Moving Transparent Statistics Forward at CHI

Matthew Kay, Steve Haroz, Shion Guha, Pierre Dragicevic, Chat Wacharamanotham

Moving Transparent Statistics Forward at CHI

Matthew Kay

University of Michigan

Ann Arbor, MI, USA

mjskay@umich.edu

Steve Haroz

Northwestern University

Chicago, IL, USA

stats@steveharoz.com

Shion Guha

Marquette University

Milwaukee, WI, USA

shion.guha@marquette.edu

Pierre Dragicevic

Inria

Orsay, France

pierre.dragicevic@inria.fr

Chat Wacharamanotham

University of Zurich

Zurich, Switzerland

chat@ifi.uzh.ch

Abstract

Transparent statistics is a philosophy of statistical reporting whose purpose is scientific advancement rather than persuasion. We ran a SIG at CHI 2016 to discuss problems and limitations in statistical practices in HCI and options for moving the field towards clearer and more reliable ways of writing about experiments, and received an overwhelming response. This SIG resulted in rough drafts of reviewer guidelines, resources for authors, and other suggestions for advancing a vision of *transparent statistics* within the field; this year, we propose a concentrated one-day writing workshop to develop those documents into a polished state with input from a diverse cross-section of the CHI community.

Author Keywords

Statistics; methodology; user studies; quantitative methods; transparent statistics.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI)

Background

Human-computer interaction (HCI) is a large, multidisciplinary field drawing on a variety of statistical approaches. However, many of our existing practices have drawn increasing criticism, such as an

overreliance on null-hypothesis significance testing (NHST), a lack of replication and meta-analysis, and few studies published with data or study materials. These practices have been criticized within both HCI [7,3,6,9] and related fields [13,4,14], and have even reached the popular press in coverage of the *replication crisis* in social science.

The purpose of this workshop is to build upon the enthusiasm at the CHI 2016 Special Interest Group on Transparent Statistics [10] in order to develop—*at the workshop*—a set of guidelines and recommendations to help improve the transparency and quality of statistical reports in the field.

The ethos of the workshop will not be to admonish or shame researchers for their existing practices. The multifaceted nature of HCI means that we will always need to remain open to a range of practices, and a fixed set of DOs and DON'Ts would be both too brittle to change over time and too restrictive in the face of the various ways of generating knowledge in our field. Instead, we propose to advance a vision of *transparent statistical communication* for the field of HCI. We refer to *transparent statistics* as a philosophy of statistical reporting whose purpose is to advance scientific knowledge rather than to persuade. Whatever the methods used, we (the workshop participants) can provide guidance on how to make the communication of empirical findings more transparent, how to facilitate reproduction and replication of work, and how to make evaluation of work (e.g., by peer reviewers) easier and more fair. Hand in hand with improving authors' statistical practice, we also aim to provide clearer guidelines for HCI reviewers on how to fairly evaluate statistical claims.

To that end, workshop participants will develop specific recommendations and artifacts---outlined below---to help bring about change within the HCI community. We will develop these artifacts with a *more is more* philosophy: what can authors do to improve the transparency of their communication? What can a reviewer do to encourage transparency? What changes to the review process might encourage transparency and incentivize researchers? In this way we hope to avoid the time-honored tradition of admonishing researchers for doing statistics poorly, and instead encourage them—and guide them—to do better.

From the enthusiasm generated at the CHI 2016 Special Interest Group on Transparent Statistics, we have already signed up 80+ members to a transparent statistics in HCI mailing list [11] and have begun drafting a set of documents:

1. **Reviewer guidelines** giving high-level guidance and strategies for reviewing statistical reports in CHI submissions, and a more technical **FAQ** answering specific questions about which practices should be encouraged.

2. **Resources** for authors, such as exemplary papers whose statistical methods promote transparent communication.

3. **Suggestions** for practical changes to move the field forward. For the workshop we will provide initial suggestions (intended to seed brainstorming, not to be exhaustive), such as:

- Changes/additions to CHI reviewing processes, like the inclusion of badges for papers with open data, materials, or preregistered material. These could be

patterned after (or even use) badges from the Open Science Framework (OSF) [1]

- Voluntary pre-registration of analyses for papers
- MOOCs or other courses for authors and reviewers
- Suggested graduate curricula for HCI PhDs
- Adding new metadata in PCS, such as experimental meta-data similar to that collected manually by Caine [2] to facilitate tracking changes in statistical practice in the field over time.

As discussed further below in the section "Post-Workshop Plans", these documents are meant to be shared publicly once finalized and we will engage the CHI Executive Committee and/or Program Chairs to discuss the feasibility of the suggested changes to the review process, and the possibility of incorporating some of the recommendations into official documents like the SIGCHI reviewer guidelines.

While we have been encouraged by discussion on the mailing list, distributed collaboration presents challenges to engaging the full community in developing and completing the artifacts. While some discussion has progressed on our mailing list, we believe that a collocated writing process will be both more efficient, more likely to engage multiple perspectives through discussion, and better able to capture disagreement and controversy (which is important for a diverse community like CHI).

Our existing documents are in an outline/draft stage (http://tinyurl.com/transparent-stats-docs), and before the work becomes too "set" in form or content, we wish to engage the community more directly in drafting the artifacts. As the purpose of the workshop will be to flesh out these artifacts, the above list is not intended to be exhaustive and the current document drafts are by no means final.

In other words, we are proposing a *working* workshop: to bring together a diverse set of perspectives from the HCI community in order to flesh out and complete the above documents in an intense one-day burst of collaborative writing and discussion. We believe that with the groundwork already laid from our previous SIG and discussions on the mailing list, we will be well-positioned to produce several documents reflecting a strong path forward for more transparent statistical communication in the field. We want these documents to reflect the diversity of practices at CHI. Further, the response to last year's SIG (which was standing-room only) and on our mailing list suggests that there is enough interest from the community to drive a successful workshop.

Organizers

The organizers of this workshop also organized the CHI 2016 Special Interest Group on Transparent Statistics last year [10,11].

Matthew Kay is an Assistant Professor at the University of Michigan School of Information. He studies the design of user-facing uncertainty in everyday sensing and prediction, such as personal informatics systems for health and applications for real-time transit prediction. He has also published work advancing the use of Bayesian statistics in VIS [8] and CHI [9]. His website is: http://www.mjskay.com.

Steve Haroz is a postdoctoral research fellow in the Psychology Department at Northwestern University. He

researches how the brain perceives and understands visually displayed information, and he has experience with the experiment design and statistical practices in both computer science and psychology. Steve also maintains a list of InfoVis publications which include statistically analyzed quantitative experiments: http://steveh.co/experiments

Shion Guha is an Assistant Professor in the Department of Mathematics, Statistics and Computer Science at Marquette University. He studies contagion of social processes in social networks, particularly privacy and more recently, algorithmic accountability, transparency and harm. He has recently published methodological papers of interest to the HCI community in GROUP'16, JASIST, and Social Media + Society. His website is: http://www.shionguha.net/

Pierre Dragicevic is a permanent research scientist at Inria since 2007, and studies information visualization (InfoVis) and HCI. He is interested in reforming statistical practice in these fields, with a focus on replacing dichotomous testing with estimation thinking. He gives regular talks (e.g., at the BELIV 2014 biannual workshop) and publishes papers [5,6] on the topic. He also maintains a Web page with reading material: http://www.aviz.fr/badstats

Chat Wacharamanotham is an Assistant Professor at the University of Zurich. He studies how scientists use statistics, both in conducting statistical analysis and in consuming statistical reports. His previous study shows that students learn statistical procedures better through a guided interactive analysis tool [15]. He can be found online at: http://zpac.ch/chat

Website
From last year's SIG, we have a website (http://transparentstatistics.org/chi2016), mailing list / Google group (https://groups.google.com/forum/#!forum/transparent-stats-hci), and a Google drive with draft documents (http://tinyurl.com/transparent-stats-docs). For the workshop, we will update the SIG website to include the workshop call for participation and instructions for how participants can prepare in advance of the workshop.

Pre-Workshop Plans
We will recruit participants through our transparent statistics Google group and mailing list (80+members), the CHI-Meta Facebook group (900 members, https://www.facebook.com/groups/834637469921428/), and various departmental mailing lists. Since we have already begun to build a community around this topic, we are confident we can attract a motivated and knowledgeable set of participants.

In addition, one of our organizers, Shion Guha, has a background in statistics and connections within that community. He has already reached out to statisticians who are interested in providing input at the workshop.

New participants not already in the Google group will be expected to join the group and review the draft documents before the workshop. They will also be asked to state which artifacts they are most interested in so that we can ensure a good balance of people at the workshop interested in working on each artifact. Because we already have draft documents in Google docs, we will also encourage participants to read through the documents (and even begin editing) in

advance of the workshop so that we can make the most of our collocated time at the conference.

Workshop Structure

The main purpose of the workshop will be to finalize the artifacts (guidelines, recommendations and lists of suggestions) started by the *transparent statistics* group. To facilitate collaborative writing, for the majority of the workshop we will break participants into 3-5 writing groups of ~3-5 people each to work on specific artifacts in separate Google documents.

The structure will be as follows:

- 20 minutes. Introduction to workshop, overview of goals and structure of workshop.
- 2 hours. Based on participants' expertise and interests (which we will balance in advance using position statements), participants will break into small groups to work on specific artifacts (e.g. a specific document or section of a document). Groups will have an organizer designated to suggest deliverables and to record ideas and controversies within each group that should be raised to full-group discussion.
- 45 minutes. Report back from groups, with a focus on topics that require broader discussion. Where differences cannot be resolved by consensus, we will focus on recording those differences in the documents in question---that is, acknowledging differences of opinion within the CHI community on best practice. In case of clearly diverging opinions on a specific question, *our reviewing guidelines will discourage CHI reviewers from rejecting a paper based on their own position concerning that question*.
- Lunch
- 2 hours. More group writing. Participants will be encouraged to rotate groups.
- 1 hour. Report back from groups (as above). Again focus on topics needing broader discussion, with an additional focus on steps needed to finalize artifacts.
- 30 minutes. Closing, discussion of concrete next steps and soliciting volunteers for particular tasks.

Post-Workshop Plans

As stated above, we plan a *working* workshop: complete (or near-complete) versions of the artifacts described in the Background will be our primary outcome. These will become artifacts available on the workshop's website and published with a DOI using an archive service like figshare (https://figshare.com/) to ensure long-term availability. Since we do not wish these guidelines to be static, we will also continue hosting and drafting future revisions of the guidelines on a collaborative editing platform, such as Google docs or a wiki. This will be especially important if not all interested CHI attendees are able to participate in the workshop due to space limits. We will also release these draft documents post-workshop to a round of public comment by posting to CHI-related mailing lists, the transparent statistics mailing list, and the CHI-Meta Facebook group.

To increase the visibility of the transparent statistics movement and of this work, we will also attempt to publish a report in a magazine that reaches a broader swath of the community, such as Interactions or Communications of the ACM.

Finally, since a few of our outcomes may involve recommendations of changes or augmentations to the review process, we will also engage the CHI Executive Committee and/or Program Chairs for future iterations of the conference to discuss the feasibility of changes to the review process, and the possibility of incorporating some of our artifacts into (or referencing those artifacts from) official documents like the SIGCHI reviewer guidelines. At least one of the CHI 2017 Subcommittee Chairs (Anne Roudaut) attended last year's SIG on transparent statistics and expressed a strong interest in the guidelines document.

As we are strongly aware of the feasibility of certain kinds of recommendations, we will aim in the workshop (and have already aimed in our initial discussions) to scope such recommendations to ideas that have a low barrier to implementation and which would not run strongly counter to CHI culture. For example, requiring pre-registration of study designs or requiring open data are recommendations that have been made in some fields that would not be feasible at CHI, at least on a short term; on the other hand, badge systems like the OSF badges [1] have been proposed in other fields, have been found to be effective at increasing transparency in published work [12], and could in principle be feasible at CHI.

Call for Participation

HCI is large and multidisciplinary, drawing on a variety of statistical practices. However, many of these existing practices have drawn increasing criticism within both HCI and related fields, including but not limited to: over-reliance on particular statistical methods, a lack of transparent reporting, a lack of replication and meta-analysis, few studies published with data or study

materials, and inadequate education in statistics. These issues have even reached the popular press in coverage of the *replication crisis* in social science.

We are running a *working* workshop to develop concrete guidelines for improving statistical practice in HCI. Participants will work in groups to flesh out guidelines for helping reviewers fairly assess statistical reports in CHI papers, concrete suggestions for changes to review processes, resources for authors, and other relevant guidelines or proposals the group wishes to advance. We will seed writing with draft guidelines developed in the wake of last years' SIG on Transparent Statistics in HCI (http://transparentstatistics.org/chi2016). However, we will not be constrained to the outline in these draft documents if discussions take us elsewhere.

These documents will be shared publicly once finalized and we will engage the CHI Executive Committee and/or Program Chairs to discuss the feasibility of suggested changes to the review process, and the possibility of incorporating some of the recommendations into official documents like the SIGCHI reviewer guidelines. We will also continue hosting and drafting future revisions of the documents on a collaborative editing platform.

We are looking for a diverse set of perspectives on quantitative methods in the HCI community to develop these documents (while qualitative methods are important to HCI, our primary focus in this workshop is on improving the use and communication of quantitative work). If you are interested in improving the state of statistical practice in HCI—whether or not you attended last year's SIG—submit a position

statement (at most 2 pages in CHI Extended Abstracts format) containing:

- A short Bio (If there are multiple authors, only include a bio for the one author who wishes to attend).
- A statement of special areas of interest (in HCI, methods, statistics and/or statistical reporting).
- A position statement on improving statistical communication or practice or a comment on our statement on transparent statistics (http://transparentstatistics.org/chi2016).
- An indication of which artifact(s) you are interested in contributing to from our draft documents (http://tinyurl.com/transparent-stats-docs), or a suggestion for another artifact you believe should be included at the workshop.

Submit to chi2017workshop@transparentstatistics.org. The final deadline for submission is **Jan 31, 2017**.

As much as we would like as many CHI researchers as possible to get involved in this initiative, we may not be able to accept all applicants due to limits in the number of attendees at CHI workshops. If this happens, we will select applications based on expertise, diversity of perspectives in quantitative methods, diversity of HCI domains, and diversity of interest in specific artifacts (to ensure we have people interested in working on each artifact). **Our goal is that the participants reflect the diversity in quantitative methods, statistics, and domains across HCI.**

The purpose of the position statement is to help us determine *specific* participants, so **we encourage one author per position statement**. Position statements with multiple authors should clearly indicate which author will attend. If your position statement is accepted, the attending author must register for both the workshop and at least one day of the conference.

References

1. Ben B. Blohowiak and others (Center for Open Science). 2013. Badges to Acknowledge Open Practices (Wiki). https://osf.io/tvyxz/wiki/home/

2. Kelly Caine. 2016. Local Standards for Sample Size at CHI. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16, ACM Press, 981–992. http://doi.org/10.1145/2858036.2858498

3. Paul Cairns. 2007. HCI... not as it should be: inferential statistics in HCI research. In People and Computers: HCI... but not as we know it, Vol. 1. 195–201.

4. Geoff Cumming. 2013. The new statistics why and how. Psychological science.

5. Pierre Dragicevic, Fanny Chevalier, and Stephane Huot. 2014. Running an HCI experiment in multiple parallel universes. In CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14). ACM, New York, NY, USA, 607-618.

6. Pierre Dragicevic. 2016. Fair Statistical Communication in HCI. In Modern Statistical Methods for HCI, J. Robertson and M.C. Kaptein (Eds.). Springer. http://tinyurl.com/fairstats-author

7. Maurits Kaptein and Judy Robertson. 2012. Rethinking statistical analysis methods for CHI. In CHI 2012.

8. Matthew Kay and Jeffrey Heer. 2016. Beyond Weber's Law: A Second Look at Ranking Visualizations of Correlation. IEEE Transactions on Visualization and Computer Graphics 22, 1: 469–478. http://doi.org/10.1109/TVCG.2015.2467671

9. Matthew Kay, Gregory Nelson, and Eric Hekler. 2016. Researcher-centered design of statistics: Why Bayesian statistics better fit the culture and incentives of HCI. In CHI 2016.

10. Matthew Kay, Steve Haroz, Shion Guha, and Pierre Dragicevic. 2016. Special Interest Group on Transparent Statistics in HCI. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16). ACM, New York, NY, USA, 1081-1084. http://transparentstatistics.org/chi2016/chi-2016-sig.pdf

11. Matthew Kay, Steve Haroz, Shion Guha, and Pierre Dragicevic. 2016. Special Interest Group on Transparent Statistics in HCI Web Page. http://transparentstatistics.org/chi2016

12. Mallory C. Kidwell, Ljiljana B. Lazarević, Erica Baranski, et al. 2016. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. PLoS Biology 14, 5: 1–15. http://doi.org/10.1371/journal.pbio.1002456

13. Rex B Kline, American Psychological Association, and others. 2004. Beyond significance testing: Reforming data analysis methods in behavioral research.

14. Open Science Collaboration and others. 2015. Estimating the reproducibility of psychological science. Science 349, 6251, aac4716.

15. Chat Wacharamanotham, Krishna Subramanian, Sarah Theres Völkel, and Jan Borchers. 2015. Statsplorer: Guiding Novices in Statistical Analysis. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15, ACM Press, 2693–2702. http://doi.org/10.1145/2702123.2702347