Proceedings Paper:

Ballou, Nicholas, Warriar, Vivek R and Deterding, Christoph Sebastian orcid.org/0000-0003-0033-2104 (2021) Are You Open?:A Content Analysis of Transparency and Openness Guidelines in HCI Journals. In: CHI Conference on Human Factors in Computing Systems (CHI '21),. ACM

https://doi.org/10.1145/3411764.3445584

Are You Open? A Content Analysis of Transparency and Openness Guidelines in HCI Journals

Nick Ballou
Queen Mary University of London
n.b.ballou@qmul.ac.uk

Vivek R Warriar
Queen Mary University of London
v.r.warriar@qmul.ac.uk

Sebastian Deterding
University of York
sebastian.deterding@york.ac.uk

ABSTRACT

Within the wider open science reform movement, HCI researchers are actively debating how to foster transparency in their own field. Publication venues play a crucial role in instituting open science practices, especially journals, whose procedures arguably lend themselves better to them than conferences. Yet we know little about how much HCI journals presently support open science practices. We identified the 51 most frequently published-in journals by recent CHI first authors and coded them according to the Transparency and Openness Promotion guidelines, a high-profile standard of evaluating editorial practices. Results indicate that journals in our sample currently do not set or specify clear openness and transparency standards. Out of a maximum of 29, the modal score was 0 (mean = 2.5, SD = 3.6, max = 15). We discuss potential reasons, the aptness of natural science-based guidelines for HCI, and next steps for the HCI community in furthering openness and transparency.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; **Empirical studies in HCI**.

KEYWORDS

editorial practice, Transparency and Openness Promotion guidelines, open science

ACM Reference Format:
Nick Ballou, Vivek R Warriar, and Sebastian Deterding. 2021. Are You Open? A Content Analysis of Transparency and Openness Guidelines in HCI Journals. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3411764.3445584

1 INTRODUCTION

Recent years have seen major efforts to evaluate and improve the credibility of published research. Catalyzed by the publication of highly improbable results (e.g., [3]) and meta-scientific research on topics such as questionable research practices [20], misconduct and fraud [47], and publication bias [11], a growing number of researchers have become aware of systems and practices that undermine the trustworthiness of the research literature.

The situation, according to some, is dire. One study demonstrated that as much as half of published research could be false [27]. Another similarly shows that with a combination of common statistical practices like choosing between more than one dependent variable, optionally controlling for covariates, and collecting more data after a non-significant result, false positive rates can rise as high as 61% for a nominal $p < .05$ level of significance [50]. A project to replicate 100 important findings in psychology found that only 36 replication studies yielded significant effects, compared to 97 of the original studies [41]. On average, effect sizes in the replication studies were approximately half of the original studies.

Under the labels "open science" and "science reform", numerous researchers and organizations are seeking to address these issues by introducing new norms, standards, and practices such as increased sharing of data, materials, and code; preregistration to distinguish confirmatory and exploratory analyses; and registered reports to ensure that well-designed experiments producing null results still constitute publishable findings. The science reform movement can be seen as a systematic attempt at promoting openness and transparency across research culture, summarized in Figure 1. Platforms like the Open Science Framework (https://osf.io) and Dataverse (https://dataverse.org) strive to make open research practices like preregistration and data sharing *possible* and *easy*, while collaborative efforts and outreach work like the ReproducibiliTea journal club [42] or CHI special interest group meetings [28] currently work to make them *normative*. We argue that publishers can play a crucial role in *rewarding*—and ultimately *requiring*—that authors adopt open practices. It is this topic that forms the basis of the current paper.

1.1 Open Science and HCI

Recognition of the above problems and calls to action have permeated human-computer interaction (HCI) [8, 12, 55, 58] and connected fields, including computer science [7], health informatics [9], graphics and visualization [4, 30], and computing education [1].

Notable community efforts include the RepliCHI workshop series [58] deliberating the form, fit, and value of replications for HCI research, recently reiterated by Cockburn and colleagues [8], and the CHI Transparent Statistics events [28], which have resulted in recommendations for revising the author submission guidelines of CHI (https://transparentstatistics.org/). A good number of HCI and CSCW researchers have been looking into socio-technical, design, and usability barriers and solutions to the adoption of open science practices and tools [15, 16, 32, 44, 45, 52]. Others have empirically assessed the status of openness and transparency in the

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445584

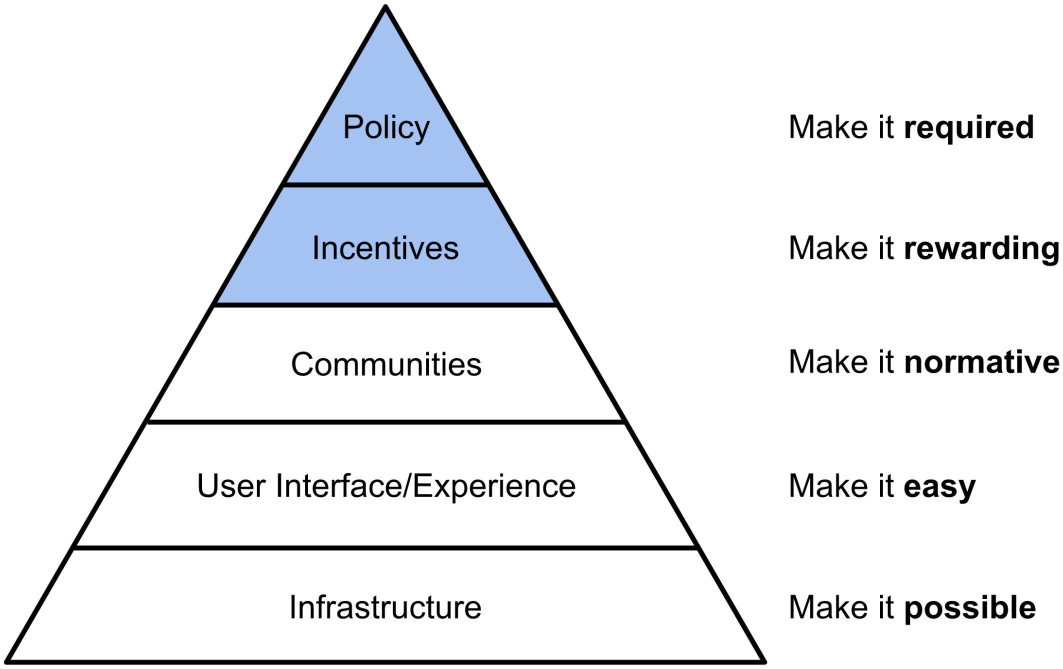

Figure 1: Hierarchy of research culture factors affecting the adoption of openness and transparency standards, with focus of current paper highlighted in blue. Adapted from [39].

HCI community. Thus, Hornbaek et al. [25] found that only 3% of HCI articles across 4 outlets in 2014 were replications, and many of these accidental. Vornhagen and colleagues [53] investigated current practices involving transparency and null hypothesis significance testing in papers at the CHI sister conference CHI PLAY. Only 20% of the studied CHI PLAY papers included detailed materials (questionnaires, software, etc), less than 5% were accompanied by open data, and none of them shared their analysis code. Similarly low (<3%) rates of data sharing were found in another study of CHI 2016 and 2017 papers [12]. A self-report survey study of CHI authors from 2018–19 offers a slightly more promising outlook, with approximately 20% of authors sharing raw data [55]. While the latter study is subject to self-selection bias, it may indicate that rates of sharing have increased somewhat in recent years.

One continuing strand in this work is reflection on the aptness of open science norms, standards, and practices for HCI research. "Open science" is construed in a variety of ways across research communities, variously concerned with e.g. increasing public access to science, redressing democratic inequalities in knowledge access, or making scientific collaboration more efficient [14]. In the context of the replication crisis and responding reform movement, "open science" is chiefly construed as a set of (sometimes quite technical) standards and practices geared towards furthering Mertonian norms of science [35] and preventing so-called questionable research practices that are seen to undermine the reliability, validity, and trustworthiness of academic research [39, 40]. This movement is largely grounded in scientific paradigms that aim to develop general theories through quantitative and hypothetico-deductive research; by extension, it tends to align with a broadly realist/post-positivist philosophy of science. Specifically, the open science reform movement is concerned with previously unacknowledged "researcher degrees of freedom" and incentives that lead researchers to (ab)use hypothetico-deductive experimental and statistical methods in ways that generate spurious positive results and under-report actual negative findings. Open science norms and standards guide researchers to publicly "lock in" their hypotheses and research designs and honestly and fully report them, including all data and analysis procedures, such that their work manifests an actual hypothesis test, and other researchers can scrutinise it in each step.

HCI, in contrast, often engages in describing and understanding novel phenomena, generating theory, exploring new solutions and design spaces, engineering work, or applied, critical, or qualitative research. The HCI community includes many researchers that pursue e.g. designerly or qualitative research paradigms and/or subscribe to pragmatist, constructivist, or other philosophies of science. Many of the underlying assumptions and concerns of realist/post-positivist hypothetico-deductive research do not transfer into these families of research, and vice versa. This diversity notwithstanding, as Cockburn and others have pointed out, good portions of HCI research *are* engaged in quantitative, hypothetico-deductive research [7, 8] and/or would benefit from engaging in large incremental research programs developing general theories [31]. Furthermore, openness and transparency standards hold many benefits for HCI across research paradigms that go beyond increasing the trustworthiness of quantitative, hypothetico-deductive work [58].

Hence, there have been some community efforts within computer science and HCI to institute transparency and openness standards.

The most robust movement in this area has arguably been the ACM Artifact Review and Badging policy (https://www.acm.org/publications/policies/artifact-review-and-badging-current), which is however not mandated and has not been adopted by e.g. CHI. Other efforts so far have mostly fizzled out: the RepliCHI community attempted to establish a dedicated publication stream for replications that became a one-time "RepliCHI award" at CHI 2013. The Transparent Statistics in HCI community has developed a proposal for amending CHI guides for authors and reviewers that have not been adopted (https://transparentstatistics.org). The CHI 2018 SIG on Transparency and Openness Promotion guidelines [6] set out to develop a community recommendation to SIGCHI, which (based on its publicly available information) has not progressed.

One institutional reason for this lack of adoption may be that HCI and computer science more widely strongly rely on conference proceedings [54], which is often justified with their high speed: conferences use comparably short 'one-shot' review processes in which practices like pre-registration would be hard to unfeasible to implement (see e.g. the debate documented in [6]).

That said, the HCI publication ecosystem also features a healthy share of journals, often for more integrative or extended work building on conference publications. Six of the ten most highly cited HCI venues in the most recent Google Scholar ranking are journals [21]. However, comparatively little attention has been paid to the role of HCI journal publishers in open science. As important as conferences and CHI are to HCI as a field, a large portion of authors' research will end up being published in journals. Journals also arguably have fewer constraints on adopting open science standards than conferences with annual schedules. For these reasons, we sought to investigate the extent to which journals preferred by members of the CHI community support open research practices. This focus on journals also offers context to previous work, as the majority of empirical assessments around open science in HCI have focused on individual researchers rather than institutions.

1.2 Transparency and Openness Promotion Guidelines

A good starting point for assessing the transparency and openness of HCI journals are the Transparency and Openness Promotion Guidelines (TOP); established in 2015 by a committee of researchers, editors, and funding agency representatives [40], they are now widely adopted, with over 5,000 signing journals and scholarly organizations. They contain eight modular standards for transparent publishing practices, each with three levels of increasing stringency: the respective standard is explicitly disclosed in journal guidelines (level 1), required for accepted articles (level 2), or adherence to the standard is verified before publication (level 3). The TOP Guidelines explicitly neither mandate nor recommend that all journals implement all standards at the highest level (a common misconception): their modularity and levels are designed to allow different journals to choose a configuration fitting the needs and constraints of the disciplines they serve. Conveniently for the purposes of assessment, the TOP Guidelines go alongside a TOP Factor metric, a summary statistic on how strongly a journal adopts the standards, complete with a rubric for coding. This metric translates the 8 guidelines into 10 standards and gives a journal 0-3 points for each standard, usually aligned with the three levels of stringency, apart from standard 10, Badging, where journals can receive a maximum of 2 points, for a maximum TOP Factor of 29. Certain standards are conceptually similar, and have been grouped for description here.

1 Citation standards. Citation standards refer to the citation of external data sets used by the authors of a given publication, as well as the possibility of storing one's own data such that it can be cited as a separate object. The Joint Declaration of Data Citation Principles argues that "Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications" [22]. This has become increasingly important as science continues to shift toward greater computational complexity, larger data, and more extensive collaborations, making datasets in many cases too large to be published alongside a single paper.

In order to ensure that stored data are FAIR (Findable, Accessible, Interoperable, and Reusable [57]), it is necessary for journals to clearly articulate standards for citing data. Within the TOP standards, journals receive one point for describing how and when authors should cite of data with clear examples, two points for requiring that data cited adheres to these citation standards, and three points for verifying that data citations follow guidelines and lead to persistent and usable data.

2 Data transparency, 3 Analytical method/code transparency, and 4 Research materials transparency. Standards 2–4 describe the sharing of data, analysis materials (statistical code, programs), and research materials like questionnaires, which are all considered vital in order to both reproduce the results in a manuscript as well as replicate (parts of) the work in any future study. At level 1, journals simply require authors to disclose in an availability statement for each of these components (or one statement referencing all of them) whether each is available and under what conditions. At level 2, journals require that these are made publicly available unless the authors provide a compelling ethical or legal conflict that precludes sharing. At level 3, these are required to be publicly available, and will be checked by reviewers or a data manager to confirm or computationally reproduce results before publication.

5 Design and analysis (reporting) transparency. Reporting transparency refers to whether journals have systems detailing minimal levels of description that authors must provide for their studies. A variety of guidelines exist for particular types of studies (e.g., CONSORT for clinical trials [49], PRISMA for systematic reviews and meta-analyses [36, 43]), which seek to ensure that readers are provided with sufficient detail to evaluate the comprehensiveness of the method and results, and that future researchers could replicate the study.

Reporting guidelines vary significantly from field to field, and from method to method within a field, and thus it may not be feasible for journals to exhaustively articulate reporting standards for each type of article they accept. Instead, 1 point is awarded if journals reference particular reporting guidelines that they recommend authors follow, 2 points are awarded if journals require that authors adhere to these reporting standards, and 3 points are awarded if journals review and enforce articles to ensure that reporting standards are met.

6 Study (pre)registration and 7 Analysis plan (pre)registration. Preregistration of studies involves registering the study design, variables, and treatment conditions in advance of conducting the study, in order to make clear what the original goals of the research were and to differentiate confirmatory from exploratory research, thereby reducing HARKing (Hypothesizing After Results are Known). Including an analysis plan involves specifying a sequence of analyses or the statistical model that will be reported. Analysis plan preregistration nearly always supersedes study preregistration, as in order to prospectively register the statistical models one intends to run, authors must describe the data that will be collected and the design of the study that will produce that data.

In the TOP standards, journals receive one point for stating or indicating that the work was preregistered, two points for verifying that the work adheres to its preregistration plan and ensures that exploratory results are clearly differentiated from confirmatory ones, and three points for requiring that confirmatory research is always preregistered.

8 Replication and 9 Publication Bias. Replication and publication bias standards intend to combat the biases raised above against insufficiently novel and null results. One powerful initiative has been the rise of registered reports [5], a publishing format in which study designs are peer-reviewed before being conducted, and if deemed informative, are given in-principle acceptance, meaning that they will be published regardless of the outcomes. The differences are stark: a recent study found that non-registered reports in psychology contained 96% positive results, while only 44% of the results from registered reports were positive [48].

Journals receive one point for explicitly accepting or encouraging the submission of replication studies (standard 8) and stating that novelty and statistical significance are not criteria for publication decisions (standard 9). They receive two points respectively for reviewing replication and novel studies blinded to results, and three points for offering registered reports for each type of study.

10 Badges. The final category is somewhat more peripheral, and concerns whether journals award badges to articles that engage in some of the open practices above. The three badges used by the Center for Open Science are Preregistration, Open Data, and Open Materials. The limited existing evidence suggests that awarding badges is associated with increased rates of data sharing [29]. A journal scores one point for awarding one or two badges, and a maximum of two points for awarding all three badges.

1.3 Research Question and Goals

Our research question was the following: *To what extent do the journals most commonly published in by the CHI community support or require openness through their publishing policies*? As indicated above, we chose to focus in this study on journal articles and exclude conference proceedings, because 1) there is some empirical work on HCI conferences but none on journals to our knowledge, 2) the TOP factor guidelines currently target journals, and 3) because the expanded timeline and opportunities for authors to revise and resubmit manuscripts allows for more straightforward support of certain guidelines like preregistration. This should not be taken to indicate that TOP guidelines have no value for conferences; rather, adapting TOP to conferences such as CHI is a challenging but valuable topic that we will return to in the discussion.

We hope the investigation of this question will achieve three goals:

- Inform authors publishing HCI research about an alternative metric to impact factors when evaluating candidate journals for submissions
- Highlight editorial practices that may be relevant to establishing the trustworthiness of research for readers
- Draw attention toward opportunities for greater openness in published HCI research, and encourage editors and editorial boards to implement more transparent practices at their journals

2 METHOD

The population of interest in our study is journals used by the HCI community. However, journal database classifications—including in HCI—are not necessarily reliable [56]. For instance, on informal inspection, we found that 20 of the 50 most cited HCI-classified journals on Scimago are not HCI-focal on face value. To better reflect HCI publication behavior than current databases, we chose instead to construct our journal sample by analyzing the publication history of recent CHI first authors (see e.g. [46] for a similar approach).

To establish the outlets in which CHI authors publish most frequently, we used R's *RScopus* [38] package to query the Scopus API [13], as previous work found Scopus to be a comprehensive source of publication and citation information in HCI research [34]. We began by collecting a list of all CHI papers from 2016 to present, totalling 4,676 documents. We chose this cutoff date both because the TOP Factor framework was first released in 2015 and to reflect journal preferences of the recent CHI community. From the list of papers, we extracted each unique researcher who had first-authored or co-authored one or more papers in CHI since 2016. This list consisted of 10,213 authors, of whom 3,305 were listed as first author on one or more CHI papers. Of the first authors, 48% report an affiliation in North America, 35% Europe, 13% Asia, 3% Oceania, and 1% Africa.

For each researcher, we then queried Scopus again for their full publication history, using their Scopus-provided Author ID to avoid name ambiguity and limiting results to journal articles only. This resulted in a total of 109,745 articles, of which 83,436 were unique (the difference being the result of journal articles authored by more than one CHI author). Finally, for each article, we extracted the name of the publication and calculated the frequency with which each journal appeared.

2.1 Journal Selection

Results revealed that CHI coauthors publish in a much more diffuse and field-diverse list of journals than CHI first authors. CHI authors published in a total of 9,650 unique journals, while CHI first authors published in 2,665 unique journals. Although 29 of the top 50 journals overlapped across lists, the top 50 journals for CHI first authors account for 39.0% of those authors' total publications, whereas the top 50 journals for all CHI authors account for only 20.0%. In order to have a more focused and discipline-specific list, we chose to evaluate the journals most frequently published in by recent CHI first authors only.

In the original list of 50, 7 entries did not meet our inclusion criteria, namely that journals must be (1) peer-reviewed, (2) accept unsolicited submissions at any time, without requiring attendance or presentation at a conference or event, (3) publish empirical findings, and (4) have author guidelines in English. A further two were duplicates of others already appearing on the list due to name changes at the journal. These nine entries were each replaced by the next-highest journal on the list (see supplementary materials for details). The final two entries on the list were tied, and thus both were included, leading to our final list of 51. We compared our resulting journal sample with Google Scholar's 2020 top HCI venues by h-index, which features 12 conferences and eight journals; all eight journals appeared in our journal sample, indicating convergent validity of our sampling method.

Standard	**1st cycle**		**2nd cycle**	
	κ	**IRA**	κ	**IRA**
Data citation	.68	.73	.72	.76
Data transparency	.64	.87	.49	.85
Analytical methods transparency	.71	.91		
Materials transparency	.96	.98		
Reporting guidelines	.16	.30	.10	.78
Study preregistration	.24	.89	.54	.94
Analysis preregistration	.66	.98		
Replication	-.06	.84	-.05	.87
Publication bias	.80	.98		
Badges	1.0	1.0		
Overall		.85		.90

Table 1: Reliability and agreement scores after each round of coding. κ = interrater reliability using Cohen's quadratic weighted κ, IRA = interrater agreement as raw percentage.

2.2 Coding

Two raters began by coding a test set of five random journals from the list. Coding consisted of visiting the journal's homepage and reading all documentation on the submission process, author guidelines, editorial policies, article types, and any other related pages. Key search strings (e.g., repl*, reg*, data*) were used to double check the presence or absence of policies for particular standards. Raters then met to discuss discrepancies in the initial codes, resulting primarily in conceptual clarification of each standard. Each author then separately coded the remaining 46 journals.

We used the R package *irr* [19] to calculate both interrater agreement (IRA) as raw percentage and interrater reliability using quadratic weighted Cohen's κ. After the first round of coding, overall agreement was 85%, and reliability and agreement were acceptable to excellent for 5 TOP standards: analytical methods transparency, materials transparency, analysis preregistration, publication bias, and badges (Table 1).

For the other 5 TOP standards, metrics were poorer, particularly for reporting standards (both reliability and agreement) and replication (with reliability indicating slight systematic *dis*agreement). Reliability scores should be interpreted with caution, however; due to extremely low variability in scores for some of the standards, a small number of disagreements had a substantial impact on the results.

The two raters thus met to discuss discrepancies. Results-blind discussion about coding strategies revealed one fundamental misinterpretation of the role of reporting guidelines, prompting retraining with the TOP guideline materials for that standard. Discussion also exposed two major ambiguities: first, the degree to which publisher guidelines (intended to provide information about all of the publisher's journals) should only be considered, and second, whether particular guidelines such as pre-registration should be counted if they are described only in the context of clinical randomized controlled trials (RCTs). We clarified our coding scheme to specify 1) that publisher guidelines are only considered if the author guidelines for the particular journal explicitly refer to or link to these pages, and 2) that scores would be awarded based on whether the journal articulated standards for at least one type of study (thus counting RCTs).

Raters then recoded those five standards with the revised coding scheme. After the second cycle, overall agreement improved to 90%. Agreement dramatically improved for reporting guidelines, but this was accompanied by a slight decrease in the already-low reliability score due to an substantial decrease in variation. Reliability and agreement improved moderately for study preregistration and slightly for data citation, but decreased slightly for data transparency. Replication scores were largely unchanged, with high agreement but effectively at chance levels (with reliability near-zero).

The remaining 44 (9.6%) disagreements were resolved through discussion among the two raters. Disagreements tended to be the result of unclear information from journals (sometimes hidden via one or more nested links), remaining ambiguities in the coding scheme, and human error. Multiple disagreements often had the same cause (e.g., a particular publisher for whom all journals had a broken link), further magnifying the effect of relatively minor differences in coding.

Results of the coding process reflect both limitations in the TOP rubric and the convolutions of author guidelines. Although high agreement scores indicate that disagreements will not meaningfully affect overall results, standards with low reliability scores should be interpreted with some caution. However, the nature of the disagreements was insightful, and contributing factors will be addressed in the discussion section below. For each score, we have documented the passage (where these exist) that led to the decision. All of these as well as the codebook specifying which rules were introduced at each stage of the coding process can be found in the supplemental materials (https://osf.io/ck7em/).

3 RESULTS

Figure 2 shows bar charts for each of the TOP standards. The mean TOP score across all top journals was 2.5 (mode = 0, median = 1, max = 15, SD = 3.6). The standard with the highest average score was data citation, where 51% (26/51) of journals achieved a score

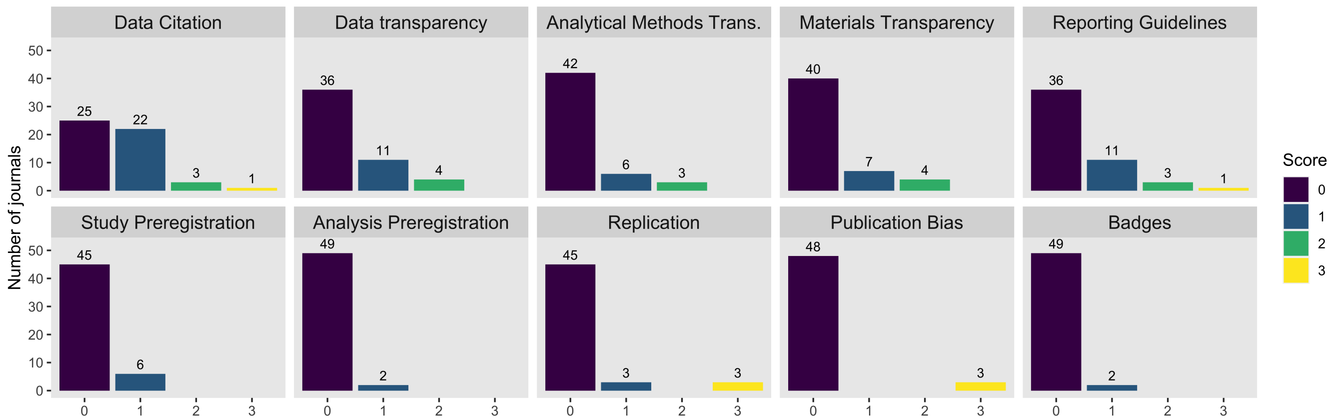

Figure 2: Bar plots showing frequency in raw numbers of each code (0, 1, 2, 3) for each of the 10 TOP standards.

of 1 or above. The standard with the lowest average was analysis preregistration, for which only 4% (2/51) received a non-zero score.

Table 2 shows the results for each journal. Only 6 journals received scores above 4; these were *Frontiers in Psychology* (15), *PLOS One* (14), *Multimodal Technologies and Interaction* (12), *Sensors* (11), *Scientific Reports* (9), and the *Journal of Medical Internet Research* (8). Notable is that 2 of these top 6 scores—those of *Scientific Reports* and *PLOS One*—were obtained by the two domain-general journals in the sample, which publish HCI research alongside research from a variety of other domains. A third high-scoring journal, *Frontiers in Psychology*, also has a notably wide remit.

Large publishers tended to receive the same or similar scores for all their journals (e.g., all IEEE journals scored 0; 4/5 Taylor & Francis journals received a score of 2, with the last one receiving a 3; ACM journals received four 0s and a 1).

4 DISCUSSION

Our results show that journals published in by recent CHI first authors do not currently set or specify encompassing openness and transparency guidelines as articulated in the TOP Guidelines. Data citation policies appeared in about half of journals, followed by reporting guidelines and data/methods/materials transparency, for which about one-third of journals had policies. Policies regarding the remaining TOP standards, including preregistration, replication studies, and publication bias, are highly uncommon.

Perhaps more worrying is that our scores represent a generous interpretation of the underlying rubrics. During coding and when resolving disagreements, we attempted to give the benefit of the doubt to journal policies whenever possible, which included cases of policies behind broken links (Taylor & Francis journals, 4 occurrences); reporting and registration guidelines that were specified only for clinical trials, which do not form a substantial proportion of HCI articles (4 occurrences); references to manuals like the American Psychological Association Publication Manual [2] or the International Committee of Medical Journal Editors guidelines [26] that describe both formatting and reporting without clear indications of which aspects of these were recommended or required (3 occurrences); and internally inconsistent guidelines stating that an availability statement is required in one section, and encouraged in another (1 occurrence).

As noted, reporting standards and transparency standards for data, analytical methods and materials were moderately prevalent, with an average of a third of journals (9–25) featuring some guidelines. This is concerning as nearly all publishers of the journals examined here are members of the Committee on Publication Ethics (COPE). One of COPE's 10 core practices is Data and Reproducibility, stating "Journals should include policies on data availability and encourage the use of reporting guidelines and registration of clinical trials and other study designs according to standard practice in their discipline" [10]. In other words: the studied HCI journals in the majority do not live up to the data and reproducibility standards their own publishers have subscribed to. While other COPE policies regarding misconduct and fraud were found nearly universally in author guidelines, COPE membership is evidently not sufficient to guarantee even a statement of data availability (TOP score of 1), much less a requirement that data be shared when possible (TOP score of 2) in most of the journals we studied.

4.1 Usability and Accessibility: An Opportunity for HCI Researchers

Subjectively, we found during the coding process that guidelines that did exist were difficult to find and follow, often spread across a number of pages, documents, and FAQ sections, and were sometimes incomplete or even self-contradicting. It is likely that many—perhaps most—authors submitting papers to these journals will only scan these documents and may miss crucial information about transparency. This is all the more disconcerting for outlets from a field—HCI—that considers usability and accessibility key concerns.

With this in mind, HCI has the opportunity to play a vital role not just in reforming their own field, but in supporting open science interfaces and tools, and assisting publishers, developers, and other creators with the design and creation of maximally usable and accessible systems (make it *easy*, Figure 1). Members of the CHI community have already begun to do this. Pu et al. [45] report an

Journal Name	Publisher	Citation	Data Trans.	Analytical Method Trans.	Materials Trans.	Reporting Trans.	Study prereg	Analysis prereg	Replication	Publication Bias	Badges	TOP Factor
IEEE Trans Vis Comput Graph	IEEE	0	0	0	0	0	0	0	0	0	0	0
Int J Hum Comput Stud	Elsevier	1	0	0	0	0	0	0	0	0	0	1
ACM Trans Comput Hum Interact	ACM	0	0	0	0	1	0	0	0	0	0	1
Pers Ubiquitous Comput	Springer	0	1	1	1	0	0	0	0	0	0	3
Interact Comput	BCS	3	0	0	0	0	0	0	0	0	0	3
IEEE Pervasive Comput	IEEE	0	0	0	0	0	0	0	0	0	0	0
Comput Hum Behav	Elsevier	1	0	0	0	1	0	0	0	0	0	2
Int J Hum Comput Interact	T&F	1	1	0	0	0	0	0	0	0	0	2
J Med Internet Res	JMIR	1	0	0	1	1	1	0	0	3	1	8
Hum Comput Interact	T&F	1	1	0	0	0	0	0	0	0	0	2
Comput Graph Forum	Wiley	0	0	0	0	0	0	0	0	0	0	0
IEEE Comput Graph Appl	IEEE	0	0	0	0	0	0	0	0	0	0	0
Behav Inform Technol	T&F	1	1	0	0	0	1	0	0	0	0	3
PLOS One	PLOS	1	2	0	2	3	0	0	3	3	0	14
ACM Trans Graph	ACM	0	0	0	0	0	0	0	0	0	0	0
IEEE Software	IEEE	0	0	0	0	0	0	0	0	0	0	0
ACM Trans Access Comput	ACM	0	0	0	0	0	0	0	0	0	0	0
Univers Access Inform Soc	Springer	0	1	1	1	0	0	0	0	0	0	3
First Monday	FMEG	0	0	0	0	0	0	0	0	0	0	0
Comput Supp Coop Work	Springer	0	0	0	0	0	0	0	0	0	0	0
Comput Educ	Elsevier	1	0	0	0	0	0	0	0	0	0	1
Multimed Tools Appl	Springer	0	1	1	1	0	0	0	0	0	0	3
J Assoc Inf Sci Technol	Wiley	2	0	0	0	0	0	0	0	0	0	2
New Media Soc	SAGE	0	0	0	0	1	0	0	0	0	0	1
Sci Rep	Nature	1	1	1	1	2	1	1	1	0	0	9
Front Psychol	Frontiers	1	2	2	2	2	0	0	3	3	0	15
Entertain Comput	Elsevier	1	0	0	0	0	0	0	0	0	0	1
Int J Child Comput Interact	Elsevier	1	0	0	0	0	0	0	0	0	0	1
ACM Trans Interact Intell Syst	ACM	0	0	0	0	0	0	0	0	0	0	0
IEEE Trans Haptics	IEEE	0	0	0	0	0	0	0	0	0	0	0
Int J Mob Hum Comput Interact	IGI Global	1	0	0	0	0	0	0	0	0	0	0
Inf Process Manag	Elsevier	0	0	0	0	1	0	0	0	0	0	1
Comput Graph	Elsevier	1	0	0	0	0	0	0	0	0	1	2
J Vis Lang Comput	Elsevier	1	0	0	0	0	0	0	0	0	0	1
IEEE Access	IEEE	0	0	0	0	0	0	0	0	0	0	0
J Multimodal User Interfaces	Springer	1	1	1	1	0	0	0	0	0	0	4
IEEE Multimedia	IEEE	0	0	0	0	0	0	0	0	0	0	0
Multimodal Technol Interact	MDPI	2	2	2	2	1	1	1	1	0	0	12
Digit Creativ	T&F	1	1	0	0	0	0	0	0	0	0	2
Presence: Virtual Aug Real	MIT Press	0	0	0	0	1	0	0	0	0	0	1
Sensors	MDPI	2	2	2	2	1	1	0	1	0	0	11
User Model User-adapt Interact	Springer	0	1	1	1	0	0	0	0	0	0	3
IEEE Trans Affect Comput	IEEE	0	0	0	0	0	0	0	0	0	0	0
J Am Med Inform Assoc	AMIA/OUP	1	0	0	0	0	0	0	0	0	0	1
Int J Med Inform	Elsevier	1	0	0	0	2	1	0	0	0	0	4
Appl Ergon	Elsevier	1	0	0	0	0	0	0	0	0	0	1
Hum Factors	SAGE	0	0	0	0	1	0	0	3	0	0	4
Inf Soc	T&F	1	1	0	0	0	0	0	0	0	0	2
Inf Vis	SAGE	1	0	0	0	1	0	0	0	0	0	2
ACM Trans Appl Percept	ACM	0	0	0	0	0	0	0	0	0	0	0
Games Cult	SAGE	0	0	0	0	1	0	0	0	0	0	1

Table 2: Scores for each of the 10 standards for all 51 journals. TOP factor scores in the rightmost column are the sum of scores for each module. ISO-4 journal abbreviations are used for space reasons.

interview study addressing users' purposes for using preregistration and whether preregistration templates align with those goals, Fernando & Kuznetsov [17] discuss opportunities and challenges associated with open science/open source hardware, and Feger et al. [15, 16] outline motivation and design considerations for systems supporting reproducibility.

4.2 A Path Forward for TOP Guidelines in HCI

The applicability of the TOP Guidelines for HCI is subject to debate. A document assembled during the CHI 2018 special interest group meeting on TOP standards discusses various pros and cons of CHI adopting each of the guidelines [6]. Participants noted a variety of barriers to the implementation: the increased burden on reviewers, challenges with anonymization, the need to protect participants and commercial confidentiality, and certain types of HCI studies—e.g., engineering/artifact development or qualitative research—for which some or all TOP standards do not neatly apply or make sense.

Similar challenges are present in the adaptation of TOP guidelines to conference proceedings. It is unclear, for example, how a registered report—which may need to be submitted and reviewed months or years before data collection and analysis are concluded—could be possible in a conference with a narrow annual review and publication cycle. Nonetheless, we encourage those in the HCI community who see value in the openness practices articulated in TOP to consider how these adaptations might work, and use TOP as a starting—but not ending—point. Using registered reports as an example, a conference could publish the protocol for a registered report in a separate (e.g., late-breaking work) track one year, and commit to publishing the results in a following year's full-paper proceedings; this aligns with arguments for reducing the size of publishable units of research output [18]. Recent moves of conferences like CSCW or CHI Play to a journal format via PACMHCI [33] suggest that more time-extensive review and revision cycles can work in HCI conferences, which makes it substantially easier to implement certain TOP standards (e.g., giving reviewers sufficient time to reproduce findings with shared data/analytical methods/material; see TOP standards 2–4).

We acknowledge these challenges to the development of openness in HCI and its conference-heavy ecosystem. However, we wish to emphasise again that it is neither necessary or advisable to require all transparency and openness practices promoted in the science reform movement and articulated in the TOP Guidelines in all research: the TOP Guidelines are intentionally designed to be fitted to disciplinary needs. In a sense, they are a forcing function for thinking through which aspects of transparency and openness promotion would apply to one's field, and transparently reporting that one has done so. Standards can be implemented alongside clear delineations of when and how they may not apply. Also, standards are not fixed but in an ongoing development: work is already being done to extend and amend transparency and openness standards to serve the concerns and quality standards of non-experimental and non-hypothetico-deductive work, including qualitative studies [23, 24] and computational research [51]. If HCI researchers find that e.g. certain TOP Guidelines do not readily apply to their kind of research, this invites them to contribute to reflecting and developing standards and good practice of openness and transparency that do benefit their work.

All that said, we believe that for the sizeable group of HCI researchers whose work fits the quantitative, hypothetico-deductive paradigm, our findings point to clear opportunities at each stage of the publishing process to enact change. First, researchers can reward journals that support transparency by submitting their work to these outlets (make it *rewarding*). While impact factors, citation metrics, and disciplinary prestige are likely to continue to be influential for the foreseeable future, authors can augment their decisions with alternative metrics like the TOP factor.

Though it may be a thankless task for the time being, peer reviewers can push for greater transparency, for example asking for data, code, and materials when possible, and attempting to reproduce results [37]. This will contribute to both error detection as well as creating a *normative* expectation from authors, who may later be more inclined to make vetted artifacts public. Both authors and reviewers may, in the absence of clear guidance from journals, use checklists of the sort proposed by Simmons et al. [50] to improve the quality of reporting.

Finally, editors and publishers can institute clearer standards of transparency in their journals, which in many cases would mean to live up to the COPE core practices they already subscribe to. It is clear that to receive 'full points' on the TOP factor would require a significant investment of resources by publishers, editors, and peer reviewers (who may be ultimately responsible for verifying adherence to analysis plans or reproducing results, for example), which likely requires reforming publishing economics as well. We do not expect that these changes are possible overnight. And again, we do not expect that a TOP 'full score' would be apt for all kinds of HCI research: standards could be specified where necessary for the kind of work submitted. Nonetheless, journals can make a positive change immediately simply by adding clear statements about these topics to their author guidelines, which can be adapted from TOP's example wordings for each standard and level https://osf.io/9f6gx/).

4.3 Limitations

Low reliability scores for certain standards (reporting guidelines, replication) are a notable limitation of this study. The overall high agreement scores indicate that discrepancies would have minimal effects on the total distribution of scores; however, one should be cautious when interpreting individual journals' scores for the low-reliability TOP standards. Our work is therefore better understood as a snapshot of the editorial policies of the sample as a whole, rather than an indictment of particular journals or publishers. We recommend authors carefully review journal guidelines in light of their own priorities, and that future research use more coding rounds or additional raters.

While we believe our list of journals is a good representation of the CHI community's publishing outlet preferences, some HCI communities, for example those working in different regions, languages, or combinations of disciplines, may be not well represented by our sample. We also recognise that some journals in our sample (e.g. PLOS One, IEEE Computer Graphics and Applications) arguably are not HCI-focal, limiting the construct validity of our sample. That said, we subsampled just those 10 journals with "Human/Child-Computer", "Computer-Human", "Ergonomics", and "Human Factors" in their title, and the thrust of our findings doesn't change: their mean TOP score is 1.8. Finally, it is likely that journal preferences have changed since our sampling period; HCI researchers may have gradually begun to select journals with more open and transparent practices in recent years, as the issues surrounding research trustworthiness continue to attract attention.

5 CONCLUSION

In this study, we showed that the majority of journals most frequently published in by recent CHI first authors do not currently set clear openness and transparency standards as articulated by the TOP Guidelines. Editorial policies regarding replication, (pre)registration, and publication bias are especially uncommon, appearing in less than 12% of journals. We reflect upon the suitability of current transparency guidelines like TOP for HCI, and argue that while ongoing adjustments are necessary to better suit the field, this can occur in parallel with efforts to increase transparency and openness. We lay out opportunities for authors, reviewers, and editors to enact positive change and ultimately improve the trustworthiness of HCI research. At a minimum, HCI researchers across methodological paradigms and philosophies of science should be able to rely on their peers working in a realist, quantitative, hypothetico-deductive paradigm to live up to their own standards and best practices.

DATA AVAILABILITY

The data, codebook, and analysis scripts used in the paper are available on the Open Science Framework (https://osf.io/ck7em/).

FUNDING

This work was supported by the EPSRC Centre for Doctoral Training in Intelligent Games & Games Intelligence (IGGI) [EP/S022325/1], the EPSRC/AHRC Centre for Doctoral Training in Media and Arts Technology [EP/L01632X/1] and the Digital Creativity Labs, funded by EPSRC/AHRC/Innovate UK [EP/M023265/1].

ACKNOWLEDGMENTS

The authors would like to thank all reviewers for valuable feedback.

REFERENCES

[1] Alireza Ahadi, Arto Hellas, Petri Ihantola, Ari Korhonen, and Andrew Petersen. 2016. Replication in Computing Education Research: Researcher Attitudes and Experiences. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research - Koli Calling '16*. ACM Press, Koli, Finland, 2–11. https://doi.org/10.1145/2999541.2999554

[2] American Psychological Association (Ed.). 2019. *Publication Manual of the American Psychological Association* (seventh edition ed.). American Psychological Association, Washington.
[3] Daryl J. Bem. 2011. Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect. *Journal of Personality and Social Psychology* 100, 3 (2011), 407–425. https://doi.org/10.1037/a0021524
[4] Nicolas Bonneel, David Coeurjolly, Julie Digne, and Nicolas Mellado. 2020. Code Replicability in Computer Graphics. *ACM Transactions on Graphics* 39, 4 (July 2020), 1–8. https://doi.org/10.1145/3386569.3392413
[5] Christopher D. Chambers. 2013. Registered Reports: A New Publishing Initiative at Cortex. *Cortex* 49, 3 (March 2013), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016
[6] Lewis L. Chuang and Ulrike Pfeil. 2018. Transparency and Openness Promotion Guidelines for HCI. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–4. https://doi.org/10.1145/3170427.3185377 Debate document available at https://osf.io/9wrdz/.
[7] Andy Cockburn, Pierre Dragicevic, Lonni Besançon, and Carl Gutwin. 2020. Threats of a Replication Crisis in Empirical Computer Science. *Commun. ACM* 63, 8 (July 2020), 70–79. https://doi.org/10.1145/3360311
[8] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. HARK No More: On the Preregistration of CHI Experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–12. https://doi.org/10.1145/3173574.3173715
[9] Enrico Coiera, Elske Ammenwerth, Andrew Georgiou, and Farah Magrabi. 2018. Does Health Informatics Have a Replication Crisis? *Journal of the American Medical Informatics Association* 25, 8 (Aug. 2018), 963–968. https://doi.org/10.1093/jamia/ocy028
[10] Committee on Publication Ethics. Accessed 16 Sep 2020. Core Practices.
[11] Kerry Dwan, Douglas G. Altman, Juan A. Arnaiz, Jill Bloom, An-Wen Chan, Eugenia Cronin, Evelyne Decullier, Philippa J. Easterbrook, Erik Von Elm, Carrol Gamble, Davina Ghersi, John P. A. Ioannidis, John Simes, and Paula R. Williamson. 2008. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. *PLoS ONE* 3, 8 (Aug. 2008), e3081. https://doi.org/10.1371/journal.pone.0003081
[12] Florian Echtler and Maximilian Häußler. 2018. Open Source, Open Science, and the Replication Crisis in HCI. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–8. https://doi.org/10.1145/3170427.3188395
[13] Elsevier. Searches conducted Aug 27 2020. Scopus. Elsevier. http://api.elsevier.com and http://www.scopus.com. Accessed 2020-08-27.
[14] Benedikt Fecher and Sascha Friesike. 2014. Open Science: One Term, Five Schools of Thought. In *Opening Science*, Sönke Bartling and Sascha Friesike (Eds.). Springer International Publishing, Cham, 17–47. https://doi.org/10.1007/978-3-319-00026-8_2
[15] Sebastian S. Feger, Sünje Dallmeier-Tiessen, Albrecht Schmidt, and Pałel W. Woźniak. 2019. Designing for Reproducibility: A Qualitative Study of Challenges and Opportunities in High Energy Physics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–14. https://doi.org/10.1145/3290605.3300685
[16] Sebastian S. Feger, Sünje Dallmeier-Tiessen, Paweł W. Woźniak, and Albrecht Schmidt. 2019. The Role of HCI in Reproducible Science: Understanding, Supporting and Motivating Core Practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–6. https://doi.org/10.1145/3290607.3312905
[17] Piyum Fernando and Stacey Kuznetsov. 2020. OScH in the Wild: Dissemination of Open Science Hardware and Implications for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376659
[18] Alexandra Freeman. 2019. Octopus: A Revolution in Scientific Publishing. In *Oral Presentations*. BMJ Publishing Group Ltd, Oxford, A29.2–A30. https://doi.org/10.1136/bmjebm-2019-EBMLive.56
[19] Matthias Gamer, Jim Lemon, Ian Fellows, and Puspendra Singh. 2019. Irr: Various Coefficients of Interrater Reliability and Agreement.
[20] Stephen L. George. 2016. Research Misconduct and Data Fraud in Clinical Trials: Prevalence and Causal Factors. *International Journal of Clinical Oncology* 21, 1 (Feb. 2016), 15–21. https://doi.org/10.1007/s10147-015-0887-3
[21] Google. Accessed 15 Sep 2020. Google Scholar Top Publications: Human-Computer Interaction.
[22] Data Citation Synthesis Group. 2014. *Joint Declaration of Data Citation Principles*. Technical Report. Force11. https://doi.org/10.25490/A97F-EGYK
[23] Tamarinde Laura Haven, Timothy M. Errington, Kristian Gleditsch, Leonie van Grootel, Alan M. Jacobs, Florian Kern, Rafael Piñeiro, Fernando Rosenblatt, and Lidwine Mokkink. 2020. *Preregistering Qualitative Research: A Delphi Study*. Preprint. SocArXiv. https://doi.org/10.31235/osf.io/pz9jr
[24] Tamarinde L Haven and Dr. Leonie Van Grootel. 2019. Preregistering Qualitative Research. *Accountability in Research* 26, 3 (April 2019), 229–244. https://doi.org/10.1080/08989621.2019.1580147
[25] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is Once Enough?: On the Extent and Content of Replications in Human-Computer Interaction. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*. ACM Press, Toronto, Ontario, Canada, 3523–3532. https://doi.org/10.1145/2556288.2557004
[26] International Committee of Medical Journal Editors. 2020. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. http://www.icmje.org/icmje-recommendations.pdf.
[27] John P. A. Ioannidis. 2005. Why Most Published Research Findings Are False. *PLoS Medicine* 2, 8 (Aug. 2005), e124. https://doi.org/10.1371/journal.pmed.0020124
[28] Matthew Kay, Steve Haroz, Shion Guha, and Pierre Dragicevic. 2016. Special Interest Group on Transparent Statistics in HCI. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*. ACM Press, San Jose, California, USA, 1081–1084. https://doi.org/10.1145/2851581.2886442
[29] Mallory C. Kidwell, Ljiljana B. Lazarević, Erica Baranski, Tom E. Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, Curtis Kennett, Agnieszka Slowik, Carina Sonnleitner, Chelsey Hess-Holden, Timothy M. Errington, Susann Fiedler, and Brian A. Nosek. 2016. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology* 14, 5 (May 2016), e1002456. https://doi.org/10.1371/journal.pbio.1002456
[30] Robert Kosara and Steve Haroz. 2018. Skipping the Replication Crisis in Visualization: Threats to Study Validity and How to Address Them : Position Paper. In *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*. IEEE, Berlin, Germany, 102–107. https://doi.org/10.1109/BELIV.2018.8634392
[31] Vassilis Kostakos. 2015. The Big Hole in HCI Research. *Interactions* 22, 2 (Feb. 2015), 48–51. https://doi.org/10.1145/2729103
[32] Yang Liu, Tim Althoff, and Jeffrey Heer. 2020. Paths Explored, Paths Omitted, Paths Obscured: Decision Points & Selective Reporting in End-to-End Data Analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376533
[33] Regan Mandryk. 2020. PACM Statement.
[34] Lokman I. Meho and Yvonne Rogers. 2008. Citation Counting, Citation Ranking, and h-Index of Human-Computer Interaction Researchers: A Comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology* 59, 11 (2008), 1711–1726. https://doi.org/10.1002/asi.20874
[35] Robert K. Merton. 1973. *The Sociology of Science. Theoretical and Empirical Investigations*. University of Chicago Press, Chicago, London.
[36] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and The PRISMA Group. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine* 6, 7 (July 2009), e1000097. https://doi.org/10.1371/journal.pmed.1000097
[37] Richard D. Morey, Christopher D. Chambers, Peter J. Etchells, Christine R. Harris, Rink Hoekstra, Daniël Lakens, Stephan Lewandowsky, Candice Coker Morey, Daniel P. Newman, Felix D. Schönbrodt, Wolf Vanpaemel, Eric-Jan Wagenmakers, and Rolf A. Zwaan. 2016. The Peer Reviewers' Openness Initiative: Incentivizing Open Research Practices through Peer Review. *Royal Society Open Science* 3, 1 (Jan. 2016), 150547. https://doi.org/10.1098/rsos.150547
[38] John Muschelli. 2019. Rscopus: Scopus Database 'API' Interface.
[39] Brian A. Nosek. 2019. Improving Transparency and Reproducibility of Research.
[40] B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. L. Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, and T. Yarkoni. 2015. Promoting an Open Research Culture. *Science* 348, 6242 (June 2015), 1422–1425. https://doi.org/10.1126/science.aab2374
[41] Open Science Collaboration. 2015. Estimating the Reproducibility of Psychological Science. *Science* 349, 6251 (Aug. 2015), aac4716–aac4716. https://doi.org/10.1126/science.aac4716
[42] Amy Orben. 2019. A Journal Club to Fix Science. *Nature* 573, 7775 (Sept. 2019), 465–465. https://doi.org/10.1038/d41586-019-02842-8
[43] Matthew James Page, Joanne McKenzie, Patrick Bossuyt, Isabelle Boutron, Tammy Hoffmann, cindy d mulrow, Larissa Shamseer, Jennifer Tetzlaff, Elie Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy Grimshaw, Asbjørn Hróbjartsson, Manoj Mathew Lalu, Tianjing Li, Elizabeth Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley Stewart, James Thomas, Andrea Tricco, Vivian Andrea Welch, Penny Whiting, and David Moher. 2020. *The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews*. Preprint. MetaArXiv. https://doi.org/10.31222/osf.io/v7gm2
[44] Irene V. Pasquetto, Ashley E. Sands, Peter T. Darch, and Christine L. Borgman. 2016. Open Data in Scientific Settings: From Policy to Practice. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 1585–1596. https://doi.org/10.1145/2858036.2858543
[45] Xiaoying Pu, Licheng Zhu, Matthew Kay, and Frederick Conrad. 2019. Designing for Preregistration: A User-Centered Perspective. In *Extended Abstracts of the*

2019 CHI Conference on Human Factors in Computing Systems. ACM, Glasgow Scotland Uk, 1–6. https://doi.org/10.1145/3290607.3312862

[46] Steven M. Quiring. 2007. Trends in Publication Outlets of Geographer-Climatologists. *The Professional Geographer* 59, 3 (Aug. 2007), 357–364. https://doi.org/10.1111/j.1467-9272.2007.00618.x

[47] Stuart Ritchie. 2020. *Science Fictions: Exposing Fraud, Bias, Negligence and Hype in Science*. Bodley Head, London.

[48] Anne M. Scheel, Mitchell Schijen, and Daniel Lakens. 2020. *An Excess of Positive Results: Comparing the Standard Psychology Literature with Registered Reports*. Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/p6e9c

[49] Kenneth F Schulz, Douglas G Altman, David Moher, and CONSORT Group. 2010. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. *BMJ* 340, mar23 1 (March 2010), c332–c332. https://doi.org/10.1136/bmj.c332

[50] Joseph P Simmons, Leif D Nelson, and Uri Simonsohn. 2011. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science* 22, 11 (Nov. 2011), 1359–1366. https://doi.org/10.1177/0956797611417632

[51] Victoria Stodden and Sheila Miguez. 2014. Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. *Journal of Open Research Software* 2, 1 (July 2014), e21. https://doi.org/10.5334/jors.ay

[52] Theresa Velden. 2013. Explaining Field Differences in Openness and Sharing in Scientific Communities. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13*. ACM Press, San Antonio, Texas, USA, 445. https://doi.org/10.1145/2441776.2441827

[53] Jan Benjamin Vornhagen, April Tyack, and Elisa D Mekler. 2020. *Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency*. Preprint. Open Science Framework. https://doi.org/10.31219/osf.io/58wzp

[54] George Vrettas and Mark Sanderson. 2015. Conferences versus journals in computer science. *Journal of the Association for Information Science and Technology* 66, 12 (2015), 2674–2684.

[55] Chat Wacharamanotham, Lukas Eisenring, Steve Haroz, and Florian Echtler. 2020. Transparency of CHI Research Artifacts: Results of a Self-Reported Survey. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376448

[56] Qi Wang and Ludo Waltman. 2016. Large-Scale Analysis of the Accuracy of the Journal Classification Systems of Web of Science and Scopus. *Journal of Informetrics* 10 (2016), 347–364. https://doi.org/10.1016/j.joi.2016.02.003

[57] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3, 1 (Dec. 2016), 160018. https://doi.org/10.1038/sdata.2016.18

[58] Max L. L. Wilson, Paul Resnick, David Coyle, and Ed H. Chi. 2013. RepliCHI: The Workshop. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*. ACM Press, Paris, France, 3159. https://doi.org/10.1145/2468356.2479636