Evaluation of editors’ abilities to predict the citation potential of research manuscripts submitted to The BMJ: a cohort study

Sara Schroter, 1 Wim E J Weber, 1 Elizabeth Loder, 1 Jack Wilkinson, 2 Jamie J Kirkham 2

ABSTRACT

OBJECTIVE
To evaluate the ability of The BMJ editors to predict the number of times submitted research manuscripts will be cited.

DESIGN
Cohort study.

SETTING
Manuscripts submitted to The BMJ, reviewed, and subsequently scheduled for discussion at a prepublication meeting between 27 August 2015 and 29 December 2016.

PARTICIPANTS
10 BMJ research team editors.

MAIN OUTCOME MEASURES
Reviewed manuscripts were rated independently by attending editors for citation potential in the year of first publication plus the next year: no citations, below average (10 citations), average (10-17 citations), or high (>17 citations). Predicted citations were subsequently compared with actual citations extracted from Web of Science (WOS).

RESULTS
Of 534 manuscripts reviewed, 505 were published as full length articles (219 in The BMJ) by end of 2019 and indexed in WOS, 22 were unpublished, and one abstract was withdrawn. Among the 505 manuscripts, the median (IQR [range]) number of citations in the year of publication plus the following year was 9 (4-17 [0-150]); 277 (55%) manuscripts were cited <10 times, 105 (21%) were cited 10-17 times, and 123 (24%) cited >17 times. Manuscripts accepted by The BMJ were cited more highly (median 12 (3-24) citations) than those rejected (median 7 (3-12) citations). For all 10 editors, predicted ratings tended to increase in line with actual citations, but with considerable variation within categories; nine failed to identify the correct citation category for >50% (range 31%-52%) of manuscripts, and κ ranged between 0.01 to 0.19 for agreement between predicted and actual categories. Editors more often rated papers that achieved high actual citation counts as having low citation potential than the reverse. Collectively, the mean percentage of editors predicting the correct citation category was 43%, and for 160 (32%) manuscripts at least 50% of editors predicted the right category.

CONCLUSIONS
Editors weren’t good at estimating the citation potential of manuscripts individually or as a group; there is no wisdom of the crowd when it comes to BMJ editors.

Introduction
“Impact factor mania” is a “debilitating and destructive epidemic.”1-5 One of the criticisms of the impact factor, which has become a proxy for the relative importance of a journal, is that it is easy to game.6-8 A former editor of JAMA, George Lundberg, admitted he embarked on a deliberate strategy to increase the impact factor of his journal in the 1980s.9 If this can be done deliberately, this would suggest that journal editors can select papers that will attract numerous citations.6 The BMJ’s impact factor rose from 5 in 2000 to 96 in 2021, suggesting that our editors are quite brilliant in predicting citation potential of unpublished research. This would be highly unusual.9

Although many studies have identified factors influencing citability, accurately predicting citation numbers has proved difficult.10-19 Few studies have successfully used information available before publication for this purpose.19 Recently machine learning techniques have been used to find patterns that predict citation count with some success.20 21

To try to prove our own brilliance scientifically (before we are replaced by machine learning) is a challenge few editors would be able to resist, and we were no different. We thus set out to test our ability to predict citations of unpublished research papers submitted to The BMJ.

Methods
At The BMJ, research papers with favourable external peer reviews are discussed at a weekly decision-making committee meeting (supplementary fi g 1). Between 27 August 2015 and 29 December 2016, research team editors were invited to independently assess the citation potential of manuscripts due to be discussed at these committee meetings. Editors were masked to the ratings given by others. The impact factor for The BMJ at the start of the survey was 17.445. While
this calculation includes non-research articles and a different citation period, we used it as an average value for the number of citations to BMJ research papers. Editors were told to indicate how many citations they thought each manuscript would generate in the year it was first published plus the first calendar year after publication. Editors chose from the following categories: no citations; below The BMJ average (<10 citations in the year of publication plus one year); around The BMJ average (10-17 citations in the year of publication plus one year); or more than The BMJ average (>17 citations in the year of publication plus one year). Our sample size was based on when we thought editors were getting bored with this weekly request.

Study participants
A self selected sample of 10 (6 men, 4 women) fiercely competitive, paid BMJ research editors of varying experience working across several specialties. The most important core members of the research team participated, but we can't reveal who they were. We excluded the paid statistical advisors attending meetings in case they were better than us.

Identification of subsequent published articles and citations
We knew which articles were published in The BMJ. For articles rejected from The BMJ, we pasted key parts of the submitted manuscript title, author names, or elements of the submitted abstract into Google's search engine to try to match submissions with publications. If there was no obvious match, we searched authors’ institutional affiliation pages, conducted author searches in Web of Science (WOS), or searched by ORCID and trial registry number where available. If the first extractor could not find a publication, a second extractor tried, and final decisions were based on discussion and agreement. All searches were completed by 10 May 2022, when we extracted citation data from Web of Science Core Collection (all editions, 2009 to present).

Inclusion and exclusion criteria
If more than one version of an article was discussed, we included only the first version discussed. We excluded research submissions submitted for The BMJ's Christmas issue as these papers are always brilliant. We also excluded submissions for which we could not find a publication, or resulted in publications not indexed in WOS, or were not published as full length journal articles, or were published after 2019, and one submission that was published as an abstract and withdrawn.

Statistical analysis
Primary analysis
The category “no citations” was used sparingly, so we pooled the categories “no citations” and “below average” citations for analysis. For each editor, we calculated the number of manuscripts they assigned to the correct citation category. We calculated κ statistics (95% confidence intervals) for each editor, using Fleiss-Cohen weights,22 as implemented in the R package vcd.23 We calculated how often each editor’s classification was “extremely wrong” (a highly cited paper was estimated to have low citation potential or vice versa). We compared editors’ ability to predict citation potential for articles published in The BMJ versus elsewhere. We assessed accuracy separately for Research Methods and Reporting (RMR) articles, as these are typically highly cited. To assess the editors as a group, we calculated the mean percentage of editors identifying the correct category per manuscript and the number and percentage of manuscripts for which at least 50% of the editors were correct.

Secondary analysis
Additional unplanned analyses included repeating the primary analyses using cut-offs of >30 and >50 citations to look at ability to predict the citation potential of the most highly cited articles and using the outcome of citations in the first two years after publication to reflect the impact factor calculation.

Patient and public involvement
One of the criticisms of the impact factor is that it does not measure clinical relevance.24 We agree, and we doubt that there are many patients losing sleep over citations of research papers. Thus, no patients were specifically involved in defining the research hypothesis or the outcome measures, nor were they involved in the design and implementation of the study.

Results
Sample
Of 534 manuscripts discussed at 69 meetings, 23 manuscripts were excluded because we could find no subsequent publication (n=18), or only an abstract or preprint (n=3), or a substantially different article (n=1), or a withdrawn abstract (n=1) (see supplementary fig 2). The 511 retrieved manuscript-publication pairs were screened for eligibility, and a further six were excluded as the journal of publication was not indexed in the WOS Core Collection (n=4) or it was published after 2019 (n=2). Editors rated different numbers of manuscripts (range 35-496; median 239 [IQR 74-388]) depending on meeting attendance. Among the 505 eligible manuscript-publication pairs, 219 (43%) were published in The BMJ. Accepted manuscripts (median 12 [IQR 7-24]) generated more citations than rejected manuscripts (median 7 [IQR 3-12]). The 22 unpublished manuscripts received 108 ratings; 57 (53%) low, 30 (28%) average, and 21 (19%) high ratings.

Ability to estimate actual citations
For the 505 papers, the median (IQR [range]) number of actual citations in the year of publication plus the following year was 9 (4-17 [0-150]). A total of 277 (55%) manuscripts generated <10 citations, 105
Nine of 10 editors were unable to identify the correct citation category for more than 50% (range 31%-54%) of manuscripts (supplementary fig 3). Agreement between estimated and actual categories for all editors ranged from $\kappa=0.01$ to 0.19. Editors did not frequently rate papers with low actual citation counts as having high citation potential; more commonly highly cited papers were rated as having low citation potential (supplementary table 1, supplementary fig 4).

For BMJ and non-BMJ articles, estimates were very imprecise for some editors because of the small sample size. Some estimates were less than zero (treated as zero), and $\kappa$ ranged from 0 to 0.19 for BMJ and 0 to 0.11 for non-BMJ articles. None of the editors predicted the citation category correctly for 50% or more of The BMJ articles (35% to 45%). For non-BMJ articles, 2/10 editors rated 50% or more correctly (range 26% to 63%). Accuracy of estimates for RMR articles are reported in the supplementary file.

For the collective assessment of editors across all 505 manuscripts, the mean percentage of editors predicting the correct citation category was 43%, and for only 160 (32%) manuscripts at least 50% of editors predicted the right category (table 1). Editors were collectively better at predicting the citation potential for manuscripts that would not be highly cited and manuscripts that would later be rejected by The BMJ and RMRs. Collective estimated citation potential was higher for manuscripts accepted for publication in The BMJ compared with those published elsewhere (table 2).

Secondary analysis
When the primary analysis was repeated using the outcome of citations in the first two calendar years after the year of publication, the results were similar (see supplementary material). Editors were poor at predicting the citation potential category for the most highly cited manuscripts (those generating >30 citations or >50 citations). Fifty five manuscripts (11%) received more than 30 citations, and editors generally rated these as “average” or “low” (range 50% to 95% incorrect). A substantial proportion of these manuscripts were rated as “low” (range across editors 15% to 51%). As there were only 20 papers receiving...

Table 1 | Editors’ collective assessment of manuscripts

<table>
<thead>
<tr>
<th>Citation category:</th>
<th>Mean percentage of correct ratings per manuscript</th>
<th>No (%) of manuscripts where ≥50% of editors predicted correct category</th>
</tr>
</thead>
<tbody>
<tr>
<td>All manuscripts (n=505)</td>
<td>43%</td>
<td>160 (32%)</td>
</tr>
<tr>
<td>Manuscripts with low citations (n=277)</td>
<td>55%</td>
<td>133 (48%)</td>
</tr>
<tr>
<td>Manuscripts with average citations (n=105)</td>
<td>34%</td>
<td>17 (16%)</td>
</tr>
<tr>
<td>Manuscripts with high citations (n=123)</td>
<td>23%</td>
<td>10 (8%)</td>
</tr>
<tr>
<td>Manuscript decision:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accepted in BMJ (n=219)</td>
<td>36%</td>
<td>64 (20%)</td>
</tr>
<tr>
<td>Rejected by BMJ (n=286)</td>
<td>47%</td>
<td>116 (41%)</td>
</tr>
<tr>
<td>Type of manuscript:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Methods and Reporting manuscripts (n=32)</td>
<td>47%</td>
<td>13 (41%)</td>
</tr>
<tr>
<td>Research manuscripts (n=473)</td>
<td>42%</td>
<td>147 (31%)</td>
</tr>
<tr>
<td>Superstar manuscripts:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With &gt;30 citations (n=55)</td>
<td>31%</td>
<td>8 (15%)</td>
</tr>
<tr>
<td>With &gt;50 citations (n=20)</td>
<td>39%</td>
<td>6 (30%)</td>
</tr>
</tbody>
</table>
>50 citations, there were only a few ratings per editor (median 9 (IQR 3.75-14)). All editors except one rated most of these incorrectly (that is, estimated that they would receive <17 citations) (supplementary table 2).

Discussion

Principal findings
Editors can be motivated to publish highly citable manuscripts because these inflate impact factors and editors’ sense of self importance, but this can bias which articles get published, where, and when. The accuracy of our editors’ predictions for citation potential was generally poor for papers rejected and accepted for publication in The BMJ, and only one editor consistently identified “superstar” manuscripts. Editors did predict the citation potential of manuscripts they went on to accept as higher than those they rejected. While this could indicate that predicted citation potential influences editorial decisions, it could also reflect the methodological quality and importance of the accepted articles. Editors were not good at predicting citation potential either alone or as a committee (no wisdom of the crowd when it comes to editors).

Comparison with other studies
If we had bothered to do a literature review before we started, we wouldn't have needed to do this study as others had already shown we reject many highly citable papers. Consistent with other journals following the fate of their manuscripts, we found that those rejected and published elsewhere were cited significantly less than those we accepted. Of course, the visibility and perceived quality associated with publications in The BMJ may have contributed to the higher citations. Computational modelling and machine learning are developed in 2017, so we can usually get away with having inadequate training. Article and authorship characteristics may have influenced editors’ predictions and decision making. Numerous studies have shown that characteristics of manuscripts (such as study quality, study design, sample size, setting, topic, specialty, characteristics of results, funding, number of references), authorship (names, number of authors, previous articles published, etc), and journals can influence citations.

Strengths and limitations of this study
Our study has a number of potential limitations. Firstly, identification of published articles can be difficult as article titles often change substantially from submission, but we made use of all the additional submission data authors needlessly have to upload to our database on first submission to help with the searches. Secondly, as citation data are given per calendar year, as with the impact factor calculation, papers published earlier in the year have a citation advantage. Thirdly, papers rejected at editorial meetings may have had less citation potential as they were probably mainly published in journals with lower impact factors and later when the topics may have been of lower interest. Fourthly, The BMJ has a large international audience, a high impact factor, and a low acceptance rate, which may also have influenced citations received and the generalisability of our results. Finally, our primary analyses were based on a different citation window to the impact factor calculation, but, as acknowledged by the creator of the impact factor, the metric could just as well be based on the previous year’s articles alone.

Policy implications
Our new editor in chief should consider hiring different editors if he wants to publish more highly cited papers, but there is no evidence that these people exist. Editorial decision making is a complex process, yet editors, like peer reviewers, rarely bother to seek training. Core competencies for biomedical editors, a useful framework for assessing editors, were only developed in 2017, so we can usually get away with having inadequate training. Article and authorship characteristics may have influenced editors’ predictions and decision making. Numerous studies have shown that characteristics of manuscripts (such as study quality, study design, sample size, setting, topic, specialty, characteristics of results, funding, number of references), authorship (names, number of authors, previous articles published, etc), and journals can influence citations.

Conclusions
Most editors tended to be on the cautious side; they did not often rate papers with low actual citation counts as having high citation potential but more commonly rated papers that were highly cited as having low citation potential. We see this as a good thing as The BMJ editors do try to focus on the quality of manuscripts and the importance of the content for their readership rather than be swayed by impact factor mania. That said, we probably can’t resist the temptation to unmask the data and seek the advice only from the single editor who could predict correct citation categories more than 50% of the time.

Thanks to Nilanee Nehrujee for assistance with data collection and The BMJ research editors, who are unnamed to avoid embarrassing them, for participating. We thank Professor Robin Ferner, who kindly acted as the handling editor for this manuscript. Contributors: SS, WW, EL designed the study. SS collected the data. JW analysed the data with support from JJK. All authors interpreted the data. All authors contributed to the writing of the manuscript and approved the final version to be published. SS is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. Funding: This project received no additional funding. Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare:

Table 2 | Number (percentage) of ratings in the low, average, and high prediction citation categories for manuscripts published in The BMJ, published elsewhere, and not published

<table>
<thead>
<tr>
<th>Predicted citation category</th>
<th>Published BMJ (n=219)</th>
<th>Published non-BMJ (n=286)</th>
<th>Unpublished (n=22)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>361 (36%)</td>
<td>817 (60%)</td>
<td>57 (25%)</td>
</tr>
<tr>
<td>Average</td>
<td>430 (42%)</td>
<td>398 (29%)</td>
<td>30 (28%)</td>
</tr>
<tr>
<td>High</td>
<td>223 (22%)</td>
<td>151 (11%)</td>
<td>21 (19%)</td>
</tr>
</tbody>
</table>

no support from any organisation for the submitted work; SS, WEW, and EL are employed by or seconded to The BMJ; JW holds statistical or methodological editor roles for Cochrane Gynaecology and Fertility, BIOMG (International Journal of Obstetrics and Gynaecology), Reproduction and Fertility, and Fertility and Sterility, no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: We did not seek ethical approval for the study. As we were too busy navel gazing we forgot to seek ethical approval, but we probably didn’t need it anyway.

Data sharing: The anonymised dataset and code can be made available to researchers on request to the corresponding author.

Transparency: The lead author (SS) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported. All important aspects of the study have been omitted, and that any discrepancies from the study as planned (and, if relevant, registered) have been explained. All errors are the fault of her coauthors. As we realised that this evaluation of our capabilities might backfire spectacularly, we did not want to draw unnecessary attention to it and decided not to register the study.

Dissemination to participants and related patient and public communities: We will tell our colleagues, who participated in the survey, the results, but anonymised, and we have to work with each other in the future. We have presented results at a conference, and the response was hysterical laughter, precluding any more attempts at dissemination.

Provenance and peer review: Not commissioned, externally peer reviewed. Because members of BMJ staff were involved in the conduct of this research and writing the paper, assessment and peer review have been carried out entirely by external advisers. No member of BMJ staff has been involved in making the decision on this paper. We have been carried out entirely by external advisers. No member of BMJ staff has been involved in making the decision on this paper. We have presented results at a conference, and the response was hysterical laughter, precluding any more attempts at dissemination.

We will tell our colleagues, who participated in the survey, the results, but anonymised, to prevent animosity, as these are all ambitious individuals, and we have to work with each other in the future. We have presented results at a conference, and the response was hysterical laughter, precluding any more attempts at dissemination.


Bryan G, Fletcher J, Kale R. Do you need to be an editor to accept or reject research papers sent to the BMJ? Abstract presented at the Fifth International Congress on Peer Review and Biomedical Publication, September 16-18, 2005, Chicago, Illinois.


Supplementary material: Fig 1, The BMJ review process; fig 2, flow diagram; table 1, analysis of the extreme disagreements for the 10 editors; table 2, analysis of the accuracy of predicting superstar manuscripts (n=20); fig 3, proportion of articles with incorrect and correct ratings by editor; fig 4, proportion of articles with extreme disagreements