Assessing the scientific content of predatory journals

Authors: David Moher[1,2,†, ††, *], Larissa Shamseer[1,2,††], Kelly D Cobey[1,2,3,††], Manoj M Lalu[1,2,4,5], James Galipeau[1], Marc T Avey[1,2], The STudy Reporting In Predatory journals (STRIP) Group[‡]

[‡]STRIP Group members: Nadera Ahmadzai[1], Mostafa Alabousi[6], Pauline Barbeau[1], Andrew Beck[1], Raymond Daniel[1], Robert Frank[6], Mona Ghannad[1,7,8], Candyce Hamel[1], Mona Hersi[1], Brian Hutton[1,2], Inga Isupov[9], Trevor A McGrath[6], Matthew DF McInnes[1,2,9], Matthew J Page[10,11], Misty Pratt[1], Kusala Pussegoda[1], Beverley Shea[1,2], Anubhav Srivastava[12], Adrienne Stevens[1,13], Kednapa Thavorn[1,2], Sasha van Katwyk[1], Roxanne Ward[1], Dianna Wolfe[1], Fatemeh Yazdi[1], Ashley M Yu[6], Hedyeh Ziai[6]

Affiliations

[1] Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada

[2] School of Epidemiology, Public Health and Preventive Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada

[3] School of Natural Sciences, Department of Psychology, University of Stirling, Stirling, Scotland

[4] Regenerative Medicine Program, Ottawa Hospital Research Institute, Ottawa, Canada.

[5] The Department of Anesthesiology and Pain Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada

[6] Undergraduate Medical Education, Faculty of Medicine, University of Ottawa, Ottawa, Canada

[7] Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

[8] Centre de Recherche Épidémiologie et Statistique Sorbonne Paris Cité (CRESS-UMR1153) Inserm, Université Paris Descartes, Centre d'épidémiologie Clinique, Paris, France

[9] Department of Radiology, Faculty of Medicine, University of Ottawa, Ottawa, Canada

[10] School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

[11] School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

[12] David R. Cheriton School of Computer Science, Faculty of Mathematics, University of Waterloo, Waterloo, Canada

[13] Translational Research in Biomedicine (TRIBE) Graduate Program, University of Split School of Medicine, Split, Croatia

[†] Visiting Scholar, Meta-Research Innovation Centre at Stanford (METRICS), Stanford University, Stanford, California, United States of America

[††] Joint first authorship

[*] Corresponding author. Email: dmoher@ohri.ca

ABSTRACT

Entities that have become known as 'predatory' journals and publishers are permeating the world of scholarly publishing, yet little is known about the papers they publish. We examined a cross-section of 1907 human and animal biomedical studies, recording their study designs, epidemiological and reporting characteristics. In our sample more than two million humans and over eight thousand animals were included in predatory publications. Only 40% of studies report having ethics approval. Of the 17% of articles reporting their funding source, the US National Institutes of Health was most frequently named. Corresponding authors were most often from India (511/1907, 26.8%) and the US (288/1907, 15.1%). The reporting quality of work reported in our sample was poor and worse than contemporaneous samples from the legitimate literature. Many studies were missing key methodological details and findings. Our results raise important ethical concerns since research in predatory journals is difficult to identify and not indexed in scientifically curated biomedical databases. Funders and academic institutions need to develop explicit policies to drive grantees and prospective authors away from these entities.

BACKGROUND

Entities that have become known as 'predatory' publishers and journals have ensconced themselves as part of the fabric of academic publishing[1]. These entities are becoming increasingly common and are now publishing articles at an alarming rate[2]. They are thought to be exploiting the open-access publishing model and the 'publish or perish' mantra in academia. They promote themselves as 'open-access', providing peer review, and having an impact factor, while boasting article processing charges (APCs) that are a fraction of the cost of legitimate open-access journals (<$150 USD[3]). Their listed editors and staff often do not have a legitimate association with the journal or seem to be contrived[3]. In addition, these outlets fail to use copyright typical of open access or to provide critical information about journals' ethics policies[3].

Defining 'Predatory' Journals

- The term 'predatory' journal was coined and popularized by librarian Jeffrey Beall[1] to describe publishers and journals lacking the ethics, integrity, and operating standards of their recognized, legitimate counterparts
- Until 15 January 2017, Jeffrey Beall maintained the website *Scholarly Open Access*, listing "potential, possible, or probable predatory" publishers and journals here: www.scholarlyoa.com
- These lists collectively became known as "Beall's List"
- The process Beall used to designate journals and publishers as 'predatory' was not evidence based or clear
- Our understanding is that:
  - Predatory journals were identified as Beall encountered them
  - Journals had to meet *some* of the >50 criteria pre-specified by Beall.
  - Judgment as 'predatory' against eligibility criteria was made unilaterally by Beall

While much has been written about the questionable scientific and business practices of predatory journals[4,5], we know little about their content. For example, do they publish clinical research, such as randomized trials and/or preclinical research, such as in vivo studies? For the research that is published what is the impact of apparent non-existent peer review and editorial oversight[6]? These questions and others are important because the content from these entities is permeating the scholarly record. Some publications from these journals, while not typically indexed in scientifically curated bibliographic databases (e.g., Medline), are being picked up by non-discriminating sources such as Google Scholar or PubMed Central [i.e., repository for National Institutes of Health (NIH)-funded research and other publicly accessible full-text research].

The ethical issues of predatory publishing entities are twofold. If research is being published in predatory journals, it is unlikely be effectively scrutinized by peer review[6], and, perhaps more significant, important research may never reach end users including patients, clinicians, or policymakers, and may never be included in future research (e.g. in systematic reviews) because it cannot be easily identified. In this research, we set out to: 1) Characterize what biomedical research is being published in predatory journals, and 2) Determine the completeness of reporting of the most common clinical study designs.

METHODS

We evaluated a cross section of biomedical research published in predatory journals using the following methods.

Research Team

Thirty-two people were involved in the project, led by a core group (DM, LS, KDC, JG, ML, MTA). The team was experienced in extracting information from a broad spectrum of primary clinical and preclinical studies and assessing the quality of such research. Several members of the team also had specific expertise examining predatory journals and publishers. The protocol was deposited in the Open Science Framework (OSF) (https://osf.io/r2gj6/) as well as the University of Ottawa Open Access repository (http://www.ruor.uottawa.ca/handle/10393/34253) prior to data extraction.

Team Training

Prior to conducting the study, all team members involved in the extraction process were required to participate in a minimum two-hour training session, given by three members of the core group (DM, LS, KDC). The training included a brief overview of predatory journals and the project, a detailed discussion and review of the designs of clinical and preclinical studies, and an exercise in categorizing five different types of clinical and preclinical study designs identified by the core team. The training ended with an overview of the project management online software program used to perform the work, DistillerSR (Evidence Partners, Ottawa, Canada). The research team subsequently met for information sessions throughout the duration of the project and the core

group was available for any additional questions, in person or via e-mail. Here we describe the methods used, including some minor amendments from those described in our protocol.

Journal Sampling Strategy

We used the *Scholarly Open Access* website (http://scholarlyoa.com) as the source of journals from which to obtain articles to include in this study. Content of this website was removed on 15 January 2017 (reason unknown)[7], however, archived versions of the website and lists are available through https://archive.org/. Within the website, two lists of 'potential, possible, or probable predatory scholarly open-access' publishers had been maintained: 1) single journal publishers (e.g. 'standalone' journals) and 2) multiple journal publishers (e.g. publishers publishing more than one journal). For the list of single journal publishers, we based our selection of journals on a previous study, with modifications as described below; the process of journal selection is described elsewhere[3]. For the list of multiple journal publishers, the process and methods of journal selection is fully described below.

Single publisher journals

In a previous study[3], of 397 single journal publishers listed on the Scholarly Open Access website in July 2014 (http://web.archive.org/web/20140706180853/http://scholarlyoa.com/individual-journals/), 81 were deemed as having a fully biomedical scope, according to Medline's journal selection criteria https://www.nlm.nih.gov/pubs/factsheets/jsel.html. For the current study, from the 81 journals, we selected only those evaluated as part of the previous study, leaving us with 45 biomedical journals. If, at the time of article retrieval, any of these journals contained no content or listed articles of which none could be retrieved, we replaced them with another journal (from the pool of 36 remaining biomedical journals).

Multiple Journal Publishers

As of 17 December 2015, 923 publishers were listed on the *Scholarly Open Access* website list of multiple journal publishers (http://web.archive.org/web/20141129175402/http://scholarlyoa.com/publishers/?). We recorded this list of publisher names and URLs in Microsoft Excel (version 14.7.1, Microsoft Corporation,

Seattle, Washington, USA) and selected a 20% sample of publishers (n = 185) using a random number generator. One team member (A. Srivastava) visited each publisher's website and obtained the name and URL for all listed journals. Five team members (DM, LS, KDC, JG, MTA) screened the resulting list of journals to determine which were biomedical. To do this, two assessors independently scrutinized each journal's website to find the listed scope of each journal, and applied the Medline journal selection criteria (as above) to determine which were biomedical journals. Assessors noted whether each journal's scope was fully or partially biomedical. Disagreements in screening were resolved by third party arbitration. Of journals identified as fully biomedical, we randomly selected 200 (using a random number generator) for inclusion in our sample. Journals in this sample that did not list any articles, or listed articles of which none could be retrieved, were replaced with another randomly selected fully biomedical journal to maintain the sample of 200 journals.

Identifying Articles for Inclusion

For the 245 journals we identified (see Fig.1A), one of four team members (PB, MG, LS, II) visited each journal's website and sought PDFs of the published articles. Up to 25 of the most recently published articles were downloaded. Based on previous experience examining presumed predatory journals, we anticipated that some of the journals would not list their content chronologically. For such journals, we identified articles using the following approach:

1) Where articles were organized according to study design, they were obtained in the following preferential order: clinical trials, systematic reviews, observational studies, preclinical studies, diagnostic accuracy studies, case reports, non-systematic reviews, and opinions/commentaries/editorials. As such, if there were ≥25 clinical trials, in addition to other study types, we obtained only 25 clinical trials, starting with the most recently published if a chronology was apparent within the study design grouping;
2) If studies were ordered by topic area, we selected up to 25 articles from what we believed to be biomedical topics. For example, if the journal published on biology and pharmacy, and articles were sectioned under these two headings, we first retrieved up to 25 articles from the section containing pharmacy articles (i.e., direct biomedical topic) prior to retrieving any remaining articles from the biology section. If studies were listed by topic area, but multiple biomedical topics were apparent (e.g., psychiatry and neurology), we retrieved articles from

the first relevant category, moving to the next only if the first category contained fewer than 25 articles; and

3) Where articles were not ordered by date of publication, study design, or topic area, we obtained up to 25 articles in the order they appeared on the journal website. When a journal was encountered that published no articles, we excluded the journal from our sample and replaced it with another, as described above.

The process of identifying articles for extraction of characteristics is depicted in Fig.1B. We aimed to obtain a sample of articles describing clinical or preclinical (*in-vivo*) research. Two assessors independently screened each article to identify those describing research in humans or whole animals (i.e., *in-vivo*). Methods research and non-research articles such as protocols, editorials, commentaries, opinion pieces, and letters to the editor were excluded. Non-biomedical research, such as social science research, and research on agriculture, livestock, and wildlife were also excluded.

Extraction of Epidemiological Characteristics

The following information was extracted from each included article, where available: the corresponding author's name (or last author if no corresponding author was indicated), email address, academic affiliation, and affiliated country; journal and publisher location (country); month and year of publication; number of authors; whether ethics approval was obtained; funder name; study design (as stated by authors); sample size or number of included participants. We also made an assessment of the broad ICD-10 classification of disease being reported (http://apps.who.int/classifications/icd10/browse/2016/en), the type of funders (government, industry for profit, non-profit organization, academic), study design of each included article (author-stated design; and reviewer-assessed design determined using a combination of approaches from the following tools: Penelope (http://www.peneloperesearch.com/equatorwizard/) and a study classification algorithm developed by the Agency for Healthcare Research and Quality's Evidence-based Practice Centre program (AHRQ-EPC).

Five members of the core group (LS, KDC, JG, ML, MTA) piloted the data extraction form on multiple articles and refined it until there was agreement on at least 80% of items across assessors. Following this pilot exercise, members of the larger team received specific training on the form, and two team members independently completed data extraction for each article. Disagreements were resolved by third party arbitration.

Completeness of Reporting Assessment

For feasibility, we assessed a purposeful selection of our sample for reporting completeness. For articles reporting multiple preclinical studies, we assessed the first *in-vivo* study reported in the article. In addition, due to their more direct impact on health care, we also assessed the completeness of reporting for the following seven clinical research designs: systematic reviews or meta-analyses of clinical trials; randomized controlled trials; non-randomized clinical trials; case-control studies; observational cohort studies; cross sectional studies; and diagnostic test accuracy studies. For primary clinical research studies, we excluded those reporting on multiple studies (e.g. two different cohort studies) or multiple designs (e.g. case report and cohort study in a single article or nested studies).

When available, the author-stated study design was used as the default for study design categorization. When authors did not state a study design, two independent reviewers assessed the design for categorization. For these studies, prior to evaluating reporting, assessors were asked to confirm whether they agreed with the designated study design. If not, a third verification of study design was undertaken by one of 3 assessors (DM, LS, or KT).

We used modified reporting guideline checklists to guide the assessment of completeness of reporting (S1). The final version of each checklist used in this project is provided in S2 to S9. For each study, checklist items were assessed as 'not reported' (no information was present about the particular item), 'completely reported' (all components of the item were addressed somewhere in the article text), or 'partially reported' (some but not all details of an item was provided somewhere in the text).

Five members of the core team piloted the reporting assessment forms using multiple articles for each study type form and refined the forms until there was agreement on at least 80% of items across assessors. Following this pilot exercise, members of the larger team were given a 1.5-hour training session on how to assess reporting using the forms. Team members were also provided with examples of good reporting taken from the relevant reporting guidelines documents. Two team members independently assessed the completeness of reporting for each study. Disagreements were resolved by third party arbitration.

Data Management

Data were managed using a combination of tools. The list of journals and publishers obtained from the *Scholarly Open Access* website were recorded in Microsoft Excel, which was also used to track and extract journal names and URLs from publishers. The PDFs of journal articles were manually retrieved from journal websites (at the time of writing, we are unaware of any scientifically curated bibliographic database indexing content from these journals) and uploaded to DistillerSR. We used DistillerSR to facilitate screening of biomedical journals and articles for inclusion, for extraction of epidemiological characteristics for each article, and for assessing the completeness of reporting of included studies.

De-duplication

We began this study under the assumption that the publishers and journals (and articles within them) from the two lists we obtained from the *Scholarly Open Access* website were unique. Since we are unaware of any database that indexes these journals and the articles they publish, bibliographic citations for articles were unavailable to us, preventing the use of traditional de-duplication methods (e.g., using reference management software). During data extraction, cleaning, and analysis, we encountered several instances of duplicate publications. As such, we carried out a post-hoc investigation into duplicates at both the journal and article level.

For all 245 included journals, we used Microsoft Excel's duplicate detection function to identify whether any journal names were duplicated. For each duplicated name, we checked whether the URL was also duplicated. If it was, the journal was deemed a duplicate and the latter instance (higher reference number in our sample) was removed. For journals with the same name but

different URLs, we visited each URL and compared the websites. By this process, we found 11 pairs of journals with the same name and identical looking websites but with unique URLs from two different publishers (e.g., Advance Journals and BioMed Research). The latter instance (i.e. higher reference number in our sample) of each duplicate journal was removed. Associated articles from duplicate journals were also identified and removed. At the article level, we used extracted data to facilitate duplicate detection. In Excel, we used the duplicate detection function to identify duplicate corresponding (or last) author names. For each set of duplicates, we assessed whether the extracted email address and the extracted number of authors were the same. If both of these conditions were true, we visited the article PDFs and visually inspected them to determine whether they were reporting on the same study. For each duplicate detected we excluded the instance with the assigned higher reference number in our sample. We identified one journal (Reviews of Progress) listed in both the single and multiple journal publisher lists. Results of these duplicate detection processes are incorporated into Fig.1A and Fig.1B.

Data Analysis

Data were exported from DistillerSR into Microsoft Excel, where they were cleaned. One of five assessors (LS, KDC, JG, ML, MTA) resolved discrepancies in the data entered into open text boxes. Two people (JG and ML) independently categorized data collected on funding sources into five categories (government, academic, industry, unfunded explicitly stated, funding source not stated; discrepancies resolved by consensus) by running a Google search of each named funder and making a judgment about category assignment. The analysis was carried out using Stata/IC (version 13.1, StataCorp, College Station, Texas, USA). Fig. 3 was created using QGIS (version 2.18, QGIS Development Team).

Descriptive summary statistics were used to analyze the epidemiological characteristics. Specifically, we calculated the median and interquartile range for continuous data items and proportions for dichotomous items. Categorical items were reported as counts within each category. Completeness of reporting was summarized as the proportion of articles (within a given study design) reporting within each response category (yes, no, partial) for each checklist item. These findings were depicted via radar diagrams, which present the proportion of articles

completely reporting, and both completely and partially reporting, each checklist item for a given study design.

Differences Between Protocol and Study

The methods above describe the steps that were followed to complete this study. Below is a listing of departures from the planned methods and analyses described in our protocol: https://osf.io/6tr54/

1. We de-duplicated articles manually using the method described above (De-duplication)
2. For feasibility, we included only full-text articles; abstracts and conference proceedings were excluded.
3. Due to the large number of case reports and case series we encountered, we categorized their study design as such (rather than labeling them 'other').
4. We collected the number of included participants for all research studies, rather than just those included in the reporting assessment.
5. We collected the funder name for any affiliated funding including investigator funding, research chairs, etc. (i.e., not just for project-related funding), as it was not always possible to distinguish between types of funding.
6. We (LS, DM, or KT) made a systematic verification/determination of study designs if not reported by study authors and if there were discrepancies between assessors in the categorization of study design; all assessors performed an additional verification of the designated study design as encountered during the reporting assessment.
7. Since the majority of reporting items were judged as "Partially Reported" across study designs, we added a layer representing "Yes" + "Partial" in addition to "Yes" (originally planned) to the radar diagrams depicting reporting assessments.
8. We categorized author location by 2014 World Bank (Organization for Economic Co-operation and Development) income level.
9. We did not explore differences between single-publisher journals and multiple-publisher journals as planned; we did not feel there was a sufficient basis for doing so.
10. We did not examine differences in epidemiological characteristics between level of reported funding (yes/no) and reports of research approved by ethics committees (yes/no)

as planned as there were too few articles reporting this information for comparisons to be meaningful.

11. At the request of the publishing journal (*Nature*) we carried out the following:
    a. We confirmed that studies naming NIH as a funder of the research were likely to have true affiliations with NIH. We did this by searching for the study key words,authors, or grant numbers in the NIH funding database, NIH RePORTER.
    b. We checked whether NIH-funded studies and US authors were concentrated in particular journals. To do this, we performed a cross tabulation of articles with authors from the US (Table S11) or with NIH-funding (Table S12) against the publishing journal.
    c. We counted the number of distinct corresponding authors at D.Y. Patil University, University of Tehran, Univeristy of Texas, and Harvard University. (Table S13)
    d. We summarized the number of articles with the same corresponding (or last) author. (Table S14)

Data Availability

Materials for this project, including the screening and data extraction forms used, and the raw data for this study are available on the OSF (https://osf.io/uydvr/).

RESULTS

Of 397 single-journal predatory publishers and 923 multiple-journal predatory publishers we initially identified, after exclusions of duplicates, we included 236 biomedical journals and 3702 articles (Fig.1A; Fig.1B). Very few journals (59/236; 25.0%) reported their location (Table 1). Of those that did, the primary location was India (19/59, 32.2%), followed by the US (17/59; 28.8%). Nearly 30% of the total journals we sought articles from had no published articles listed (99/343) and were replaced by another journal. Journals contained a median of 19 articles (of a maximum of 25 we sought), we were only able to obtain our pre-set maximum of 25 articles for 42% (100/236) of our sample. We determined that 1907/3702 (51.5%) of the identified articles reported on primary biomedical research, or were systematic reviews of humans or animals. We evaluated the epidemiological characteristics of these articles (Fig.1B). We found that 1499 (1499/1907; 78.6%) reported on a single clinical study using a single design; 201 (201/1907;

10.5%) reported preclinical *in vivo* studies. Articles in our sample were published between 2009 and 2016; the median year of publication was 2015 (Table 2). There was a median of four authors per article.

Publishing in predatory journals is a global phenomenon. Approximately one-sixth of the corresponding authors were from the US (288/1907; 15.1%) and just over a quarter were from India (511/1907, 26.8%) (Fig.2). More than half of the authors were from high or upper middle-income countries (57%). Authors were from a variety of institutions. The most frequently named author-affiliated institutions were from India (3 institutions), USA (2 institutions), Nigeria (1 institution), and Iran (1 institution) (Table 2).

Across biomedical human and animal studies, approximately a third reported obtaining ethics approval (724/1817; 39.8%; Table 3). More than half of these studies did not contain any information about whether they obtained ethics approval (1076/1817; 59.2%), while a few explicitly noted they did not have ethical approval (17/1817; 0.9%). There were 345 different funders reported in 323 publications; 1397 out of the 1907 (73.3%) publications did not report any information about funding, while just 187 out of 1907 (9.8%) noted they were not funded. The US NIH was the most prevalent funder in our sample (41/323; 12.7%). Two of the other largest funders were based in India (University Grants Commission: 15/323, 4.6%; Indian Council of Medical Research: 8/323, 2.5%). Funders were primarily academic (e.g., academic: 124/345, 35.9%) or government agencies (122/345, 35.4%).

1556 (81.6%) of studies we examined reported on clinical research, the remainder were preclinical (Table 4). The majority of primary clinical studies did not include a sample size calculation. Data collected from more than 2 million people were included in our sample of articles (Table 5). There were 283 *in-vivo* experiments reported in 191 publications (Table 6). Overall, 8414 animals were evaluated in these reports.

Assessment of Completeness of Reporting

We evaluated the completeness of reporting of all 201 *in vivo* preclinical studies and 861 clinical reports (Fig.1B; Table 4). The radar diagrams presented in Fig.3 illustrate the assessed

completeness of reporting for eight distinct study designs against pre-specified reporting guidelines and checklist items. For the 94 RCTs examined (Fig.3A; S10.A), most of the items pertaining to the trial design, methods, and results were reported less than 40% of the time. Study objective(s) and the numbers of participants randomized were reported in 98.9% and 80.9% of trials, respectively. In the 44 non-randomized trials (Fig.3B, S10.B), the items on study objective and numbers of participants allocated were also the most frequently reported (88.6% and 86.4%, respectively). Complete reporting of the remaining checklist items ranged from just 2.3–36.4%.

For the 679 observational studies examined (Fig.3C - 3E; S10.C – S10.E), the study objective(s), some results items, and study conclusions were most frequently reported across designs, with the remaining items tending to be incompletely reported across publications. For the 23 diagnostic accuracy studies examined (Fig.3F), the study objective, participant results, interpretations, and implications had the highest frequency of complete reporting (all >78%). The remaining items were reported in less than half of the studies examined (range: 0.0%-47.8%). For the 21 systematic reviews examined (Fig.3G), with the exception of reporting study design in the title or abstract and providing interpretation in the discussion (76.2% and 66.7%, respectively), most of the items were incompletely reported (e.g. ≤33%).

For the 201 preclinical *in-vivo* studies examined (Fig.3H), there was substantial variability in reporting across checklist items. For example, the title/abstract, objective, and control group were reported in ≥92% of articles. In contrast, blinding, eligibility, and adverse events were reported in <6% of articles.

DISCUSSION

Our results reveal some disturbing findings. Seventeen percent of the articles were based on funded research. The clinical research articles in our sample included more than two million participants; 8414 animals were included in the *in-vivo* publications. Extrapolation across all journals presumed to be predatory, and which publish biomedical research, suggests that more than 52 million humans and animals may have been involved in these publications (Calculation S1). That such a substantial amount of data may never be available for use by others poses major ethical dilemmas. For example, for the nearly 8000 participants who participated in the

randomized trials likely did so thinking their participation, while potentially beneficial to them personally, would also contribute to future patients[8]. Given that these papers are scientifically weak and difficult to identify, it is not likely that this knowledge will be useful or available to others[9].

Based on our findings, extrapolating to all biomedical research published in these entities, at least 18,000 funded research studies are possibly tucked away in these non-indexed, scientifically questionable journals (Calculation S2), including research supported by NIH (the top funder of research in our sample). To what extent funders know that the work they are funding, often using taxpayer revenues, is being used towards publications in predatory journals, is unknown. A cursory examination of the websites of the 10 most common funders reported in our study identified just one funder (University Grants Commission, India) that provided guidance regarding journal quality. Researchers have recently summarized the predatory journal situation in India[10]. It appears that few funders have explicit policies against the use of their funds towards publication in predatory journals.

The completeness of reporting of selected study types was poor. Particularly deficient were details about study methods, results, and study registration (where assessed). This is wasteful[11]. While research published in legitimate journals is not particularly well reported either[12–14] the absolute state of reporting in predatory journals is substantively worse based on our findings. For example, trials published in legitimate journals completely report two 'bias protection' characteristics of randomized trials, sequence generation and allocation concealment, 45% and 30% of the time, respectively[15]. In predatory journals, these numbers drop substantially to 23% and 10%, respectively. A recent evaluation of reporting across 300 Medline-indexed systematic reviews found that 70% reported assessing the risk of bias[16], whereas only 9.5% of reviews in this study did so.

Such inadequate reporting in predatory entities is likely the result of several factors, including what appears to be non-existent or very poor peer review and editorial oversight[6]. For example, one of the included randomized trials was titled, "Renaissance of Art of Non Descent Vaginal Hysterectomy" (http://bit.ly/2k3LTsj). For this uninformative title (i.e., lacking description of the

study's population, interventions, outcomes, and design) to pass peer review and editorial approval suggests that these journals lack a basic understanding of clinical research. While there is no strong evidence to support the effectiveness of peer review[17] we identified numerous examples of what we think amounts to a lack of editorial oversight and/or peer review. One such example is our finding that only 38% of articles mentioned they received approval from an ethics committee. It is unclear how any scientific editor, and ultimately an editor-in-chief, would agree to publish research that has not been ethically approved and fully documented in every publication. Such behavior is outside the recommendations of best practice advocated by international editorial groups, such as the Committee on Publication Ethics.

Limitations

As we set out to complete a descriptive study without a formal hypothesis, we did not calculate a formal sample size for this study. Due to fiscal constraints, we decided *a priori* to include a convenience sample of 25 articles from each included journal. We feel our sample of 1907 articles is large enough for the results to be generalized.

We included case reports and case series in the numerator and denominator of studies reporting ethics information. We recognize that case reports may not require ethics approval and may warrant exclusion from this equation. However, since there is no standard or consensus for whether ethics approval (versus consent) is required for case reports, we decided not to omit case reports or case series from this equation.

The tools we used to categorize study designs may be sub-optimal. There is no single validated tool available for this purpose. Even among experts, study design classification is a complex process[18]. For this reason, we carried out a triple-verification process for study design classification. We did not include every type of research design, particularly from preclinical research, in our evaluation of reporting. Our focus was clinical research because it is potentially most relevant to patient care. We also chose to include *in vivo* animal reports because they often have translational appeal for possible first-in-human experimentation.

We used reporting guideline checklists to rate the completeness of reporting of selected research types. These checklists are not specifically validated for use in this capacity; however, their long-standing use for this purpose makes our research comparable to evaluations of reporting in other studies. Because of the subjective nature of judging the completeness of reporting, we recognize that our findings for this aspect of our study may not be straightforward to reproduce. We also acknowledge that our sample of articles may contain undetected duplicates. As stated, the method of de-duplication was limited due to the nature of predatory journals not being indexed in databases (to our knowledge). We have made our training materials and assessment forms available (https://osf.io/y6hw2/) to enable replication of our methods.

Implications

Recent evidence suggests that predatory journals and publishers are gaining momentum[2] and will continue to pose substantial problems to the legitimate scientific record and elsewhere unless action is taken against them. Funders and academic institutions should consider introducing policies to explicitly direct grantees and prospective authors away from these entities while maintaining their open access mandates. Publications from these entities also pose downstream consequences. Faculty appointment committees and promotion and tenure committees may not be well equipped to detect predatory publications included in CVs. A recent study by Bagues *et al*., showed that 5% of Italian researchers had presumed predatory publications on their CV's, with an increasing trajectory[19]. Some researchers are likely deliberately enhancing the size of their CVs with predatory journal publications, while others likely have more pragmatic reasons (e.g. low APCs), while others may be being duped into submitting. Will predatory journal publications inadvertently contribute to the promotion and tenure of faculty? How will these publications impact grant application committees' assessments? At least one presumed predatory publisher already promotes the fact that they have published work funded by the NIH (http://www.jacobspublishers.com/index.php/nih-article-edition). Will the results of research published in these journals be used to help inform best practice, particularly in jurisdictions where the research was conducted? Perhaps universities working collaboratively with funders should consider developing joint policies and guidance about the hazards of predatory journals and publishers and the need to avoid them all at costs. Such information needs to be displayed prominently on the websites of universities, research institutions, funders, and others.

References and Notes:

1. Beall, J. Predatory publishers are corrupting open access. *Nature*. **489,** 179 (2012).
2. Shen, C. & Bjork, B.-C. 'Predatory' open access: a longitudinal study of article volumes and market characteristics. *BMC Medicine* **13,** 230 (2015).
3. Shamseer, L. *et al*. Potential predatory and legitimate biomedical journals: can you tell the difference? A cross-sectional comparison. *BMC Medicine* **15,** 28 (2017).
4. Clark, J. & Smith, R. Firm action needed on predatory journals. *BMJ* **350,** h210 (2015).
5. Moher, D. & Moher, E. Stop predatory publishers Now: Act collaboratively. *Annals of Internal Medicine* **164,** (2016).
6. Bohannon, J. Who's afraid of peer review. *Science*. **342,** 60–65 (2013).
7. Beall, J. What I learned from predatory publishers. *Biochemia Medica* **27,** 273–278 (2017).
8. Groves, T. Enhancing the quality and transparency of health research. *BMJ* **337,** a718 (2008).
9. Chan, A.-W. A. *et al*. Increasing value and reducing waste: addressing inaccessible research. *Lancet* **383,** 257–66 (2014).
10. Seethapathy, G. S., Santhosh Kumar, J. U. & Hareesha, A. S. India's scientific publication in predatory journals: need for regulating quality of Indian science and education. *Current Science* **111,** 1759–64 (2016).
11. Glasziou, P. *et al*. Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet* **383,** 267–276 (2014).
12. Glasziou, P., Meats, E., Heneghan, C. & Shepperd, S. What is missing from descriptions of treatment in trials and reviews? *BMJ* **336,** 1472–74 (2008).
13. Fung, A. E., Palanki, R., Bakri, S. J., Depperschmidt, E. & Gibson, A. Applying the CONSORT and STROBE Statements to Evaluate the Reporting Quality of Neovascular Age-related Macular Degeneration Studies. *Ophthalmology* **116,** 286–296.e4 (2009).
14. Hopewell, S., Dutton, S., Yu, L.-M., Chan, A.-W. & Altman, D. G. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* **340,** c723 (2010).
15. Turner, L., Shamseer, L., Altman, D. G., Schulz, K. F. & Moher, D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled

trials published in medical journals? A Cochrane reviewa. *Systematic Reviews* **1,** 60 (2012).

16. Page, M. J. *et al*. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLOS Med* **13,** e1002028 (2016).
17. Jefferson, T., Rudin, M., Brodney, F. S. & Davidoff, F. Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database Syst Rev* MR000016┬¼ (2007).
18. Hartling, L., Bond, K., Santaguida, P. L., Viswanathan, M. & Dryden, D. M. Testing a tool for the classification of study designs in systematic reviews of interventions and exposures showed moderate reliability and low accuracy. *Journal of clinical epidemiology* **64,** 861–71 (2011).
19. Bagues, M., Sylos-Labini, M. & Zinovyeva, N. *A walk on the wild side: an investigation into the quantity and quality of `predatory' publications in Italian academia. LEM Papers Series* (Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, 2016).
20. Hopewell, S. *et al*. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* **371,** 281–3 (2008).
21. STROBE Group. Draft STROBE checklist for abstracts. (2011). Available at: http://www.strobe-statement.org/fileadmin/Strobe/uploads/checklists/STROBE_checklist_conference_abstract_DRAFT.pdf.
22. Beller, E. M. *et al*. PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts. *PLoS medicine* **10,** e1001419 (2013).
23. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS biology* **8,** e1000412 (2010).
24. NIH. Principles and guidelines for reporting preclinical research. *National Institutes of Health website - About NIH* (2014). Available at: http://www.nih.gov/about/reporting-preclinical-research.htm. (Accessed: 11th February 2017)

Supplementary Materials: Supplementary materials are found at the end of this manuscript.

Acknowledgments: We would like to thank Robert J. Willson for his assistance in preparing Fig.2; and Chantelle Garritty and Felipe Langer for their contributions to the protocol. DM is funded through a University Research Chair, University of Ottawa. MML was supported by a Canadian Heart and Stroke Foundation Fellowship and is currently by receives salary support from the Ottawa Hospital Anesthesiology Associates Alternate Funds Association via the Ottawa Hospital Department of Anesthesiology, Ottawa, Ontario, Canada. BH is funded by a New Investigator Awards from CIHR/DSEN. MJP is supported by an Australian National Health and Medical Research Council Early Career Fellowship (1088535).

Author Contributions: We used ICMJE criteria to determine authorship. DM, LS, and KDC are joint first authors. The study was conceived of by DM, LS, KDC, and JG. DM, LS, KDC, JG, ML, MTA drafted the initial protocol, with input and approval from the vast majority of the STRIP team. LS and KDC managed the research project. All authors carried out the research. LS, KDC, JG, ML, and MTA cleaned the data. LS and KDC analyzed the data, prepared tables and figures, and drafted the methods and results section. DM drafted the initial manuscript, and all authors provided feedback and gave their approval prior to submission. CRediT Taxonomy for authorship can be found in OSF: https://osf.io/t4frg/.

Author Information: The authors declare the following competing interests: BH receives consultancy fees from Cornerstone Research Group for methodologic advice related to the conduct of systematic reviews and meta-analysis. Correspondence should be addressed to **D.M.** (dmoher@ohri.ca).

Table 1. Journal Information

Item	Measure	Result
Primary location of journal	Top 5 countries (n, %)[i]	India (19, 32.2%) USA (17, 28.8%) Canada (4, 6.8%) Iran (3, 5.1%) UK, Nigeria, Bulgaria (2, 3.4% each) Not stated (177, 75.0%)
Number of articles available (up to max of 25)	Median (range)	19 (1–25)

[i] Denominator: n = 59 journals

Table 2. Article Demographics[ii]

Item	Measure	Result
Publication Year[iii]	Median (range)	2015 (2009–2016)
Number of Authors	Median (IQR) (range)	4 (3–6) (1–34)
Country of Corresponding or Last author	Top 5 listed (n, %)	India (511, 26.8%) USA (288, 15.1%) Nigeria (99, 5.2%) Iran (82, 4.3%) Japan (75, 3.9%) Not stated (26, 1.4%)
Income Level of Author Country[iv]	High income Upper Middle Income Lower Middle Income Lower Income	725 (38.5%) 342 (18.2%) 788 (41.9%) 26 (1.4%)
Academic Institutions of Corresponding or Last author	Top 5 listed (n, %)	D. Y. Patil University, India (20, 1.0%) Manipal University, India (15, 0.8%) University of Tehran, Iran (14, 0.7%) University of Texas, USA (11, 0.6%) University of Port Harcourt, Nigeria; Harvard University, USA; Bangalore Medical College and Research Institute, India (9, 0.5%) Not stated (15, 0.8%)

[ii] denominator: n = 1907 (human or animal biomedical research)
[iii] not reported for 7 articles
[iv] denominator: n=1881 (articles with author location reported)

Table 3. Ethics and Funding[v]

Item	Measure	Result
Ethics Approval[vi]	Yes (%) No (%) Not reported (%)	724 (39.8%) 17 (0.9%) 1076 (59.2%)
ICD-10	Top 5 (n, %)	Neoplasms (incl. cancers, carcinomas, tumors) (275, 14.4%) Infections and parasitic diseases (231, 12.1%) Endocrine, nutritional, and metabolic disease (178, 9.3%) Circulatory system (161, 8.4%) Nervous system (138, 7.2%)
Funding	Funded (%) Not Funded (%) Not Reported (%)	323 (16.9%) 187 (9.8%) 1397 (73.3%)
Number of Funders	n	345
Funder Name	Top 5 (n, % of articles)	National Institutes of Health, USA (41, 12.7%) University Grants Commission, India (15, 4.6%) Indian council of Medical Research (8, 2.5%) Ministry of Education, Culture, Sports, Science and Technology, Japan; National Natural Science Foundation, China (7, 2.2%) Tehran University of Medical Sciences, Iran (6, 1.9%)
Funder Type	Academic Government Industry Not-for profit Can't tell[vii]	124 (35.9%) 122 (35.4%) 28 (8.1%) 52 (15.1%) 19 (5.5%)

[v]denominator: n = 1907 (human or animal biomedical research)
[vi] denominator: n = 1817 (ethics assumed not applicable for systematic reviews (n=21) and cell line work (n=69));
[vii] Not enough information provided or we were unable to determine the type of funder based on a Google search of the funder name and country.

Table 4. Article Type

Item		n (%)
Research Type	Clinical[viii] Preclinical (*in vivo*) Cell or tissue work	1556 (81.6%) 201 (10.6%) 150 (7.8%)
Clinical Research Design	Randomized Controlled Trial	94 (6.0%)
	Non-randomized Control Trial	44 (2.8%)
	Cross Sectional	443 (28.5%)
	Observational Cohort	180 (11.6%)
	Case Control	56 (3.6%)
	Diagnostic Accuracy	23 (1.5%)
	Systematic Reviews of Interventions	21 (1.3%)
	Case Report or Series	488 (31.4%)
	Qualitative Research	34 (2.2%)
	Other[ix]	173 (11.1%)

[viii] Primary study or clinical systematic review

[ix] Includes articles reporting more than one study or design (n = 57); and N-of-1 studies, other experimental studies, other observational studies, systematic reviews of observational studies, undetermined designs.

Table 5. Sample Size Calculations and Number of Included Participants

Study Design	Sample size calculation reported	Number of reports	Number of participants
Randomized Controlled Trial[x]	Yes No, but # of participants reported Not reported	26 (27.7%) 68 (72.3%) 0 (0%)	7,780
Non-randomized Controlled Trial	Yes No, but # of participants reported Not reported	7 (15.9%) 36 (81.8%) 1 (2.3%)	5,480
Observational Cohort	Yes No, but # of participants reported Not reported	6 (3.3%) 172 (95.6%) 2 (1.1%)	801,849
Case Control	Yes No, but # of participants reported Not reported	1 (2.0%) 46 (90.2%) 4 (7.8%)	10,483
Cross Sectional	Yes No, but # of participants reported Not reported	40 (9.0%) 386 (87.1%) 17 (3.8%)	1,259,170
Diagnostic Test Accuracy	Yes No, but # of participants reported Not reported	0 23 (100.0%) 0	4048
Systematic Reviews	Number of participants in included studies reported[xi] Not reported, but determined	3 (14.3%) 18 (85.7%)	41,697
Case Report or Series[xii]	(Not collected)		2193
Other[xiii]	(Not collected)		8555
Overall number of participants	n		2,141,255

[x] Proportion of studies reporting sample size calculation
[xi] Sample size calculation not relevant
[xii] based on 487 case reports or series reporting this information
[xiii] based on 155 studies reporting this information

Table 6. Preclincial studies: Number of experiments and animals

Preclinical Studies		
Number of experiments overall	n	283[xiv]
Number of experiments per article	Median (range)	1 (1–18)
Sample size calculation	Yes No, but # of animals reported Not reported	3 159 39
Number of animals	n	8414[xv]

[xiv] Based on 201 studies reporting this information

[xv] Based on 162 studies reporting sample size of number of animals

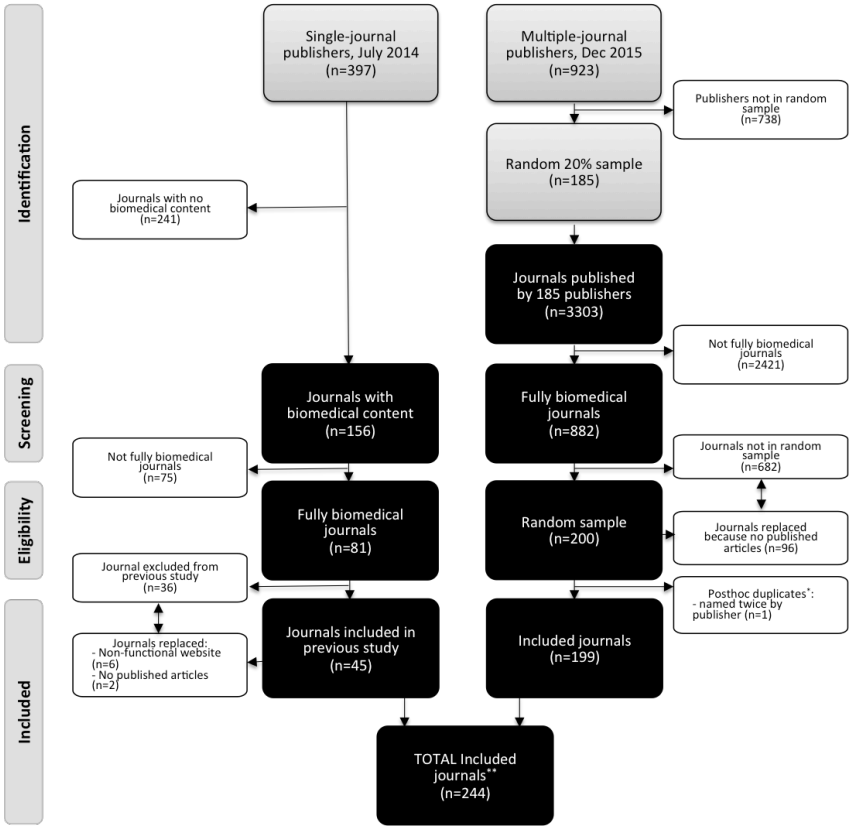


Fig. 1A. Flow of the publisher (grey) and journal (black) identification and selection process. *contains 44 duplicate journals identified post-hoc due to publishers listing same journal twice on their website (n = 33), and one publisher masking as two separate publishers (i.e., different publisher name, but links to same journals) (n = 11).** 8 journals were excluded (post-hoc) due to duplicate articles identified during data extraction

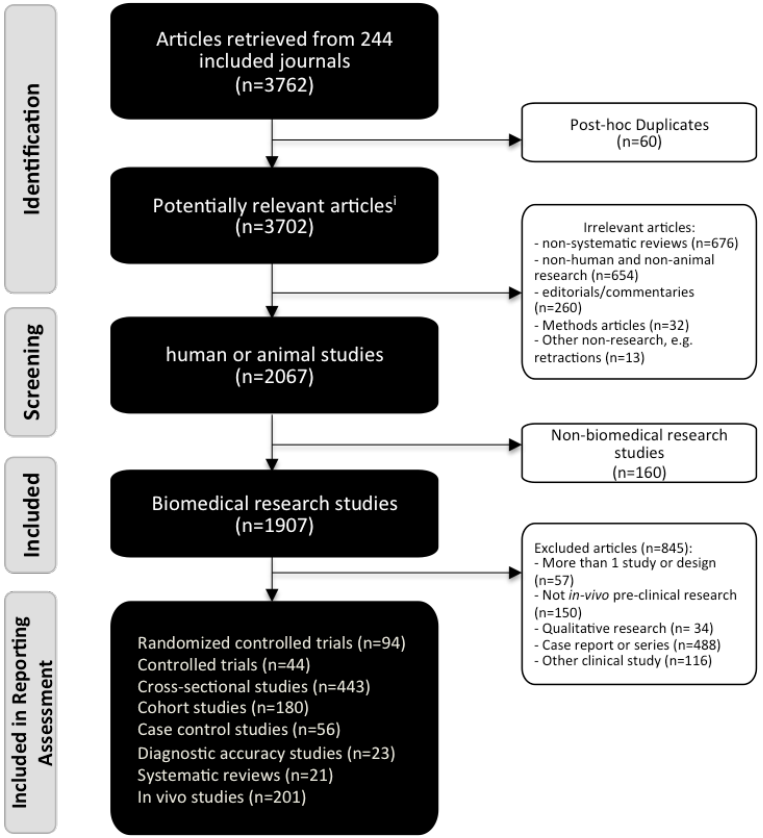


Fig. 1B. Flow of article identification and selection and inclusion in reporting assessment

[i] 1026 articles obtained from single-journal publishers, 2676 articles obtained from multiple-journal publishers

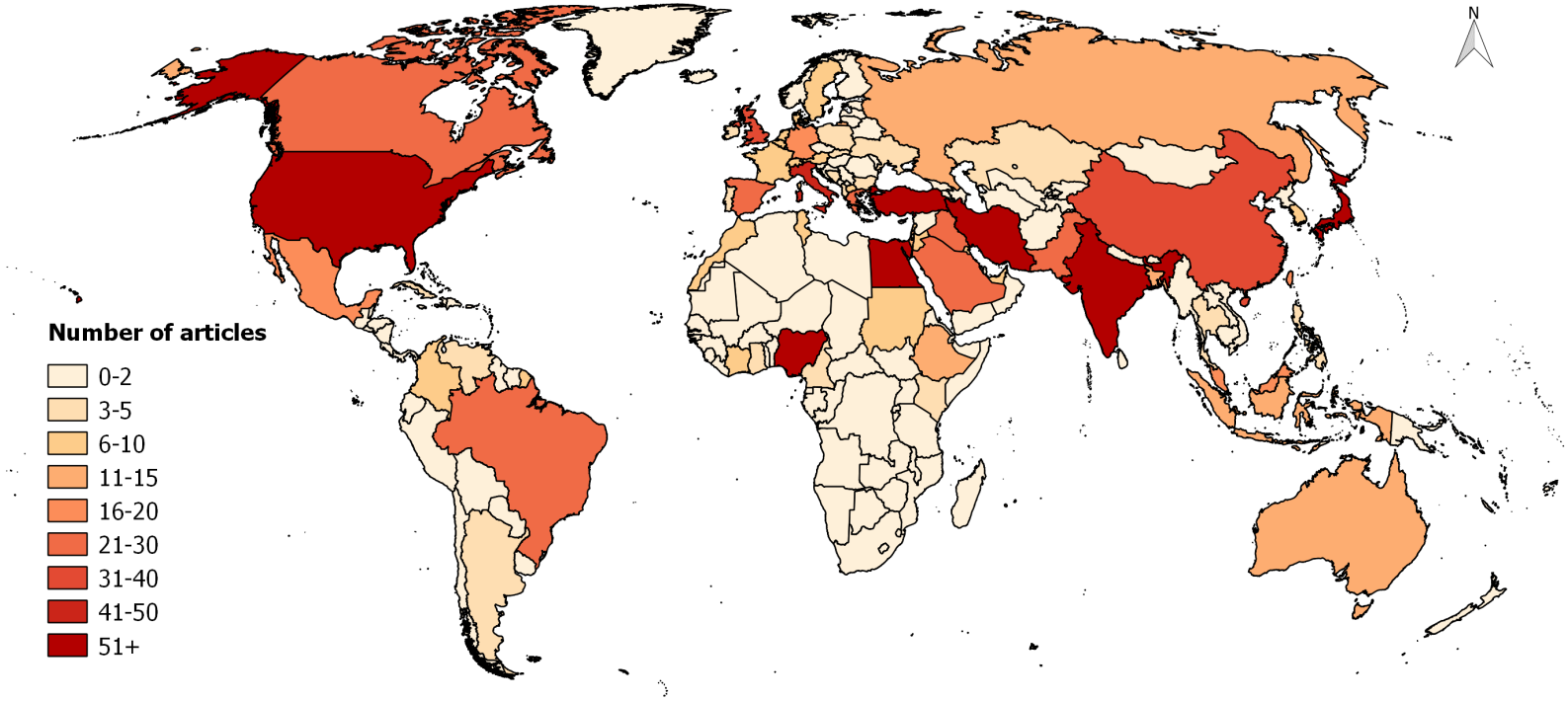
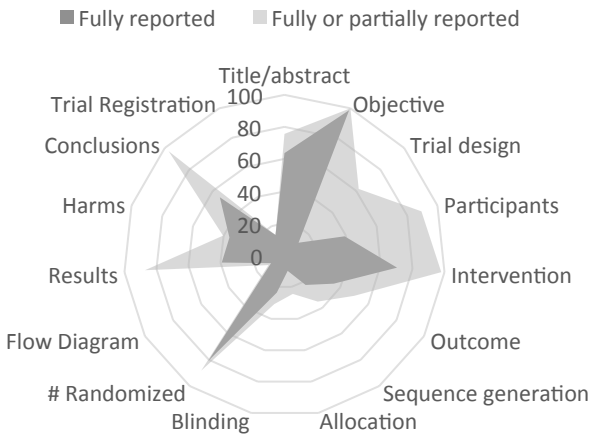
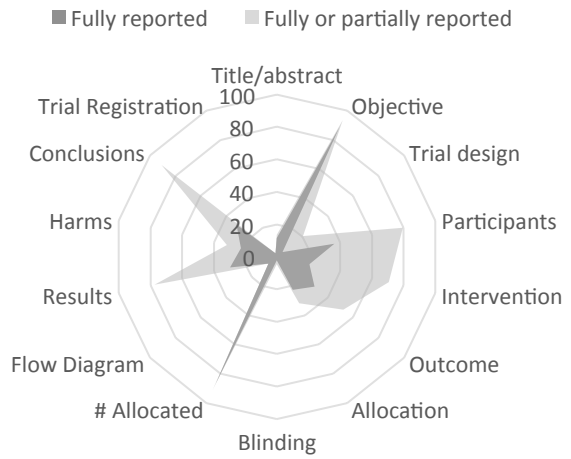


Fig. 2. Country location of corresponding or last author

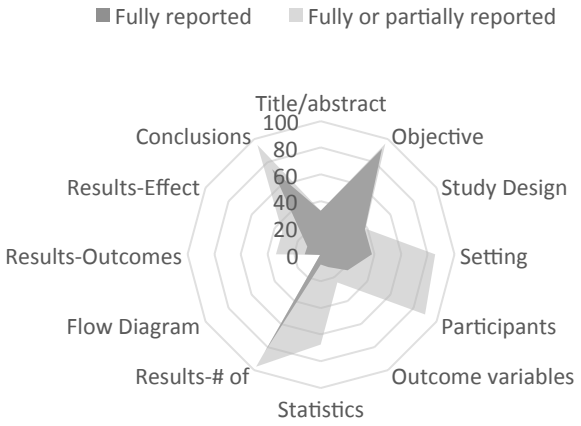
3A. Randomized Controlled Trials (N=94)



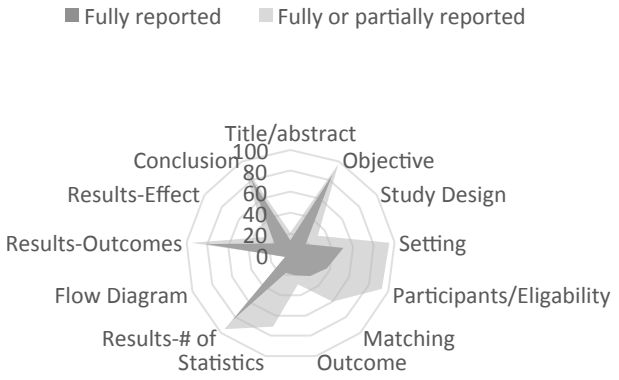
3B. Non-randomized Controlled Trials (N=44)



3C. Cross-Sectional Studies (N=443)



3D. Cohort Studies (N=180)



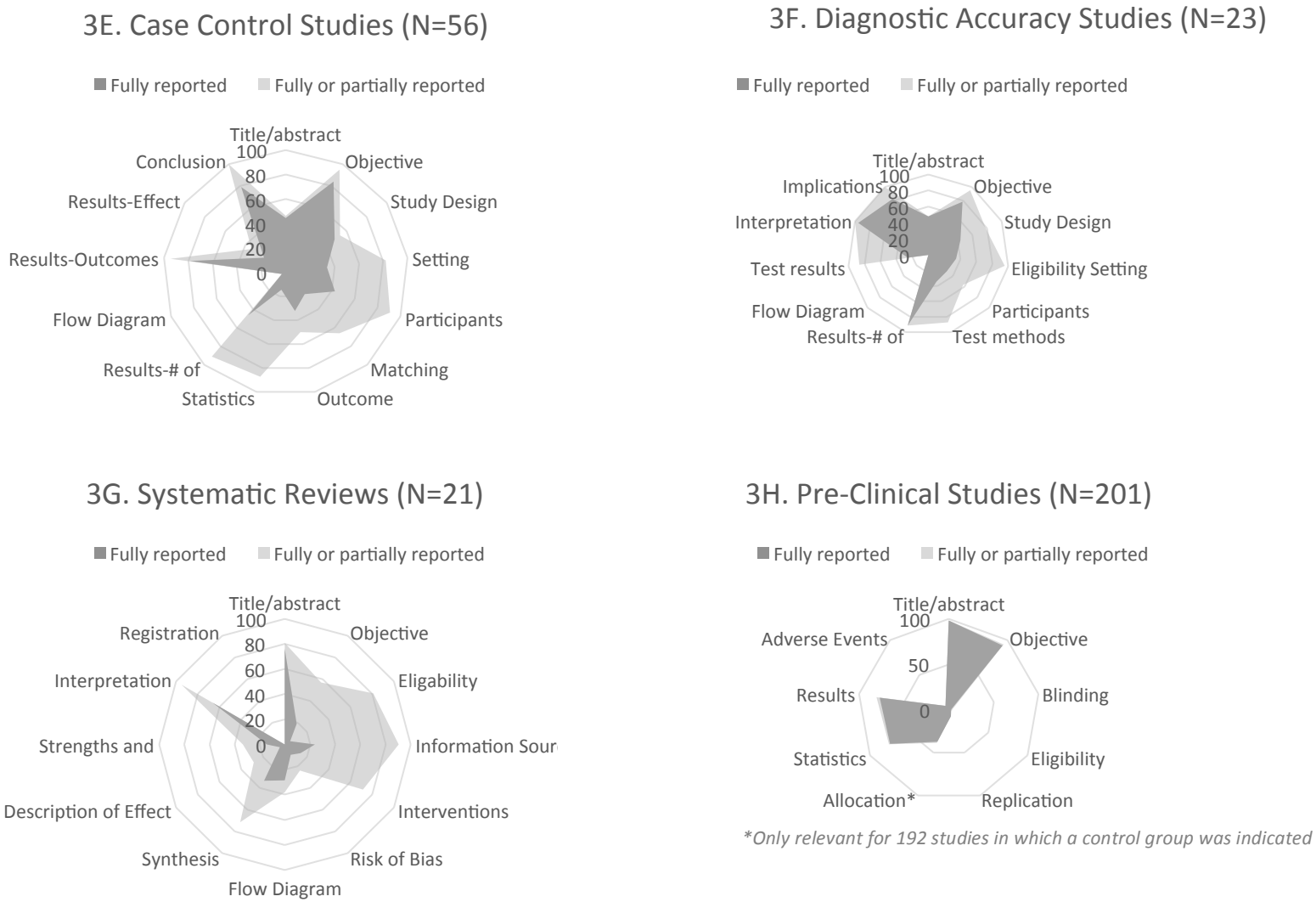


Fig. 3. Frequency of complete reporting for each (modified) reporting guideline item, for each study design group. Both the frequency of 'fully reporting' (dark grey) and 'fully or partially reporting' (light grey) are presented. For details on corresponding reporting items, please see Supplementary Materials.

Supplementary Materials:

Tables S1-S14

Calculations S1 and S2

Table S1. Reporting guideline checklists used for included study designs

Study design	Reporting Checklist Used
Randomized trial	CONSORT for journal and conference abstracts[20]
Non-randomized trial	CONSORT for journal and conference abstracts[20] (modified)
Case Control	STROBE for journal and conference abstracts (draft version)[21]
Observational Cohort	STROBE for journal and conference abstracts (draft version)[21]
Cross Sectional	STROBE for journal and conference abstracts (draft version)[21]
Diagnostic test Accuracy	STARD for journal and conference abstracts (draft) Bossuyt, P. Draft STARD for abstracts (personal communication). 2016.
Systematic Review or Meta-Analysis	PRISMA for journal and conference abstracts[22]
Preclinical *in-vivo*	modified selection of items from: ARRIVE reporting guideline checklist[23] and NIH Reproducibility guidelines[24]

Table S2. Randomized controlled trial reporting assessment items

Checklist Item	Yes	No	Partially Reported
Title/abstract: Is the trial described as 'randomized' in the title or abstract?			
Objective: Is the specific objective or hypothesis of the trial stated?			
Trial Design: Is a description of the trial design (e.g. parallel, cluster, non-inferiority), including allocation ratio provided?			
Participants: Are eligibility criteria for participants and the settings where the data were collected described?			
Interventions: Are the interventions for each group, with sufficient details to allow replication, including how and when they were actually administered stated?			
Outcome: Is there a clearly defined primary outcome, including details of how and when it/they were assessed?			
Sequence generation: Is the method used to generate the random allocation sequence described?			
Allocation concealment: Is the mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned, described?			
Blinding (masking): If blinding was done, is there a description of who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how? If blinding was not done, is a rationale provided?			
Numbers randomized: Is the number of participants randomized to each group reported?			
Flow Diagram: Is there a flow diagram?			
Outcome: For the primary outcome, is a result for each group as well as the estimated effect size (e.g. the size of the difference between groups) and its precision (such as 95% confidence interval) reported?			
Harms: Are important adverse events or side effects reported or if none are reported, is it mentioned that there were none?			
Conclusions: Is there an interpretation of findings that is consistent with results, balancing benefits and			

harms, and considering other relevant evidence?			
Trial Registration: Is a registration number and name of the trial register provided?			

Table S3. Non-randomized controlled trial reporting assessment items

Checklist Item	Yes	No	Partially Reported
Title/abstract: Does the title or abstract describe a specific non-randomized study design? (e.g. non-randomized, quasi-experimental, controlled clinical trial, controlled before-after)			
Objective: Is the specific objective or hypothesis of the trial stated?			
Trial Design: Is a description of the trial design (e.g. parallel, cluster, non-inferiority), including allocation ratio provided?			
Participants: Are eligibility criteria for participants and the settings where the data were collected described?			
Interventions: Are the interventions for each group, with sufficient details to allow replication, including how and when they were actually administered stated?			
Outcome: Is there a clearly defined primary outcome, including details of how and when it/they were assessed?			
Sequence generation: Is the method used to generate the non-random allocation sequence described? (e.g. alternation [alternating between two interventions], rotation [cycling through more than two interventions])			
Blinding (masking): If blinding was done, is there a description of who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how? If blinding was not done, is a rationale provided?			
Numbers allocated: Is the number of participants assigned to each group reported?			
Flow Diagram: Is there a flow diagram?			
Outcome: For the primary outcome, is a result for each group as well as the estimated effect size (e.g. the size of the difference between groups) and its precision (such as 95% confidence interval) reported?			
Harms: Are important adverse events or side effects reported or if none are reported, is it mentioned that there were none?			
Conclusions: Is there an interpretation of findings that is consistent with results, balancing benefits and			

harms, and considering other relevant evidence?			
Trial Registration: Is a registration number and name of the trial register provided?			

Table S4. Cross sectional study reporting assessment items

Checklist Item	Yes	No	Partially Reported
Title/abstract: Is the study's design (e.g. cross-sectional) is indicated in the title or abstract?			
Objectives: Are specific objectives or hypothesis stated?			
Study Design: Is there a statement of the study design (e.g. cross-sectional) in the methods section?			
Setting: Is there a description of the study setting, follow-up dates or dates at which the outcome events occurred/were present?			
Participants: Cross-sectional study: Are the i) eligibility criteria, and ii) major sources and methods of selection of participants described?			
Variables: Is there a clearly defined primary outcome/endpoint/dependent variable?			
Statistical methods: Are statistical methods, including those used to control for confounding, described?			
Results - Participants: Are the number of subjects at the beginning AND end of the study reported?			
Flow Diagram: is a flow diagram reported? (may be in an appendix)			
Main Results: Are the numbers of outcome events or summary measures reported for each study group?			
Main Results: Are unadjusted effect estimates (e.g. relative risk) and, if applicable, confounder-adjusted effect estimates and their precision (e.g., 95% confidence intervals) reported?			
Conclusion: Is there a general interpretation of study results referencing study objectives?			

Table S5. Cohort study reporting assessment items

Checklist Item	Yes	No	Partially Reported
Title/abstract: Is the study's design (e.g. cohort) is indicated in the title or abstract?			
Objectives: Are specific objectives or hypothesis stated?			
Study Design: Is there a statement of the study design (e.g cohort) in the methods section?			
Setting: Is there a description of the study setting, follow-up dates or dates at which the outcome events occurred/were present?			
Participants: Cohort study: Are the i) eligibility criteria, ii) the sources and methods of selection of participants, and ii) methods of follow-up described?			
Participants: Cohort study: For matched studies, are matching criteria and the number of exposed (e.g. smoking) and unexposed (non-smoking) subjects reported?			
Variables: Is there a clearly defined primary outcome/endpoint/dependent variable?			
Statistical methods: Are statistical methods, including those used to control for confounding, described?			
Results - Participants: Are the number of subjects at the beginning AND end of the study reported?			
Flow Diagram: is a flow diagram reported? (may be in an appendix)			
Main Results: Are the numbers of outcome events or summary measures over time reported for each study group?			
Main Results: Are unadjusted effect estimates (e.g. relative risk) and, if applicable, confounder-adjusted effect estimates and their precision (e.g., 95% confidence intervals) reported?			
Conclusion: Is there a general interpretation of study results referencing study objectives?			

Table S6. Case control study reporting assessment items

Checklist Item	Yes	No	Partially Reported
Title/abstract: Is the study's design (e.g. case-control) is indicated in the title or abstract?			
Objectives: Are specific objectives or hypothesis stated?			
Study Design: Is there a statement of the study design (e.g. case control) in the methods section?			
Setting: Is there a description of the study setting, follow-up dates or dates at which the outcome events occurred/were present?			
Participants: Case-control: Are the i) eligibility criteria, and ii) the major sources and methods of case ascertainment and control selection described?			
Participants: Case-control study: For matched studies, are matching criteria and the number of controls per case reported?			
Variables: Is there a clearly defined primary outcome/endpoint/dependent variable?			
Statistical methods: Are statistical methods, including those used to control for confounding, described?			
Results - Participants: Are the number of subjects at the beginning AND end of the study reported?			
Flow Diagram: is a flow diagram reported? (may be in an appendix)			
Main Results: Are the numbers in each exposure category (e.g. cancer patients) or summary measures of exposure reported?			
Main Results: Are unadjusted effect estimates (e.g. relative risk) and, if applicable, confounder-adjusted effect estimates and their precision (e.g., 95% confidence intervals) reported?			
Conclusion: Is there a general interpretation of study results referencing study objectives?			

Table S7. Diagnostic accuracy study reporting assessment items

Checklist Item	Yes	No	Partially Reported
Title/abstract: Does the title or abstract indicate that this is a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or area under the ROC curve)?			
Background and Objectives: Are study objectives or hypotheses stated?			
Study Design: Is it clear whether the study is prospective (i.e. data collection was planned before the index test and reference standard were performed) or retrospective (i.e. data collection was planned after the index test and reference standard were performed)?			
Participants: Are i) eligibility criteria for participants and ii) the settings where the data were collected, described?			
Participants: Is it clear whether participants formed a consecutive, random, or convenience series?			
Test methods: Is a *replicable* description of the index test and reference standard provided? (i.e. enough detail that either can be replicated)			
Results: Are the number of participants with and without the target condition who were included in the analysis, reported?			
Flow diagram: Is there a flow diagram?			
Test Results: Are estimates of diagnostic accuracy (e.g. area under the curve/ROC, likelihood ratios, odds ratios) and their precision (e.g. 95% confidence intervals) reported?			
Discussion: Is there a general interpretation of the results?			
Discussion: Are implications for practice, including the intended use of the index test, described?			

Table S8. Systematic review reporting assessment items

Checklist Item	Yes	No	Partially Reported
Title/abstract: Is this review identified as either: a systematic review, meta-analysis, or both in the title or abstract?			
Objectives: Is there an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)?			
Eligibility criteria: Are eligibility criteria reported, including the following study characteristics: i) PICOS (study design) ii) length of follow-up; and the following report characteristics: i) years considered, ii) language, iii) publication status?			
Information Sources: Are all information sources included in the search (e.g., databases with dates of coverage, contact with study authors to identify additional studies) and the date last searched reported?			
Risk of bias: Are methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information was used in any data synthesis, described?			
Flow diagram: Is there a flow diagram?			
Synthesis of results: For all outcomes considered (benefits or harms), are simple summary data for each intervention group reported for each study?			
Description of the effect: For all outcomes considered (benefits or harms), are effect estimates and confidence intervals, ideally with a forest plot, reported for each study?			
Strengths and Limitations of evidence: Are the limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias) reported? (or stated that there were none)			
Interpretation: Do the authors report a general interpretation of the results in context of other research and important implications for future research?			
Registration: Is a registration number and registry name reported?			

Table S9. Preclinical (*in-vivo*) reporting assessment items

If more than 1 in vivo study is reported, please only assess the first one encountered in the methods section.

Checklist Item	Yes	No	Partially Reported
Title/abstract: Does the title or abstract indicate that this is a non-human animal study?			
Objective: Are the study objectives/hypotheses clearly stated?			
Blinding: Is it clear whether the following groups were blinded: experimenters and/or caregivers, outcome assessors, or those analysing data?			
Eligibility Criteria: Are inclusion/exclusion criteria for data/samples/subjects clearly stated?			
Replication: Is replication, reproduction, or repetition of in vivo experiments (e.g. done multiple times) described?			
Were control and experimental groups used in this experiment?			
If Yes to above: Allocation: was the method of allocation to groups, including randomization, stated?			
Statistical Analyses: Are details of the statistical methods used for each analysis provided?			
First Result: Were summary estimates (e.g measures of central tendency), *for each group, if applicable,* with measures of variance (e.g. error bars defined as SD, SEM, or CI) reported for the first reported *in vivo* result/outcome in the results section?			
Adverse Events: Were any adverse events described (or was it stated that there were none)?			

Table S10.A-H. Reporting assessment summary data for the eight designs assessed.

A. Reporting assessment of randomized controlled trials

Item	Is the item reported?			Total	% yes	% yes and partial
	No	Partial	Yes			
Title/abstract	23	11	60	94	63.8	75.5
Objective	0	1	93	94	98.9	100.0
Trial design	36	47	11	94	11.7	61.7
Participants	10	47	37	94	39.4	89.4
Intervention	2	26	66	94	70.2	97.9
Outcome	47	14	33	94	35.1	50.0
Sequence	61	12	21	94	22.3	35.1
Allocation	71	14	9	94	9.6	24.5
Blinding	65	7	22	94	23.4	30.9
# Randomized	11	7	76	94	80.9	88.3
Flow Diagram	83	2	9	94	9.6	11.7
Results	12	45	37	94	39.4	87.2
Harms	56	4	34	94	36.2	40.4
Conclusions	3	40	51	94	54.3	96.8
Trail Registration	81	0	13	94	13.8	13.8

B. Reporting assessment of non-randomized trials

Item	Is the item reported?			Total	% yes	% yes and partial
	No	Partial	Yes			
Title/abstract	38	1	5	44	11.4	13.6
Objective	3	2	39	44	88.6	93.2
Trial design	35	7	2	44	4.5	20.5
Participants	9	19	16	44	36.4	79.5
Intervention	13	22	9	44	20.5	70.5
Outcome	21	10	13	44	29.5	52.3
Allocation	30	4	10	44	22.7	31.8
Blinding	42	1	1	44	2.3	4.5
# Allocated	4	2	38	44	86.4	90.9
Flow Diagram	41	0	3	44	6.8	6.8
Results	10	21	13	44	29.5	77.3
Harms	30	4	10	44	22.7	31.8
Conclusions	4	26	14	44	31.8	90.9
Trail Registration	43	0	1	44	2.3	2.3

C. Reporting assessment of cross-sectional studies

Item	Is the item reported?			Total	% yes	% yes and partial
	No	Partial	Yes			
Title/abstract	299	0	144	443	32.5	32.5
Objective	17	21	405	443	91.4	96.2
Study Design	272	2	169	443	38.1	38.6
Setting	64	211	168	443	37.9	85.6
Participants	44	296	103	443	23.3	90.1
Outcome variables	336	58	49	443	11.1	24.2
Statistics	144	266	33	443	7.4	67.5
Results-Participants	13	57	373	443	84.2	97.1
Flow Diagram	440	0	3	443	0.7	0.7
Results-Outcomes	294	96	53	443	12.0	33.6
Results-Effect Estimates	294	96	53	443	12.0	33.6
Conclusions	22	95	326	443	73.6	95.0

D. Reporting assessment of cohort studies

Item	Is the item reported?			Total	% yes	%yes and partial
	No	Partial	Yes			
Title/abstract	142	14	24	180	13.3	21.1
Objective	4	20	156	180	86.7	97.8
Study Design	123	26	31	180	17.2	31.7
Setting	9	79	92	180	51.1	95.0
Participants/Eligibility	12	101	67	180	37.2	93.3
Matching	71	59	50	180	27.8	60.6
Outcome	127	16	37	180	20.6	29.4
Statistics	52	95	33	180	18.3	71.1
Results-Participants	8	20	152	180	84.4	95.6
Flow Diagram	170	0	10	180	5.6	5.6
Results Outcomes	9	24	147	180	81.7	95.0
Results-Effect Estimates	127	19	34	180	18.9	29.4
Conclusion	4	28	148	180	82.2	97.8

E. Reporting assessment of case control studies

Item	Is the item reported?			Total	% yes	% yes and partial
	No	Partial	Yes			
Title/abstract	30	1	25	56	44.6	46.4
Objective	3	6	47	56	83.9	94.6
Study Design	26	3	27	56	48.2	53.6
Setting	10	27	19	56	33.9	82.1
Participants	5	27	24	56	42.9	91.1
Matching	19	24	13	56	23.2	66.1
Outcome	28	10	18	56	32.1	50.0
Statistics	7	41	8	56	14.3	87.5
Results-Participants	5	25	26	56	46.4	91.1
Flow Diagram	54	0	2	56	3.6	3.6
Results Exposures	3	9	44	56	78.6	94.6
Results -Effect Estimates	37	7	12	56	21.4	33.9
Conclusion	0	12	44	56	78.6	100.0

F. Reporting assessment of systematic reviews

Item	Is the item reported?			Total	% yes	% yes and partial
	No	Partial	Yes			
Title/abstract	4	1	16	21	76.2	81.0
Objective	9	8	4	21	19.0	57.1
Eligibility	4	16	1	21	4.8	81.0
Information Sources	2	14	5	21	23.8	90.5
Interventions	6	12	3	21	14.3	71.4
Risk of Bias	16	3	2	21	9.5	23.8
Flow Diagram	13	2	6	21	28.6	38.1
Synthesis	6	8	7	21	33.3	71.4
Description of Effect	15	5	1	21	4.8	28.6
Strengths and Limitations of Effect	14	4	3	21	14.3	33.3
Interpretation	1	6	14	21	66.7	95.2
Registration	21	0	0	21	0.0	0.0

G. Reporting assessment of diagnostic accuracy studies

Item	Is the item reported?			Total	% yes	% yes and partial
	No	Partial	Yes			
Title/abstract	12	0	11	23	47.8	47.8
Objective	1	4	18	23	78.3	95.7
Study Design	5	8	10	23	43.5	78.3
Eligibility Setting	1	14	8	23	34.8	95.7
Participants	10	6	7	23	30.4	56.5
Test methods	3	12	8	23	34.8	87.0
Results- Participants	2	0	21	23	91.3	91.3
Flow Diagram	23	0	0	23	0.0	0.0
Test results	3	14	6	23	26.1	87.0
Interpretation	0	1	22	23	95.7	100.0
Implications	0	4	19	23	82.6	100.0

H. Reporting assessment of preclinical studies

Item	Is the item reported?			Total	% yes	% yes and partial
	No	Partial	Yes			
Title/abstract	4	0	197	201	98.0	98.0
Objective	12	4	185	201	92.0	94.0
Blinding	193	2	6	201	3.0	4.0
Eligibility	195	1	5	201	2.5	3.0
Replication	186	0	15	201	7.5	7.5
Allocation[i]	118	2	72	192	37.5	38.5
Statistics	49	2	150	201	74.6	75.6
Results	39	7	155	201	77.1	80.6
Adverse Events	189	0	12	201	6.0	6.0

[i] Only relevant for 192 studies in which a control group was used

Table S11: Journals publishing US-affiliated authors (Nature requested)

Reference Number and Journal Name	Freq.	Percent
185_012_40 Journal of Surgery	13	4.51
2229_153_34 The Open Neuroimaging Journal	11	3.82
181_012_36 Journal of Pediatrics & Child Care	8	2.78
840_050_03 American Journal of Cancer	8	2.78
176_012_31 Journal of Ocular Biology	7	2.43
1963_136_01 Journal of Clinical Medicine Research	7	2.43
2232_153_37 The Open Nutrition Journal	7	2.43
1967_136_05 World Journal of Nephrology and Urology	6	2.08
1970_136_08 Cardiology Research	6	2.08
791_048_18 Jacobs Journal of Gynecology and Obstetrics	6	2.08
147_012_02 International Journal of Otorhinolaryngology	5	1.74
2225_153_30 The Open Medical Informatics Journal	5	1.74
586_021_08 Diabetes Research	5	1.74
596_021_18 Neuro	5	1.74
776_048_03 Jacobs Journal of Allergy and Immunology	5	1.74
790_048_17 Jacobs Journal of Gerontology	5	1.74
803_048_30 Jacobs Journal of Otolaryngology	5	1.74
1243_083_22 Neonatology & Clinical Pediatrics	4	1.39
157_012_12 Journal of Clinical & Medical Case Reports	4	1.39
1639_107_09 Gastroenterology & Hepatology: Open Access	4	1.39
1648_107_18 Journal of Psychology & Clinical Psychiatry	4	1.39
1667_107_37 Hematology & Transfusion International Journal	4	1.39
182_012_37 Journal of Pharmaceutics & Pharmacology	4	1.39
2171_152_13 Heart Health	4	1.39
2175_152_17 International Journal of Cancer Research and Molecular Mechanisms	4	1.39
2227_153_32 Open Medicine Journal	4	1.39
781_048_08 Jacobs Journal of Clinical Case Reports	4	1.39
785_048_12 Jacobs Journal of Emergency Medicine	4	1.39
811_048_38 Jacobs Journal of Sports Medicine	4	1.39
995_069_13 Journal of Cancer Science & Therapy (JCST)	4	1.39
1076_074_11 Obesity & Control Therapies: Open Access	3	1.04
1260_083_39 Obesity & Weight Loss	3	1.04
150_012_05 Journal of Andrology & Gynaecology	3	1.04
1646_107_16 Advances in Obesity, Weight Management & Control	3	1.04
177_012_32 Journal of Oncobiomarkers	3	1.04
179_012_34 Journal of Orthopedics & Rheumatology	3	1.04
2176_152_18 International Journal of Dentistry and Oral Health	3	1.04
2177_152_19 International Journal of Endocrinology and Metabolic Disorders	3	1.04
2193_152_35 Surgery: Open Access	3	1.04

612_021_34 Surgical Research	3	1.04
778_048_05 Jacobs Journal of Anesthesiology and Research	3	1.04
798_048_025 Jacobs Journal of Neurology and Neuroscience	3	1.04
808_048_35 Jacobs Journal of Pulmonology	3	1.04
812_048_39 Jacobs Journal of Surgery	3	1.04
1228_083_07 Anesthesia & Clinical care	2	0.69
1296_088_04 Athens Journal of Health	2	0.69
154_012_09 Journal of Cancer Sciences..	2	0.69
1671_107_41 MOJ Addiction Medicine & Therapy	2	0.69
170_012_25 Journal of Integrative Medicine & Therapy	2	0.69
1846_125_01 Neurological Research and Therapy: Open Access	2	0.69
190_012_45 Journal of Vaccine & Immunotechnology	2	0.69
1921_128_7 Open Journal of Pharmacology	2	0.69
1965_136_03 Journal of Neurology Research	2	0.69
2170_152_12 Gastric Disorders and Therapy	2	0.69
2205_153_10 The Open Cardiovascular Medicine Journal	2	0.69
2458_159_02 Archive of Neuroscience	2	0.69
2720_168_25 International Journal of Cancer Research	2	0.69
4013_213_01 International Journal of Health Research	2	0.69
4031_231_01 Journal of Medical Research and Practice	2	0.69
590_021_12 Gynecology and Obstetrics Research	2	0.69
774_048_01 Jacobs Journal of Addiction and Therapy	2	0.69
792_048_19 Jacobs Journal of Hematology	2	0.69
7_003_4 Advances in Modern Oncology Research	2	0.69
801_048_28 Jacobs Journal of Ophthalmology	2	0.69
802_048_29 Jacobs Journal of Orthopedics and Rheumatology	2	0.69
842_050_05 American Journal of Cancer Biology	2	0.69
917_054_03 Journal Of Advances In Allergy & Immunologic Diseases	2	0.69
988_069_06 Journal of Glycomics & Lipidomics (JGL)	2	0.69
993_069_11 Journal of Bioanalysis & Biomedicine (JBABM)	2	0.69
1061_073_15 International Journal of Advanced Nursing Studies	1	0.35
1099_074_34 SOJ Microbiology & Infectious Diseases	1	0.35
1222_083_01 Infectious & Non Infectious Diseases	1	0.35
1229_083_08 AIDS Clinical Research & STDs	1	0.35
1230_083_09 Cancer Biology and Treatment	1	0.35
1257_083_36 Nephrology & Renal Therapy	1	0.35
1258_083_37 Otolaryngology, Head & Neck Surgery	1	0.35
1644_107_14 MOJ Orthopedics & Rheumatology	1	0.35
1690_107_60 MOJ Women's Health	1	0.35
1755_111_34 International Journal of Clinical Medicine Research	1	0.35
180_012_35 Journal of Parkinson’s disease and Alzheimer's disease	1	0.35
1848_125_03 Heart Health : Open Access	1	0.35
1858_125_13 Clinical Research and Development : Open Access	1	0.35
1877_125_32 Aperito Journal of Oral Health and Dentistry	1	0.35

188_012_43 Journal of Transplantation & Stem Cell Biology	1	0.35
2164_152_06 Clinical Anesthesia and Management	1	0.35
2169_152_11 Epidemiology and Public Health Reviews	1	0.35
2181_152_23 International Journal of Nephrology and Kidney Failure	1	0.35
2194_152_36 Transplantation Research Journal	1	0.35
2457_159_01 Anesthesiology and Pain Medicine	1	0.35
3343_183_06 JOURNAL OF MULTIDISCIPLINARY PATHOLOGY	1	0.35
4003_203_01 Global Journal of Medicine and Public Health	1	0.35
4012_212_01 International Journal of Drug Development and Research	1	0.35
4016_216_01 International Journal of Medical Research & Health Sciences	1	0.35
601_021_23 Otolaryngology - Open Journal (OTLOJ)	1	0.35
608_021_30 Pulmonary Research and Respiratory Medicine	1	0.35
799_048_26 Jacobs Journal of Nursing and Care	1	0.35
813_048_40 Jacobs Journal of Vaccines and Vaccination	1	0.35
986_069_04 Journal of Antivirals & Antiretrovirals (JAA)	1	0.35
987_069_05 Journal of Data Mining in Genomics & Proteomics (JDMGP)	1	0.35
TOTAL	288	

Table S12: Journals publishing NIH–funded studies (Nature requested)

Number of articles	Journal Name
4	International Journal of Cancer Research and Molecular Mechanisms
4	Journal of Cancer Science & Therapy (JCST)
3	The Open Neuroimaging Journal
2	World Journal of Nephrology and Urology
2	Open Medicine Journal
2	Diabetes Research
2	Jacobs Journal of Allergy and Immunology
1	Control Therapies: Open Access
1	SOJ Microbiology & Infectious Diseases
1	International Journal of Otorhinolaryngology
1	Journal of Integrative Medicine & Therapy
1	Journal of Ocular Biology
1	Journal of Orthopedics & Rheumatology
1	Journal of Health, Medicine and Nursing
1	182_012_37 Journal of Pharmaceutics & Pharmacology
1	Journal of Vaccine & Immunotechnology
1	Open Journal of Pharmacology
1	The Open Medical Informatics Journal
1	The Open Nutrition Journal
1	Archive of Neuroscience
1	Neuro
1	Advances in Modern Oncology Research
1	Jacobs Journal of Addiction and Therapy
1	Jacobs Journal of Anesthesiology and Research
1	Jacobs Journal of Ophthalmology
1	American Journal of Cancer Biology
1	Journal of Antivirals & Antiretrovirals (JAA)
1	Journal of Data Mining in Genomics & Proteomics (JDMGP)
1	Journal of Glycomics & Lipidomics (JGL)

Table S13: Number of publications by authors from select top institutions (Nature requested)

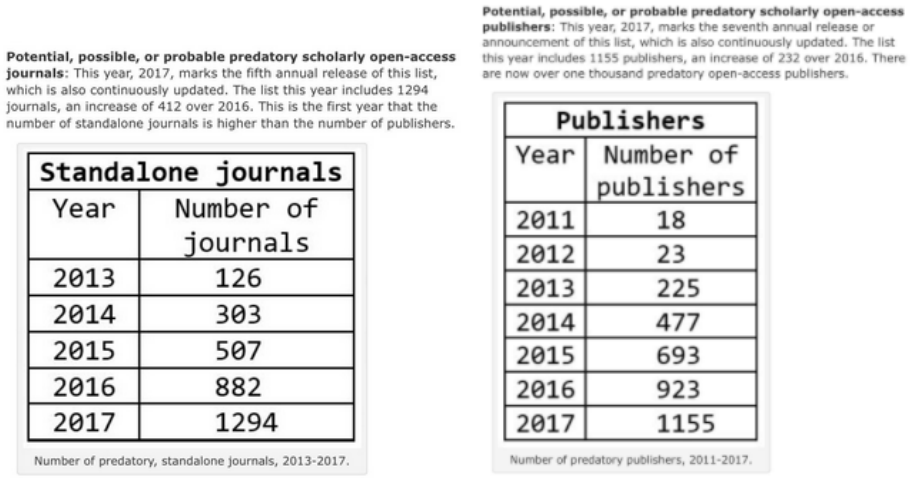
University	Number of authors	Number of publications	Description of occurrences
D. Y. Patil University	14	20	1 author published 6 unique articles (4 in the same journal: International Journal of Medical and Clinical Research); 2 authors published two unique articles each; 10 authors published 1 unique article each
University of Tehran	14	14	14 authors published 1 unique article each
University of Texas	8	11	1 author published 3 unique articles in a single journal: Journal of Pediatrics & Child Care; 7 authors published 1 unique article each
Harvard University	7	9	2 authors published 2 unique articles each in: The Open Neuroimaging Journal and Journal of Surgery; 5 authors published 1 unique article each

Table S14: Number of duplicate authors in sample (Nature requested)

Number of articles with same author name/email address	Number of occurrences
2	85
3	21
4	6
5	2
6	1

Calculation S1. Estimation of the number of human and animal participants reported in biomedical studies published in predatory journals.

The calculations below are based on the 2017 estimations of predatory journals provided on a blog by Jeffrey Beall: https://scholarlyoa.com/2017/01/03/bealls-list-of-predatory-publishers-2017/. We have provided a screenshot of those numbers below:



Potential, possible, or probable predatory scholarly open-access journals: This year, 2017, marks the fifth annual release of this list, which is also continuously updated. The list this year includes 1294 journals, an increase of 412 over 2016. This is the first year that the number of standalone journals is higher than the number of publishers.

Standalone journals	
Year	Number of journals
2013	126
2014	303
2015	507
2016	882
2017	1294

Number of predatory, standalone journals, 2013-2017.

Potential, possible, or probable predatory scholarly open-access publishers: This year, 2017, marks the seventh annual release or announcement of this list, which is also continuously updated. The list this year includes 1155 publishers, an increase of 232 over 2016. There are now over one thousand predatory open-access publishers.

Publishers	
Year	Number of publishers
2011	18
2012	23
2013	225
2014	477
2015	693
2016	923
2017	1155

Number of predatory publishers, 2011-2017.

Estimation of the number of biomedical predatory journals (2017)

	single journal publishers	multiple journal publishers
Number of biomedical journals	81	882 x 5 = 4,410
Number of journals in sample	397	3,303 x 5 = 16,515
% biomed journals	20.40%	26.70%
Average number of journals/publisher	1	16,515 journals /923 publishers = 17.89
Number of publishers (2017 estimate)	1294	1155
Number of journals (2017 estimate)	1,294 publishers x 1 journal = 1,294	1,155 publishers x 17.89 journals = 20663
Number of biomed journals (2017 estimate)	20.40% x 1,294 publishers = **264**	26.70% x 20663 = **5,517**

Estimation of the number of participants included in biomedical research published in biomedical predatory journals.

Number of participants in biomed research/journal (single and multiple combined)	2,141,255 humans 8,414 animals = 2,149,669
Number of journals in our sample (excluding duplicates)	236
% of participants/journal	2,149,669/236 = 9,109
Potential number of participants over 6-year period	9,109 x (264+5,517) **= 52,657,781**

Calculation S2. Estimation of the number of funded studies in biomedical predatory journals

Number of biomed journals (2017 estimate)	264 + 5,517 = 5,781
Average number of articles/journal	19
Number of articles across biomedical predatory journals	19 x 5,781 = 109,839
Proportion of studies receiving funding/journal	17%
Number of funded biomedical studies	109,839 x 0.17 = **18,673**