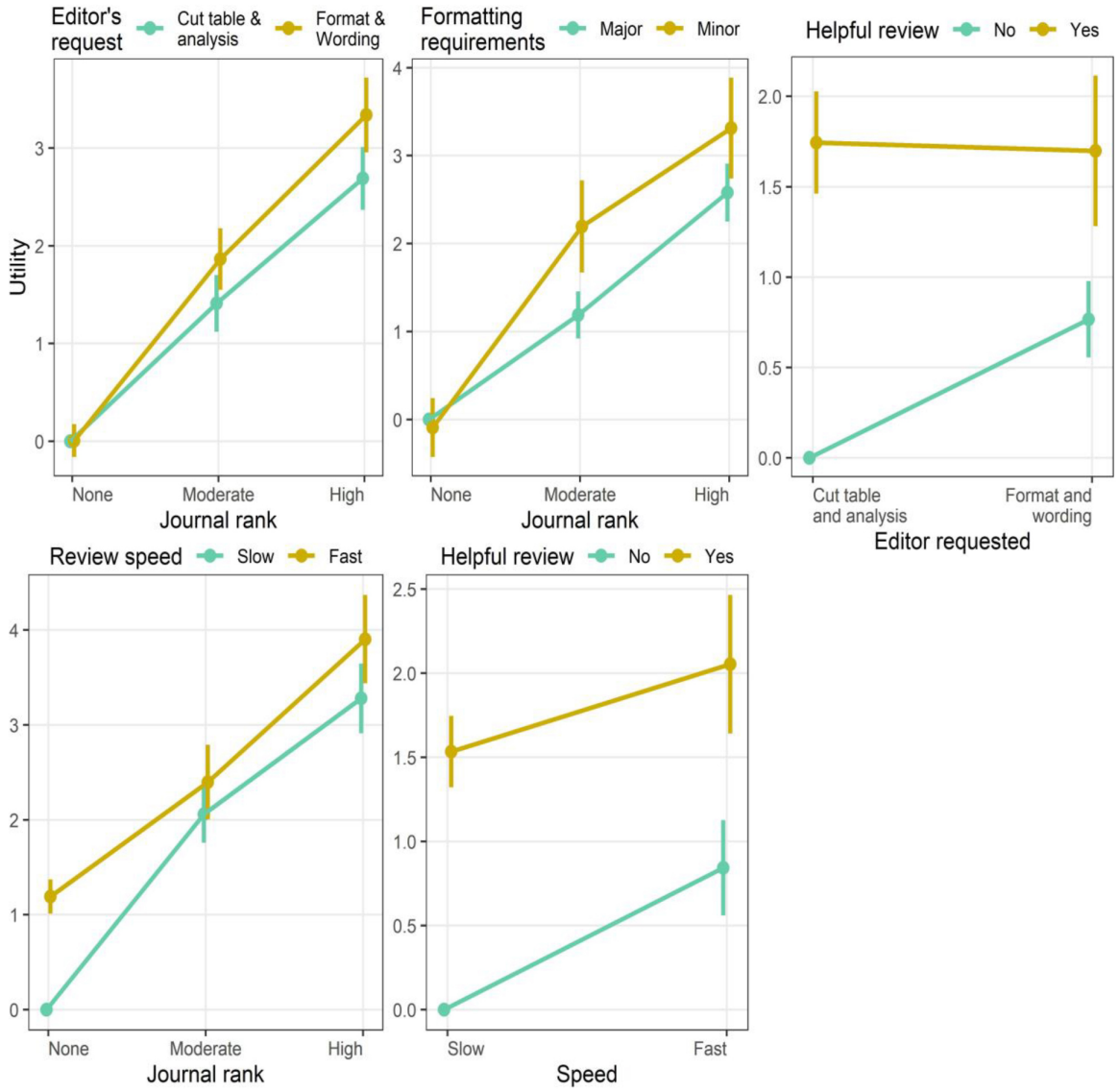

Figure 5. Utility estimates for the five planned interactions

Note: The dots are the means and the vertical lines are 95% confidence intervals. The reference group is the left-most level on the x-axis with the green line.

Discussion

Health and medical researchers had the strongest preference for the journal impact factor over any other attribute tested. This was both a desire for high impact factors and an aversion to journals without an impact factor. The importance of impact factor for researchers has been called an obsession (Müller and de Rijcke, 2017; Onstad and Sime, 2024), a mania (Casadevall and Fang, 2014), and a game that encourages questionable practices (Falagas and Alexiou, 2008). Major international initiatives have sought to reduce the influence of impact factors, such as DORA in 2012 (https://sfdora.org/), and RESQUE (Gärtner *et al*., 2024) and COARA in 2022 (https://coara.eu/). Despite these initiatives and extensive debate on the negative consequences of using impact factors to evaluate researchers, the highest possible impact factor is a target for many researchers.

A focus group participant framed impact factors as useful for 'quantifying my academic abilities', while another survey participant revealed a target threshold, 'I've been told if it isn't in an impact factor over 10 it doesn't matter/count'. Some researchers are using impact factors to verify

the quality of their work, which was also found in an investigation of how biomedical researchers use impact factors (Rushforth and de Rijcke, 2015). However, the same study found that other researchers were more ambivalent about impact factors and our latent class results had two groups where the impact factor was not the most important (Figure 4). More research on why researchers prioritize impact factors would be useful for the ongoing debate on improving research culture (Field *et al*., 2024).

Journals must be indexed for three years to obtain an impact factor, but some respondents interpreted a journal without an impact factor as predatory rather than new, as stated by a survey participant, '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'. This indicates that having an impact factor is a quality marker for some researchers and is probably why predatory journals falsify impact factors (Elmore and Weston, 2020). It also has implications for legitimate journals that lose their impact factor, such as the journal *eLife* because of its alternative peer-review system (Elifesciences 2024).

Researchers with more publications and more years of experience had a stronger preference for higher impact factors (Figure 3). This could be because some early career researchers have yet to understand the importance of impact factors. Another explanation is a survivorship effect, as researchers with high impact factor publications have an advantage in employment and promotion (Pitt and Mewburn, 2016; McKiernan *et al.,* 2019), while researchers with papers in less prestigious journals are outcompeted (Smaldino and McElreath, 2016).

Some survey respondents commented that they could not understand how a paper in a high impact factor journal could not be useful for their promotion or fellowship, which was a combination in the discrete choice tasks. This illustrates the power of the impact factor, as it trumps the content of the paper (Casadevall and Fang, 2014). A recent survey shows how the content of papers is commonly neglected by grant and hiring committees, since more than half used journal impact factor to assess quality (Hrynaszkiewicz *et al.*, 2024). When fellowship and hiring committees make career-changing decisions based on impact factors, a clear signal is sent to researchers to prioritize impact factors over content. A researcher in our interviews appeared comfortable with being assessed based on impact factors: 'People have to quantify me by something. So impact factor is a very important way to do that.' However, a focus group participant recognized that impact factors are usually meaningless when considering real-world impact: 'I've been working together with senior executives in the government and federal government. They don't care about that [journal impact factor], they only want you to give them a half-page summary.'

A focus group participant gave a perspective on impact factors that was pragmatic and confessional: 'Considering and admitting for everybody, for various reasons, usually go for a top ranked journal in its field, and everything, and some of that will be purely mercenary, because that's what's required.' Personal values are ceded to the reward systems that use impact factors and/or journal ranking. We aimed to distinguish researchers with a stronger focus on system requirements by asking if they had a target number of publications per year and 53% had a personal and/or institutional target. However, having a target did not greatly alter the researchers' preferences (Figure 3). Potentially, most researchers are 'playing the game' and the preference for journal ranking remains high regardless of the desired number of publications (Chapman *et al*., 2019).

A surprising result was the lack of difference in researchers' preferences for papers that are useful for promotion by experience and publication numbers (Figure 3). This could be because the competition for funding and promotion never ends and researchers are always looking to earn academic currency. Tenured or retired professors may be under less pressure (Niles *et al*., 2020) and a professor from the focus groups commented: 'I am the least strategic person when it comes to publishing, but I think that also comes with seniority as I have no need to ever write a promotion application again!'

The survey participants were randomly assigned to a scenario in which their hypothetical article had not yet been submitted to a journal or had already been desk-rejected twice (Box 1). This was raised in the focus groups, with comments including: 'But then, after many rejections, right?

You just want to get it out.' However, in the survey, the previous rejections did not have an effect as the researchers' preferences were remarkably similar (Figure 3). Researchers' preferences may be impervious to rejection, as the logical approach is to continue to pursue the highest impact factor possible. Preferences may change with more than two rejections or if the rejections were after peer review rather than desk rejections.

The lowest utility was for an editor's request for formatting changes, compared with cutting a table and analysis. On average, researchers preferred not to cut their analysis, but this was less of a priority than the impact factor, formatting at the submission stage or the speed of peer review. In the latent class analysis, the group with the strongest preference for the impact factor had a surprising preference for cutting results (Figure 4), showing a willingness to compromise on their evidence to be published in prestigious journals (Casadevall and Fang, 2014). This compromise was also discussed in our focus groups as a likely trade-off during the peer-review process: 'I certainly have examples where I have cut things out of papers to try and get something published.' Cutting results has also been discussed in the literature; for example, 'Academics who play the "publish or perish" game have a strong incentive to … accept all "suggestions" by the referees even if one knows that they are misleading or even incorrect' (Frey *et al*., 2009), and how during peer review 'authors … remove ideas and insights that they believe in from their work' (Eisen *et al.,* 2022). To the best of our knowledge, our survey is the first to show this compromise empirically. An important implication is that the journals with the highest impact factors potentially have the most partial evidence, as researchers are more willing to 'hold their nose' to satisfy editors of influential journals (Maggio *et al.,* 2024). One could argue that the journals were correct and that the cuts improved the paper. However, the scenario we presented to researchers was 'you believe it [your paper] is good quality' and the cut was 1,000 out of 4,000 words and included a table. Some researchers potentially rationalized this compromise, thinking that the removed results could be included in a supplement, but this relegates their findings to the whim of an editor (Schmid, 2017). The willingness of researchers to select results when negotiating with high impact journals may partly explain why papers in high impact journals have a low reliability (Brembs, 2018).

An interesting finding from the focus groups and the survey is that the researchers showed a relatively strong preference for helpful reviews and were willing to wait longer for helpful reviews. For example, an interview respondent said, 'If there's something that can improve them [my papers], I want them to be improved'. The preference for helpful reviews did not change by the researchers' experience or number of publications (Figure 3), so it was not restricted to early career researchers. The latent group analysis showed that the second largest group preferred a helpful review (Figure 4). The relatively strong preference for helpful reviews shows clear support for peer review, as many researchers value the expertise of their peers. Similarly, an international survey on peer review found that 93% disagree with the claim that peer review is unnecessary and 85% believe that peer review benefits scientific communication (Ware, 2008). Another international survey of authors found that the usefulness of reviewers' feedback was an important factor for authors' journal choice (Rowley *et al.,* 2022).

Related studies

Previous studies have examined researchers' publication preferences using hypothetical journal choices. As with our results, the journal's impact factor dominated preferences, more important than the journal's editorial board, the journal's standing among peers, the quality of reviews, the waiting time for reviews and the probability of being accepted (Rousseau and Rousseau, 2012). Journal prestige, described using 'journal level', was also the most important attribute for junior authors in a conjoint analysis that compared the prestige of the journal, the number of authors, the order of the authors and the investment of time of researchers (Krasnova *et al.,* 2014). A choice-set survey found that researchers were willing to trade citations for a more prestigious journal (Salandra *et al.,* 2021). A discrete choice experiment examined the metrics academics use when choosing papers to read

(Lemke *et al.,* 2021). There were clear preferences for citation counts, followed by journal impact factor and download counts.

Limitations

Our discrete choice experiment was hypothetical and examined stated preferences rather than revealed preferences. Our scenario told the respondents to assume that they would make all the decisions, but the reality of multi-author teams is that decisions may be a consensus or compromise of the authors' competing priorities and experiences.

The low response rate (8.5%) reduces our ability to generalize and creates a non-response bias. Our approach email included words such as 'journal' and 'publishing' and so may have appeared similar to the nuisance journal requests that researchers regularly receive and may have been automatically or manually deleted. Our survey respondents could be more engaged with the publication process than the general population. We found a difference in response rate by country, hence our results over-represent some countries.

The interviews and focus groups used to select the final attributes were all conducted with Australian researchers and therefore the wording of the attributes and levels may have been less meaningful to some respondents. We took an international approach and included researchers from any country; however, there will be differences among countries in the pressure to publish in high impact journals. Hence, our average results shown here may not be generalizable to many countries. However, the 'publish or perish' mantra applies across the research world, and the pressure to publish was commonly cited as a key problem in a recent international survey of biomedical researchers' opinions on research reproducibility (Cobey *et al.,* 2024).

We included only health and medical researchers, and hence we cannot generalize our results to other fields of science. Health research is one of the largest in terms of funding (Gibbons, 2023) and number of publications (National Science Board/National Science Foundation, 2024), and has multiple journals with extremely high journal impact factors (some over 100). Therefore, it could be relatively competitive compared to other fields, meaning that researchers are more willing to compromise to succeed.

Acknowledgements

The focus groups and interviews were approved by the Queensland University of Technology Human Research Ethics committee (18 April 2023, number: LR 2023-6685-13695). The survey was approved by the Queensland University of Technology Human Research Ethics committee (5 March 2024, number: LR 2024-8188-18148). This work received funding from an internal grant from the Centre for Healthcare Transformation at Queensland University of Technology.

References

Althouse, B. *et al.* (2008) 'Differences in impact factor across fields and over time', *Journal of the American Society for Information Science and Technology*, 60, 1, pp.27–34. https://doi.org/0.1002/asi.20936

Anderson, M. *et al.* (2007) 'The perverse effects of competition on scientists' work and relationships', *Science and Engineering Ethics*, 13, 4, pp.437–61. https://doi.org/10.1007/s11948-007-9042-5

Barnett, A. (2024) 'Code and data for a discrete choice experiment of authors' preferences'. https://doi.org/10.5281/zenodo.12814359; https://github.com/agbarnett/publication_preferences

Barnett, A. *et al.* (2024) 'Study protocol: a discrete choice experiment to examine researchers' publication preferences: an international cross-sectional survey'. https://doi.org/10.17605/OSF.IO/P9GUJ

Binswanger, M. (2013) 'Excellence by nonsense: the competition for publications in modern science' in Bartling, S. and Friesike, S. (eds) *Opening Science*, Springer, Cham, pp.49–72. https://doi.org/10.1007/978-3-319-00026-8_3

Bohorquez, N., Malatzky, C., McPhail, S., Kularatna, S. *et al.* (2024a) 'Enhancing health preferences research: guidelines for qualitative attribute development in stated preference studies', Center for Open Science (OSF), available at https://osf.io/g9jbt (accessed August 2025).

Bohorquez, N., Malatzky, C., McPhail, S., Mitchell, R. *et al.* (2024b) 'Attribute development in health-related discrete choice experiments: a systematic review of qualitative methods and techniques to inform quantitative instruments', *Value in Health*, 27, 6. https://doi.org/10.1016/j.jval.2024.05.014

Brembs, B. (2018) 'Prestigious science journals struggle to reach even average reliability', *Frontiers in Human Neuroscience*, 12. https://doi.org/10.3389/fnhum.2018.00037

Casadevall, A. and Fang, F. (2014) 'Causes for the persistence of impact factor mania', *mBio*, 5, 2. https://doi.org/10.1128/mbio.00064-14

Chapman, C. *et al.* (2019) 'Games academics play and their consequences: how authorship, h-index and journal impact factors are shaping the future of academia', *Proceedings of the Royal Society B*, 286, 1916, paper 20192047. https://doi.org/10.1098/rspb.2019.2047

Cobey, K. *et al.* (2024) 'Biomedical researchers' perspectives on the reproducibility of research', *PLoS Biology*, 22, 11, pp.1–15. https://doi.org/10.1371/journal.pbio.3002870

Dawes, M. *et al.* (2005) 'Sicily statement on evidence-based practice', *BMC Medical Education*, 5, 1. https://doi.org/10.1186/1472-6920-5-1

Eisen, M. *et al.* (2022) 'Peer review without gatekeeping', *eLife*, 11. https://doi.org/10.7554/elife.83889

Elifesciences (2024) 'The eLife Model: an update on progress following changes in Web of Science indexing status', available at https://elifesciences.org/inside-elife/c11c6101/the-elife-model-an-update-on-progress-following-changes-in-web-of-science-indexing-status (accessed August 2025).

Elmore, S. and Weston, E. (2020) 'Predatory journals: what they are and how to avoid them', *Toxicologic Pathology*, 48, 4, pp.607–10. https://doi.org/10.1177/0192623320920209

Falagas, M. and Alexiou, V. (2008) 'The top-ten in journal impact factor manipulation', *Archivum Immunologiae et Therapiae Experimentalis*, 56, 4, pp.223–6. https://doi.org/10.1007/s00005-008-0024-5

Field, S. *et al.* (2024) 'Exploring the dimensions of responsible research systems and cultures: a scoping review', *Royal Society Open Science*, 11, 1. https://doi.org/10.1098/rsos.230624

Frey, B., Eichenberger, R. and Frey, R. (2009) 'Editorial ruminations: publishing *Kyklos*', *Kyklos*, 62, 2, pp.151–60. https://doi.org/10.1111/j.1467-6435.2009.00428.x

Galárraga, O. *et al.* (2020) 'iSAY (incentives for South African youth): stated preferences of young people living with HIV', *Social Science & Medicine*, 265, paper 113333. https://doi.org/10.1016/j.socscimed.2020.113333

Gärtner, A., Leising, D. and Schönbrodt, F. (2024) 'Towards responsible research assessment: how to reward research quality', *PLoS Biology*, 22, 2, e3002553. https://doi.org/10.1371/journal.pbio.3002553

Génova, G., Astudillo, H. and Fraga, A. (2015) 'The scientometric bubble considered harmful', *Science and Engineering Ethics*, 22, 1, pp.227–35. https://doi.org/10.1007/s11948-015-9632-6

Gibbons, M. (2023) 'R&D expenditures at US universities increased by $8 billion in FY 2022', National Center for Science and Engineering Statistics, available at https://ncses.nsf.gov/pubs/nsf24307 (accessed August 2025).

Gonzalez, J. (2019) 'A guide to measuring and interpreting attribute importance', *The Patient: Patient-Centered Outcomes Research*, 12, 3, pp.287–95. https://doi.org/10.1007/s40271-019-00360-3

Gonzalez, J. M., Johnson, F. R. and Finkelstein, E. (2024). 'To pool or not to pool: Accounting for task non-attendance in subgroup analysis', *Journal of Choice Modelling*, 51, p. 100487. https://doi.org/10.1016/j.jocm.2024.100487

Greenberg, S. (2009) 'How citation distortions create unfounded authority: analysis of a citation network', *BMJ*, 339, b2680. https://doi.org/10.1136/bmj.b2680

Greene, W. and Hensher, D. (2003) 'A latent class model for discrete choice analysis: contrasts with mixed logit', *Transportation Research Part B: Methodological*, 37, 8, pp.681–98. https://doi.org/10.1016/S0191-2615(02)00046-2

Hensher, D. (2010) 'Hypothetical bias, choice experiments and willingness to pay', *Transportation Research Part B: Methodological*, 44, 6, pp.735–52. https://doi.org/10.1016/j.trb.2009.12.012

Hrynaszkiewicz, I., Novich, B., Harney, J. and Kierme, V. (2024) 'A survey of how biology researchers assess credibility when serving on grant and hiring committees', available at https://osf.io/preprints/metaarxiv/ht836_v1 (accessed October 2025).

Koren, G. (1991) 'Bias against negative studies in newspaper reports of medical research', *JAMA: Journal of the American Medical Association*, 266, 13, p.1824. https://doi.org/10.1001/jama.1991.03470130104037

Krasnova, H. *et al.* (2014) 'Publication trade-offs for junior scholars in IS: conjoint analysis of preferences for quality, first authorship, collaboration, and time', Proceedings of the International Conference on Information Systems (ICIS). https://aisel.aisnet.org/icis2013/proceedings/GeneralISTopics/6/

Lancaster, G. *et al.* (2010) 'Trials in primary care: statistical issues in the design, conduct and evaluation of complex interventions', *Statistical Methods in Medical Research*, 19, 4, pp.349–77. https://doi.org/10.1177/0962280209359883

Lancsar, E. and Louviere, J. (2008) 'Conducting discrete choice experiments to inform healthcare decision making: a user's guide', *PharmacoEconomics*, 26, 8, pp.661–77. https://doi.org/10.2165/00019053-200826080-00004

Leighton, J. (2017) *Using Think-Aloud Interviews and Cognitive Labs in Educational Research. Understanding Qualitative Research*, Oxford University Press, Oxford.

Lemke, S., Mazarakis, A. and Peters, I. (2021) 'Conjoint analysis of researchers' hidden preferences for bibliometrics, altmetrics, and usage metrics', *Journal of the Association for Information Science and Technology*, 72, 6, pp.777–92. https://doi.org/10.1002/asi.24445

Maggio, L. *et al.* (2024) '"The best home for this paper": a qualitative study of how authors select where to submit manuscripts', *Perspectives on Medical Education*, 13, 1. https://doi.org/10.5334/pme.1517

McKiernan, E. *et al.* (2019) 'Use of the journal impact factor in academic review, promotion, and tenure evaluations', *eLife*, 8. https://doi.org/10.7554/elife.47338

Merton, R. and Storer, N. (1973) *The Sociology of Science: Theoretical and Empirical Investigations*, Phoenix Books, Chicago.

Müller, R. and de Rijcke, S. (2017) 'Thinking with indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences', *Research Evaluation*, 26, 3, pp.157–68. https://doi.org/10.1093/reseval/rvx023

National Science Board/National Science Foundation (2024) *Publications Output: US Trends and International Comparisons*. https://ncses.nsf.gov/pubs/nsb202333/ (accessed November 2024).

Niles, M. *et al.* (2020) 'Why we publish where we do: faculty publishing values and their relationship to review, promotion and tenure expectations', *PLoS One*, 15, 3, pp.1–15. https://doi.org/10.1371/journal.pone.0228914

Onstad, D. and Sime, K. (2024) 'The ethical and social effects of the obsession over journal impact factor', *Annals of the Entomological Society of America*, 117, 3, pp.160–2. https://doi.org/10.1093/aesa/saae013

Özdemir, S. *et al.* (2009) 'Who pays attention in stated-choice surveys?', *Health Economics*, 19, 1, pp.111–18. https://doi.org/10.1002/hec.1452

Pitt, R. and Mewburn, I. (2016) 'Academic superheroes? A critical analysis of academic job descriptions', *Journal of Higher Education Policy and Management*, 38, 1, pp.88–101. https://doi.org/10.1080/1360080X.2015.1126896

Publons (2018) *Global State of Peer Review Series*, Clarivate Analytics. https://publons.com/static/Publons-Global-State-Of-Peer-Review-2018.pdf.

ReedJohnson, F. *et al.* (2013) 'Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force', *Value in Health*, 16, 1, pp.3–13. https://doi.org/10.1016/j.jval.2012.08.2223

Rice, D. *et al.* (2020) 'Academic criteria for promotion and tenure in biomedical sciences faculties: cross sectional analysis of international sample of universities', *BMJ*, 369. https://doi.org/10.1136/bmj.m2081

Rousseau, S. and Rousseau, R. (2012) 'Interactions between journal attributes and authors' willingness to wait for editorial decisions', *Journal of the American Society for Information Science and Technology*, 63, 6, pp.1213–25. https://doi.org/10.1002/asi.22637

Rowley, J. *et al.* (2022) 'Factors influencing researchers' journal selection decisions', *Journal of Information Science*, 48.3, pp.321–35. https://doi.org/10.1177/0165551520958591

Rushforth, A. and de Rijcke, S. (2015) 'Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands', *Minerva*, 53, 2, pp.117–39. https://doi.org/10.1007/s11024-015-9274-5

Salandra, R., Salter, A. and Walker, J. (2021) 'Are academics willing to forgo citations to publish in high-status journals? Examining preferences for 4* and 4-rated journal publication among UK business and management academics', *British Journal of Management*, 33, 3, pp.1254–70. https://doi.org/10.1111/1467-8551.12510

Sayers, E. *et al.* (2021) 'Database resources of the national center for biotechnology information', *Nucleic Acids Research*, 50, D1, pp.D20–D26. http://dx.doi.org/10.1093/nar/gkab1112

Schimanski, L. and Alperin, J. (2018) 'The evaluation of scholarship in academic promotion and tenure processes: past, present, and future', *F1000Research*, 7, paper 1605. https://doi.org/10.12688/f1000research.16493.1

Schmid, S. (2017) 'Five years post-DORA: promoting best practices for research assessment', *Molecular Biology of the Cell*, 28, 22, pp.2941–4. https://doi.org/10.1091/mbc.e17-08-0534

Schroter, S. *et al.* (2022) 'Evaluation of editors' abilities to predict the citation potential of research manuscripts submitted to the BMJ: a cohort study', *BMJ*, 379. https://doi.org/10.1136/bmj-2022-073880

Smaldino, P. and McElreath, R. (2016) 'The natural selection of bad science', *Royal Society Open Science*, 3, 9, paper 160384. https://doi.org/10.1098/rsos.160384

Ware, M. (2008) 'Peer review in scholarly journals: perspective of the scholarly community – results from an international study', *Information Services & Use*, 28, 2, pp.109–12.

Appendices

Appendix 1. Details of the discrete choice experiment design

An additional file includes details on the literature review, focus groups and interviews, and thinking aloud exercise. It is available at https://osf.io/gjch7.

Appendix 2. Participant information sheet and survey questions

The online version of the participant information sheet is available at https://osf.io/p9guj/wiki/home/. A PDF version of the survey is available at https://osf.io/j7mce. The survey was delivered online using Qualtrics. The survey questions differed by two questions between the pilot and final survey as we altered the question that aimed to examine researcher's publishing expectations. This is because for the original statement – 'My department's or research group's expectations with respect to publishing are reasonable' – 81% responded Agree or Strongly Agree, creating limited variance between respondents. Hence, in the main survey we asked researchers if they had an annual publication target and what it was.

Appendix 3. Survey responses over time

Figure A1.

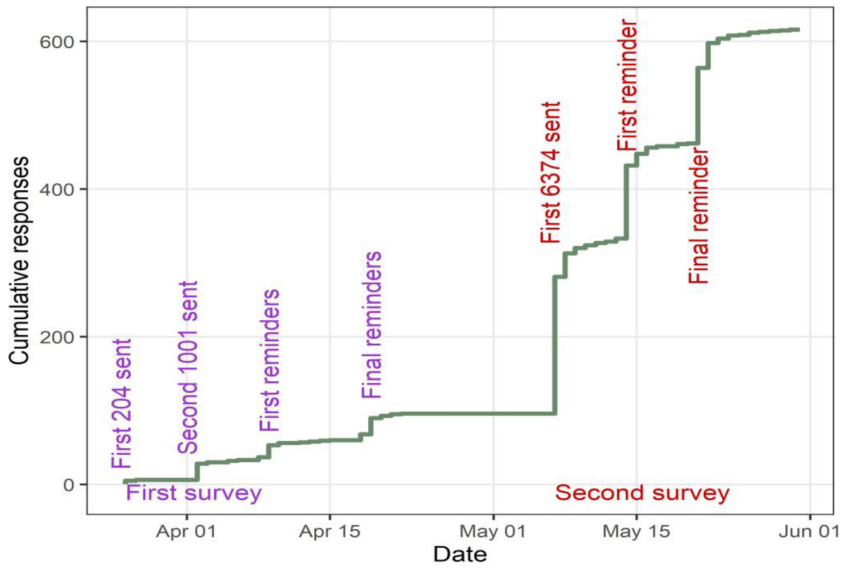

Figure A1. Cumulative number of survey responses over time for the pilot and final design

Appendix 4. Classification tree predicting survey response

We used a classification tree to predict the response to the survey (yes/no) based on the researchers' email domain (a proxy for the country, for example, au = Australia), and whether the researcher's affiliation mentioned the words "Hospital", "Dentist*" or "University". The classification tree had three leaves with a cross-validated error of 0.990 and a standard error of 0.034. The tree used only the email domain, but found a relatively large difference in response proportions. We present the results as a table instead of a plotted tree, as the number of email domains makes the plot cluttered (Table A1).

Table A1. Results of the classification tree for predicting survey response using email domain

Email domains	Number of researchers (%)	Response proportion
126, 163, ai, al, alumni.cern, bd, bg, cc, cn, co, cu, cz, de, ec, ehu.eus, et, eu, for.paris, ge, gh, gm, gr, health.mil, hotmail, hr, in, iq, ir, is, ivi.int, jp, ke, kr, lb, lt, lv, ma, mt, mx, net, nhs, om, pe, pt, qa, qq, rs, ru, sc, sg, si, sina, sy, tn, tw, ua, us, vn, who.int, yahoo, yuhs.ac	2487 (34%)	0.03
ar, at, be, br, ca, cat, com, edu, e.g., es, fi, fr, gmail, gov, hk, it, nl, no, org, pk, pl, sa, se, th, tr	4109 (56%)	0.09
ae, au, ch, cl, cy, dk, ee, gu, hu, id, i.e., il, jo, lk, lu, mw, my, ng, np, nz, ro, to, tropmedres.ac, ug, uk, za	788 (11%)	0.21

Appendix 5. Item-missing data

The amount of missing data by question number is shown in Figure A2. The missing data patterns are clustered by similarity. The question numbers are presented in order. There is evidence of survey fatigue as the percentage of missing answers increases from left to right.

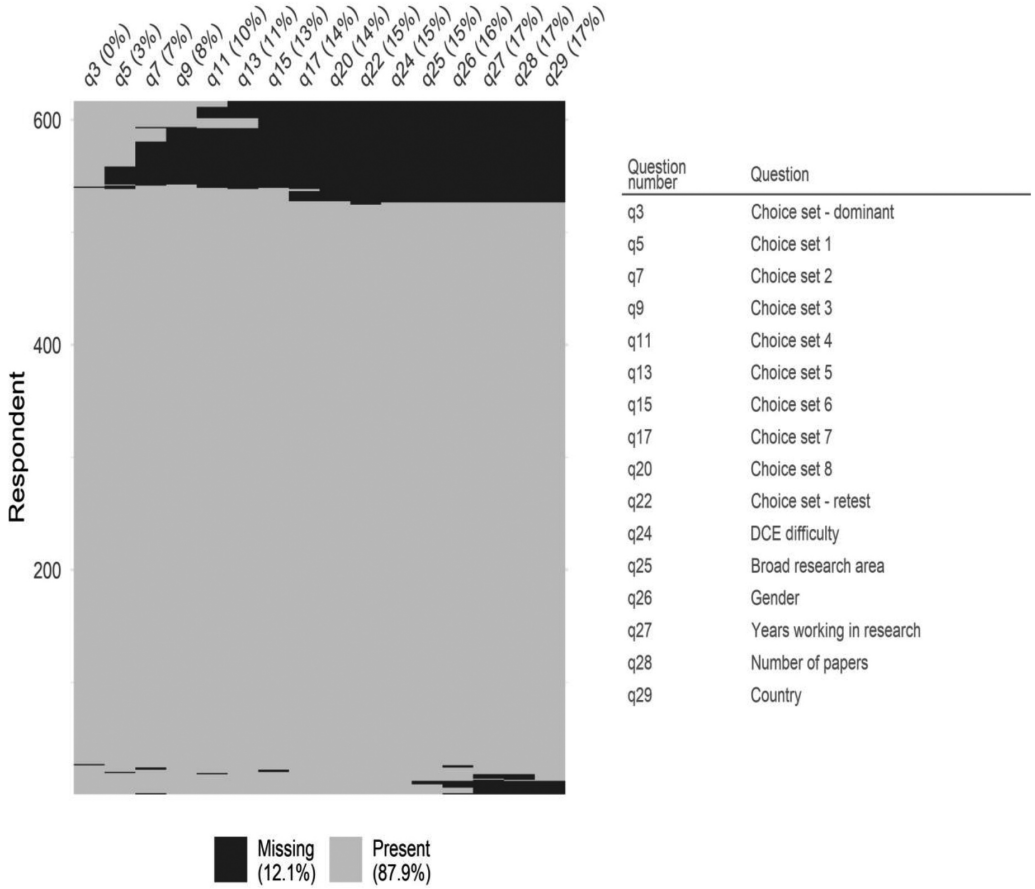

Figure A2. Item missing data for the 616 survey responses

Note: The column headings show the question number and percent missing. The panel on the right shows the questions for each question number.

Appendix 6. Respondents' countries

Table A2

Table A2. Number and % of responses by country

Country	Number	%	Country	Number	%
United States	78	15	Bangladesh	3	1
United Kingdom	55	11	Ethiopia	3	1
Australia	49	10	Indonesia	3	1
Italy	20	4	Malawi	3	1
Brazil	19	4	New Zealand	3	1
Spain	19	4	Nigeria	3	1
China	18	4	Saudi Arabia	3	1
India	14	3	Greece	2	<1
Tu¨rkiye	14	3	Kenya	2	<1
Canada	13	3	Nepal	2	<1
France	13	3	Pakistan	2	<1
Germany	13	3	Romania	2	<1
Netherlands	13	3	United Arab Emirates	2	<1
Switzerland	13	3	Argentina	1	<1
Japan	12	2	Austria	1	<1
Sweden	11	2	Benin	1	<1
Ireland	9	2	Croatia	1	<1
Denmark	8	2	Cyprus	1	<1
Egypt	8	2	Estonia	1	<1
Norway	7	1	Hungary	1	<1
Colombia	6	1	Jordan	1	<1
Iran	6	1	Kuwait	1	<1
South Africa	6	1	Luxembourg	1	<1
Finland	5	1	Namibia	1	<1
Israel	5	1	Panama	1	<1
Korea, South	5	1	Portugal	1	<1
Belgium	4	1	Russia	1	<1
Chile	4	1	Singapore	1	<1
Malaysia	4	1	Tanzania	1	<1
Mexico	4	1	Thailand	1	<1
Poland	4	1	Uganda	1	<1
Taiwan	4	1			

Note: There were 63 countries in total.

Appendix 7. Attribute interactions

Table A3

Table A3. Utility estimates and 95% confidence intervals for the planned interactions between attributes

Attribute interaction	Utility difference	95% CI	p-value
High rank × Editor's request	0.64	0.23 to 1.06	0.0025
Moderate rank × Editor's request	0.45	0.04 to 0.86	0.0323
High rank × Formatting requirements	0.83	0.37 to 1.28	<0.001
Moderate rank × Formatting requirements	1.10	0.67 to 1.53	<0.001
High rank × Speed	–0.57	–1.00 to –0.13	0.0104
Moderate rank × Speed	–0.85	–1.29 to –0.42	<0.001
Helpful review × Speed	–0.32	–0.63 to –0.02	0.0384
Editor's request × Helpful review	–0.81	–1.15 to –0.48	<0.001

Note: The interactions are plotted in Figure 5. This table shows only the interaction terms and not the main effects. These results help judge the null hypothesis of whether there was no interaction.