Criteria Definition for Research Ethics, Openness, and Transparency

KAVOUS SALEHZADEH NIKSIRAT, University of Lausanne, Switzerland
LAHARI GOSWAMI, University of Lausanne, Switzerland
POOJA S. B. RAO, University of Lausanne, Switzerland
JAMES TYLER, University of Lausanne, Switzerland
ALESSANDRO SILACCI, University of Lausanne, Switzerland and School of Management of Fribourg, HES-SO University of Applied Sciences and Arts Western Switzerland, Switzerland
SADIQ ALIYU, University of Lausanne, Switzerland
ANNIKA AEBLI, University of Lausanne, Switzerland
CHAT WACHARAMANOTHAM, Swansea University, United Kingdom
MAURO CHERUBINI, University of Lausanne, Switzerland

This document is supplementary material for the paper ***"Changes in Research Ethics, Openness, and Transparency in Empirical Studies between CHI 2017 and CHI 2022"*** published at ACM CHI 2023. The document provides the rationale behind each criterion in detail with additional citations. We hope that knowing the rationale will better encourage practices related to research ethics, openness, and transparency.

The original paper and paper and all supplementary materials are freely available at https://doi.org/10.17605/osf.io/n25d6.

ACM Reference Format:
Kavous Salehzadeh Niksirat, Lahari Goswami, Pooja S. B. Rao, James Tyler, Alessandro Silacci, Sadiq Aliyu, Annika Aebli, Chat Wacharamanotham, and Mauro Cherubini. 2023. Criteria Definition for Research Ethics, Openness, and Transparency. 1, 1 (March 2023), 15 pages. https://doi.org/10.1145/3544548.3580848

1 INTRODUCTION

We identify the dimensions of 'good research' practices related to research ethics, transparency, and openness, and operationalize them into criteria that are measurable in published HCI materials. We identified practices related to research ethics, openness, research transparency, and reporting transparency. Accordingly, we defined 40 specific criteria across these practices.

These criteria are explained in Tables 1–6. Each criterion has a code name (e.g., IRB). In our assessment process, we checked papers for relevant criteria and, when included, their supplementary materials also. To identify the existence of a criterion in a paper, we used keyword matching to highlight candidate sentences. Given that different terms might have different meanings based on a specific context, we carefully checked the paragraphs that contained those sentences. For others, we thoroughly inspect their supplementary materials to check whether a given material is shared. Finally,

Authors' addresses: Kavous Salehzadeh Niksirat, kavous.salehzadehniksirat@unil.ch, University of Lausanne, Switzerland; Lahari Goswami, lahari.goswami@unil.ch, University of Lausanne, Switzerland; Pooja S. B. Rao, pooja.rao@unil.ch, University of Lausanne, Switzerland; James Tyler, james.tyler@unil.ch, University of Lausanne, Switzerland; Alessandro Silacci, alessandro.silacci@unil.ch, University of Lausanne, Switzerland and School of Management of Fribourg, HES-SO University of Applied Sciences and Arts Western Switzerland, Switzerland; Sadiq Aliyu, sadiq.aliyu@unil.ch, University of Lausanne, Switzerland; Annika Aebli, nikaaebli@gmail.com, University of Lausanne, Switzerland; Chat Wacharamanotham, chat@acm.org, Swansea University, United Kingdom; Mauro Cherubini, mauro.cherubini@unil.ch, University of Lausanne, Switzerland.

2023. Manuscript submitted to ACM

Table 1. Criteria definition for research ethics

CODE, criterion [Justification]	Filtering keyword examples	Inspection notes	Subset†
IRB Did the study receive approval from an institutional review board? [28, section 15.3.3]	IRB, institutional review board, ethical approval	—	—
CONSENT (reported) Was written consent obtained from study participants? [28, section 15.3.4]	consent, consent form, sign	Checking papers for the report of consent collection	Papers with human participants
CONSENT (form shared) Do supplementary materials include the consent form? [28, section 15.3.4]	consent, consent form, sign	Checking supplementary materials for a shared consent form	Papers with human participants
STUDY-COMPENSATION Was participants' compensation explained in the paper? [28, section 15.2.3]	paid, compensated, USD	Checking the type and amount of compensation	Papers with human participants
ANON Was any data anonymization used? [48, section 1]	pseudonymized, k-anonymity, identity	—	—
FACE-PHOTO Are the facial photos in the paper shared with consent, or if not, are the participants' privacy protected? [10, 39]	permission, consent	Checking the figures with participant face/body	Papers with participant's photo
VULNERABLE Were any ethical measures taken to support vulnerable participants? [35, Ten Lessons][45]	minority, impairment, immigrants	Inspect methodology if any measure were taken and described at all	Papers with vulnerable participants
ANIMAL Were any ethical measures taken to support animals? [16, Item 14]	animal computer interaction, pet, dog	Inspect methodology if any measure were taken and described at all	Papers with animal participants

†A blank cell indicates that the criterion is applicable to all empirical studies.

not all criteria are relevant for all papers. Therefore, denominator subsets for each criterion filter the papers based on their relevance. For instance, we only reviewed qualitative and mixed-methods papers and disregarded quantitative studies for criteria related to qualitative studies (e.g., sharing an interview guide). The spreadsheet version of the criteria is provided in Sup. 4 to allow researchers to sort the list according to criteria category (i.e., to use from the reviewers' perspective) or research phase (i.e., to use from the authors' perspective). The full list keywords is provided in Sup. 5.

2 PRACTICES RELATED TO RESEARCH ETHICS

National laws, state laws, and institutional regulations often form the basis for ethical guidelines. Also, different science communities can have domain-specific codes of ethics such as ACM Code of Ethics and Professional Conduct and IEEE

Code of Ethics. These regulations support researchers in respecting participants' dignity and privacy, protecting them against mental or physical health risks. We identified seven criteria for research ethics including four categories of general practices, privacy-related practices, specific practices for vulnerable populations, and ethics for animals. The outline summary of practices related to research ethics is presented in Table 1.

2.1 General Practices for Research Ethics

Acquiring ethical approval (**IRB**) before conducting studies with participants is a necessary step for researchers as it ensures safeguarding the participants [38][28, section 15.3.3]. Researchers in their ethics submission should anticipate potential risks (e.g., discrimination, disclosure of private information, power imbalance) and note countermeasures to avoid such ethical issues [25, 36]. Nowadays, receiving ethical approval is required by ethics committees in many universities and research institutes.

It is necessary to collect participants' consent for participating in the study before conducting them (**CONSENT**) [28, section 15.3.4]. Such consent should inform participants about the study's goal, possible risks and benefits, the measure for protecting participants' data, and their rights during and after the experiment (e.g., to withdraw from the experiment and request for deletion of the data). The consent should be explicit and without any pressure (e.g., power imbalance, colleague pressure), and ideally should be written than verbal.

From an ethical point of view, it is vital to compensate participants for the amount of time and effort they dedicated to the experiments (**STUDY-COMPENSATION**) [28, section 15.2.3]. Participants should either receive monetary benefits such as cash/bank payments, gift cards, and lotteries or immediate benefits from the study. Immediate benefits can include education through attending a workshop or getting privileged use of particular services that help them improve their well-being. Being transparent about incentives is also crucial from a replicability point of view as its amount and type can be important factors for replication studies [32].

2.2 Privacy

All types of data including participants' personal data (e.g., demographics and characteristics) and experimental data should be kept secure and anonymized to protect participants' privacy (**ANON**) [48, section 1]. Different practices can be applied to data such as pseudonymization [1, 30] or ensuring k-anonymity [49]. In the papers, we searched for evidence of whether any anonymization techniques were applied.

Sharing facial photos without consent could incur harm to people present in the photos [10]. Given the user-centric nature of HCI research, it is custom to use participants' photos as figures in the papers to depict how they interacted with specific technology or what the experimental setup was. In such cases, asking for participants' permission before using their faces in the photos is necessary. In our reviews, we checked if the authors declared consent collection before sharing photos in the paper (**FACE-PHOTO**). Avoidance of violating photo privacy is possible through obfuscation (e.g., blurring or masking) [20], or by using photos that does not fully show the face of participants (e.g., photos with a VR headset on the eyes) [1].

2.3 Research Ethics with Vulnerable Populations

Vulnerable populations are people who are more at risk of being harmed and unable to protect their interests, such as racial and ethnic minorities, gender specific minorities, and those with chronic health conditions and severe mental illnesses. Researchers must be particularly cautious when working with these groups [35, 45]. Such participants should be informed, educated, and protected from risks and damages. They might need to be accompanied by health care

Table 2. Criteria definition for practices related to openness

CODE, criterion [Justification]	Location(s) to Inspect*	Filtering keyword examples	Inspection notes	Subset†
PAYWALL-ACMDL Is the paper in ACM DL available as open access? [4]	ACM DL website	—	Checking the label above the paper's title on the ACM DL page pointed by the DOI.	—
FREE-PDF-EXTERN Is the paper PDF available on external platforms other than ACM DL? [22]	Google Scholar	—	Searching for links of the paper's PDF in Google Scholar (except the main ACM DL link)	—
EXTRA Are any research artifacts beyond the paper provided anywhere? [44]	Paper, ACM DL website, external websites linked from the paper	Open Science Framework, GitHub, supplementary	(1) Checking the paper page (as linked by its DOI) on the ACM DL. (2) Checking the link of repositories in the paper, and inspecting the repository page. (3) Checking the appendix of the paper.	—
EXTRA-EXIST Do all provided research artifacts exist at the location specified in the paper? [47]	The page of the external repository or ACM DL website	—	Checking their existence in the supplementary materials and in the URLs provided in the paper.	Papers that meets the EXTRA criterion
EXTRA-FAIR Do any of the locations of provided research artifacts satisfy the FAIR principle? [47]	The page of the external repository or ACM DL website	—	Checking if the location meets the FAIR Principles	Papers that meets the EXTRA criterion

*A blank cell indicates that the only location to inspect is the paper itself. A non-blank cell lists the specific locations. †A blank cell indicates that the criterion is applicable to all empirical studies.

professionals or family members (**VULNERABLE**). Moreover, we considered part of papers with student participants as a vulnerable population. It is common in HCI to recruit university students for experiments. However, when the study is about learning context or happens in a classroom environment, there can likely be imbalanced power dynamics between the participants (students) and researchers (teachers). Therefore, we decided to consider such cases as vulnerable populations. In our review, we check if any additional ethical measures were reported to protect the well-being of the concerned vulnerable population, other than following the *general practices* (e.g., IRB, CONSENT). An example of such a measure is for study participants with limited consent capacity, researchers obtain written informed consent from their legal representative or caregiver prior to the study, and the participant provides assent during the study.

2.4 Research Ethics for Animals

Animal-Computer Interaction (ACI) is an emerging field in HCI, with ACI papers appearing most frequently at the International Conference on ACI. Nevertheless, CHI has included some of these in recent years. Although the risk of experiments in the field of ACI is lower compared to other fields such as biology and medicine, it is important to be mindful of animal welfare (**ANIMAL**). We used ARRIVE guidelines 2.0 for assessing animal ethics [16, Item 14].

3 PRACTICES RELATED TO OPENNESS

We review openness: a practice of sharing research reports, findings, and extra materials and making sure that such materials are accessible to the public, free of charge [14, 41]. Openness is a recent movement, and many research institutions and funding bodies advocate for it. Following open-access practices is essential for widespread accessibility to expanding knowledge, as independent researchers and researchers from low-income countries can keep themselves updated with recent scientific advances. Table 2 summarizes the criteria.

3.1 Open Publication

ACM provides open-access publication options for authors (i.e., gold open access) [4]. Thus, the readers can access the papers without facing paywall barriers.[1] To publish open-access, authors had to pay a fee, mostly by their institutions or funding agencies. Nevertheless, this could be difficult for some researchers without such support. In our review, we checked if the paper publishing was with either "Open Access" or "Public Access" label (**PAYWALL-ACMDL**).

Some papers may not be open access on the ACM DL platform, but researchers may make authors' versions available publicly (**FREE-PDF-EXTERN**). They could use a personal or institution website, or paper sharing platforms such as ResearchGate, HAL, or arXiv. ACM permits authors to post the "author's version" of the paper on their homepage or institutional websites [4]. This is called green open access. Indeed, green and gold open access complement each other during transition periods [19]. On the other hand, ACM explicitly prohibits sharing on commercial social networking websites such as ResearchGate [4] due to the potential for copyright infringement [22].We do not advocate the practice of sharing papers on commercial social networking websites, but we assess this as an existing practice to understand the level of accessibility via different platforms.

3.2 Sharing Supplementary Materials

Given the page limits in CHI papers,[2] for transparent science, researchers would share extra research artifacts as supplementary materials [7]. The most standard practice is using the ACM DL and submitting supplementary materials besides the paper PDF. Another option for sharing is to use external platforms that promote data sharing on open and collaborative frameworks, such as OSF or Zenodo. The last option is to use the appendix section at the end of the paper. In this review, we checked if CHI authors used one of these practices for sharing research artifacts (**EXTRA**).[3]

We also check if artifact sharing was properly implemented. Thus, we assess whether the content referred to or promised by the authors was available (**EXTRA-EXIST**) or if there was any missing data.

Finally, we assess if the archived research artifacts were publicly accessible (**EXTRA-FAIR**). To this end, first, we check if the used repositories are compatible with FAIR principles. FAIR principles refer to being Findable (i.e., by having unique identifiers), Accessible (i.e., by not being locked), Interoperable (i.e., by providing ReadMe files to clarify the structure), and Reusable (i.e., by providing metadata that can support readers to understand the data and reuse it) [47]. Platforms such as OSF or Zenodo are FAIR compatible, but some researchers might use incompatible platforms. For example, they might use a personal website for sharing data. Another example of a non-compatible platform commonly used among HCI researchers is GitHub. GitHub is not accessible because repositories are deletable and not findable

[1]Note that SIGCHI, through a dedicated website, allows readers to access the papers published in SIG-sponsored proceedings regardless of if the papers are open access or not [3]. This system has been effective since 2014. Nevertheless, most researchers might not know about this website.
[2]Note that even though CHI 2022 has no strict page limits, authors are encouraged to be concise and that the number of papers should be proportional to the contribution of each paper.
[3]EXTRA criterion actually reflects the transparency aspect, but we included it here because it explains the denominator for the EXTRA-EXIST and EXTRA-FAIR criteria. In the transparency subsection (see 4), there are criteria that are in finer granularity by type of research artifacts.

because it lacks a unique and persistent identifier. The advantage of sharing on ACM DL is that it is FAIR compatible. Where researchers should ensure interoperability and reusability, ACM DL does support findability and accessibility.[4] Second, we check if the papers that use appendices were eventually free perpetually (i.e., labeled as "open" or "public" on the ACM DL page of the paper). We checked this because if a paper is not publicly accessible, the readers that face a paywall barrier cannot access its research artifacts. Papers that could satisfy either the first or the second condition were labeled as "Yes."

4 PRACTICES RELATED TO TRANSPARENCY

Transparency is one of the most important practices essential for replication and reproducibility. To ensure transparent research, researchers can share the important elements of their study such as study materials, data collection and analysis procedure, collected raw or processed data, and experimental artifacts such as software tested within the study. The majority of the criteria related to transparency were informed based on the TOP Guidelines [29, 31] and an earlier research artifact taxonomy Wacharamanotham et al. [43, Fig. 2]. Tables 3 and 4 summarized the criteria for transparency.

4.1 Preregistration

This is the practice of registering a study design before data collection (ideally) or data analysis phases (**PREREG**) [11, 31]. Preregistration usually requires submitting researchers' plans and decisions concerning sample size, independent and dependent variables, inclusion and exclusion criteria, and data analysis. The critical feature of preregistration is time stamps which cannot change after registration. Preregistration is particularly important for confirmatory studies that do hypothesis testing. In this case, they can help avoid HARKing (i.e., Hypothesising After the Results are Known) [11, 23]. Preregistration can even serve exploratory and qualitative research, where researchers can register their initial beliefs and perceptions about their study [21] - potentially avoiding later biases. The most commonly used services for preregistration are `OSF registries` and `AsPredicted`.

4.2 Sharing Study Materials

Researchers can share stimuli used or tested in the studies [43] (**SHARE-STIMULI**). Such stimuli can be visual or auditory materials presented to participants to elicit their responses, such as storyboards used in surveys for scenario testing or deck of cards used in participatory design. This criterion excludes interview questionnaires as explained next.

One of the commonly used metrics in HCI research is questionnaires or surveys. Questionnaires can be well-established scales or questions designed or adopted by researchers for a specific context. The questionnaires can be deployed either online (a.k.a surveys) or in the lab. Sharing questionnaires is important for replicability, as it allows researchers to use identical questions in their replication study (**SHARE-SURVEY**) [43, taxonomy]. Given that some researchers might use several questionnaires but only share a few, we labeled such cases as "partially." We also counted pre-study questionnaires such as demographic and screener questionnaires.

Interview protocol is a document that includes the list of questions asked during the interview, and it may include instructions for probing questions and how to start or wrap up the interview. Sharing interview protocol (**SHARE-INTERVIEW-GUIDE**) can help understand how the qualitative responses are elicited [43, taxonomy]. This gesture is important not for replication per se but also for cross-sectional studies to execute identical interviews with participants

[4]Note that even for papers without open access, ACM provides supplementary materials as freely accessible [2].

Table 3. Criteria definition for transparency-related practices (Part 1)

CODE, criterion [Justification]	Location(s) to Inspect*	Filtering keyword examples	Inspection notes	Subset†
PREREG Was the study preregistered? [11, 31]	Paper, the page of the preregistration platform	preregister, AsPredicted, OSF Registries	Checking the link of preregistration platform in the paper, and inspecting the page	—
SHARE-STIMULI Are study stimuli (except survey questionnaires) archived? [43, taxonomy]	Appendix, supplementary materials	—	Checking additional materials	Paper with human participants
SHARE-SURVEY Are questionnaires or surveys archived? [43, taxonomy]	Paper, Appendix, supplementary materials	questionnaire, online survey, post-test survey	Checking keywords in the paper and searching for additional materials	Papers that used questionnaires or surveys
SHARE-INTERVIEW-GUIDE Is interview guide archived? [43, taxonomy]	Paper, Appendix, supplementary materials	interview protocol, interview questions, interview guide	Checking keywords in the paper and searching for additional materials	Qualitative papers that used interviews
SHARE-STUDY-PROTOCOL Is the study protocol archived? [33, Publicly Accessible Study Protocol]	Appendix, supplementary materials	experiment protocol, procedure, checklist	Checking keywords in the paper and searching for additional materials	—
JUSTIFY-N-QUAL Was the sample size justified (qualitative studies)? [8]	—	saturation, sample size, theoretical sampling	—	Qualitative or mixed papers
JUSTIFY-N-QUAN Was the sample size justified (quantitative studies)? [26, 34]	—	power analysis, sample size, G*power	—	Quantitative or mixed papers
DEMOGRAPHICS Was the demographic information of the participants described? [18]	—	background, characteristics, demographic	—	Paper with human participants
CONDITION-ASSIGNMENT Did the study properly explain study design (e.g., grouping, IDVs)? [42, section 6.3]	—	between-subject, independent variable, condition	—	Quantitative or mixed papers (with experiments)

*A blank cell indicates that the only location to inspect is the paper itself. A non-blank cell lists the specific locations. †A blank cell indicates that the criterion is applicable to all empirical studies.

from different ethnicities, races, and communities to better understand the similarities and differences. Not necessarily all qualitative studies have interviews. Many researchers report anecdotes from conversation analysis, ethnographic studies, etc. We considered structured, semi-structured, and non-structured interviews.

Researchers usually report their experimental procedure in the method sections of their paper. While such a narrative can help to understand the procedure, it might omit some necessary details for replication [33]. Sharing study protocol—written before data collection—improves the credibility of research, and it facilitated replication as it includes fine-grained study details. Also, it might also be helpful to compare it to the published paper to identify reporting biases [33]. Thus, we searched for any detailed instructions, such as a checklist document that explains all practical steps in a very detailed approach that can help to replicate the experiment (**SHARE-STUDY-PROTOCOL**). We labeled the criterion as 'Yes' if it provides a complete and detailed explanation, "partially" if it shares only part of the procedure.

4.3 Practices related to Participants

Justifying sample size before a study occurs is a well-known practice that explains how much the findings collected with a given sample size can be generalizable. The practice is more solid for quantitative research (**JUSTIFY-N-QUAN**) as researchers can conduct power analysis and calculate the minimum required sample size to acquire significant findings with a specific level of power and acceptable effect size [26, 34]. Support for these processes can come from software such as G*Power and other innovations for sample size computation [46]. For qualitative research (**JUSTIFY-N-QUAL**), the most common justification for sample size is saturation [8] where researchers recruit participants until they reach saturation in their qualitative analysis (i.e., participants are no longer revealing new discussion topics). The second most common approach is to rely on previous studies, with researchers referring to a previous article. Lastly, the researchers might mention practical limitations or logistics.

To be able to replicate a user study, it is essential to know its participants' characteristics (**DEMOGRAPHICS**) [18]. Recruiting participants of different ages, gender, sexual orientations, ethnicities, and socioeconomic statuses can cause substantial changes in the replication study results compared with the original one. Thus, sharing participant information is necessary to be as transparent as possible. At the same time, there is no standard on the extent of explanation there should be for participants' details. While some researchers can use supplementary materials and provide fine-grained information about their participants, they must be mindful of ethical restrictions (e.g., sharing personally identifiable information about their participants). Even in some cases where participants are anonymized, providing their background information can help others to infer their identities, particularly if they are from low-population communities and possess rare background characteristics.

4.4 Study Design

The next family of criteria is related to study design (**CONDITION-ASSIGNMENT**). For reproducibility, It is essential to describe the experimental design, such as the number of experimental conditions, within-subject, between-subject design, or mixed design. In particular, for quantitative studies, it is necessary to clearly explain independent and dependent variables [42, section 6.3].

4.5 Data Analysis

Sharing data analysis procedures is a critical step to support reproducing the analysis (**SHARE-ANALYSIS-CODE**) [43, taxonomy][31]. We assessed if CHI authors reported and shared the analysis process and if shared any script. For qualitative studies, we were interested to see if CHI authors reported the approach they use to analyze their qualitative data (**SPECIFY-QUAL-ANALYSIS**) such as grounded theory, thematic analysis, or in vivo coding. We checked if CHI authors mentioned the name of the approach and if they explained the procedure, even briefly.

Table 4. Criteria definition for transparency-related practices (Part 2)

CODE, criterion [Justification]	Location(s) to Inspect*	Filtering keyword examples	Inspection notes	Subset†
SPECIFY-QUAL-ANALYSIS Is qualitative data analysis approach named or explicitly described? [43, taxonomy][31]	Paper, Appendix, supplementary materials	grounded theory, open coding, thematic analysis	—	Qualitative or mixed papers
SHARE-ANALYSIS-CODE Is quantitative data analysis code shared? [43, taxonomy][31]	Paper, Appendix, supplementary materials	R script, SPSS code	—	Quantitative or mixed papers
QUAL-DATA-RAW Is raw qualitative data shared? [43, taxonomy][31]	Paper, Appendix, supplementary materials	interview transcripts, observatory field notes, diary entries	—	Qualitative or mixed papers
QUAL-DATA-PROCESSED Is processed qualitative data shared? [43, taxonomy][31]	Paper, Appendix, supplementary materials	codebook	—	Qualitative or mixed papers
QUAN-DATA-RAW Is raw quantitative data shared? [43, taxonomy][31]	Paper, Appendix, supplementary materials	raw data, log, timestamp	In addition to checking the paper, inspect data to see if it is raw or processed	Quantitative or mixed papers
QUAN-DATA-PROCESSED Is processed quantitative data shared? [43, taxonomy][31]	Paper, Appendix, supplementary materials	processed data, anonymized data, dataset	In addition to checking the paper, inspect data to see if it is raw or processed	Quantitative or mixed papers
SHARE-SOFTWARE Is the source code of the software shared? [43, taxonomy]	Supplementary materials	source code, prototype, Docker	—	Papers with artifact as one of the contribution
SHARE-HARDWARE Is the code of the hardware shared? [43, taxonomy]	Supplementary materials	blueprint, 3D design, open hardware	—	Papers with artifact as one of the contribution
SHARE-SKETCH Is any hand-drawn sketch shared?	Supplementary materials	sketch, drawing, mental model	—	—

*A blank cell indicates that the only location to inspect is the paper itself. A non-blank cell lists the specific locations. †A blank cell indicates that the criterion is applicable to all empirical studies.

4.6 Data Sharing

The next series of practices are related to data sharing. We assessed data sharing practices for both qualitative and quantitative studies. We also checked if the shared material was raw or processed data [43, taxonomy][31]. We first checked if papers with qualitative studies shared any raw data such as interview transcripts, interview notes, or observatory field notes (QUAL-DATA-RAW). It can also be known as selective data, as the collection is at the researchers' discretion. The necessity of sharing such material is under debate by qualitative researchers [40], and it might have ethical implications for participants [17]. Next, among papers with qualitative studies, we searched if they shared their codebook or any document that shows annotations on qualitative data and the categorization of topics, known as themes (**QUAL-DATA-PROCESSED**). We searched supplementary materials of quantitative studies to find raw quantitative data (QUAN-DATA-RAW). Such data could be non-selective data, such as log data collected through software/tools without the researchers' active involvement has been mentioned. Finally, we searched supplementary materials of quantitative studies to find processed and selective quantitative data (**QUAN-DATA-PROCESSED**). We did not consider sharing quantitative results (i.e., aggregated data) as processed quantitative data sharing.

4.7 Sharing Artifacts

The following criteria are related to sharing research and experimental artifacts, including software, hardware, and sketch [43, taxonomy]. Hardware and software can be either an artifact built to be tested in a user study or made as the study's outcome. We checked if papers shared the source code of the software they developed (**SHARE-SOFTWARE**). We searched among supplementary materials to see if shared hardware code or schematics (**SHARE-HARDWARE**). Lastly, we checked if papers shared any hand-drawn sketches (**SHARE-SKETCH**). These can be, for example, sketches drawn during a participatory design or participants' mental models, and participants or researchers can draw them.

5 PRACTICES RELATED TO REPORTING

Recently many guidelines have emerged on properly reporting scientific studies (e.g., [37]). We recognized several criteria for reporting findings of HCI studies. Most of our criteria are related to reporting quantitative data, while one is about reporting qualitative data. For quantitative criteria, we assessed if the CHI authors properly used and reported the statistical tests. We classified the statistical analysis into four categories [24]: frequentist hypothesis test, frequentist estimation with uncertainty, Bayesian hypothesis test, and Bayesian estimation with uncertainty.[5] The reporting criteria are listed in Tables 5 and 6.

5.1 Reporting Null Hypothesis Significance Testing (NHST) Results

We checked if papers reported a central tendency of data (e.g., mean, median, or mode) and its variability (e.g., standard deviation, standard error, quartiles, min/max) [13, see Guideline 5-6][27]. We checked this criterion (**STAT-DESCRIPTIVE**) separately for non-categorical and categorical variables. In contrast, the standard reporting approach for categorical variables is reporting count, frequency, or proportion without a variability measure. Presenting the central tendency and variability of data in each group (or condition) allows the readers to judge the simple effect size (if not reported). These group statistics also enable future meta-analysis (whereas reporting only statistical tests will not).

We checked if CHI authors clearly described the tests that they used (**STAT-CLEAR-PROCEDURE**) [13, Guideline 3]. A clear description of the method is necessary for reproducing the analysis with the same raw data. It also allows subsequent studies to use the results in their planning and comparison.

Next, we assessed if CHI authors checked and reported the distribution of the data before deciding to run a parametric or non-parametric test (**STAT-NORMALITY**) [5, 9, 27]. Reporting the assessment of statistical assumptions allows readers to determine whether the chosen statistical approach is suitable. We also checked if CHI authors used any statistical

[5]Given that none of the articles in our sample uses Bayesian statistics, we omitted the instructions for Bayesian statistics—which could have been drawn from a section in the SAMPL guideline from the field of medicine [27].

Table 5. Criteria definition for practices related to transparency in reporting (Part 1)

CODE, criterion [Justification]	Filtering keyword examples	Inspection notes	Subset
STAT-DESCRIPTIVE (central tendency) For each key dependent variable on the interval or ratio scale, were their sample central tendency reported? [13, see Guideline 5-6][27]	mean, average, M	—	Quantitative or mixed papers with frequentist statistics
STAT-DESCRIPTIVE (variability) For each key dependent variable was their sample variability reported? [13, see Guideline 5-6][27]	SD, SE, quartile	—	Quantitative or mixed papers with frequentist statistics
STAT-DESCRIPTIVE (categorical data) Were their sample reported for each key dependent variable on the normal or ordinal scale? [13, see Guideline 5-6][27]	median, mod, N	—	Quantitative or mixed papers with frequentist statistics
STAT-CLEAR-PROCEDURE Is the statistical procedure for data analysis clearly named? [13, Guideline 3]	ANOVA, Mann-Whitney, Chi-square test	—	Quantitative or mixed papers with frequentist statistics
STAT-NORMALITY When the normality assumption is required by the statistical procedure, was the assumption assessed? [5, 9, 27, 42]	Normality, parametric, Shapiro-Wilk	—	Papers that use statistical tests that require this assumption
STAT-OTHER-ASSUMPTIONS When the statistical procedure requires additional assumptions, were they assessed? [9, section 4.2] [27, 42]	homogeneity of variance, sphericity, multicollinearity assumption	—	Papers that use statistical tests that require this assumption

assumptions for frequentist statistics (STAT-OTHER-ASSUMPTIONS). The most common statistical assumptions are homogeneity of variance or sphericity, used for t-test and ANOVA.

We checked if CHI authors reported the main values necessary for reporting each test (**STAT-PARAMETERS**). We considered the degree of freedom, the test value (e.g., t-value, F-value), and the p-value, as essential elements for readers to determine the statistical validity of a report [27]. The degree of freedom indicates the number of data points (i.e., the number of independent observations), and it can be used to infer the sample size if the test type is known [42]. Given that the same statistics using known t or F values will not yield a unique p-value, the degree of freedom is a necessary parameter for determining the statistical distribution from which the p-value is drawn. The degree of freedom is not necessarily equal to the number of participants. For example, in an experiment where each participant performs multiple repetitions of the same conditions, the analyst could choose to model each repetition individually or use the average of the values from each participant. Choosing the latter will result in a smaller degree of freedom [27]. p-values are the main indicator of statistical significance and they should be reported either precisely [13] or with the level of significance (e.g., $p < .05$, $p < .01$, $p < .001$) [6].

Where the aforementioned variables are essential for reporting, reporting further values such as effect size and confidence interval is appropriate and can allow readers to make a richer interpretation of the results. Reporting effect

Table 6. Criteria definition for practices related to transparency in reporting (Part 2)

CODE, criterion [Justification]	Filtering keyword examples	Inspection notes	Subset
STAT-PARAMETERS (df) Were degree of freedom reported? [27, 42]	—	Check the reported test results, for example, for $t(15) = 2.20$ degrees of freedom = 15	Quantitative or mixed papers with frequentist statistics
STAT-PARAMETERS (test value) Were the test statistic and all test parameters reported? (e.g., F-value) [27, 42]	—	Check the reported test results, for example, for $t(15) = 2.20$ Test statistic: 2.20	Quantitative or mixed papers with frequentist statistics
STAT-PARAMETERS (p-value) Were p-value reported? [27, 42]	—	Check the reported p-values in text and tables	Quantitative or mixed papers with frequentist statistics
STAT-EFFECT-SIZE For the effects that were tested, were effect sizes reported? [27, 42, 50]	Effect size, partial η^2, Cohen's d	—	Quantitative or mixed papers with frequentist statistics
STAT-CI For the effects that were tested, were their confidence intervals reported? [13, 42]	95% CI, confidence interval, bootstrapped CI	—	Quantitative or mixed papers with frequentist statistics
ESTIMATES-INTERVAL Were interval estimates reported? [15, Tips 15 & 18]	95% CI, confidence interval	—	Quantitative or mixed papers with estimation statistics
ESTIMATES-VIS-UNCERTAINTY Was the uncertainty of the effect visualized? [15, Tip 16]	—	Checking result figures if uncertainty was plotted using confidence interval	Quantitative or mixed papers with estimation statistics
QUAL-INTERVIEW-REPORT Did the study properly report themes and quotes? [28, section 8.10.4]	quotes, themes, categories	Irrespective of the keywords, carefully inspecting the results section	Qualitative or mixed papers that used interviews

size (STAT-EFFECT-SIZE) is important as it helps readers to understand the difference between groups and to understand if the significant finding is practical (i.e., large effect size) [27, 50]. Based on the type of test, reporting effect size can occur in different ways. For example, where specific measures such as Pearson r in a correlation test, r^2 in a regression test, η_p^2 in ANOVA can indicate effect size, the effect size can also be computed in a more general manner such as using Cohen's d. Confidence intervals are a way of reporting the degree of uncertainty for the findings (STAT-CI) [13]. For instance, 95% CIs are the upper and lower range that a population parameter will fall with 95% probability. Confidence intervals allow readers to make richer interpretations beyond the dichotomous (significant or not). The upper or lower bound can be interpreted separately based on the research question. The best practice is to report p-values, 95% CIs, and effect size together.

5.2 Reporting Estimation Results

The next two criteria are for the articles that use estimation statistics. We extracted these criteria according to guidelines provided by Dragicevic [15] for estimation statistics. First, we checked reporting Interval Estimates for inferences (**ESTIMATES-INTERVAL**). Applicable intervals could be the confidence intervals or Bayesian credible interval or predictive interval. To meet this criterion, the interval must lend itself for inference. For example, consider a between-subject experiment that compares two conditions. The confidence interval of the mean difference can be directly used for inference. Alternatively, the confidence interval of samples are indirectly applicable using the overlap rule [12]. However, if this experiment is within-subjects, the overlap rule no longer applies. Therefore, only the confidence interval of the difference could be used for inference. Lastly, the common interval level is 95%. Other levels such 90% or 99% are acceptable only if justified [15].

The use of figures is encouraged rather than making textual reports. Dragicevic [15] described the best practices for confidence intervals. Thus, we checked if the papers used any graphics to visualize their estimates (**ESTIMATES-VIS-UNCERTAINTY**).

5.3 Reporting Interview Results

This criterion (**QUAL-INTERVIEW-REPORT**) assesses whether the qualitative results from interviews were analyzed and reported. Papers that meet this criterion at least report a summary of the findings into topics or themes. We focus only on interviews because they are prevalent and frequently used with other data collection methods. Researchers might omit interview results when other data are more salient or cherry-pick a few quote to report. Although these characteristics could have been applied to results from open questions in surveys, we decided not to include them in this criterion. The non-interactive nature of surveys do not guarantee adequately rich data for researchers to analyze.

REFERENCES

[1] Jacob Abbott, Haley MacLeod, Novia Nurain, Gustave Ekobe, and Sameer Patil. 2019. Local Standards for Anonymization Practices in Health, Wellness, Accessibility, and Aging Research at CHI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300692

[2] ACM. 2019. ACM Policy on Submission, Hosting, Access, and Ownership of Digital Artifacts. https://www.acm.org/publications/policies/digital-artifacts

[3] ACM. 2019. Free Public Access to SIG-Sponsored Proceedings around the Time of the Event (OpenSurround, OpenTOC). https://www.acm.org/publications/policies/free-access

[4] ACM. 2022. Open Access Publication & ACM. https://www.acm.org/publications/openaccess

[5] Balazs Aczel, Barnabas Szaszi, Alexandra Sarafoglou, Zoltan Kekecs, Šimon Kucharský, Daniel Benjamin, Christopher D. Chambers, Agneta Fisher, Andrew Gelman, Morton A. Gernsbacher, John P. Ioannidis, Eric Johnson, Kai Jonas, Stavroula Kousta, Scott O. Lilienfeld, D. Stephen Lindsay, Candice C. Morey, Marcus Munafò, Benjamin R. Newell, Harold Pashler, David R. Shanks, Daniel J. Simons, Jelte M. Wicherts, Dolores Albarracin, Nicole D. Anderson, John Antonakis, Hal R. Arkes, Mitja D. Back, George C. Banks, Christopher Beevers, Andrew A. Bennett, Wiebke Bleidorn, Ty W. Boyer, Cristina Cacciari, Alice S. Carter, Joseph Cesario, Charles Clifton, Ronán M. Conroy, Mike Cortese, Fiammetta Cosci, Nelson Cowan, Jarret Crawford, Eveline A. Crone, John Curtin, Randall Engle, Simon Farrell, Pasco Fearon, Mark Fichman, Willem Frankenhuis, Alexandra M. Freund, M. Gareth Gaskell, Roger Giner-Sorolla, Don P. Green, Robert L. Greene, Lisa L. Harlow, Fernando Hoces de la Guardia, Derek Isaacowitz, Janet Kolodner, Debra Lieberman, Gordon D. Logan, Wendy B. Mendes, Lea Moersdorf, Brendan Nyhan, Jeffrey Pollack, Christopher Sullivan, Simine Vazire, and Eric-Jan Wagenmakers. 2020. A Consensus-Based Transparency Checklist. *Nature Human Behaviour* 4, 1 (Jan. 2020), 4–6. https://doi.org/10.1038/s41562-019-0772-6

[6] APA. 2020. *Publication Manual of the American Psychological Association, Seventh Edition.* American Psychological Association (APA), IL, USA. https://apastyle.apa.org/products/publication-manual-7th-edition

[7] Ángel Borrego and Francesc Garcia. 2013. Provision of Supplementary Materials in Library and Information Science Scholarly Journals. *Aslib Proceedings: New Information Perspectives* 65, 5 (Jan. 2013), 503–514. https://doi.org/10.1108/AP-10-2012-0083

[8] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 981–992. https://doi.org/10.1145/2858036.2858498

[9] Paul Cairns. 2007. HCI... Not as It Should Be: Inferential Statistics in HCI Research. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but Not as We Know It - Volume 1 (BCS-HCI '07)*. BCS Learning & Development Ltd., Swindon, GBR, 195–201.

[10] Mauro Cherubini, Kavous Salehzadeh Niksirat, Marc-Olivier Boldi, Henri Keopraseuth, Jose M. Such, and Kévin Huguenin. 2021. When Forcing Collaboration Is the Most Sensible Choice: Desirability of Precautionary and Dissuasive Mechanisms to Manage Multiparty Privacy Conflicts. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 53:1–53:36. https://doi.org/10.1145/3449127

[11] 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)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173715

[12] Geoff Cumming and Sue Finch. 2005. Inference by Eye: Confidence Intervals and How to Read Pictures of Data. *American Psychologist* 60 (2005), 170–180. https://doi.org/10.1037/0003-066X.60.2.170

[13] Douglas Curran-Everett and Dale J. Benos. 2004. Guidelines for Reporting Statistics in Journals Published by the American Physiological Society. *Physiological Genomics* 18, 3 (Aug. 2004), 249–251. https://doi.org/10.1152/physiolgenomics.00155.2004

[14] Giovanni Destro Bisol, Paolo Anagnostou, Marco Capocasa, Silvia Bencivelli, Andrea Cerroni, Jorge Contreras, Neela Enke, Bernardino Fantini, Pietro Greco, Catherine Heeney, Daniela Luzi, Paolo Manghi, Deborah Mascalzoni, Jennifer Molloy, Fabio Parenti, Jelte Wicherts, and Geoffrey Boulton. 2014. Perspectives on Open Science and Scientific Data Sharing:An Interdisciplinary Workshop. *Journal of anthropological sciences = Rivista di antropologia: JASS* 92 (2014), 179–200. https://doi.org/10.4436/JASS.92006

[15] Pierre Dragicevic. 2016. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*, Judy Robertson and Maurits Kaptein (Eds.). Springer International Publishing, Cham, 291–330. https://doi.org/10.1007/978-3-319-26633-6_13

[16] Nathalie Percie du Sert, Amrita Ahluwalia, Sabina Alam, Marc T. Avey, Monya Baker, William J. Browne, Alejandra Clark, Innes C. Cuthill, Ulrich Dirnagl, Michael Emerson, Paul Garner, Stephen T. Holgate, David W. Howells, Viki Hurst, Natasha A. Karp, Stanley E. Lazic, Katie Lidster, Catriona J. MacCallum, Malcolm Macleod, Esther J. Pearl, Ole H. Petersen, Frances Rawle, Penny Reynolds, Kieron Rooney, Emily S. Sena, Shai D. Silberberg, Thomas Steckler, and Hanno Würbel. 2020. Reporting Animal Research: Explanation and Elaboration for the ARRIVE Guidelines 2.0. *PLOS Biology* 18, 7 (July 2020), e3000411. https://doi.org/10.1371/journal.pbio.3000411

[17] Casey Fiesler, Christopher Frauenberger, Michael Muller, Jessica Vitak, and Michael Zimmer. 2022. Research Ethics in HCI: A SIGCHI Community Discussion. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3491101.3516400

[18] John Furler, Parker Magin, Marie Pirotta, and Mieke van Driel. 2012. Participant Demographics Reported in "Table 1" of Randomised Controlled Trials: A Case of "Inverse Evidence"? *International Journal for Equity in Health* 11, 1 (March 2012), 14. https://doi.org/10.1186/1475-9276-11-14

[19] Jean-Claude Guédon. 2004. The "Green" and "Gold" Roads to Open Access: The Case for Mixing and Matching. *Serials Review* 30, 4 (Jan. 2004), 315–328. https://doi.org/10.1016/j.serrev.2004.09.005

[20] Rakibul Hasan, Yifang Li, Eman Hassan, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. 2019. Can Privacy Be Satisfying? On Improving Viewer Satisfaction for Privacy-Enhanced Photos Using Aesthetic Transforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300597

[21] Tamarinde L. Haven, Timothy M. Errington, Kristian Skrede Gleditsch, Leonie van Grootel, Alan M. Jacobs, Florian G. Kern, Rafael Piñeiro, Fernando Rosenblatt, and Lidwine B. Mokkink. 2020. Preregistering Qualitative Research: A Delphi Study. *International Journal of Qualitative Methods* 19 (Jan. 2020), 1–13. https://doi.org/10.1177/1609406920976417

[22] Hamid R. Jamali. 2017. Copyright Compliance and Infringement in ResearchGate Full-Text Journal Articles. *Scientometrics* 112, 1 (July 2017), 241–254. https://doi.org/10.1007/s11192-017-2291-4

[23] Norbert L. Kerr. 1998. HARKing: Hypothesizing After the Results Are Known. *Personality and Social Psychology Review* 2, 3 (Aug. 1998), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

[24] John K. Kruschke and Torrin M. Liddell. 2018. The Bayesian New Statistics: Hypothesis Testing, Estimation, Meta-Analysis, and Power Analysis from a Bayesian Perspective. *Psychonomic Bulletin & Review* 25, 1 (Feb. 2018), 178–206. https://doi.org/10.3758/s13423-016-1221-4

[25] Steinar Kvale. 2007. *Doing Interviews*. SAGE Publications, Ltd, 1 Oliver's Yard, 55 City Road, London England EC1Y 1SP United Kingdom. https://doi.org/10.4135/9781849208963

[26] Daniël Lakens. 2022. Sample Size Justification. *Collabra: Psychology* 8, 1 (March 2022), 33267. https://doi.org/10.1525/collabra.33267

[27] Tom Lang and Douglas Altman. 2016. Statistical Analyses and Methods in the Published Literature: The SAMPL Guidelines. *Medical Writing* 25 (Sept. 2016), 31–36. https://journal.emwa.org/statistics/statistical-analyses-and-methods-in-the-published-literature-the-sampl-guidelines/

[28] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction - 2nd Edition*. Morgan Kaufmann, MA, USA. https://www.elsevier.com/books/research-methods-in-human-computer-interaction/lazar/978-0-12-805390-4

[29] David Thomas Mellor, Jolene Esposito, Alexander C. DeHaven, Victoria Stodden, and Olivia Lowrey. 2022. TOP Guidelines. https://www.cos.io/initiatives/top-guidelines

[30] Thomas Neubauer and Johannes Heurix. 2011. A Methodology for the Pseudonymization of Medical Data. *International Journal of Medical Informatics* 80, 3 (March 2011), 190–204. https://doi.org/10.1016/j.ijmedinf.2010.10.016

[31] 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. Levy 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

[32] Jessica Pater, Amanda Coupe, Rachel Pfafman, Chanda Phelan, Tammy Toscos, and Maia Jacobs. 2021. Standardizing Reporting of Participant Compensation in HCI: A Systematic Literature Review and Recommendations for the Field. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3411764.3445734

[33] George Peat, Richard D. Riley, Peter Croft, Katherine I. Morley, Panayiotis A. Kyzas, Karel G. M. Moons, Pablo Perel, Ewout W. Steyerberg, Sara Schroter, Douglas G. Altman, Harry Hemingway, and for the PROGRESS Group. 2014. Improving the Transparency of Prognosis Research: The Role of Reporting, Data Sharing, Registration, and Protocols. *PLOS Medicine* 11, 7 (July 2014), e1001671. https://doi.org/10.1371/journal.pmed.1001671

[34] Marco Perugini, Marcello Gallucci, and Giulio Costantini. 2018. A Practical Primer To Power Analysis for Simple Experimental Designs. *International Review of Social Psychology* 31, 1 (July 2018), 20. https://doi.org/10.5334/irsp.181

[35] Camille R. Quinn. 2015. General Considerations for Research with Vulnerable Populations: Ten Lessons for Success. *Health & Justice* 3, 1 (Jan. 2015), 1. https://doi.org/10.1186/s40352-014-0013-z

[36] Michael Quinn Patton. 2022. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. https://us.sagepub.com/en-us/nam/qualitative-research-evaluation-methods/book232962

[37] Simon Schwab, Perrine Janiaud, Michael Dayan, Valentin Amrhein, Radoslaw Panczak, Patricia M. Palagi, Lars G. Hemkens, Meike Ramon, Nicolas Rothen, Stephen Senn, Eva Furrer, and Leonhard Held. 2022. Ten Simple Rules for Good Research Practice. *PLOS Computational Biology* 18, 6 (June 2022), e1010139. https://doi.org/10.1371/journal.pcbi.1010139

[38] Loan E. Sieber and Martin B. Tolich. 2022. Planning Ethically Responsible Research. https://us.sagepub.com/en-us/nam/planning-ethically-responsible-research/book237233

[39] Jose M. Such, Joel Porter, Sören Preibusch, and Adam Joinson. 2017. Photo Privacy Conflicts in Social Media: A Large-scale Empirical Study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3821–3832. https://doi.org/10.1145/3025453.3025668

[40] Poorna Talkad Sukumar, Ignacio Avellino, Christian Remy, Michael A. DeVito, Tawanna R. Dillahunt, Joanna McGrenere, and Max L. Wilson. 2020. Transparency in Qualitative Research: Increasing Fairness in the CHI Review Process. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3334480.3381066

[41] Ruben Vicente-Saez and Clara Martinez-Fuentes. 2018. Open Science Now: A Systematic Literature Review for an Integrated Definition. *Journal of Business Research* 88 (July 2018), 428–436. https://doi.org/10.1016/j.jbusres.2017.12.043

[42] Jan B. Vornhagen, April Tyack, and Elisa D. Mekler. 2020. Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, New York, NY, USA, 4–18. https://doi.org/10.1145/3410404.3414229

[43] 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*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376448

[44] Chat Wacharamanotham, Fumeng Yang, Xiaoying Pu, Abhraneel Sarma, and Lace Padilla. 2022. Transparent Practices for Quantitative Empirical Research. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3491101.3503760

[45] Ashley Marie Walker, Yaxing Yao, Christine Geeng, Roberto Hoyle, and Pamela Wisniewski. 2019. Moving beyond 'One Size Fits All': Research Considerations for Working with Vulnerable Populations. *Interactions* 26, 6 (Oct. 2019), 34–39. https://doi.org/10.1145/3358904

[46] Xiaoyi Wang, Alexander Eiselmayer, Wendy E. Mackay, Kasper Hornbaek, and Chat Wacharamanotham. 2021. Argus: Interactive a Priori Power Analysis. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (Feb. 2021), 432–442. https://doi.org/10.1109/TVCG.2020.3028894

[47] 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 (March 2016), 160018. https://doi.org/10.1038/sdata.2016.18

[48] Günter Wilms. 2019. Guide on Good Data Protection Practice in Research. https://www.eui.eu/documents/servicesadmin/deanofstudies/researchethics/guide-data-protection-research.pdf

[49] Raymond Chi-Wing Wong, Jiuyong Li, Ada Wai-Chee Fu, and Ke Wang. 2006. (α, k)-Anonymity: An Enhanced k-Anonymity Model for Privacy Preserving Data Publishing. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. Association for Computing Machinery, New York, NY, USA, 754–759. https://doi.org/10.1145/1150402.1150499

[50] Koji Yatani. 2016. Effect Sizes and Power Analysis in HCI. In *Modern Statistical Methods for HCI*, Judy Robertson and Maurits Kaptein (Eds.). Springer International Publishing, Cham, 87–110. https://doi.org/10.1007/978-3-319-26633-6_5