What Do We Do? Lessons Learned from Conducting Systematic Reviews to Improve HCI Dissemination

Zaidat Ibrahim
Indiana University Bloomington
Bloomington, Indiana, USA

Pegah Karimi
Indiana University–Purdue University Indianapolis
Indianapolis, Indiana, USA

Aqueasha Martin-Hammond
Indiana University–Purdue University Indianapolis
Indianapolis, Indiana, USA

Christina Harrington
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Katie Siek
Indiana University Bloomington
Bloomington, Indiana, USA

ABSTRACT

Systematic reviews are essential in helping researchers address pre-defined research questions through explicit, methodical, and reproducible techniques for identifying studies and comprehensively synthesizing their findings. We highlight our experiences conducting two systematic review studies in HCI: (1) women's reproductive health research in HCI and (2) the intersection of identity and older adults in health research. We identify patterns and lessons that can be applied to enhance the reporting and communication of our research. While these lessons may not be universally applicable, they provide HCI researchers with the opportunity for introspection regarding how we convey our findings to the broader research community. Additionally, these lessons contribute to upholding transparency and integrity in our work, rendering it more long-lasting and beneficial for secondary purposes, like literature reviews and study replication. We provide recommendations and, where feasible, good examples of how to effectively report participants' demographics and study methodology in our HCI work.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

HCI Research, Best Practices, Systematic Literature Reviews, Meta Analysis, Research Methods, Lessons Learned

ACM Reference Format:
Zaidat Ibrahim, Pegah Karimi, Aqueasha Martin-Hammond, Christina Harrington, and Katie Siek. 2024. What Do We Do? Lessons Learned from Conducting Systematic Reviews to Improve HCI Dissemination. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3613905.3637117

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05.
https://doi.org/10.1145/3613905.3637117

1 INTRODUCTION

Systematic reviews play an important role in enhancing researchers' comprehension of existing research landscapes, identifying interconnected research networks, specific problems being addressed, and knowledge gaps. Although HCI researchers create comprehensive related work sections with a solid foundation for addressing their research inquiries, systematic reviews offer a distinct opportunity to expand our typical related work sections by incorporating a broader spectrum of intra- and interdisciplinary work that extends beyond the boundaries of current citation clusters and networks. Systematic reviews published at CHI span from more local and introspective reviews about study sizes and compensation [3, 19, 24] to broadly defined reviews e.g. investigating trust in social media [34]. Stefanidi et al. [28] reported a 40% increase in systematic review publications since 2005; however, they cautioned that owing to the diverse nature of methodologies for conducting literature reviews, there is no consensus on shared reporting standards in HCI.

Leveraging Stefanidi et.al [28]'s work, we searched the ACM Digital Library for "Systematic Review" and filtered by "SIGCHI" for sponsor and "Research Article" for the content type. The search query yielded 340 results from 2009 through July 2023. We screened the titles and abstracts of these papers and found only 35 systematic review papers, as shown in Figure 1. The earliest systematic reviews were published in 2014 at CHI [17], EATIS [23] and ITS [25]. Similar to Stefanidi et al. [28]'s finding of a rising trend in the publication of literature review papers at CHI and DIS, we also noted that of our 35 systematic reviews, 17 were published at CHI, and 5 were published at DIS, constituting 63% of all systematic reviews published since 2014. This indicates a growing interest in systematic reviews within the SIGCHI community, with a steady increase.

Rogers et al. [22] argue that the growing volume of yearly publications makes it progressively challenging to stay current on the literature, thus underscoring the heightened necessity for systematic reviews. Rogers et al. [22] further highlight that *"the papers most cited are cited quite a bit more often than the average paper, while the number of citations papers received per year is declining overall [...] it is getting more difficult for new ideas to break through and shake up established ones in HCI."* [22]. Our case studies builds on these previous works by highlighting the obstacles we encountered as researchers in our systematic review work and contributing suggestions for best practices we could adopt as a research community.

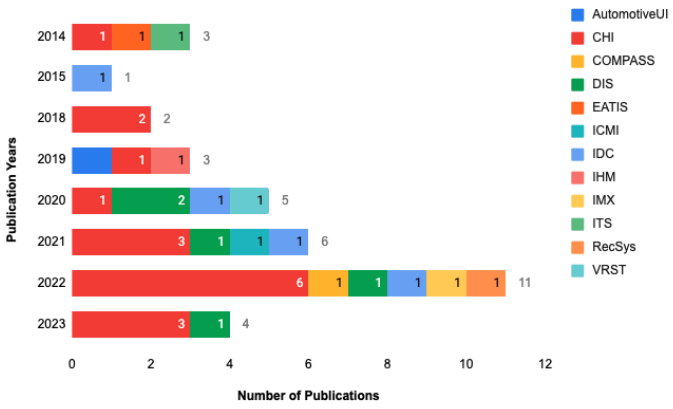

Figure 1: Systematic Reviews Published in SIGCHI Conferences 2014-2023

We do not aim to be prescriptive in our lessons learned; instead, we hope to spark collaborative conversations on how to report our primary research to better support science and future knowledge synthesis through literature reviews.

2 BACKGROUND ON SYSTEMATIC REVIEW

Systematic reviews were adopted from the field of medicine, particularly Evidence-Based Medicine (EBM), where it is considered a "gold standard" for enabling decision-making grounded in evidence [22]. The first formal semblance of research synthesis appeared in 1975 under the term 'meta-analysis' [4, 5]. Subsequently, Cochrane Collaboration–a network of researchers, clinicians, and other healthcare professionals was established and dedicated to producing systematic reviews of healthcare interventions [5, 27]. Ever since, other fields, including HCI, have embraced systematic reviews as a well-established method for synthesizing and summarizing research transparently and systematically. In 1999, the QUOROM Statement was established as a guideline for reporting meta-analyses of randomized controlled trials [20], which was updated to PRISMA [20] in 2009. Researchers use PRISMA flow charts, as shown in Figure 2, to communicate their search and screening processes. As systematic reviews continue to gain traction within the CHI community, it is crucial to acknowledge that this approach comes with several challenges: labor-intensive efforts and significant delays between completing a review to its publication date. In our case studies, presented in subsequent sections, we highlight the additional challenges we encountered.

2.1 Positionality and Acknowledgment

We have been formally trained in a small subset of disciplines - specifically computer science, interaction design, and public health. The HCI community typically publishes qualitative studies and mixed-method evaluations of the design and use of sociotechnical systems. We acknowledge that the HCI community is a multi-, inter-, and intra-disciplinary community with diverse epistemologies. Thus, we recognize that study designs, collaborations, participant context, community partnerships, legal and ethical considerations, and funding mechanisms impact how results are reported. Thus, in this case study, we do not mean to unduly critique studies after the fact, but instead use current reporting traditions as an example of the challenges encountered to synthesize and understand our research afterward.

3 CASE STUDIES

We provide a brief overview of our experiences conducting two systematic reviews on different corpora of published papers. In each case, 2-3 researchers worked for months to identify publications, review publications, and extract data for the systematic review. We were motivated to conduct a systematic review on HCI-oriented work in reproductive health because, in developing the related work section, we were struck by how little information was provided about the participants. For example, one's religion and cultural context would impact how one would use a reproductive health-oriented system (e.g., Muslim people cannot fast during menstruation, but are expected to track how many days they miss to make it up later). Thus, we set out to understand what information was reported about participants in reproductive health studies and what types of studies are done. We conducted a systematic review of full-research papers published in the ACM Digital Library between the years 2007 through 2023 to gain a thorough overview of HCI research studies on women's [1] reproductive health. Our overall goal was to identify, collect, and understand the direction of health-related HCI research, associated technology, and design efforts that directly engaged with persons who experience some aspects of the female reproductive cycle.

3.1 Case 1: Systematic Review on Women's Reproductive Health Research in HCI

The Search and Screen Process. We brainstormed and iterated on nine search terms that were related to the reproductive cycle[29]: *"women's health"*,*"puberty"*,*"fertility"*, *"menstruation"*,*"pregnancy"*, *"miscarriage"*, *"abortion"*,*"lactation"* and *"menopause."* . It is important to emphasize the researchers' efforts in refining search terms to ensure their relevance in capturing pertinent papers. Two researchers engaged in multiple iterations of the search terms before finalizing the selection of the nine search terms above. For example, we opted not to include the terms "period," "birth," and "conception" due to their ambiguity. "Period" brought papers referring to "time period"; "conception" yielded the "conception of an idea" [30] or "misconception" [32]; while "Birth" in a paper referenced "Birth Certificate" [26], or "Birth date" [16]. Instead of "period", we opted for "menstruation" as a suitable search term. We found that papers including "conception" and "birth" in the context of the reproductive cycle also encompassed terms such as "fertility" or "pregnancy."

Inclusion and Exclusion Criteria. For our inclusion criteria, we considered papers published between 2007 and 2023 in English that directly engaged with participants, discussed technologies for reproductive health, or contributed towards future designs of such technology. We excluded papers that were not peer-reviewed, did not directly engage with individuals experiencing the reproductive cycle, concentrated on technology for caregivers and healthcare professionals, or lacked direct engagement with the participants of interest.

[1]Women: Any individual assigned female at birth

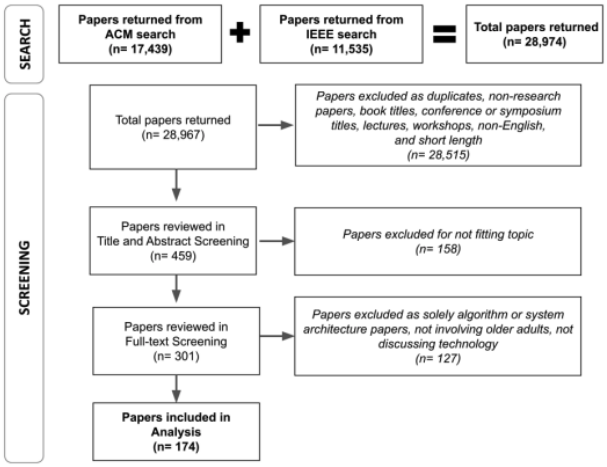

Figure 2: PRISMA Flow Chart From Case Study 2 [11]

Search Database. We used the ACM Digital Library for our review. We contacted the ACM team for permission to access downloadable proceedings from the digital library server. Upon gaining permission and access, we downloaded Extensible Markup Language (XML files) from the ACM servers. Our lab created a Python script to convert the XML into a curated Comma-Separated values (CSV) file based on search terms. The script returned a CSV output of papers that matched our search term criteria. Our research team then took on the burden of manually downloading papers from 2017 through 2023. We searched and considered papers from all 24 SIGCHI-sponsored conferences [2]. Upon review and filtering through our eligibility criteria, we finalized our corpus for the systematic review.

Challenges Encountered. In addition to creating our own Python script to help curate and iterate upon the publication corpus, we encountered two additional challenges: limited data accessibility and search precision issues. As of the summer of 2022, the most recent data available for download from the ACM server only covered publications up to 2017. This limitation restricts access to more recent publications and burdened our research team with manually downloading papers. The search feature on the ACM Digital Library can yield results that may not be directly related to the search term. Currently, there is no efficient way to filter search results, specifically by keywords present in the "title" and "abstract," leading to potential confusion. Our research team worked around it by conducting two rounds of screening of titles and abstracts and multiple rounds of discussions among the research team.

3.2 Case 2: Systematic Review on the Intersection of Identity and Older Adults in Health Research

The second case study focuses on a systematic review conducted to understand the identity characteristics of older adults engaged in pervasive and ubiquitous health technology design research. The review identified the characteristics of the older adults involved, the methods used, and the impact of including older adults with different identity characteristics on participation and design [11]. To conduct the review, the research team searched and reviewed full papers published in the ACM digital and IEEE Xplore digital libraries. The team chose these databases because they include computing literature emphasizing human-computer interaction and design.

3.2.1 The Search and Screen Process. The search included papers published between January 2008 and December 2020. The research team included four reviewers, two of whom were involved throughout the entire review process. The process consisted of two main phases: search and screening. During the search phase, the research team met several times to develop a common set of search terms to retrieve all potentially relevant papers for each of the two databases. After reaching a consensus on search terms, the team also met with the subject librarians at each institution to learn potential methods for retrieving metadata and corresponding papers from each library. The team used the available public-facing advanced query tools for each digital library to retrieve potentially relevant papers. Following the search phase, the research team conducted title and abstract screening and full-paper screening of papers to remove those that did not meet inclusion criteria. The full search and screen process is described in detail in the paper [11] and in Figure 2.

Inclusion and Exclusion Criteria. Inclusion criteria required papers to fit within the designated date range, focus on technology designs primarily used by older adults as opposed to for monitoring, diagnosis, or treatment, be published in English, and discuss a type of health technology. Exclusion criteria removed articles that were not full-research papers, focused on health or wellness, or studying the design of a technology used directly by older adults [11].

Search Database. For the IEEE Explore digital library, the research team downloaded search results from the public-facing advanced query tools in batches of 2000 results as comma-separated value (CSV) spreadsheets that included paper metadata (e.g., title, abstract, number of pages). IEEE allows the download of up to 2,000 papers' metadata at a time; therefore, the research team manually batches 2,000 articles to retrieve the total 11,535 papers returned by the IEEE query. ACM allowed for a more specific search, but there was no easy approach for getting the paper meta-data or PDFs through the public-facing interface. Because of this, the team created a workaround for downloading paper meta-data and PDF files using the returned query results. After completing the title and abstract screening for full-paper screening, the team manually downloaded papers from IEEE and ACM. Given the size of the dataset of papers (n=28,515) returned from the initial search, the team elected only to download PDFs of papers that made it through the title and abstract screening process.

Challenges Encountered. The research team encountered several challenges conducting the review. One of the team's first challenges was related to the differences in the advanced search options available between the two databases. Both databases provided an advanced query-builder function, allowing users to build queries with a form-like interface. However, IEEE offered a query command allowing users to enter the command directly and run it against the

[2]Full list of SIGHI-Sponsored Conferences can be found here: https://sigchi.org/conferences/conference-history/

database. While the team used the same query for both databases, they had to adjust the query to the available options in each form. The initial search query also returned more than 10,000 potentially relevant papers in both databases. However, the database limited the number of papers that could be downloaded in the CVS file containing metadata. Therefore, results needed to be downloaded in batches, requiring more effort to organize and combine search results. There was also no easy way of doing batch downloads of PDFs of full papers needed for the full paper screening. Therefore, the team downloaded the PDFs manually. The team later found out that many of the papers in the initial search did not fit one or more of the inclusion criteria; however, one of the reasons for the large number of papers was due to keywords such as *"aging"* or *"senior,"* which could have more than one meaning.

While the CHI community awaits updates from ACM regarding data curation, there is an opportunity for researchers to come together and focus on adopting best practices for presenting their research and findings. By collectively improving how research is presented and reported, researchers can enhance the transparency and accessibility of their work, mitigating some of the challenges associated with systematic reviews in HCI.

4 LESSONS LEARNED

Through our systematic review processes, we encountered papers where we had to carefully read and discuss research articles at length. While we found some ways of reporting results to be challenging during screening, we also encountered papers that effectively conveyed the findings of their research. We acknowledge that previously, some SIGCHI-sponsored publication venues had paper length limits for submission (CSCW removed page limits in 2013 and CHI removed page limits in 2021) [10], therefore limiting researchers from reporting with as much clarity. In light of these observations, our lessons learned address the topic of reporting HCI findings. We offer insights into best practices for reporting: where possible, we showcase instances of exemplary reporting, identify areas where improvements can be made and considerations for how we can enhance communication of our findings. We structured our lessons learned in three broad categories: search process & keywords, participants' demographics, and study methodology. We present a summary of our recommendations in Table 1.

4.1 Search Process and Keywords

Most systematic reviews begin with the keyword and search process. Refining the search process is very important. Search query yields a mixture of relevant and irrelevant papers mostly because of how a word is used. For example, "Conception" means different things in design papers versus reproductive health papers. Certain terms can have multiple meanings depending on the community, even within the computing discipline. For example, in HCI, the term "aging" often refers to research focused on design for older adults or those 60-65 years of age or older. Yet, as we found in one of our case studies, within other computing disciplines, such as those focused on energy, power systems, or smart grid, "aging" often refers to battery degradation. In Case Study 1, we removed "birth" from our search term. While this helped filter a larger chunk of irrelevant papers using "birth" for other than the reproductive aspects, the limitation is that we could have potentially missed out on any paper that had "birth" but never referred to other stages of the reproductive cycle, e.g., pregnancy or fertility. On the other hand, in Case Study 2, the team elected to keep the terms "aging" and "senior," resulting in a large dataset of papers to screen.
Recommendation: The search process could employ tools for Word Sense Disambiguation (WSD) to decipher meanings for each word researchers intend to include in the systematic reviews. Authors should reflect on the significance of the keywords we incorporate within our paper's *"Keywords"* section. If possible, seek feedback from other researchers to help select unique, community-supported search terms. For example, researchers could review past publications in the area and suggest future keywords in systematic reviews, workshops, or special interest groups. In the long term, this effort holds promise for supporting ACM's Digital Library, particularly if the ACM chooses to integrate keyword-based filtering into its search functionality. We acknowledge that the regular language approach afforded by simple keyword matching has limitations. Nevertheless, keyword matching is immediately familiar to almost any person. More complex approaches: context-free grammars, semantic parsing, machine learning, or the entire stack of natural language processing approaches could be powerful pieces to explore for the researcher interested in advanced approaches.

4.2 Participant Demographics

Here, we discuss how participant demographics are conveyed, comprising participant groups, sample size, and demographic characteristics, such as age and gender.

4.2.1 Who are the participants? Occasionally, the composition of the participant pool is not easily apparent. For example, a study abstract and introduction may have indicated that it exclusively involved older adults in its participant count. Still, upon closer examination of the paper, it became evident that the study included two distinct participant categories: older adults and their caregivers, both giving feedback on a design. In these cases, both participant categories were documented as participants in the review synthesis. We observed that papers that identified participants often made clear references in the method sections and described the role of each type of participant. For example, Foong et al.[9] include a section in the Methods titled "Client and Volunteer Participants," signaling that both were participants in their study. Within this section, they discuss inclusion criteria and demographics for both types of participants to explain who their participants were.
Recommendation: Convey the study's participants, including multiple participant groups, and specify which participant groups have been analyzed and reported in paper abstracts and introduction. In the methods section, thoroughly and comprehensively answer the questions: *"Who are participants in the study?"* and "Which participant group(s) are we reporting in this paper?"

4.2.2 Sample Size. Sometimes, the study's participant sample size appeared unclear or ambiguous. For instance, studies using focus groups as a methodology sometimes indicate ranges of participant numbers (e.g., 4-6 participants in each of 4 focus groups), leading to ambiguity in the total number of participants (e.g., it could be

Table 1: Summary Table: Recommendations For Improving Our HCI Reporting

When should researchers consider?	Recommendation
When selecting keywords for publication	**Reflect on the words used in the keywords section of the paper** • Iterate on keywords to ensure that keywords are not ambiguous and that they properly communicate the full research study • Look at similar papers to identify keywords used • Coordinate with researchers in the area to identify keywords
Study design and publication writing	**Provide comprehensive information on participants demographics** • Clearly identify who are participants; if multiple participant groups are included, identify which groups and how many are in each • Account for how many participants started and participated in each method. If participant data was cleaned or data were not included in the analysis, provide a justification • Exercise caution when reporting age-related and gender-related data, especially in multi-study papers with diverse populations **Enhance the clarity of study methods** • Employ visuals to enhance the presentation of a complex, multi-phase study • Specify the location (e.g., country, state, province) in which the study was conducted • Provide details on ethics board approvals • Provide transparent and clear information about participant compensation (e.g., how much; mechanism for compensation) • Specify study durations: (i) overall study timeframe, (ii) participant engagement period, and (iii) individual phase durations in multi-part studies

between 16-24 participants). Since the HCI community has not defined systematic review processes yet, even between the two case studies, we handled participant ranges differently. In case study 1, researchers took the average of the min and max range (e.g., leading to an estimate of 20 participants) as the number of participants for data synthesis. Whereas researchers in case study 2 filtered out studies that provided a range for their data synthesis reporting. Through our research, we observed good examples of reporting participant sample sizes in Wilcox et al. [33] and Iriguchi et al. [12]. Both papers offered transparency regarding age distribution, demographic composition, and gender diversity. More recent publications, such as Wilcox et al. [33], include the participant sample size in their abstract– *"We present a cross-cultural diary study with 64 transgender(trans) and non-binary adults in..."* therefore enhancing the clarity of participant information in the study. Iriguchi et al. [12] took their reporting a step further by specifying the number of participants enrolled and highlighting how many were included in their findings. For instance, a research study may enroll 50 participants, but only 40 completed the study and thus were analyzed and reported. Iriguchi et al. [12] reports as follows *"Fifty-nine female participants participated in the experiments. We analyzed the data of 23 pre-menopausal and 20 post-menopausal participants. [...]. We excluded 9 participants in perimenopause, four whose menstrual status was unknown, and three who had had a hysterectomy."*
Recommendation: Explicitly state the total number of participants, including those involved in focus groups. Then, highlight the number of participants whose data were analyzed and reported and state the reasons behind excluding specific participants' data (if any) from both analysis and reporting.

4.2.3 Demographic Details on Age and Gender Distribution. When reporting participant demographics, particularly regarding age and gender distribution, we encounter specific challenges stemming from the methods employed in data collection. To illustrate, age data is frequently gathered using age ranges. Although this practice does not present an immediate issue, it becomes more intricate when a study involves multi-study populations, each with its own set of age ranges. For example, in Tuli et al. [31], various research methods, including online surveys, semi-structured interviews, and focus groups, were utilized with distinct groups such as young adults, parents, teachers, social workers, and health professionals. Another instance is seen in the study conducted by Kilias et al. [13], which focused on pregnant people where three different research methods were employed: an exploratory workshop, interviews, and the Prenatal Yoga Multi-Sensorial Environment. While the age range was reported for the exploratory workshop, we wished we knew the age of participants within the other two methods. We also observed that gender distribution is often overlooked in women's health research. This omission may be attributed, in part, to the predominant focus of many researchers on cisgender women when investigating reproductive health issues. Although researchers such as Pyle et. al [21] are beginning to fill this gap.

Similarly, in research involving older adults, differences in reporting age range can be partly due to differences in how "older adult" or "senior" are defined in different parts of the world; however,

definitions may not be included in the paper. For example, one multi-country study aimed to develop a mobile interface to support older adults with visual disabilities with medication [8]. While the interview participants were generally described as older adults at least 50 or older, some interviewees were what some may consider younger adults (34 years old). We would have liked to know why people younger than 50 were included or how many people under 50 were included. We recognize that part of the challenge is that papers involving multi-part studies were most likely done throughout a researcher's dissertation or research agenda, where researchers iteratively improve upon their study design based on what they learn and hone research, writing, and reporting skills.

Recommendation: We encourage extra diligence when reporting age-related information, particularly in cases involving multiple studies and diverse populations within a single paper. In instances where age-related data collection may have been sub-optimal in certain aspects of the multi-part studies, we encourage research to highlight this as a limitation. In scenarios where multiple studies were conducted, and age data was collected, it is essential to report these age data for all the studies. As studies increasingly encompass a broader range of gender identities, we also urge greater care in reporting this information, emphasizing the need for a clear, comprehensive, and transparent presentation of both age and gender-related data.

4.3 Study Methodology

In more recent publications, we observed many papers conducting multi-part studies or incorporating multiple methods to address their research questions. Our investigation of study methodology concentrates on three aspects: the reporting of multi-part studies, disclosure of the study location or population density, and the ethical considerations surrounding compensation. We explore each of these topics in the subsequent subsections.

4.3.1 Large Multi-Part Studies. As an increasing number of studies embrace multiple methods and phases in their research endeavors, it becomes challenging and complex to follow each part and sometimes makes it difficult to understand what was done, when, and why. In the process of conducting Case Study 1 for multi-part papers, we employed a collaborative approach by having multiple researchers read and engage in discussions about each section. We visualized this process on a whiteboard, facilitating a comprehensive understanding of the chronology and intricacies within each phase of multi-part studies.

We observed good examples of reporting in Bolesnikov et al. [1], in their study exploring the intersection of wearable technology and queer expression practices. The researchers employed two distinct methods across two sequential phases of their research and effectively conveyed the comprehensive study methodology through the use of visual aids (see Figure 3 of [1]). The study methodology diagram included participant numbers, what activities occurred within phases 1 and 2 of the study, and what events happened simultaneously. This visual aid not only enhances readers' comprehension but also has the potential to support the replicability of their approach in other research.

Recommendation: When reporting a large multi-part study, it's worth considering the use of visual aids, pictorials, or illustrations to enhance the methods section. We recognize that this might be constrained if page length is limited.

4.3.2 Study Location. Study location comprises the geographical location (country) where the research was carried out or where participants were observed and the population density of the specific area or locality from which the participants were drawn. While this information may not directly impact the primary objectives of the research, it holds considerable value for other researchers who may require the paper for systematic or scoping reviews, where access to additional data and context is vital for a comprehensive understanding of the study's relevance and applicability. Furthermore, this information gives us insight into the context and resources available to participants - which can impact their lived experiences and interactions with sociotechnical systems.

A promising trend we saw in the systematic reviews was researchers conducting studies with participants in more than one country. A WEIRD [15] example is from Kresney et al. [14] where they borrowed from the medical informatics community [2] and created a "Table 1" that differentiated the demographics based on two countries - the United States and the United Kingdom. One challenge with these cross-country comparative studies is the granularity is still large; thus, the context and resources available to participants are not as specific.

Understanding one's context and location is a particular challenge when looking at studies that rely on Twitter/𝕏 data (e.g., [6, 7]). We recognize that Twitter/𝕏 location data may not have been scraped or available if the users turned off their location data. This issue raised a crucial point for discussion within our systematic review teams. Ultimately, we created a new label ("online") for study location if information was not provided.

Recommendation: In our role as researchers, we advocate for transparent reporting of study locations, including details about the regions and countries involved. When describing the study methodology, researchers can adopt a systematic review approach by posing essential questions such as, "How can we comprehensively report the study's location?" In cases where studies utilize scraped data from platforms like Twitter/𝕏, the research team should engage in discussions about how to present Twitter/𝕏 data in a manner that supports ad hoc studies while prioritizing participant privacy and safety.

4.3.3 Study Duration. While many papers report the duration of studies, we sometimes stumbled upon research papers that report duration in a more qualitative manner (e.g., "[participants] performed self-monitoring activities during a specific period of time (from weeks up to months) until childbirth"). We acknowledge that researchers report their findings based on their own study goals and contributions; thus, in these more qualitative reportings, their goals may not have been specifically on the activity or relied on time commitments. We note, however, that without this type of data it makes literature synthesis and reproducibility challenging. In these cases, we simply reported duration as "unspecified," as we were not certain how many weeks or months. We appreciated Nurain et al. [18]'s use of visuals in a research study exploring older adults' tracking practices to communicate the study duration. The researchers included a study overview diagram highlighting (see Figure 1 in [18]) the duration of each part of the study.

Recommendation: In reporting duration, we encourage researchers to consider the different aspects that make up the study duration– the duration from start to finish of the entire study, the duration/time involvement in which participants engaged with the study, and in the case of a multi-part study, the duration of each phase of the study. We encourage comprehensive information on duration. In fact, this could also be captured within the visual aid we suggested in section 4.3.1, where we highlighted incorporating a visual aid to support study methodology.

4.3.4 Compensation and Ethics Board Approval. In our review, we encountered papers that lacked details concerning ethical board approvals and the status of participant compensation. In some cases where compensation was mentioned, the specific form or method of compensation remained unclear. For our systematic review synthesis and analysis in Case Study 1, in computing compensation methods for studies, we categorized the compensation methods using compensation vehicles recommended by Pater et al. [19].
Recommendation: We encourage researchers to specify the information on ethics board approvals, the status of participant compensation, and methods or forms of compensation. We also recommend adopting guidelines proposed in Pater et al. [19]

5 CONCLUSION

Systematic reviews provide researchers with the ability to synthesize literature in a specific area to identify current understandings and knowledge gaps. Human-computer Interaction researchers have increasingly been using systematic reviews to examine our own practices and subcommunities. In our two case studies, we highlight challenges we encountered trying to conduct systematic reviews. Based on our experiences, we provide 8 recommendations to the CHI community, encompassing improvements on how we report on study demographics, methodology, location, duration, compensation, ethics board review, and categorization by keywords.

REFERENCES

[1] Adrian Bolesnikov, Karen Anne Cochrane, and Audrey Girouard. 2023. Wearable Identities: Understanding Wearables' Potential for Supporting the Expression of Queer Identities. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 393, 19 pages. https://doi.org/10.1145/3544548.3581327

[2] J Brender, Jan Talmon, Elske Ammenwerth, Jytte Brender, Nicolette de Keizer, Pirkko Nykänen, Michael Rigby, and E. Ammenwerth. 2013. STARE-HI–Statement on reporting of evaluation studies in Health Informatics. *Applied Clinical Informatics* 4 (01 2013), 331–358. https://doi.org/10.4338/ACI-2013-04-RA-0024

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

[4] Iain Chalmers, Larry V Hedges, and Harris Cooper. 2002. A brief history of research synthesis. *Evaluation & the health professions* 25, 1 (2002), 12–37.

[5] Mike Clarke. 2016. History of evidence synthesis to assess treatment effects: Personal reflections on something that is very much alive. *Journal of the Royal Society of Medicine* 109, 4 (2016), 154–163.

[6] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Major Life Changes and Behavioral Markers in Social Media: Case of Childbirth. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1431–1442. https://doi.org/10.1145/2441776.2441937

[7] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting Postpartum Changes in Emotion and Behavior via Social Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3267–3276. https://doi.org/10.1145/2470654.2466447

[8] Mari Ervasti, Minna Isomursu, and Igone Idigoras Leibar. 2011. Touch- and audio-based medication management service concept for vision impaired older people. In *2011 IEEE International Conference on RFID-Technologies and Applications*. IEEE, 244–251. https://doi.org/10.1109/RFID-TA.2011.6068645

[9] Pin Sym Foong, Shengdong Zhao, Felicia Tan, and Joseph Jay Williams. 2018. Harvesting Caregiving Knowledge: Design Considerations for Integrating Volunteer Input in Dementia Care. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18, Paper 79)*. Association for Computing Machinery, New York, NY, USA, 1–12.

[10] R Stuart Geiger. 2019. The Rise and Fall of the Note: Changing Paper Lengths in ACM CSCW, 2000-2018. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 222 (nov 2019), 10 pages. https://doi.org/10.1145/3359324

[11] Christina Harrington, Aqueasha Martin-Hammond, and Kirsten E Bray. 2022. Examining Identity as a Variable of Health Technology Research for Older Adults: A Systematic Review. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 265, 24 pages. https://doi.org/10.1145/3491102.3517621

[12] Mayuko Iriguchi, Hiroki Koda, and Nobuo Masataka. 2018. Colour Perception Characteristics of Women in Menopause. In *Proceedings of the 2018 International Joint Workshop on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia* (Yokohama, Japan) *(MMArt&ACM'18)*. Association for Computing Machinery, New York, NY, USA, 20–25. https://doi.org/10.1145/3209693.3209694

[13] Ida Kilias Svenningsen and Teresa Almeida. 2021. Embodying the Emotional Pregnancy: Design Explorations of a Prenatal Yoga Multi-Sensorial Environment. In *Proceedings of the 32nd Australian Conference on Human-Computer Interaction* (Sydney, NSW, Australia) *(OzCHI '20)*. Association for Computing Machinery, New York, NY, USA, 189–197. https://doi.org/10.1145/3441000.3441044

[14] K. Cassie Kresnye, Mona Y. Alqassim, Briana Hollins, Lucia Guerra-Reyes, Maria K. Wolters, and Katie A. Siek. 2021. What to Expect When You Are No Longer Expecting: Information Needs of Women Who Experienced a Miscarriage. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare* (Atlanta, GA, USA) *(PervasiveHealth '20)*. Association for Computing Machinery, New York, NY, USA, 85–96. https://doi.org/10.1145/3421937.3421995

[15] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 143, 14 pages. https://doi.org/10.1145/3411764.3445488

[16] Pamela J. Ludford, Reid Priedhorsky, Ken Reily, and Loren Terveen. 2007. Capturing, Sharing, and Using Local Place Information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1235–1244. https://doi.org/10.1145/1240624.1240811

[17] Elisa D. Mekler, Julia Ayumi Bopp, Alexandre N. Tuch, and Klaus Opwis. 2014. A Systematic Review of Quantitative Studies on the Enjoyment of Digital Entertainment Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 927–936. https://doi.org/10.1145/2556288.2557078

[18] Novia Nurain and Chia-Fang Chung. 2023. "I Left My Legacy, Told My Story": Understanding Older Adults' Tracking Practices to Promote Active Aging. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 459–475. https://doi.org/10.1145/3563657.3596083

[19] 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 141, 16 pages. https://doi.org/10.1145/3411764.3445734

[20] PRISMA. 2023. History & Development of PRISMA. Retrieved October 3, 2023 from http://prisma-statement.org/PRISMAStatement/HistoryAndDevelopment

[21] Cassidy Pyle, Lee Roosevelt, Ashley Lacombe-Duncan, and Nazanin Andalibi. 2021. LGBTQ Persons' Pregnancy Loss Disclosures to Known Ties on Social Media: Disclosure Decisions and Ideal Disclosure Environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[22] Katja Rogers and Katie Seaborn. 2023. The Systematic Review-Lution: A Manifesto to Promote Rigour and Inclusivity in Research Synthesis. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 424, 11 pages. https://doi.org/10.1145/3544549.3582733

[23] Carolina Salvador, Arturo Nakasone, and Jose Antonio Pow-Sang. 2014. A Systematic Review of Usability Techniques in Agile Methodologies. In *Proceedings of the 7th Euro American Conference on Telematics and Information Systems* (Valparaiso, Chile) *(EATIS '14)*. Association for Computing Machinery, New York, NY,

USA, Article 17, 6 pages. https://doi.org/10.1145/2590651.2590668

[24] Shruti Sannon and Andrea Forte. 2022. Privacy Research with Marginalized Groups: What We Know, What's Needed, and What's Next. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 455 (nov 2022), 33 pages. https://doi.org/10.1145/3555556

[25] Zahra Shakeri Hossein Abad, Craig Anslow, and Frank Maurer. 2014. Multi Surface Interactions with Geospatial Data: A Systematic Review. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces* (Dresden, Germany) *(ITS '14)*. Association for Computing Machinery, New York, NY, USA, 69–78. https://doi.org/10.1145/2669485.2669505

[26] Supriya Singh, Anuja Cabraal, Catherine Demosthenous, Gunela Astbrink, and Michele Furlong. 2007. Password Sharing: Implications for Security Design Based on Social Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 895–904. https://doi.org/10.1145/1240624.1240759

[27] Mark Starr, Iain Chalmers, Mike Clarke, and Andrew D Oxman. 2009. The origins, evolution, and future of The Cochrane Database of Systematic Reviews. *International journal of technology assessment in health care* 25, S1 (2009), 182–195.

[28] Evropi Stefanidi, Marit Bentvelzen, Paweł W. Woźniak, Thomas Kosch, Mikołaj P. Woźniak, Thomas Mildner, Stefan Schneegass, Heiko Müller, and Jasmin Niess. 2023. Literature Reviews in HCI: A Review of Reviews. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 509, 24 pages. https://doi.org/10.1145/3544548.3581332

[29] Yuko Takeda. 2010. Understanding the Life Stages of Women to Enhance Your Practice. *Japan Medical Association Journal* 53 (09 2010).

[30] Hao Tan, Yi Zhu, and Jianghong Zhao. 2012. Development of an Automotive User Interface Design Knowledge System. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Portsmouth, New Hampshire) *(AutomotiveUI '12)*. Association for Computing Machinery, New York, NY, USA, 201–208. https://doi.org/10.1145/2390256.2390291

[31] Anupriya Tuli, Shruti Dalvi, Neha Kumar, and Pushpendra Singh. 2019. "It's a Girl Thing": Examining Challenges and Opportunities around Menstrual Health Education in India. *ACM Trans. Comput.-Hum. Interact.* 26, 5, Article 29 (jul 2019), 24 pages. https://doi.org/10.1145/3325282

[32] Joe Tullio, Anind K. Dey, Jason Chalecki, and James Fogarty. 2007. How It Works: A Field Study of Non-Technical Users Interacting with an Intelligent System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 31–40. https://doi.org/10.1145/1240624.1240630

[33] Lauren Wilcox, Renee Shelby, Rajesh Veeraraghavan, Oliver L. Haimson, Gabriela Cruz Erickson, Michael Turken, and Rebecca Gulotta. 2023. Infrastructuring Care: How Trans and Non-Binary People Meet Health and Well-Being Needs through Technology. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 489, 17 pages. https://doi.org/10.1145/3544548.3581040

[34] Yixuan Zhang, Joseph D Gaggiano, Nutchanon Yongsatianchot, Nurul M Suhaimi, Miso Kim, Yifan Sun, Jacqueline Griffin, and Andrea G Parker. 2023. What Do We Mean When We Talk about Trust in Social Media? A Systematic Review. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 670, 22 pages. https://doi.org/10.1145/3544548.3581019