REMEMBERING STEPHEN FIENBERG

ALEKSANDRA B. SLAVKOVIĆ AND LARS VILHUBER

Guest Editor
Penn State University
e-mail address: sesa@stats.psu.edu

Managing Editor
Cornell University
e-mail address: lars.vilhuber@cornell.edu

ABSTRACT. Stephen Fienberg passed away in December of 2016. Among many, many other things, he was a co-founder of the Journal of Privacy and Confidentiality. This special issue of the Journal brings together many of his former collaborators, doing what Stephen wanted them to do in this forum: write about the interface of statistics, confidentiality protection, computer science, and many other related topics.

INTRODUCTION

Stephen Fienberg passed away in December of 2016. Among his numerous contributions to the research, education and the practice of statistics, social sciences and machine learning, he was a co-founder of this journal, and to the end, editor-in-chief of it. His absence led to a significant hiatus in the activities of the Journal. This is thus a special issue in two ways.

First, this issue relaunches the Journal, and the editorials by Dwork (2018b) and Vilhuber (2018) in this issue address what the editorial team has accomplished in the past year, and will aim to accomplish in the next few years.

> While we were compiling this issue, we learned with extraordinary sadness that Joyce Fienberg, Steve's widow, had been brutally murdered during the Tree of Life massacre in Pittsburgh, PA on October 27, 2018. We struggle to comprehend. Our sincerest condolences go out to Steve and Joyce's family. This issue is in honor of Steve, but it is dedicated to Joyce, Steve and Joyce's immediate and extended family, and to all of those whose lives were touched by their generosity.
>
> *Lars and Sesa*

Second, this issue is in honor of Steve, and his impact on methodology and practice of privacy and confidentiality. Many others have written about Steve's life, both during his last years (Straf and Tanur 2013; Behseta and Slavković 2013), and after he had passed away

(Carnegie Mellon University 2016; Erosheva and Slavković 2017; Mejia 2017; Wasserman 2017; Rubin 2018; Carriquiry, Reid, and Slavković, forthcoming). Steve was a visionary when it came to interdisciplinary statistical education, research and outreach, and he greatly valued development of statistical methodology for privacy protection and disclosure avoidance. Together with students and collaborators he has produced more than 70 technical papers, 15 discussion and editorial pieces, and has given many congressional and government testimonies on this topic (e.g., Fienberg and Sanil 1997; Fienberg and Steele 1998; Duncan et al. 2001; Trottini and Fienberg 2002; Slavković and Fienberg 2004; Fienberg and Slavković 2005; Fienberg 2006; Fienberg and Slavkovic 2008; Dobra et al. 2009; Fienberg, Nardi, and Slavković 2009; Fienberg, Rinaldo, and Yang 2010; Hall and Fienberg 2010; Fienberg 2011; Hall, Fienberg, and Nardi 2011; Yu et al. 2014b; Yu et al. 2014a; Steorts et al. 2014; Wang, Lei, and E. Fienberg 2015; Wang, Lei, and Fienberg 2016; Lei et al. 2018).

His earliest contribution came in the mid-90s (Fienberg 1994) when he proposed a bootstrap-like approach for creating synthetic data files, similar to the current synthetic data methodology that relies on multiple imputation. He had been an advocate for the importance of framing data privacy as a statistical problem that requires treating both the data utility and the disclosure risk as random variables and the transparency of masking procedures in order to achieve the right statistical inference, not the individual re-identification. This line of thinking led to his pivotal role in establishing interactions between the statisticians, practitioners and those from the computer science community working on privacy from 2003 on. Some fun bits about the first workshop at Bertinoro, Italy have been captured in Slavković (2013), including Steve's role in saving the workshop and setting the path for a development of what we refer to nowadays as a "formal (statistical) privacy methodology", and co-founding of this Journal in 2009 (see also Dwork 2018a). Steve clearly articulated some of the principles that today's formal data privacy protection methodology depends on, and he was instrumental in bringing together approaches from algorithmic and cryptographic computer science with notions of statistical data utility. This Journal remains a testament to his interest in the topic, and with the articles in this issue, we honor that interest, while relaunching the Journal.

To mark the relaunch, we have brought together many of his former collaborators. We have asked them to do what Steve wanted them to do in this forum: write about the interface of statistics, confidentiality protection, computer science, and many other related topics. The articles in this issue are the response to that very special call for papers.

We also asked them to contribute some reminiscences about an event, an episode, a memory of their time with Steve, or how Steve influenced their lives. You will find ten of these personal notes interspersed with the scientific articles, including the <u>one on the next page</u>, or you can also read them separately, see the table of contents at https://doi.org/10.29012/jpc.v8i1.

1. Articles in this Issue

Some of the articles in this issue specifically target topics dear to Steve. Dwork and Ullman (2018) call out a problem identified by Steve: the tension between interactive, exploratory analysis and differential privacy. Both they and Amitai and Reiter (2018) propose solutions based on partitioning the dataset. Amitai and Reiter (2018) in particular, propose a differentially private algorithm for reporting posterior probabilities and posterior quantiles of linear regression coefficients.

As we write this editorial in 2018, a debate is ongoing about how to preserve researcher access to microdata collected by US statistical agencies, including the American Community Survey (ACS), while providing stronger privacy guarantees (see e.g. Abowd and Schmutte, forthcoming; Abowd 2016). Shlomo (2018) provides a summary of the past privacy protection as used by statistical agencies, and discusses where the field is heading. Kinney and Karr (2018) contribute to this discussion, by comparing a specific use of the current ACS for the computation of a Comparable Wage Index (CWI) using both public-use and confidential data. They find comparable results , although the results diverge as the geographic areas for which the CWI is computed become less populated. Li et al. (2018) in turn propose a differentially-private way to provide high-dimensional synthetic histograms, of the type that might be used to construct future CWI.

When differential privacy is used to protect databases, valid statistical inferences still need to be drawn. In theory and practice, the true sampling distributions of the statistics are often approximated by Gaussian distributions. These approximations, however, need to be adjusted when the underlying data are perturbed by DP. Wang et al. (2018) lay out how to generate valid approximating distributions for differentially private statistics.

Acknowledgment

Sesa Slavković
Lars Vilhuber

As readers of this Journal know, I paid my tribute to Steve Fienberg in my 2016 Julius Shiskin Lecture:

"Finally, I would like to acknowledge the role of Stephen Fienberg of Carnegie Mellon University. I'm sure almost everyone in this auditorium can cite a path-breaking contribution of Steve's that had a major impact on statistics and the federal statistical system. I want to highlight the foresight that he had in gathering researchers from the SDL community and the emerging computer science data-privacy community in Bertinoro, Italy, in 2005. This is where I first met Cynthia Dwork and the team of young cryptographers who were shattering the received wisdom in SDL with methods that Steve recognized as revolutionary. I'm also going to spend much of this lecture on those methods. The last time Steve and I talked about this, at this year's JSM, he confided to me that our big mistake was that "we did not grow the community fast enough." I hope this lecture helps solve that problem too." (JPC, 2017, Vol. 7, No. 3 DOI: 10.29012/jpc.v7i3.404)

To which I would now add, that I hope this volume in his honor also expands the community of scholars working on these important issues. As compelling as the cryptographers are, privacy-preserving data analysis must have equal participation from domain scientists, technologists, and statisticians. Good science and strong privacy protections do compete for the same scarce resource (the information in confidential databases), but efficient, workable solutions require input from all these specialists.

John M. Abowd
DOI: 10.29012/jpc.707

References

Abowd, John M. 2016. "How Will Statistical Agencies Operate When All Data Are Private?" *Journal of Privacy and Confidentiality* 7 (3). doi:10.29012/jpc.v7i3.404. https://journalprivacyconfidentiality.org/index.php/jpc/article/view/404.

Abowd, John M., and Ian M. Schmutte. Forthcoming. "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices." *American Economic Review.*

Amitai, Gilad, and Jerome P. Reiter. 2018. "Differentially private summaries for linear regression coefficients." *Journal of Privacy and Confidentiality* 8 (1). doi:10.29012/jpc.683. https://doi.org/10.29012/jpc.683.

Behseta, Sam, and Aleksandra Slavković. 2013. "Interview with Steve Fienberg." *Chance*, no. November. Visited on 09/12/2018. https://chance.amstat.org/2013/11/interview/.

Carnegie Mellon University. 2016. *Stephen E. Fienberg, 1942-2016.* Visited on 09/12/2018. http://www.cmu.edu/news/stories/archives/2016/december/obituary-fienberg.html.

Carriquiry, Alicia L., Nancy Reid, and Aleksandra B. Slavković. Forthcoming. "Stephen Elliott Fienberg 1942-2016, Founding Editor of the Annual Review of Statistics and Its Application." *Annual Review of Statistics and Its Application.*

Dobra, Adrian, Stephen E Fienberg, Alessandro Rinaldo, Aleksandra Slavkovic, and Yi Zhou. 2009. "Algebraic statistics and contingency table problems: Log-linear models, likelihood estimation, and disclosure limitation." In *Emerging applications of algebraic geometry*, 63–88. Springer.

Duncan, George T, Stephen E Fienberg, Rammaya Krishnan, Rema Padman, Stephen F Roehrig, et al. 2001. "Disclosure limitation methods and information loss for tabular data." *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*: 135–166.

Dwork, Cynthia. 2018a. "Reminiscences." *Journal of Privacy and Confidentiality* 8 (1). doi:10.29012/jpc.702. https://doi.org/10.29012/jpc.702.

— . 2018b. "The Future of the Journal of Privacy and Confidentiality." *Journal of Privacy and Confidentiality* 8 (1). doi:10.29012/jpc.705. https://doi.org/10.29012/jpc.705.

Dwork, Cynthia, and Jonathan Ullman. 2018. "The Fienberg Problem: How to Allow Meaningful Interactive Data Analysis in the Age of Differential Privacy." *Journal of Privacy and Confidentiality* 8 (1). doi:10.29012/jpc.687. https://doi.org/10.29012/jpc.687.

Erosheva, Elena, and Alexandra Slavković. 2017. "Obituary: Stephen E. Fienberg 1942–2016." *IMS Bulletin* 46:4–5.

Fienberg, Stephen E. 1994. "A radical proposal for the provision of micro-data samples and the preservation of confidentiality." *Department of statistics.*

— . 2006. "Privacy and confidentiality in an e-commerce world: Data mining, data warehousing, matching and disclosure limitation." *Statistical Science*: 143–154.

— . 2011. "What should privacy mean to you." *CHANCE* 24 (2): 7. doi:10.1007/s00144-011-0014-9. https://doi.org/10.1007/s00144-011-0014-9.

Fienberg, Stephen E, Yuval Nardi, and Aleksandra B Slavković. 2009. "Valid statistical analysis for logistic regression with multiple sources." In *Protecting persons while protecting the people*, 82–94. Springer.

Fienberg, Stephen E, Alessandro Rinaldo, and Xiaolin Yang. 2010. "Differential privacy and the risk-utility tradeoff for multi-dimensional contingency tables." In *International Conference on Privacy in Statistical Databases*, 187–199. Springer.

Fienberg, Stephen E, and Ashish P Sanil. 1997. "A Bayesian approach to data disclosure: Optimal intruder behavior for continuous data." *Journal of Official Statistics* 13 (1): 75.

Fienberg, Stephen E, and Aleksandra B Slavkovic. 2008. "A survey of statistical approaches to preserving confidentiality of contingency table entries." In *Privacy-Preserving Data Mining*, 291–312. Springer.

Fienberg, Stephen E, and Aleksandra B Slavković. 2005. "Preserving the confidentiality of categorical statistical data bases when releasing information for association rules." *Data Mining and Knowledge Discovery* 11 (2): 155–180.

Fienberg, Stephen E, and Russell J Steele. 1998. "Disclosure limitation using perturbation and related methods for categorical data." *Journal of Official Statistics* 14 (4): 485.

Hall, Rob, and Stephen E Fienberg. 2010. "Privacy-preserving record linkage." In *International conference on privacy in statistical databases*, 269–283. Springer.

Hall, Rob, Stephen E Fienberg, and Yuval Nardi. 2011. "Secure multiple linear regression based on homomorphic encryption." *Journal of Official Statistics* 27 (4): 669.

Kinney, Saki, and Alan F. Karr. 2018. "Public-Use vs. Restricted-Use: An Analysis Using the American Community Survey." *Journal of Privacy and Confidentiality* 8 (1). doi:10.29012/jpc.661. https://doi.org/10.29012/jpc.661.

Lei, Jing, Anne-Sophie Charest, Aleksandra Slavkovic, Adam Smith, and Stephen E. Fienberg. 2018. "Differentially private model selection with penalized and constrained likelihood." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181 (3): 609–633.

Li, Bai, Vishesh Karwa, Aleksandra Slavković, and Rebecca Carter Steorts. 2018. "A Privacy Preserving Algorithm to Release Sparse High-dimensional Histograms." *Journal of Privacy and Confidentiality* 8 (1). doi:10.29012/jpc.657. https://doi.org/10.29012/jpc.657.

Mejia, Robin. 2017. "Stephen E. Fienberg (1942-2016)." *Nature* 542:415. http://doi.org/10.1038/542415a.

Rubin, Donald B. 2018. "On Stephen E. Fienberg as a discussant and a friend." *Ann. Appl. Stat.* 12 (2): 683–684. doi:10.1214/17-AOAS1131SF. https://doi.org/10.1214/17-AOAS1131SF.

Shlomo, Natalie. 2018. "Statistical Disclosure Control: New Directions and Challenges." *Journal of Privacy and Confidentiality* 8 (1). doi:10.29012/jpc.684. https://doi.org/10.29012/jpc.684.

Slavković, Aleksandra B. 2013. "Steve the Matchmaker: The Marriage of Statistics and Computer Science in the World of Data Privacy." *CHANCE* 26 (4): 4–7.

Slavković, Aleksandra B, and Stephen E Fienberg. 2004. "Bounds for cell entries in two-way tables given conditional relative frequencies." In *International Workshop on Privacy in Statistical Databases*, 30–43. Springer.

Steorts, Rebecca C, Samuel L Ventura, Mauricio Sadinle, and Stephen E Fienberg. 2014. "A comparison of blocking methods for record linkage." In *International Conference on Privacy in Statistical Databases*, 253–268. Springer.

Straf, Miron L., and Judith M. Tanur. 2013. "A Conversation with Stephen E. Fienberg." *Statistical Science* 28 (3): 447–463. Visited on 09/12/2018. doi:10.1214/12-STS411. https://projecteuclid.org/euclid.ss/1377696945.

Trottini, Mario, and Stephen E Fienberg. 2002. "Modelling user uncertainty for disclosure risk and data utility." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10 (05): 511–527.

Vilhuber, Lars. 2018. "Relaunching the Journal of Privacy and Confidentiality." *Journal of Privacy and Confidentiality* 8 (1). doi:10.29012/jpc.706. https://doi.org/10.29012/jpc.706.

Wang, Yu-Xiang, Jing Lei, and Stephen E. Fienberg. 2015. "Learning with Differential Privacy: Stability, Learnability and the Sufficiency and Necessity of ERM Principle." 17.

Wang, Yu-Xiang, Jing Lei, and Stephen E Fienberg. 2016. "On-average kl-privacy and its equivalence to generalization for max-entropy mechanisms." In *International Conference on Privacy in Statistical Databases*, 121–134. Springer.

Wang, Yue, Daniel Kifer, Jaewoo Lee, and Vishesh Karwa. 2018. "Statistical Approximating Distributions Under Differential Privacy." *Journal of Privacy and Confidentiality* 8 (1). doi:10.29012/jpc.666. https://doi.org/10.29012/jpc.666.

Wasserman, Larry. 2017. "Stephen Fienberg: Superman of statistics." *Proceedings of the National Academy of Sciences of the United States of America* 114 (12): 3002–3003. Visited on 09/12/2018. doi:10.1073/pnas.1700160114. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5373399/.

Yu, Fei, Stephen E Fienberg, Aleksandra B Slavković, and Caroline Uhler. 2014a. "Scalable privacy-preserving data sharing methodology for genome-wide association studies." *Journal of biomedical informatics* 50:133–141.

Yu, Fei, Michal Rybar, Caroline Uhler, and Stephen E Fienberg. 2014b. "Differentially-private logistic regression for detecting multiple-SNP association in GWAS databases." In *International Conference on Privacy in Statistical Databases*, 170–184. Springer.