CURRICULUM VITAE

Rebecca C. Steorts

Duke University, Department of Statistical Science
Webpage: resteorts.github.io
Email: beka@stat.duke.edu

EDUCATION

2012	**Ph.D.**, Statistics, University of Florida, Gainesville, FL Advisor: Malay Ghosh Thesis: *Bayes and and Empirical Bayes Benchmarking for Small Area Estimation,* Honorable Mention (Second Place), **Leonard J. Savage Award** for the top Bayesian applied thesis internationally
2005	**M.Sc.**, Mathematical Sciences, Clemson University, Clemson, SC
2001	**B.Sc.**, Mathematics, Davidson College, Davidson, NC

PROFESSIONAL EXPERIENCE

Primary Academic Appointments

2015 – present	**Assistant Professor** Department of Statistical Science, Duke University
2012-2015	**Visiting Assistant Professor** Department of Statistics, Carnegie Mellon University Mentor: Stephen E. Fienberg

Other Affiliations

2015 – present	**Affiliated Faculty** Department of Computer Science, Department of Biostatistics and Bioinformatics, Information Initiative at Duke (iiD), and the Social Science Research Institute (SSRI) at Duke University
2021 – present	Program Lead, Census Cooperative Agreements U.S. Census Bureau Research on Record Linkage and Entity Resolution
2016 – present	Executive Director, Duke Undergraduate Machine Learning Program
2017 – present	Principal Researcher, Research Mathematical Statistician U.S. Census Bureau
2014 – present	Statistical Consultant, Human Rights Data Analysis Group (HRDAG)
2014	Data Science Consultant, Food and Agricultural Organization (FAO)

	of the United Nations
2014	Visiting Scholar at the University of Trier, Department of Economics and Social Sciences, Trier, Germany
2014	Visiting Scholar at the University of Rome "La Sapienza", Department of Methods and Models for Economics, Geography and Finance Rome, Italy
2013	Visiting Scientist in Summer at Census Program, U.S. Census Bureau, Washington D.C.

HONORS AND AWARDS

2021	Washington Statistical Society President's Invited Lecture
2021	Duke University Graduate Mentoring Award
2019	Elected Fellow of the International Statistical Institute
2017	NSF CAREER Award
2015	MIT Review Magazine's 35 Innovators Under 35 Humanitarian for software design in estimating death counts for the Syrian civil war Feature video at EmTech, Boston, MA Feature piece in MIT Review, October, 2015 Feature piece by the Human Rights Data Analysis Group (HRDAG)
2013	Honorable Mention (Second Place) for best thesis in applied methodology, Leonard J. Savage Award, International Society for Bayesian Analysis
2010-2012	United States Census Bureau Dissertation Fellowship Program

PUBLICATIONS

Peer-reviewed Publications (all after 2015 published at Duke University)

(* student or postdoctoral fellow supervised by RCS)

1. Datta, G., Ghosh, M., **Steorts, R.** and Maples, J. (2011). Bayesian Benchmarking with Applications to Small Area Estimation, *TEST*, **20**(3) 574–588, doi:10.1007/s11749-010-0218-y.

2. **Steorts, R.** and Ghosh, M. (2013). On Estimation of Mean Squared Errors of Benchmarked Empirical Bayes Estimators, *Statistica Sinica*, **23**(2) 749–767, arxiv:1304.1600, doi:10.5705/ss.2012.053.

3. Ghosh, M. and **Steorts, R.** (2013). Two-stage Bayesian Benchmarking as Applied to Small Area Estimation, *TEST*, **22**(4) 670–687, arxiv:1305.6657, doi: 10.1007/s11749-013-0338-2.

[7] 4. **Steorts, R.**, Hall, R. and Fienberg, S. (2014). SMERED: A Bayesian Approach to Graphical Record Linkage and De-duplication, **33** 922–930: *Artificial Intelligence and Statistics (AIStats)*, arxiv:1403.0211.

[5] 5. **Steorts, R.**, Ventura, S., Sadinle, M. and Fienberg, S. (2014). Blocking Comparisons for Record Linkage, *Privacy in Statistical Databases (Lecture Notes in Computer Science 8744)*, ed. J. Domingo-Ferrer, Springer, 253-268; arxiv:1407.3191, doi:10.1007/978-3-319-11257-2_20.

[6] 6. **Steorts, R.** (2015). Entity Resolution using Empirically Motivated Priors, *Bayesian Analysis*, **10**(4) 849–875, arxiv:1409.0643, doi:10.1214/15-BA965SI, **Finalist Lindley Prize.**

7. Wehbe, L.*, Ramdas, A.*, **Steorts, R.** and Shalizi, C.R. (2015). Regularized Brain Reading with Smoothing and Shrinkage Using Bayesian and Frequentist Methods, *Annals of Applied Statistics*, **9**:4 (1997-2022); arxiv:1401.6595, doi:10.1214/15-AOAS837.

[8] 8. **Steorts, R.**, Hall, R., and Fienberg, S.E. (2016). A Bayesian Approach to Graphical Record Linkage and De-duplication, *Journal of the American Statistical Association*, **111**:516 (1660-1672); arxiv:1312.4645, doi:10.1080/01621459.2015.1105807.

[4] 9. Zanella, G.*, Betancourt, B.*, Wallach, H., Miller, J., Zaidi, A.*, and **Steorts, R.** (2016). Flexible Models for Microclustering with Applications to Entity Resolution, *Neural Information Processing Systems (NIPS)*, 1417–1425, arxiv:1610.09780.

[9] 10. **Steorts. R**, Barnes, M.*, and Neiswanger, M.* (2017). Performance Bounds for Graphical Record Linkage, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 54:298–306, Editors: Aarti Singh and Jerry Zhu, arxiv:1703.02679.

11. Durante, D., Mukherjee, N.*, and **Steorts, R.** (2017). Bayesian Learning of Dynamic Multilayer Networks, *Journal of Machine Learning Research*, **18**:43 (1-29); arxiv:608.02209.

[3] 12. Chen, B.*, Shrivastava, A., and **Steorts, R.** (2018), Unique Entity Estimation with Application to the Syrian Conflict, *Annals of Applied Statistics*, **12**:2 (1039-1067); arxiv:1710.02690, doi:10.1214/18-AOAS1163, **Best Student ISAA paper**.

13. **Steorts, R.**, Tancredi, A., and Liseo, B. (2018). Generalized Bayesian Record Linkage and Regression with Exact Error Propagation. *Privacy in Statistical Databases (Lecture Notes in Computer Science 11126)*, eds. Domingo-Ferrer J., Montes F., Springer, 297-313; arxiv:1810.04808, doi:10.1007/978-3-319-99771-1_20.

14. **Steorts, R.** and Shrivastava, A. (2018). Probabilistic Blocking with An Application to the Syrian Conflict. *Privacy in Statistical Databases (Lecture Notes in Computer Science 11126)*, eds. Domingo-Ferrer J., Montes F., Springer, 297-313, Springer, 314-327; arxiv:1810.05497, doi:10.1007/978-3-319-99771-1_21.

[10] 15. Bai, L.* , Karwa. V., Slavkovic, A., and **Steorts, R.** (2018). Privacy Preserving Algorithm to Release Sparse High-dimensional Histograms, *Journal of Privacy and Confidentiality*, **8**:1; doi.org/10.29012/jpc.657.

16. Bedoya, A.D., Clement, M.E., Phelan, M., **Steorts, R.**, O'Brien, C., Goldstein, B.A. (2019). Minimal Impact of Implemented Early Warning Score and Best Practice Alert for Patient Deterioration. *Critical Care Medicine*, **47**:1 (149-55); doi.org/10.1097/ccm.0000000000003439.

17. Ghosh, Malay and **Steorts, R.** (2019). Some Variants of Constrained Estimation in Finite Population Sampling. *International Statistical Review*, **87**: 90-103, doi.org/10.1111/insr.12309.

18. Tancredi, A., **Steorts, R.**, and Liseo, B. (2020). A unified framework for de-duplication and population size estimation. *Bayesian Analysis*, **15**:2 (633-682);doi.org/10.1214/19-BA1146.

19. Lin, Qiaohui*, Betancourt*, B., Ben Goldstein, and **Steorts, R.** (2020). Prediction of Appointment No-shows using Electronic Health Records. *Journal of Applied Statistics*, **47**:7 (1220-1234) doi.org/10.1080/02664763.2019.1672631.

20. O'Brien, C. Goldstein, B., Shen, Yueqi, Phelan, M., Lambert, C., Bedoya, A. **Steorts, R.** (2020). Development, Implementation, and Evaluation of an In-Hospital, Optimized, Early Warning Score for Patient Deterioration, *MDM Policy & Practice*, **5**:1. doi.org/10.1177/2381468319899663.

21. Enamorado, T. and **Steorts, R.** (2020). Probabilistic Blocking and Distributed Bayesian Entity Resolution, *Privacy in Statistical Databases (Lecture Notes in Computer Science 12276)*, ed. Josep Domingo-Ferrer and Krishnamurty Muralidhar, 224-239; https://doi.org/10.1007/978-3-030-57521-2_16.

22. Tang, J.*, Reiter, J. and **Steorts, R.** (2020). Bayesian Modeling for Simultaneous Regression and Record Linkage, *Privacy in Statistical Databases (Lecture Notes in Computer Science 12276)*, ed. Josep Domingo-Ferrer and Krishnamurty Muralidhar, 209-223; https://doi.org/10.1007/978-3-030-57521-2_15.

23. **Steorts, R**., Schmid, T., and Tzavdis, N. (2020). Smoothing and Benchmarking for Small Area Estimation with Application to Rental Prices in Berlin, *International Statistical Review*, **88**:3 (580-598), doi.org/10.1111/insr.12373.

[1] 24. Marchant, N.*, Kaplan, A.*, Rubenstein, B., and Elzar, D., and **Steorts, R.** (2021). d-blink: Distributed End-to-End Bayesian Entity Resolution, *Journal of Computational Graphics and Statistics*, **30**:2 (406-421); arxiv:1909.06039, https://doi.org/10.1080/10618600.2020.1825451.

[2] 25. Brenda Betancourt*, Giacomo Zanella, and **Steorts, R.** (2021). Random Partition Models for Microclustering Tasks, *Journal of the American Statistical Association, Theory and Methods*, In Press, arxiv:2004.02008, https://doi.org/10.1080/01621459.2020.1841647.

26. Mosaferi, S.* Ghosh, M. and **Steorts, R.** (2021). Measurement Error Models for Small Area Estimation, *Communications and Statistics: Simulation and Computation*, In Press.

27. Binette, O.*, and **Steorts, R.** (2021). Guidance For Multiple Systems Estimation Applied to Modern Slavery in the United Kingdom. *Journal of the Royal Statistical Society, Series A*, In Press. arxiv:2112.01594.

28. Kaplan, Andee*, Betancourt, B.*, and **Steorts, R.** (2021). A Practical Approach to Proper Inference with Linked Data, *The American Statistician*, In Press, arxiv:1810.01538.

29. Binette, O.* and **Steorts, R.** (2022). (Almost) All of Entity Resolution, *Science Advances*, In Press, arxiv:2008.04443.

Invited Papers and Discussions (all after 2015 published at Duke University)

30. **Steorts, R.** and Ugarte, D.M. (2014). Discussion of "Single and Two-Stage Cross-Sectional and Time Series Benchmarking Procedures for SAE," *TEST*, **23** :680–685, arxiv:1405.6416, doi: 10.1007/s11749-014-0386-2.

31. Fienberg, S. and **Steorts, R.** (2014). Discussion of "Estimating the Distribution of Dietary Consumption Patterns," *Statistical Science*, **29** 1:95–96, arxiv:1403.0566, doi:10.1214/13-STS448.

32. Betancourt, B.* and **Steorts, R.** (2018). Bayesian Decision Making with Application to Resource Allocation, *Wiley StatsRef-Statistics Reference Online*, (eds. N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J. L. Teugels). doi:10.1002/9781118445112.stat07856.

33. Binette, O.* and **Steorts, R.** (2020). Discussion of "Multiple-systems analysis for the quantification of modern slavery: classical and Bayesian approaches", *Journal of the Royal Statistical Society, Series A*. doi.org/10.1111/rssa.12505.

34. Binette, O.* and **Steorts, R.** (2021). Modern Bayesian Entity Resolution, *Wiley StatsRef-Statistics Reference Online*, (eds. N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J. L. Teugels), In Press.

Peer-Reviewed Workshop Papers (all after 2015 published at Duke University)

35. Hall, R., **Steorts, R.** and Fienberg, S. (2012). Bayesian Parametric and Nonparametric Inference for High Dimensional Multiple Record Linkage, NIPS *Modern Nonparametric Methods in Machine Learning* workshop paper.

36. Broderick, T. and **Steorts, R.** (2014). Variational Bayes for Merging Noisy Databases, NIPS Workshops in *Advances in Variational Inference*, NIPS, arxiv:1410.4792.

37. Miller, J., Betancourt, B.*, Zaidi, A.*, Wallach, H. and **Steorts, R.** (2015). The Microclustering Problem: When the Cluster Sizes Don't Grow with the Number of Data Points, NIPS *Bayesian Nonparametrics: The Next Generation Workshop Series*, arxiv:1512.00792, **Top Five Best Workshop Papers.**

ACTIVE OR PENDING GRANTS (all after 2015 obtained at Duke University)

1. Interpretable and Scalable Entity Resolution Applied to the Decennial Census, Alfred P. Sloan Foundation, $249,909 over 12/01/2019–11/31/2022. Role: PI.
2. CAREER: Scalable Record Linkage through the Microclustering Property, NSF-SES – 1652431, $449,985, 05/15/17–04/15/22. Role: PI.

COMPLETED GRANTS (all after 2015 obtained at Duke University) – as PI

1. Big Data Analytics Applied to Tracking and Cybersecurity, Laboratory for Analytic Sciences at North Carolina State University (NCSU), $105,739 over 01/01/2017–12/31/2017. Role: PI.
2. Synthetic Data Release: The Tradeoff Between Privacy and Utility of Big Data, Laboratory for Analytic Sciences at NCSU, $85,000 over 01/01/2015–12/31/2016. Role: PI.
3. Computationally Scalable Statistical Methods for High Dimensional Record Linkage, The University of Chicago Metaknowledge Lab and the Templeton Foundation), $135,000; 2015-2016. Role: PI.
4. Posterior Prototyping: Bridging the Gap Between Record Linkage ad Regression, Laboratory for Analytic Sciences at North Carolina State University (NCSU), $99,614 over 01/01/2019–12/31/2019. Role: PI; co-PI: Brenda Betancourt (University of Florida).

COMPLETED GRANTS (all after 2015 obtained at Duke University) – as co-PI

1. Record Linkage and Privacy-Preserving Methods for Big Data, NSF-SES – 1534412, $265,579; 07/28/15–08/31/18. Role: co-PI.
2. Incorporating Dynamic Electronic Health Records Data Into a Model for Patient Deterioration, Collaborative Quantitative Approaches to Problems in the Basic and Clinical Sciences seed funding program, Duke University, $40,000, 2016-2017. Role: co-PI.

COMPLETED GRANTS (all after 2015 obtained at Duke University) – as Investigator

1. NCRN-MN:Triangle Census Research Network, NSF-SES SES-1131897, PI: Reiter, $4,087,370 over 10/01/2011– 09/30/2017. Role: Investigator.
2. MIDAS Informatics Services Group—The iSG, NIH, PIs: Wagner, Espino (University of Pittsburgh), Brown (CMU), $234,936 over 8/1/14–06/01/2015. Role: Investigator.
3. Census Research Node: Data Integration, Online Data Collection, and Privacy Protection for Census 2020, NSF SES-1130706, $256,857 over 10/1/11–06/01/2015. PIs: Fienberg/Eddy. Role: Investigator.
4. Statistics and Machine Learning for Scientific Inference, NSF DMS-1043903, $433,261 over 7/15/11–06/01/2015. PI: Kass. Role: Investigator.

SOFTWARE AND PRODUCTS ON CRAN (all after 2015 done at Duke University)

1. `clevr` (2020). Software for Clustering Evaluations Metrics. Developed by Neil Marchant and Rebecca C. Steorts. clevr. **Available on github and CRAN**.

2. `cora` (2020). Software for Entity Resolution (Data Sets). Developed by Srini Sunil, Andee Kaplan, Rebecca C. Steorts. cora. **Available on github and CRAN**.

3. `restaurant` (2020). Software for Entity Resolution (Data Sets). Developed by Srini Sunil, Andde Kaplan, and Rebecca C. Steorts. restaurant. **Available on github and CRAN**.

4. `cd` (2020). Software for Entity Resolution (Data Sets). Developed by Srini Sunil, Andee Kaplan, and Rebecca C. Steorts. cd. **Available on github and CRAN**.

5. `tlsh` (2020). Software for Blocking. Developed by Rebecca C. Steorts. tlsh. **Available on github and CRAN**.

6. `klsh` (2020). Software for Blocking. Developed by Rebecca C. Steorts. klsh. **Available on github and CRAN**.

7. `microclustr` (2020). Software for Microclustering. Developed by Brenda Betancourt, Giacomo Zanella, and Rebecca C. Steorts. microclustr. **Available on github and CRAN**.

8. `blink` (Updated 2020). Empirical Bayes Record Linkage and De-duplication `R` software. Available at github and CRAN. Updated to include the RLdata500 data set given it's deprecation from CRAN. **Available on github and CRAN**. Developed and coded by Rebecca C. Steorts.

9. `representr` (2020). Record Linkage Package for Selecting the Most Representative Record. Available at representr. Developed by Andee Kaplan, Brenda Betancourt, and Rebecca C. Steorts. **Available on github and CRAN**.

10. `italy` (Updated 2017). Software for Entity Resolution (Data Sets). Developed by Rebecca C. Steorts. Available at italy. **Available on github and CRAN**.

OTHER SOFTWARE AND PRODUCTS (all after 2015 done at Duke University)

1. `MSETools` (2021). Implements Binette and Steorts (2021). Available at MSETools.

2. `exchanger` (2021). Implements Marchant, Rubinstein, Steorts (2021). Available at exchanger.

3. `dblink` (2020). Distributed Bayesian Entity Resolution. Available at dblink. Developed by Neil Marchant and Rebecca. C. Steorts.

4. `dblinkR` (2020). Distributed Bayesian Entity Resolution for R. Available at dblinkR. Developed by Neil Marchant and Rebecca. C. Steorts.

5. `BDD` (2020). Implements Sadinle (2014). Available at BDD. Developed by Neil Marchant and Rebecca. C. Steorts.

6. `fastlink-dblink` (2020). Implements Edamorado and Steorts (2020). Available at fastlink-dblink. Developed by Ted Edamorado and Rebecca. C. Steorts.

7. `SMERED` (Updated 2016). Record Linkage and De-duplication `Java` software (*SMERED*) with post-processing software in `R`. Developed and coded by Rebecca C. Steorts and Rob Hall. `https://bitbucket.org/resteorts/smered/`.

MENTORING

*(Denotes Steorts supervised as primary advisor on at least one paper or thesis.)

Postdoctoral Advisees

Jairo Fuquene*	2019 – 2020 (UC Davis)
Andee Kaplan*	2017 – 2019 (Colorado State University)
Brenda Betancourt* (Bernstein-Forster Fellowship)	2015 – 2018 (University of Florida)
Nabanita Mukherjee*	2015 – 2016 (AbbVie)

Doctoral Advisees

Brian Kundinger* (Duke Statistical Science)	2020 – present
Olivier Binette* (Duke Statistical Science)	2019 – present
Neil Marchant* (University of Melbourne) joint with Ben Rubinstein	2018 – 2021

Graduate Student Collaboration

Neil Marchant* (Melbourne)	2018 – 2021
Sepideh Mosaferi* (ISU)	2018 – 2021
Jiurui Tang* (Duke)	2018 – 2019
Sayan Patra* (Duke)	2018 – 2019
Bai Li* (Duke)	2015 – 2018
Beidi Chen* (Rice)	2015 – 2018
Matt Barnes* (CMU)	2015 – 2017
Willie Neiswanger* (CMU)	2015 – 2017
Sam Ventura* (CMU)	2012–2014
Maurcio Sadine* (CMU)	2013–2014
Michael Pane* (CMU)	2012–2014

Doctoral thesis committee

Jiurui Tang* (Statistical Science)	2019 – 2022
Sayan Patra* (Statistical Science)	2018 – 2019
Kyle Burris (Statistical Science)	2019
Jody Heck Wortman (Statistical Science)	2016 – 2018
Ye Wang (Statistical Science)	2018
Matt Barnes* (CMU Robotics), External Member	2017– 2018
Mauricio Sadinle* (CMU Statistics)	2013–2015
Samuel Ventura* (CMU Statistics)	2013–2015
Rafael Stern (CMU Statistics)	2013–2015
Zachary Kurtz (CMU Statistics)	2012–2014

Preliminary oral committee

Brian Kundinger* (Duke Statistical Science, Chair)	2021
Olivier Binette* (Duke Statistical Science, Chair)	2021
Jiurui Tang* (Statistical Science)	2019
Lindsay Berry (Statistical Science)	2017
Luke Calkins (Pratt School of Engineering)	2017
Jody Heck Wortman (Statistical Science)	2016

Master's Thesis Advisor

Bai Li*	2015 – 2017 (Duke University PhD student)
Reuben McCreanor* (Statistical Science, MSEM)	2015 – 2017 (Survey Monkey)

Master's Project Advisor

Aasha Reddy*	2020 - 2022
Emily Gentles*	2020 - 2022
Davis Berlin	2019 – 2020 (UCLA Phd student)
Shubhi Sharma	2019 – 2020 (Stanford PhD student)
Qiaohui Lin*	2016 – 2018 (UT Austin PhD student)
Sepideh Mosaferi* (CMU Statistics)	2012, joint with Steve Fienberg (ISU PhD student)

Master Thesis Committee

Linlin Li	2021
Ziang Wang	2021
Sayan Patra*	2019
Jody Heck Wortman	2019
Sophie Yu	2017
Bai Li*, Chair	2017
Reuben McCreanor*, Chair	2017

Undergraduate Thesis Advisees

Lavonne Haong (joint with Simon Mak)	2020 – 2021 (Citibank)
Bihuan Zhuang *	2014 – 2015 (Apple)
Peter Sadosky*(CMU Statistics)	2014 – 2015 (Uber)
Emily Furnish (CMU Statistics)	2013 – 2014, joint with Steve Fienberg (W&M Law)

Undergraduate Project Advisees

Shrey Gupta	2018 – 2020
Bassim Eledath	2018 – 2020
Melody Jiang	2018 – 2019
Ritika Bharati	2017
Srinivas Sunil	2017 – 2018, joint with Andee Kaplan
Angie Shen*	2016 – 2017, joint with Ben Goldstein
Corey Vernot	2016 – 2017
Stephanie Stern (CMU Statistics)	2013 – 2014 (University of Michican MS student)
Kairavi Chahal (CMU Statistics)	2013 (American Express)
Dahiana Jiminez (CMU Statistics)	2013

Undergraduate Thesis Committee

Lavonne Haong (joint with Simon Mak), Chair	2020 – 2021 (Citibank)
Bihan Zhuang*, Chair[1]	2019
Lucy Lu (Statistical Science)	2017
Peter Sadosky* (CMU Statistics), Chair	2015
Michael Pane* (CMU Statistics)	2013

TEACHING EXPERIENCE AT DUKE

STA 602	Modern Advancements of Bayesian Methods (Spring 2022)
STA 490/690	Special Topics: Almost All of Entity Resolution (Spring 2020)
STA 790	Special Topics: Some of Entity Resolution (Fall 2020)
STA 360	Modern Advancements of Bayesian Methods (Fall 2020 – 2021)
STA 360/602	Modern Advancements of Bayesian Methods (Spring 2016 – 2019)
STA 325	Machine Learning and Data Mining (Fall 2016 – 2019)
STA 790	Special Topics: Record Linkage (Fall 2017)
STA 521	Predictive Modeling (Fall 2015)

PROFESSIONAL APPOINTMENTS AND SERVICE

EDITORIAL ACTIVITIES

Editorial Boards

Associate Editor, *Bayesian Analysis* (2022 – present)
Associate Editor, *Journal of the American Statistical Association (ACS)* (2019 – present)
Associate Editor, *Science Advances* (2019 – present)
Associate Editor, *Journal of Survey Statistics and Methodology (JSSAM)* (2018 – present)
Guest Editor, Special Issue on Record Linkage, JSSAM, (2021 – 2022).

Peer Review Activities *AIStats*; *American Statistician*; *American Economic Review*; *Annals of Applied Statistics*; *Computational Statistics and Data Analysis*; *International Conference in Machine Learning (ICML)*; *Journal of Agricultural, Biological, and Environmental Statistics*; *Journal of the American Statistical Association*; *Data Mining and Knowledge Discovery*; *Journal of Machine Learning Research*; *Journal of Official Statistics*, *Journal of Privacy and Confidentiality*; *Journal of Multivariate Analysis*; *Journal of the Royal Statistical Society, Series A*; *Journal of Survey Statistics and Methodology*; *Journal of Statistical Planning and Inference*; *Neural Information Processing Systems (NIPS)*; *Proceedings of the National Academia of Sciences of the United States*; *PLOS ONE*; *Statistical Methods and Applications*; *Statistics in Medicine*; *TEST*; *Springer Publishing, New York*; *Transactions of Knowledge and Data Engineering*.

Grant Review Panels

2019	National Science Foundation Panel, CISE
2018	National Science Foundation Panel, CISE/MMS

Adhoc Review Panels

2020	Sloan Foundation
2018, 2019	National Science Foundation and U.S. Census Bureau
2015, 2017	National Science Foundation

Conference, Workshop, and Professional Service

2021 – 2022	Program Chair of the Record Linkage Interest Group
2021 – 2022	Reviewer of SBSS Student Papers
2021	Advisor for Bayes Comp ISBA workshop on MCMC (October 2021)
2021 – 2022	Advisory Board for US Patent and Trademark Office
2020	Co-organized for Duke Undergraduate Virtual Machine Learning Day, June 2020
2020	Co-organizer for SDSS, Pittsburgh, PA
2019	Session chair and organizer, Bayes Comp, Gainesville, Florida
2018	Area Chair, Women in Machine Learning (WiML) Workshop, Montreal, Canada
2018	Organizing Committee, Workshop on Bayes, Big Data and Social Good, Marseille, France

2017–2018	Organizing Committee, IISA International Conference on Statistics
2017	Session chair, AISTATS
2012, 2014, 2016	Session organizer for topic-contributed sessions at JSM
2014 – 2015	Invited Program Committee Member: IEEE ICDM (International Conference on Data Mining) Workshop on Data Integration and Applications (DINA)
2014 – present	American Statistical Association's Committee on Scientific Freedom and Human Rights
2016 – present	Invited ICML, NIPS Program Committee
2015	Invited Program Committee Member, NIPS, Nonparametric Bayesian Workshop
2014 – present	Invited AISTATS Program Committee
2014	Organizing Committee of the Frontier of Hierarchical Modeling in Observational Studies, Complex Surveys, and Big Data: A Conference Honoring Professor Malay Ghosh; College Park, MD
2013, 2015	Session organizer for invited sessions at JSM

DEPARTMENT SERVICE

2021–2022	Duke Statistical Science PhD Graduate Admission Committee
2021 – 2022	Duke Science and Technology Departmental Faculty Search Committee
2021 – 2022	Duke Statistical Science Master's Advisory Committee
2021 – 2022	Duke Statistical Science Master's Portfolio Committee
2020 – 2021	Department of Statistical Science Masters Admission Committee
2021	Chair of Statistical Science Qualifying Exam Committee
2015 – 2020	Duke Statistical Science PhD Graduate Admission Committee
2015 – 2017	Department of Statistical Science Tenure Track Search Committee
2015 – 2017	Department of Statistical Science Seminar Series Coordinator
2015 – 2016	Department of Statistical Science Departmental Computing Committee
2013 – 2015	Co-chair of Faculty Seminars, Carnegie Mellon University, Chair 2014–2015
2013 – 2015	Co-organizer of event planning committee, Carnegie Mellon University
2012 – 2013	Committee of Graduate Admissions, Carnegie Mellon University,

UNIVERSITY SERVICE

2015 – present	A&S Department Representative, Social Science Research Institute
2016 – present	Executive Director of Duke Undergraduate Machine Learning Program
2017 – present	Faculty Advisory Board of the Duke Human Rights Center
2015 – present	Social Science Research Institute Faculty Board
2015 – present	Duke Machine Learning Seminar Committee, iiD
2015 – present	Reviewer of Data Plus Proposals, Rhodes iiD
2017 – 2018	MIDS Panel on Data Science Department of Statistics

PRESENTATIONS

Keynote Presentations

1. (2021) Washington Statistical Society President's Invited Lecture.
2. (2019) Data Sciences in an Academic Health Center setting in the 21st Century, University of Colorado, Denver, CO
3. (2018) Undergraduate Machine Learning Day, Duke University, Durham, NC
4. (2018) Federal Committee on Statistical Methodology (FCSM) and Washington Statistical Society (WSS) Workshop on Quality of Integrated Data, Bureau of Labor Statistics, Washington, DC
5. (2017) International Seminar on Data Editing, Imputation and Non Response, CIMAT, Guanajuato, Mexico
6. (2017) Bayesian Inference in Stochastic Processes, Bocconi University, Milano, Italy

Invited Short Courses and Workshops

1. (2021) Workshop on Bridging the Gap between Theory and Practice of MCMC, Bayes Comp. Joint with Pierre Jacob.
2. (2021) An Introduction to Entity Resolution. ISBA World Meeting, Kuming, China.
3. (2020) Almost All of Entity Resolution. Invited ENAR Webinar.
4. (2019) An Introduction to Modern Record Linkage. Short Course (4-hour short course). Population Dynamics and Health Program Workshop, Population Studies Center, Institute for Social Research, University of Michigan. Workshop Video with Demos.
5. (2018) An Introduction to Modern Record Linkage. Short Course (1-hour short course). Duke Undergraduate Machine Learning Day.
6. (2018) An Introduction to Modern Record Linkage. Short Course (6-hour short course). CIMAT, Guanajuato, Mexico; U.S. Census Bureau. (Joint with Andee Kaplan, Breda Betancourt, and Beidi Chen).
7. (2016) *Teaching Bayes: the Essential Parts*. Short Course (3-hour short course), ISBA World Meeting, Sardinia, Italy. Sardinia, Italy.
8. (2014) *An Introduction to Privacy and Statistical Disclosure*. Short Course at the Universite Paris Dauphine, Mathematiques, Apprentissage et Sciences Humaines (MASH), 12-hour course, Paris, France.
9. (2011) *Fundamentals and Applications of Bayesian Analysis.* Short Course (12-hour course), joint with Malay Ghosh, Novartis Oncology Global Development, Florham Park, NJ.

Invited Seminars and Conference Presentations

1. (2021) Invited Talk, EAC-ISBA 2021, Kunming, China
2. (2021) Invited Talk, CSRM, United States Census Bureau, Suitland, Maryland
3. (2021) Invited Talk, Faculty Seminar, Department of Statistics, Penn State University
4. (2021) Invited Talk, Davidson College, Davidson, NC
5. (2021) Invited Panel Discussion, USPTO Symposium on Entity Resolution, Alexandria, VA
6. (2021) Invited Talk, USPTO Symposium on Entity Resolution, Alexandria, VA
7. (2021) Invited Talk, Faculty Seminar, Dept. of Bioinformatics & Biostatistics, University of Louisville, Louisville, KY
8. (2020) Invited Talk, Privacy in Statistical Databases, Tarragona, Catalonia
9. (2020) Tutorial Session on Entity Resolution, Joint Statistical Meetings, Philadelphia, PA
10. (2020) Invited Talk, Joint Statistical Meetings, Philadelphia, PA
11. (2020) Invited Talk, Discussion and Re-joiner of "A unified framework for deduplication, ISBA
12. (2020) SDSS, Pittsburgh, PA
13. (2020) ENAR, Nashville, TN
14. (2020) Bayes Comp, University of Florida, Gainesville, Florida
15. (2019) IISA International Conference on Statistics, Mumbai, India
16. (2019) Microsoft Research, Seattle, Washington
17. (2019) Google, Seattle, Washington
18. (2019) Joint Seminar with Department of Information Sciences and Department of Statistics, University of Melbourne, Melbourne, Australia
19. (2019) Invited Discussion, ISI World Statistics Congress, Kuala Lumpur, Malaysia
20. (2019) Invited Talk, ISI World Statistics Congress, Kuala Lumpur, Malaysia
21. (2019) Conference on Current Trends in Survey Statistics, National University of Singapore, Singapore
22. (2019) Defense Science Organization (DSO), National University of Singapore, Singapore
23. (2019) Workshop on Survey Statistics, National University of Singapore, Singapore
24. (2019) La Sapienza, Department of Methods and Models for Economics, Geography and Finance, Rome, Italy
25. (2019) University of Chicago, Department of Statistics, Chicago, IL
26. (2018) Bayesian Statistics in the Big Data Era, Centre International de Rencontres Mathematiques, Marseille, France
27. (2018) Workshop in Big Data and Social Good, Centre International de Rencontres Mathematiques, Marseille, France
    https://xianblog.wordpress.com/2018/11/27/bayes-for-good/

28. (2018) Privacy in Statistical Databases (PSD), Valencia, Spain
29. (2018) Joint Statistical Meetings, Vancouver, CA
30. (2018) International Workshop on Survey Statistics and Big Data, Nancheng, China
31. (2018) International Small Area Meeting honoring Danny Pfefferman, Shanghai, China
32. (2018) IISA International Conference on Statistics, Gainesville, FL
33. (2018) Faculty Seminar, University of Michigan, Department of Biostatistics
34. (2017) IISA International Conference on Statistics, Hyderabad, India
35. (2017) Center for Survey and Research Methodology Seminar, U.S. Census Bureau, Suitland, MD
36. (2017) Faculty Seminar, University of Florida, Gainesville, FL
37. (2017) Faculty Seminar, Harvard University, Department of Statistics and Biostatistics,
38. (2017) INFORMS Healthcare, Rotterdam, Netherlands
39. (2017) Data Plus and Department of Computer Science Undergraduate Seminar, Information Initiative at Duke (iiD), Duke University
40. (2017) AISTATS, Fort Lauderdale, Florida
41. (2017) Weekly Research Meeting, Laboratory for Analytic Sciences, NC State University
42. (2016) Neural Information and Professing Systems, Barcelona, Spain
43. (2016) Faculty Seminar, Center for Language and Speech Processing, Johns Hopkins University
44. (2016) Faculty Seminar, Cambridge Biostatistics Unit
45. (2016) Isaac Newton Program: Data Linkage and Anonymization, Cambridge, UK
46. (2016) Invited Round Table Discussion, JSM, Chicago, IL
47. (2016) Invited Privacy Talk, JSM, Chicago, IL
48. (2016) Faculty Seminar, Bocconi University, Milano, Italy
49. (2016) Statistical Learning and Data Mining (SLDM) Conference, University of North Carolina at Chapel Hill, Chapel Hill, NC
50. (2016) Faculty Seminar, The University of Chicago, The Computational Institute, Chicago, IL
51. (2016) NISS Affiliates Meeting, Austin, TX
52. (2016) Late Breaking News Session, Sixth IMS-ISBA joint meeting Bayes Comp at MCM-Ski V, Lenzerheide, Switzerland
53. (2015) Spotlight presentation, Neural Information and Professing Systems, Bayesian Nonparametrics: The Next Generation Workshops, Montreal, Canada.
54. (2015) Spotlight presentation, EmTech, MIT Media Lab, Boston, MA
55. (2015) Joint Machine Learning MIT & Microsoft Research Seminar, Boston, MA

56. (2015) FOCUS Interdisciplinary Discussion Course For Duke Freshman, Durham, NC
57. (2015) Invited Session at JSM, Boston, MA
58. (2015) Faculty Seminar, University of Padua, Padua, Italy
59. (2015) ITACOSM 2015, 4th ITAlian Conference on Survey Methodology, Rome, Italy
60. (2015) IMS-Microsoft Research Workshop: Foundations of Data Science, Boston, MA
61. (2015) NCRN Spring Meeting, National Academy of Science, Washington, DC
62. (2015) Discussion of *Doing Data Science, Straight Talk from the Frontline* by Rachel Schutt, Chief Data Scientist and Senior Vice President of *NewsCorp*, Special Invited Session ENAR, Miami, FL
63. (2015) Faculty Seminar, Duke University Computer Science, Durham, NC
64. (2015) Faculty Seminar, University of North Carolina at Chapel Hill, Department of Biostatistics, Chapel Hill, NC
65. (2015) Faculty Seminar, Duke University, Department of Statistical Science, Durham, NC
66. (2015) Faculty Seminar, University of California at Berkeley, Statistics Department, Berkeley, CA
67. (2015) Faculty Seminar, University of Minnesota, Statistics Department, Minneapolis, MN
68. (2015) Faculty Seminar, University of Minnesota, Department of Biostatistics, Minneapolis, MN
69. (2015) Faculty Seminar, Pennsylvania State University, Statistics Department, State College, PA
70. (2015) Faculty Seminar, Texas A&M University, Statistics Department, College Station, TX
71. (2015) Faculty Seminar, Florida State University, Statistics Department, Tallahassee, FL
72. (2015) Faculty Seminar, North Carolina State University, Statistics Department, Raleigh, NC
73. (2014) Faculty Seminar, The Hopkins Department of Biostatistics, Balimore, MD
74. (2014) Faculty Seminar, Cornell University, Department of Statistical Science, Ithaca, NY
75. (2014) Faculty Seminar, Bayes in Paris Seminar, Universite Paris Dauphine, Paris, France
76. (2014) Faculty Seminar, University of Minnesota, Department of Biostatistics, Minneapolis, MN
77. (2014) Privacy and Statistical Databases Conference, Ibiza, Spain
78. (2014) Faculty Seminar, University of Trier, Department of Economics and Social Statistics, Trier, Germany
79. (2014) NSF-Census Research Network Annual Meeting, New York, NY
80. (2014) Small Area Estimation Conference, Poznan, Poland
81. (2014) Faculty Seminar, School of Economics, La Spienza, Universita di Roma, Roma, Italy

82. (2014) Computational Methods for Surveys and Census Data in the Social Sciences, University of Montreal, Montreal, Canada
83. (2014) Frontier of Hierarchical Modeling in Observational Studies, Complex Surveys, and Big Data: A Conference Honoring Professor Malay Ghosh, College Park, MD
84. (2014) Joint Conference in Data Mining in Business and Industry, Duke University, Durham, NC
85. (2014) Seventeenth International Conference on Artificial Intelligence and Statistics, Reykjavik, Iceland
86. (2014) Faculty Seminar, Columbia University, Department of Statistics, New York, NY
87. (2014) Faculty Seminar, Columbia University, Department of Applied Mathematics and Physics, New York, NY
88. (2014) Faculty Seminar, Iowa State University, Department of Statistics
89. (2014) Faculty Seminar, Carnegie Mellon University, Department of Statistics, Pittsburgh, PA
90. (2014) SAMSI Workshop: Censuses and Surveys, Washington, DC
91. (2013) Invited Small Area Estimation Talk, JSM, Montreal, Canada
92. (2013) 5th IMS New Researchers Conference, Montreal, Canada
93. (2013) Center for Survey and Research Methodology Seminar, U.S. Census Bureau, Suitland, MD
94. (2013) Faculty Seminar, UCLA, Los Angeles, CA
95. (2013) Faculty Seminar, Duke University, Department of Statistical Science, Durham, NC
96. (2012) Faculty Seminar, Pennsylvania State University, Department of Statistics, State College, PA
97. (2012) Faculty Seminar, Michigan State University, Department of Statistics and Probability, Lansing, MI
98. (2012) Fields Institute Symposium on the Analysis of Survey Data and Small Area Estimation in honour of the 75th Birthday of Professor J.N.K. Rao, Carleton University, Ottawa, Canada
99. (2012) Faculty Seminar, Carnegie Mellon University, Department of Statistics, Pittsburgh, PA
100. (2012) Faculty Seminar, University of Missouri, Department of Statistics, Columbia, MO
101. (2012) Faculty Seminar, Clemson University, Department of Mathematical Sciences, Clemson, SC
102. (2012) Faculty Seminar, Bucknell University, Lewisburg, PA
103. (2012) Faculty Seminar, Wake Forest University, Winston Salem, NC
104. (2012) Faculty Seminar, Williams College, Williams, MA
105. (2011) Center for Survey and Research Methodology Seminar, U.S. Census Bureau, Suitland, MD

Contributed Conference Presentations and Posters

1. (2017) BNP Poster Session, Paris, France
2. (2014) G70: A Celebration of Alan Gelfand's 70th Birthday, Duke University, Durham, NC
3. (2014) Neural Information Processing Systems, Advances in Variational Inference Workshop, Montreal, CA
4. (2014) Bayes 250 and O'Bayes Meeting, Duke University, Department of Statistical Science
5. (2014) The First Asian ISI Satellite Meeting on Small Area Estimation, Bangkok, Thailand
6. (2014) NSF–Census Research Network Poster Session, NISS Headquarters, Research Triangle Park, NC
7. (2013) New Directions in Monte Carlo Methods Workshop, University of Florida, Gainesville, FL
8. (2012) Women in Machine Learning Workshop, NIPS, Lake Tahoe, NV
9. (2012) Poster Presentation and NIPS Workshops Spotlight and Poster Presentation, NIPS, Lake Tahoe, NV
10. (2012) Contributed Small Area Estimation Talk, JSM, San Diego, CA
11. (2012) ISBA 2012 World Meeting, Kyoto, Japan
12. (2012) Faculty Seminar, University of Florida, Gainesville, FL
13. (2011) Contributed Small Area Estimation Talk, JSM, Miami, FL
14. (2010) Contributed Small Area Estimation Talk, JSM, Vancouver, BC
15. (2010) University of Florida, Department of Statistics Graduate Student Seminar, Gainesville, FL