Journals’ instructions to authors in 2017: a protocol for a cross sectional study across all disciplines

Mario Malički, 1 IJsbrand Jan Aalbersberg, 2 Lex Bouter, 3,4 Gerben ter Riet 1

1 Department of General Practice, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
2 Elsevier, Amsterdam, The Netherlands
3 Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands
4 Department of Philosophy, Faculty of Humanities, Vrije Universiteit, Amsterdam, The Netherlands

Corresponding author:
Mario Malički
Department of General Practice,
Academic Medical Center, University of Amsterdam
Amsterdam, The Netherlands
e-mail: mario.malicki@mefst.hr

Funding: This study is a part of Elsevier funded project: Fostering Transparent and Responsible Conduct of Research: What can Journals do?. Details of the project are available at the project’s data repository site: http://dx.doi.org/10.17632/53cskkwpdn.1.

Statement of interests: IJA is Senior VP of Research Integrity for Elsevier. Other authors declare no competing interests.

Reporting: This protocol was drafted using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.1

Acknowledgments: We would like to thank Ana Jerončić for advice on the sample size and randomization procedures, Adrian van der Weel, Catriona Fennell, René Bekkers, Sam Bruinsma, and Frits Rosendaal for advice on the topics to be checked within instruction to authors; Anne Consemulder and Ludo Waltman for explanations regarding Scopus List and SNIP values, and finally Nataliya Demikova, Fu Longlong, Jun Steed Huang, Chao Chen, Tengfei Tang, Lanfa Liu, and Chunxiang Cui for helping us find journals’ websites or contact information.
Abstract

Background

Scientific journals have gone through many changes, from paper-based to predominantly online content, increase in the number of authors, structured format, and lately in calls for increase in the transparency of conducting and reporting. Therefore, we sought to learn what journals, recommend to or demand from authors, by analysing a sample of journals’ instructions to authors (ItA) that is representative for all disciplines.

Methods and analysis

A cross-sectional analysis of documents: journals’ instructions to authors and journals’ scope statements, were collected, from 14 December 2017 till 24 January 2018, from a representative sample of journals indexed in Scopus (n=835). The instructions will be analysed using a combination of data extraction using regular expression matching and reading of documents with the emphasis on the recommendations regarding correction of published works, authorship, statements of interests, data sharing, detrimental research practices, obtaining ethics approvals for studies, use of reporting guidelines, statistical advice, addressing study limitations, peer review description, pre-registration of studies and replication of studies. Mentioning of these topics will be compared between journals from Arts and Humanities, Life, Social, Physical, Health, and Multidisciplinary sciences, and possible influence of the journals Source Normalized Impact Factor (SNIP) values and scientific area explored using multiple regression.

Ethics and dissemination

No ethics approval was sought for this study, as this is an analysis of publicly available data. All accompanying data, and authors’ version of the final manuscript will be shared on our project’s data repository site.
Introduction

Since their origin in the 17th century, scientific journals have gone through many changes. From unstructured abstract and manuscript format to formal structuring,² increase in the number of authors and shared (first or last) authorship,³ from paper-based to predominantly online content,⁴ development of different payment and distribution methods,⁴ as well as methods of impact measurement of manuscripts and journals.⁵ Lately, there has also been a drive towards prospective study registration⁶,⁷ publishing of manuscripts on pre-print servers before they are peer-reviewed,⁸ use of reporting guidelines to address the completeness of reporting,⁹ data sharing, conducting replication studies,¹⁰ and more emphasis on post-publication peer review.¹¹ Seen by some as ideal ways of doing research, many of these initiatives are also meant as means to foster responsible conduct of research and combat detrimental research practices.¹² However, these practices are not globally harmonized or enforced. Studies have shown that detrimental research practices still stain the scientific publishing,¹³ with up to 50 percent of conducted studies not being published,¹⁴ main outcome measures being changed from study protocols to study publications,¹⁴ hypotheses being generated after results are known,¹⁵ and improper statistical approaches being used to analyse data.¹⁶

Against this background, we sought to learn what journals, as influential players in the scientific enterprise, recommend to or demand from authors by analysing a sample of journals’ instructions to authors (ItA) that is representative for all disciplines. Additionally, we will compare recommendations or demands among journals from Arts and Humanities, Life, Social, Physical, Health, and Multidisciplinary sciences.
Methods

Study design

A cross-sectional study of documents: 1) journals’ instructions to authors and 2) journals’ scope statements, collected from the journals’ websites.

Data Sources

A list of journals was obtained from Scopus, on 26 September 2017 by downloading the Scopus Source List. From the list we excluded inactive journals, trade journals and book series, as well as journals with no 2016 Source Normalized Impact Factor (SNIP) values (a journal-level citation and impact metric). Scopus classifies journals as belonging to one of the four top-level areas (Life, Social, Physical, and Health Sciences) or assigns a Multidisciplinary (general journal) designation. Beside the four top-level areas, the journals are further classified according to Scopus’s All Science Journal Classification (ASJC) code which contains 26 scientific fields and 333 specialty subfields. The number of subfields within the fields varies from 4 to 49 (median is 11). As our goal was to compare scientific areas, we made the following adjustments to the Scopus Source List. First, after noting that some journals were missing in the Scopus multidisciplinary category (e.g. PLOS ONE and PEERJ), we added all journals in Science Citation Index Expanded – Multidisciplinary sciences (n=66) to the Scopus list of multidisciplinary journals (thus excluding them from the fields and specialty subfields assigned to them in Scopus, leaving 20,887 journals). Second, we “promoted” Arts and Humanities (classified in Scopus to belong to the top-level area of Social Sciences) to a new top-level area, thus creating the following six areas: Arts and Humanities, Life, Social, Physical, Health, and Multidisciplinary sciences. Finally, from the 20,887 journals, we selected only journals that belonged to a single top-level area (n=14,708, 70.4%).

We then conducted a proportionate stratified (based on ASJC classification and SNIP terciles) random sampling of journals in the following way:
1) Sample size needed was calculated using an 8% margin of error and 95% confidence level for each top-area using a freely available online program.

Example: In Scopus there were 3966 journals indexed as belonging exclusively to the top-level area Health sciences. Using the before mentioned program and selecting an 8% margin of error and 95% confidence level for a population of 3966 the required sample size was 145.

2) To calculate the sample of journals needed per ASJC field, we first calculated the proportion of journals belonging to each ASJC field within the top-level area, and then multiplied that proportion with the sample size obtained in step 1.

Example: The top-level area Health sciences is divided in Scopus to 5 ASJC fields: Medicine, Nursing, Veterinary, Dentistry, and Health Professions. We calculated the proportion of journals belonging exclusively to each of those fields (Medicine n=3103, 78.24%; Nursing n=149, 3.76%; Veterinary n=94, 2.37%; Dentistry n=102, 2.57%; and Health Professions n=44, 1.11%, respectively) and the proportion of journals indexed in more than one of those fields (n=474, 11.95%). We then multiplied the obtained proportions with 145 (sample size obtained in step 1). Here we give one example for Nursing journals: 0.0376 x 145 yielded 5.45 required journals.

3) As previous studies had shown that journals impact factor may play a role in the information contained within the ItAs, we wanted to sample an equal number of journals within each SNIP terciles per each ASJC field (terciles for all ASJC fields were calculated in Excel using the Percentile function with K imputed as 1/3 and 2/3 and were inclusive of the resulting number). And so, we rounded up the required number calculated in step 2 to the first number divisible by 3 (this made us surpass the calculated required sample sizes).
Example: For *Health Nursing* journals, obtained sample size in step 2 was 5.45 journals, which we rounded to 6 to allow for equal number of journals per tercile (i.e. 2 per tercile).

4) Finally, we obtained random sequences using *Random Integer Set Generator* from random.org [website](http://random.org) based on the number of journals that belonged to each tercile of the ASJC field (calculated in step 3).

Example: The 149 *Nursing* journals were divided into 3 terciles based on their SNIP values, first tercile journals were numbered 1-50, second tercile journals were numbered 51-99, and third tercile journals were numbered 100-149. The *Random Integer Set Generator* produced the 3 following sequences (numbers were ordered lowest to highest): 34 and 42, 74 and 79, 115 and 129, and so journals under those numbers were selected and their ItAs downloaded.

5) Two exceptions to the procedure above were made: 1) for the top level-area *Multidisciplinary sciences*, we included all journals indexed, as there were only 109 of them in total; 2) as we upgraded the ASJC field Arts and Humanities to a top-level area, we did not create “new” ASJC fields for it, but rather sampled only per 3 terciles.

Obtained sample sizes are available in Table 1 (the calculated sample sizes, proportions of journals per ASJC fields, and samples after rounding are available on our project’s data repository). 18
Table 1. Number of journals to be sampled.

<table>
<thead>
<tr>
<th>SNIP tercile</th>
<th>Life Sciences</th>
<th>Social Sciences</th>
<th>Physical Sciences</th>
<th>Health Sciences</th>
<th>Arts and Humanities</th>
<th>Multidisciplinary sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>47</td>
<td>51</td>
<td>54</td>
<td>51</td>
<td>44</td>
<td>37</td>
</tr>
<tr>
<td>T2</td>
<td>47</td>
<td>51</td>
<td>54</td>
<td>51</td>
<td>44</td>
<td>36</td>
</tr>
<tr>
<td>T3</td>
<td>47</td>
<td>51</td>
<td>54</td>
<td>51</td>
<td>44</td>
<td>36</td>
</tr>
<tr>
<td>Total</td>
<td>141</td>
<td>153</td>
<td>162</td>
<td>153</td>
<td>132</td>
<td>109*</td>
</tr>
</tbody>
</table>

*All journals classified as multidisciplinary in Scopus were included, as well as those listed in Science Citation Index Expanded Multidisciplinary Sciences.

Variables

As the journals were sampled from the Scopus Source List, the following variables were available: 1) quantitative variables - journals’ SNIP, SCImago Journal Rank (SJR) and CiteScore values for 2014 to 2016; 2) nominal variables – journal publisher, publishers’ country, journal classification within the top scientific area and ASJC field, indexation in Directory of Open Access Journals (DOAJ) database. Additionally, following consultations with project advisors (mentioned in the acknowledgments) and information obtained in preparation for the systematic review we are conducting as part of our project, it was agreed that we would analyse mentioning of the following topics within the journals ItAs and scope statements, and if and how they are recommended or required of authors:

1. Appeals and corrections
2. Authorship
3. Conflict of interests
4. Data sharing
5. Plagiarism and image manipulation
6. Incentives for reproducibility / replication, and publication of studies with null and negative results
7. Institutional review board / ethics committee approval
8. Use of reporting guidelines and mentioning of the EQUATOR Network
9. Addressing study limitations
10. Type of peer review (open, single, double or triple blind)
11. Pre-registration and protocols
12. Statistics
As we will store all the obtained ItAs, we may explore additional topics or additional predictors of topics covered in ItAs (e.g. journal indexation in DOAJ or publisher’s influence) in other publications or later stages. The above-mentioned topics would be analysed in the following way: 1) sentences contacting the topics would be extracted from the ItAs or scope statements using Perl (Strawberry Perl for Windows edition) regular expression matching (regex), 2) the extracted sentences and matches will then be read and mentioning of a specific topic coded. As the creation and modification of regex is an iterative process that requires reading instructions (especially those that yielded no matches) for possible not-anticipated expressions, the final version of regexes will be deposited on our project’s data repository after the data analysis is complete.\footnote{18}

*Statistical analysis*

Multiple regression will be used to explore the influence of the six top-level scientific areas and SNIP values on the mentioning of topics in ItAs. Absolute numbers, percentages, odds ratios and their corresponding 95% confidence intervals will be reported.

*Update following accessing of the journals’ websites*

To search for each journal website MMal had used Google search engine and entered the name of the journal and its ISSN or E-ISSN (available from the Scopus Source List). The search itself, orientating on the journals websites and the download or copying of journals ItAs and scopes was conducted by MMal between 14 December 2017 and 21 December 2017. During this process, it was discovered that 4 journals had ceased publishing in 2017, one journal had merged with another into a new journal, 44 had ItAs in language other than English, and for 45 journals ItAs were not available nor did the contact persons listed on the website respond to our email inquiries by 22 January 2018 (at least 3 full weeks after sending them emails). We substituted all of the former journals with others from the strata they belonged to, using the
same sampling scheme as described above (exception being the multidisciplinary journals, n=14, for which no substitute was possible as we included all journals belonging to that category). Details of the issues encountered during the download of ItAs are available on our project’s data repository. During the data extraction phase in December, in total 125 journals were contacted, of which 38 replied (30%). The template of the email sent to the journals is also available on the project’s data repository. As it proved to be more difficult than expected to find journals websites, obtain ItAs or journals’ contact information, the following additional variables, in addition to the ones listed above, were added to the study: 1) availability of scope/aim/mission of the journals on the journal website or menu/submenu bar; 2) contact information (contact menu label or contact information for the journal visible on the main web page); 3) name (or name initial) and surname of the editor(s)-in-chief of the journal; 4) contact information for the editor(s)-in-chief. Additionally, web address (i.e. Uniform Resource Locators, URLs) of the journals were stored.

Data Sharing

The full database with selected and replaced journals is available on our project’s data repository.
References: