

Joshua Snoke

RAND Corporation | Pittsburgh, PA | jsnoke@rand.org | www.joshuasnoke.com

EDUCATION

Pennsylvania State University

Ph.D., Statistics

08.2018

Graduate Minor: Social Data Analytics

Wheaton College

B.S., Mathematics and Economics

05.2013

Honors: Magna Cum Laude

RESEARCH POSITIONS

RAND Corporation

Statistician

06.2021 - Present

Associate Statistician

08.2018 - 06.2021

Summer Associate

06.2016 - 08.2016

Pennsylvania State University

Graduate Research Assistant, Department of Statistics

01.2014 - 06.2018

Graduate Research Trainee, Big Data Social Science IGERT

09.2014 - 05.2016

Administrative Data Research Centre - Scotland

Research Fellow

05.2015 - 08.2015

PEER REVIEWED PUBLICATIONS

Bowen, C. M. and **J. Snoke** (2021). Comparative Study of Differentially Private Synthetic Data Algorithms from the NIST PSCR Differential Privacy Synthetic Data Challenge. *Journal of Privacy and Confidentiality* 11 (1). <https://doi.org/10.29012/jpc.748>.

Snoke, J., G. M. Raab, B. Nowok, C. Dibben, and A. Slavković (2018). General and Specific Utility Measures for Synthetic Data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(3), pp.663-688. doi:10.1111/rssa.12358

Snoke, J., T. R. Brick, A. Slavković, and M. D. Hunter (2018). Providing Accurate Models Across Private Partitioned Data: Secure Maximum Likelihood Estimation. *The Annals of Applied Statistics*, 12(2), pp.877-914. doi:10.1214/18-AOAS1171.

Snoke, J. and A. Slavković (2018). *pMSE* Mechanism: Differentially Private Synthetic Data with Maximal Distributional Similarity. In *Privacy in Statistical Databases*. pp.138-159. Springer. doi:10.1007/978-3-319-99771-1

Snoke, J., T. R. Brick, and A. Slavković (2016). Accurate Estimation of Structural Equation Models with Remote Partitioned Data. In *Privacy in Statistical Databases*. pp.190-209. Springer. doi:10.1007/978-3-319-45381-1

ARTICLES

Snoke, J. and C. M. Bowen (2020). How Statisticians Should Grapple with Privacy in a Changing

Data Landscape. *CHANCE*, 33(4), 6-13. <https://doi.org/10.1080/09332480.2020.1847947>

Snoke, J. and C. M. Bowen (2019). Differential Privacy: What Is It? *Amstat News*. 26-28. Issue #501. <https://magazine.amstat.org/blog/2019/03/01/differentialprivacy/>

POLICY REPORTS

Peterson, S., D. Barnes-Proby, K. E. Bouskill, L. M. Davis, M. L. Mizel, B. A. Weidmer, I. Leamon, A. Mendoza-Graf, M. Strawn, **J. Snoke**, and T. E. Goode (2021). Understanding Subgroups Within the Los Angeles County Sheriff's Department: Community and Department Perceptions with Recommendations for Change. *RAND Corporation*. RRA616-1.

Matthews, M., A. R. Morral, T. L. Schell, M. Cefalu, **J. Snoke**, and R. J. Briggs (2021). Organizational Characteristics Associated with Risk of Sexual Assault and Sexual Harassment in the U.S. Army. *RAND Corporation*. RRA1013-1.

Kaufman, J. H., M. K. Diliberti, G. P. Hunter, **J. Snoke**, D. M. Grant, C. M. Setodji, and C. J. Young (2021). COVID-19 and the State of K-12 Schools: Results and Technical Documentation from the Spring 2021 American Educator Panels COVID-19 Surveys. *RAND Corporation*. RRA168-7.

Ayer, L., D. J. Schultz, M. Abbott, D. Barnes-Proby, W. Y. Chan, M. S. Dunbar, E. Hoch, H. H. Liu, M. Martineau, E. Ohana, D. Siconolfi, **J. Snoke**, C. Stevens, and V. L. Towe (2020). Mental Health Task-shifting in Community-based Organizations: Implementation, Impact, and Cost-evaluation of the Connections to Care Program. *RAND Corporation*. RR3083.

PREPRINTS

Barrientos, A. F., A. R. Williams, **J. Snoke**, and C. M. Bowen. A Feasibility Study of Differentially Private Summary Statistics and Regression Analyses for Administrative Tax Data. Preprint available at <https://arxiv.org/abs/2110.12055> *Under Review*.

Cabreros, I., **J. Snoke**, O. Osonde, I. Khan, and M. N. Elliott. Advancing Equitable Decision-Making for the Department of Defense Through Fairness in Machine Learning. *Forthcoming*.

Matthews, M., C. S. Sims, S. Robson, M. Walsh, S. Rennane, and **J. Snoke**. Physical Fitness Standards to Support Readiness and Deployability: Examination of Current Department of the Air Force Policies and Culture. *Forthcoming*.

Robson, S. M., M. C. Lytell, M. Walsh, K. C. Hall, K. M. Keller, V. Kilambi, **J. Snoke**, J. Welburn, P. Roberts, O. Hall. Air Force Enlisted Classification and Reclassification: Potential Improvements using Machine Learning and Optimization Models. *Forthcoming*.

Ororbia II, A. G., F. Linder, and **J. Snoke**. Using Neural Generative Models to Release Synthetic Twitter Corpora with Reduced Stylometric Identifiability of Users. Preprint available at <https://arxiv.org/abs/1606.01151>

GRANTS & CONTRACTS

Development and Deployment of a Validation Server for Administrative Tax Data
Funder: Urban Institute, Prime: Sloan Foundation - \$100,000 11.2020 - 10.2022

Planning and Development of a Prototype Privacy-Preserving Validation Server for Administrative Tax Data

Funder: Urban Institute, Prime: National Science Foundation - \$15,000 10.2020 - 09.2021

Comparative Study of Differentially Private Synthetic Data Algorithms and Evaluation Standards
Funder: National Institute of Standards and Technology - \$100,000 03.2020 - 11.2020

PROFESSIONAL ACTIVITIES

Professional Service

Human Subjects Protection Committee 06.2019 - Present
RAND Corporation

Committee on Privacy and Confidentiality 10.2018 - Present
American Statistical Association (ASA)

Subject Matter Expert for Differentially Private Synthetic Data Challenges 10.2018 - 05.2019
National Institute of Standards and Technology (NIST) 10.2020 - 05.2021

Academic Journal Reviewer

Annals of Applied Statistics
Biostatistics
Harvard Data Science Review
IEEE Transactions on Information Forensics & Security
Journal of Machine Learning
Journal of Privacy and Confidentiality
Journal of Research on Educational Effectiveness
Journal of the Royal Statistical Society: Series A
Journal of Survey Statistics and Methodology
Statistical Science
Statistics and Computing
Science Advances

Conference Program Committees

Symposium on Data Science and Statistics (SDSS) 06.2020
ACM-IMS Foundations of Data Science (FODS) 10.2020

FELLOWSHIPS & AWARDS

Spotlight Award 01.2021
For Leadership within the RAND Statistics Group
RAND Corporation

Dissertation Fellowship 08.2016 - 07.2017
U.S. Census Bureau - \$50,000

NSF Big Data Social Science IGERT Traineeship 09.2014 - 08.2016
Pennsylvania State University - \$60,000

Joint Statistical Meetings ASA Student Paper Award 07.2016
General and Specific Utility Measures for Synthetic Data
Social Statistics, Government Statistics, and Survey Research Methods Section

Rao Prize Conference Student Poster Award 05.2015
Estimation of Structural Equation Models for Vertically Partitioned Data
Pennsylvania State University

President's Award
Wheaton College

08.2009 - 05.2013

SELECTED PRESENTATIONS

Invited Talks

University of Pittsburgh Statistics Department 09.2021
The Statistical Underpinnings of Modern Data Privacy and Confidentiality

Society for Industrial and Applied Mathematics: Mathematics of Data Science 05.2020
Comparative Study of Differentially Private Synthetic Data Algorithms and Evaluation Standards

Simons Institute Program on Data Privacy: Foundations and Applications 03.2019
Statistical Perspectives on Differentially Private Synthetic Data

International Conference on Privacy in Statistical Databases 09.2018
pMSE Mechanism: Differentially Private Synthetic Data with Maximal Distributional Similarity

U.S. Census Bureau 06.2018
Statistical Data Privacy Methods for Increasing Research Opportunities

Joint Statistical Meetings 07.2017
Discussant. Differential Privacy in Statistical Agencies: Present and Future

International Statistical Institute World Statistics Congress 07.2017
Perspective on Utility for Synthetic Data Sets

International Conference on Privacy in Statistical Databases 09.2016
Accurate Estimation of Structural Equation Models with Remote Partitioned Data

Joint Statistical Meetings 08.2016
General and Specific Utility Measures for Synthetic Data

Workshop Talks

Penn State Statistics: Stochastic Modeling and Computational Statistics Seminar 04.2017
Secure Multiparty Maximum Likelihood Estimation with Partitioned Databases

Isaac Newton Institute Synthetic Data Workshop 11.2016
Beyond Microdata: Approaches for Generating Synthetic Tweets

TEACHING APPOINTMENTS

Pennsylvania State University

Graduate Teaching Assistant 08.2017 - 12.2017
Categorical Data Analysis

Applied Data Mining and Statistical Learning 08.2017 - 12.2017

Elementary Statistics 08.2013 - 12.2013

Wheaton College

Teaching Assistant 08.2011 - 05.2012
Intermediate Macroeconomics

Calculus II 01.2011 - 05.2011

COMPUTING SKILLS

Software

Proficient in R, Python, Git, Excel

Familiarity with SAS, STATA, Minitab, Java, SQL

Development

Contributor to R package *synthpop*