August 10, 2017

CURRICULUM VITAE

Rebecca C. Steorts

Duke University, Department of Statistical Science
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 best thesis in application
2005	**M.Sc.**, Mathematical Sciences, Clemson University, Clemson, SC
2001	**B.Sc.**, Mathematics, Davidson College, Davidson, NC

PROFESSIONAL EXPERIENCE

2015-present	**Assistant Professor** Department of Statistical Science, Duke University Affiliated faculty in the Department of Computer Science, Department of Biostatistics and Bioinformatics, Information Initiative at Duke (iiD), and the Social Science Research Institute (SSRI) at Duke
2012-2015	**Visiting Assistant Professor** Department of Statistics, Carnegie Mellon University Mentor: Stephen E. Fienberg

HONORS AND AWARDS

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.

[1] 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.

[3] 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 for 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.

[2] 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.

[6] 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 (AIStats)*, 54:298–306, Editors: Aarti Singh and Jerry Zhu, arxiv:1703.02679.

[10] 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.

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

12. **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.

13. 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.

14. Betancourt, B.* and **Steorts, R.** (2017). Bayesian Decision Making with Application to Resource Allocation, *Wiley StatsRef-Statistics Reference Online*, In Press.

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

15. 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.

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

17. 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.**

Publications Under Review and In Preparation (all work done after joining Duke University)

[7] 18. **Steorts, R**., Schmid, T., and Tzavdis, N. (2017). Smoothing and Benchmarking for Small Area Estimation with Application to Rental Prices in Berlin, *Annals of Applied Statistics*, Under Review.

[8] 19. Bai, L.* , Karwa. V., Slavkovic, A., and **Steorts, R.** (2017). Privacy Preserving Algorithm to Release Sparse High-dimensional Histograms, *Neural Information Processing Systems (NIPS)*, Under Review.

[9] 20. Chen, B., Shrivastava, A., and **Steorts, R.** (2017), Unique Entity Estimation with Application to the Syrian Conflict, *Neural Information Processing Systems (NIPS)*, Under Review.

21. Tancredi, A., **Steorts, R.**, and Liseo, B. (2017). Bayesian regression with linked and duplicated data, *Bayesian Analysis*, Under Review.

22. Shen, A.*, McCreanor, R.*, Phelan, M., O'Brien, C., **Steorts, R.**, and Goldstein, B. (2017). Developing a Prediction Model with Dynamic Electronic Health Records Data, *Journal of the Royal Statistical Society: Series A*, Under Review.

GRANTS (all after 2015 obtained at Duke University)

1. CAREER: Scalable Record Linkage through the Microclustering Property, NSF-SES – 1652431, $449,985, 05/15/17–04/15/22. Role: PI.

2. Record Linkage and Privacy-Preserving Methods for Big Data, NSF-SES – 1534412, $265,579; 07/28/15–08/31/18. Role: co-PI.

3. NCRN-MN:Triangle Census Research Network, NSF-SES SES-1131897, PI: Reiter, $4,087,370 over 10/01/2011– 09/30/2017. Role: Investigator.

4. 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.

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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 (all after 2015 done at Duke University)

1. `blink` (Updated 2017) Empirical Bayes Record Linkage and De-duplication R software. Developed and coded by Rebecca C. Steorts, Available at github and CRAN.

2. `italy` (Updated 2017). Record Linkage Database R Packages. Storing publicly available record linkage data sets on CRAN for public dissemination, Available at github and CRAN.

3. (Updated 2017) Microclustering software R, C++ software. Developed and coded by Brenda Betancourt, Jeffrey W. Miller, and Rebecca C. Steorts.

4. `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/.

5. (Updated 2015) Faculty member of Models of Infectious Disease Agent Study (MIDAS) Group (joint with Pittsburgh Bioinformatics). MIDAS scientists are creating synthetic ecosystems that endeavor to contain multiple populations useful for agent based models. Developed and coded by Shannon Gallagher, Jerzy Wieczorek, Lee Richardson, Rebecca C. Steorts, and William F. Eddy, http://www.epimodels.org/drupal/?q=node/32.

MENTORING

Postdoctoral Advisees

Andee Kaplan	2017 – present
Brenda Betancourt (Bernstein-Forster Fellowship)	2015 – present
Nabanita Mukherjee	2015 – 2016

Master's Advisees

Qiaohui Lin	2016 – present
Bai Li	2015 – present
Reuben McCreanor (Statistical Science, MSEM)	2015 – 2017
Sepideh Mosaferi (CMU Statistics)	2012, joint with Steve Fienberg

Undergraduate advisees

Angie Shen	2016 – present, joint with Ben Goldstein
Corey Vernot	2016 – 2017
Peter Sadosky	2014 – 2015
Emily Furnish (CMU Statistics)	2013 – 2014, joint with Steve Fienberg
Stephanie Stern (CMU Statistics)	2013 – 2014
Kairavi Chahal (CMU Statistics)	2013
Dahiana Jiminez (CMU Statistics)	2013

Doctoral thesis committee

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

Jody Heck Wortman (Statistical Science)	2016
Luke Calkins (Pratt School of Engineering)	2017

Master thesis committee

Bai Li (Duke Statistical Science)	2017
Reuben McCreanor (Duke Statistical Science)	2017

Undergraduate thesis committee

Lucy Lu (Statistical Science)	2017
Peter Sadosky (CMU Statistics)	2015
Michael Pane (CMU Statistics)	2013

TEACHING EXPERIENCE AT DUKE

STA 325	Machine Learning and Data Mining
STA 790	Special Topics: Record Linkage
STA 360/602	Modern Advancements of Bayesian Methods
STA 521	Predictive Modeling

PROFESSIONAL APPOINTMENTS AND SERVICE

Peer Review Activities *AIStats*; *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*; *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*; *Statistical Methods and Applications*; *Statistics in Medicine*; *TEST*; *Springer Publishing, New York*; *Transactions of Knowledge and Data Engineering*.

Grant Reviews

2015	National Science Foundation

Conference and Workshop Service

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

UNIVERSITY SERVICE

2017 – present	Faculty Advisory Board of the Duke Human Rights Center
2016 – 2017	Graduate Admission Committee, Department of Statistical Science
2015 – 2017	Tenure Track Search Committee, Department of Statistical Science
2015 – 2017	Seminar Series Coordinator, Department of Statistical Science
2015 – present	Duke Machine Learning Seminar Committee, IID
2015 – 2016	Departmental Computing Committee: Department of Statistical Science
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, Department of Statistics

PRESENTATIONS

Invited Short Courses

1. (2016) *Teaching Bayes: the Essential Parts*. Short Course (3-hour short course), ISBA World Meeting, Sardinia, Italy. Sardinia, Italy.

2. (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.

3. (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. (2018) Faculty Seminar, University of Michigan, Department of Biostatistics, (upcoming)

2. (2017) Faculty Seminar, University of Florida, Gainesville, FL (upcoming)

3. (2018) Faculty Seminar, Harvard University, Department of Statistics and Biostatistics, (upcoming)

4. (2017) International Seminar on Data Editing, Imputation and Non Response, Plenary Talk, CIMAT, Guanajuato, Mexico (upcoming)

5. (2017) INFORMS Healthcare, Rotterdam, Netherlands

6. (2017) BISP, Keynote Talk, Bocconi University, Milano, Italy

7. (2017) Data Plus and Department of Computer Science Undergraduate Seminar, Information Initiative at Duke (iiD), Duke University

8. (2017) AISTATS, Fort Lauderdale, Florida

9. (2016) Neural Information and Professing Systems, Barcelona, Spain

10. (2016) Faculty Seminar, Center for Language and Speech Processing, Johns Hopkins University

11. (2016) Faculty Seminar, Cambridge Biostatistics Unit

12. (2016) Isaac Newton Programme: Data Linkage and Anonymization, Cambridge, UK

13. (2016) Invited Round Table Discussion, JSM, Chicago, IL

14. (2016) Invited Privacy Talk, JSM, Chicago, IL

15. (2016) Faculty Seminar, Bocconi University, Milano, Italy

16. (2016) Statistical Learning and Data Mining (SLDM) Conference, University of North Carolina at Chapel Hill, Chapel Hill, NC

17. (2016) Faculty Seminar, The University of Chicago, The Computational Institute, Chicago, IL

18. (2016) NISS Affiliates Meeting, Austin, TX

19. (2016) Late Breaking News Session, Sixth IMS-ISBA joint meeting Bayes Comp at MCM-Ski V, Lenzerheide, Switzerland

20. (2015) Spotlight presentation, Neural Information and Professing Systems, Bayesian Nonparametrics: The Next Generation Workshops, Montreal, Canada.

21. (2015) Spotlight presentation, EmTech, MIT Media Lab, Boston, MA

22. (2015) Joint Machine Learning MIT & Microsoft Research Seminar, Boston, MA

23. (2015) FOCUS Interdisciplinary Discussion Course For Duke Freshman, Durham, NC

24. (2015) Invited Session at JSM, Boston, MA

25. (2015) Faculty Seminar, University of Padua, Padua, Italy

26. (2015) ITACOSM 2015, 4th ITAlian Conference on Survey Methodology, Rome, Italy

27. (2015) IMS-Microsoft Research Workshop: Foundations of Data Science, Boston, MA

28. (2015) NCRN Spring Meeting, National Academy of Science, Washington, DC

29. (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

30. (2015) Faculty Seminar, Duke University Computer Science, Durham, NC

31. (2015) Faculty Seminar, University of North Carolina at Chapel Hill, Department of Biostatistics, Chapel Hill, NC

32. (2015) Faculty Seminar, Duke University, Department of Statistical Science, Durham, NC

33. (2015) Faculty Seminar, University of California at Berkeley, Statistics Department, Berkeley, CA

34. (2015) Faculty Seminar, University of Minnesota, Statistics Department, Minneapolis, MN

35. (2015) Faculty Seminar, Pennsylvania State University, Statistics Department, State College, PA

36. (2015) Faculty Seminar, Texas A&M University, Statistics Department, College Station, TX

37. (2015) Faculty Seminar, Florida State University, Statistics Department, Tallahassee, FL

38. (2014) Faculty Seminar, The Hopkins Department of Biostatistics, Balimore, MD

39. (2014) Faculty Seminar, Cornell University, Department of Statistical Science, Ithaca, NY

40. (2014) Faculty Seminar, Bayes in Paris Seminar, Universite Paris Dauphine, Paris, France

41. (2014) Faculty Seminar, University of Minnesota, Department of Biostatistics, Minneapolis, MN

42. (2014) Privacy and Statistical Databases Conference, Ibiza, Spain

43. (2014) Faculty Seminar, University of Trier, Department of Economics and Social Statistics, Trier, Germany

44. (2014) NSF-Census Research Network Annual Meeting, New York, NY

45. (2014) Small Area Estimation Conference, Poznan, Poland

46. (2014) Faculty Seminar, School of Economics, La Spienza, Universita di Roma, Roma, Italy

47. (2014) Computational Methods for Surveys and Census Data in the Social Sciences, University of Montreal, Montreal, Canada

48. (2014) Frontier of Hierarchical Modeling in Observational Studies, Complex Surveys, and Big Data: A Conference Honoring Professor Malay Ghosh, College Park, MD

49. (2014) Joint Conference in Data Mining in Business and Industry, Duke University, Durham, NC

50. (2014) Seventeenth International Conference on Artificial Intelligence and Statistics, Reykjavik, Iceland

51. (2014) Faculty Seminar, Columbia University, Department of Statistics, New York, NY

52. (2014) Faculty Seminar, Columbia University, Department of Applied Mathematics and Physics, New York, NY

53. (2014) Faculty Seminar, Iowa State University, Department of Statistics

54. (2014) Faculty Seminar, Carnegie Mellon University, Department of Statistics, Pittsburgh, PA

55. (2014) SAMSI Workshop: Censuses and Surveys, Washington, DC

56. (2013) Invited Small Area Estimation Talk, JSM, Montreal, Canada

57. (2013) 5th IMS New Researchers Conference, Montreal, Canada

58. (2013) Center for Survey and Research Methodology Seminar, U.S. Census Bureau, Suitland, MD

59. (2013) Faculty Seminar, UCLA, Los Angeles, CA

60. (2013) Faculty Seminar, Duke University, Department of Statistical Science, Durham, NC

61. (2012) Faculty Seminar, Pennsylvania State University, Department of Statistics, State College, PA

62. (2012) Faculty Seminar, Michigan State University, Department of Statistics and Probability, Lansing, MI

63. (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

64. (2012) Faculty Seminar, Carnegie Mellon University, Department of Statistics, Pittsburgh, PA

65. (2012) Faculty Seminar, University of Missouri, Department of Statistics, Columbia, MO

66. (2012) Faculty Seminar, Clemson University, Department of Mathematical Sciences, Clemson, SC

67. (2012) Faculty Seminar, Bucknell University, Lewisburg, PA

68. (2012) Faculty Seminar, Wake Forest University, Winston Salem, NC

69. (2012) Faculty Seminar, Williams College, Williams, MA

70. (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