

## Making HCI statistics better

Dix, Alan

*License:*

None: All rights reserved

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

Dix, A 2017, Making HCI statistics better: saving the baby. in *Proceedings of the Moving Transparent Statistics Forward at CHI 2017*. Moving Transparent Statistics Forward Workshop at CHI 2017, Denver, United States, 6/05/17. <<https://transparentstatistics.org/chi2017/>>

[Link to publication on Research at Birmingham portal](#)

**Publisher Rights Statement:**

ACM allows author's manuscript archived openly on institutional repository. Eligibility for repository: Checked on 13/4/2017

**General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

**Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

---

# Making HCI statistics better: saving the baby

**Alan Dix**

HCI Centre  
School of Computer Science  
University of Birmingham  
Birmingham, B15 2TT, UK  
and  
Talis  
Birmingham, B1 3HN, UK

[alan@hcibook.com](mailto:alan@hcibook.com)

<http://alandix.com/academic/papers/chi2017-transparent-stats/>

**Abstract**

I have long had a concern over the state of statistics in HCI, so welcome a chance to be part of this workshop. I will have a special interest in advancing statistical literacy and sharing of data and meta-data.

**Author Keywords**

Statistics; human-computer interaction; evaluation; hypothesis testing; Bayesian statistics

**ACM Classification Keywords**

H.5.2. Information interfaces and presentation, User Interfaces: Evaluation/methodology, Theory and methods; G.3 PROBABILITY AND STATISTICS

**Bio**

Alan Dix is Professor at the University of Birmingham and Senior Researcher at Talis. He is well known for his textbook and research in HCI and is a member of SIGCHI Academy. However, before he was in HCI, Alan was a mathematician, including representing the UK in the International Mathematical Olympiad. In the past, he has practiced as a professional statistician and applied mathematician: including work on modelling agricultural crop sprays, medical statistics and undersea cable detection. Within HCI these skills have been applied in his foundational work on formal methods for interactive systems, the use of Bayesian

techniques in education, random sampling for visualisation of big data and uncertainty, and analysis of potential bias against human/applied areas in REF, the UK research assessment exercise. Alan is also running a course at CHI2017, "Making Sense of Statistics in HCI: From P to Bayes and Beyond".

### **Special areas of interest**

I have worked in many areas of HCI including CSCW, mobile interfaces, technical creativity, and some of the earliest work on privacy and ethical implications of intelligent data processing. More recent work includes community engagement especially in rural areas and my one thousand mile research walk around Wales, which generated substantial quantitative and qualitative open research data from blogs to biodata. In my Talis role I am interested on the use of technology in university-level education including learning analytics; that is where statistics, visualisation and big data techniques are applied within education.

### **Position Statement**

The state of knowledge and practice of statistics in HCI has been a personal concern for many years from the standard mistakes, such as treating non-significant as no effect; to more subtle issues, such as confounding the success of a prototype system with that of the principles on which it is built.

However, there is a real danger of 'throwing the baby out with the bathwater', and I am equally alarmed by some of the hype surrounding new statistics and careless application such as confidence intervals being quoted without reference to the underlying statistical distributions and sample sizes used to construct them,

and Bayesian statistics mixed with hypothesis testing in what appears to be cherry picking of techniques.

Part of the answer lies in better education (hence my statistics course at CHI this year and at CHI and other conferences some years ago), and part in better ways to structure and report methods and share data.

### **Potential Contribution**

Following on from this, and linking to my broader interests in data, I was particularly interested in the discussions about systems for sharing of data and meta-data mentioned in "Proposal for advancing transparent statistics in HCI", and also how to establish educational or author/reviewer guidance materials.

In previous writing [1] I have argued that the data gathered in HCI should be valued and reported in its own right from methodology of collection to documentation of data formats. Even if heavily anonymised or summarised, we should ideally have sufficient data published to allow alternative analyses. In particular, the availability of original data for meta-analysis would avoid the need for Bayesian summary statistics [2] or post-hoc combination of p-values.

### **References**

1. Dix, A. 2010. Human-Computer Interaction: a stable discipline, a nascent science, and the growth of the long tail. *Interacting with Computers*, 22(1) pp. 13-27. <http://www.alandix.com/academic/papers/IwC-LongFsch-HCI-2010/>.
2. Kay, M., Nelson, G., and Hekler, E. 2016. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. CHI 2016, ACM, pp. 4521-4532.