Joshua Snoke

RAND Corporation | Pittsburgh, PA | jsnoke@rand.org

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
Visiting Research Fellow 05.2015 - 08.2015

PROJECT LEADERSHIP

Funder: National Science Foundation - $125,000
Principal Investigator (Subcontract)
Creation of Synthetic Data for the Survey of Earned Doctorates 02.2024 - 10.2025

Funder: Robert Wood Johnson Foundation - $299,611
Co-Principal Investigator
Toolkit for Detecting and Correcting Bias in Healthcare Algorithms 10.2022 - 09.2024

Funder: National Science Foundation - $384,340
Principal Investigator (Subcontract)
Building a Validation Server for Administrative Tax Data 10.2020 - 03.2025

Funder: National Institutes of Health - $225,000
Principal Investigator (Subcontract)
Development of a Synthetic Survey of Health and Aging 06.2022 - 03.2024

Funder: Tableau Foundation - $21,000
Principal Investigator (Subcontract)
Equity Awareness and Best Practices in Data Privacy 04.2022 - 12.2022

Funder: Sloan Foundation - $100,000
Principal Investigator (Subcontract)
Development and Deployment of a Validation Server for Administrative Tax Data 11.2020 - 10.2022

Funder: RAND Corporation - $139,000
Co-Principal Investigator
Addressing Potential Algorithmic Bias in DoD Machine Learning Applications 10.2020 - 09.2021

Funder: National Institute of Standards and Technology - $100,000
Principal Investigator (Subcontract)
Comparative Study of Differentially Private Synthetic Data Algorithms and Evaluation Standards
03.2020 - 11.2020

PEER-REVIEW PUBLICATIONS

Snoke, J., A. Haas, S. C. Martino, and M. N. Elliott. "Differential Privacy Protections in 2020 U.S. Decennial Census Data Do Not Impede Measurement of Racial and Ethnic Disparities'.' *Medical Care Research and Review. Forthcoming.*

Williams, A. R., **J. Snoke**, C. M. Bowen, and A. F. Barrientos. "Disclosing Economists Privacy Perspectives: A Survey of American Economic Association Members on Differential Privacy and Data Fitness for Use Standards". *Harvard Data Science Review. Forthcoming.*

Snoke, J., C. M. Bowen, A. R. Williams, and A. F. Barrientos. "Incompatibilities Between Current Practices in Statistical Data Analysis and Differential Privacy". *Journal of Privacy and Confidentiality. Forthcoming.*

Snoke, J. and S. K. Kinney. "Methods for Synthetic Data Generation". Book chapter. *Forthcoming.*

Barrientos, A. F., A. R. Williams, **J. Snoke**, and C. M. Bowen (2023). "A Feasibility Study of Differentially Private Summary Statistics and Regression Analyses with Evaluations on Administrative and Survey Data". *Journal of the American Statistical Association*, pp.1-14.

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).

Snoke, J. and C. M. Bowen (2020). "How Statisticians Should Grapple with Privacy in a Changing Data Landscape". *CHANCE*, 33(4), 6-13.

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.

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.

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.

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.

PREPRINTS

Wastvedt, S., **J. Snoke**, D. Agniel, J. Lai, M. N. Elliott, S. C. Martino. "De-Biasing the Bias: Methods for Improving Disparity Assessments with Noisy Group Measurements". Available at `https://arxiv.org/abs/2402.13391`

Snoke, J., E. Meijer, D. Phillips, J. Wilkens, J. Lee. "Synthesizing Surveys with Multiple Units of Observation: An Application to the Longitudinal Aging Study in India". Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4569904`

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

POLICY REPORTS AND ARTICLES

STATISTICAL DATA PRIVACY

Bowen, C. M. and **J. Snoke** (2023). "Do No Harm Guide: Applying Equity Awareness In Data Privacy Methods. Urban Institute.

Snoke, J. and C. M. Bowen (2019). "Differential Privacy: What Is It?" *Amstat News*. 26-28. Issue #501.

EQUITABLE MACHINE LEARNING

J. Snoke, M. Walsh, J. Williams, and D. Schulker (2024). "Safe Use of Machine Learning for Air Force Human Resource Management Evaluation Framework and Use Cases." RAND Corporation. RRA1745-4.

D. Schulker, M. Walsh, A. Calkins, M. C. Graham, C. Montemayor, A. A. Robbert, S. Robson, C. M. Setodji, **J. Snoke**, J. Williams, L. A. Zhang (2024). "Leveraging Machine Learning to Improve Human Resource Management." RAND Corporation. RRA1745-1.

Cabreros, I., **J. Snoke**, O. Osonde, I. Khan, and M. N. Elliott (2023). "Advancing Equitable Decision-Making for the Department of Defense Through Fairness in Machine Learning." RAND Corporation. RRA1542-1.

WORKFORCE DEVELOPMENT

Krueger, Tracy C., S. Robson, **J. Snoke**, M. Walsh, M. Strawn, A. Atler, J. H. Campell, I. Leamon, R. Haberman, and B. Bicksler (2022). "Assessment and Selection for U.S. Air Force Special Warfare." RAND Corporation. RRA549-1.

Robson, S., M. Walsh, M. Matthews, C. S. Sims, and **J. Snoke** (2022). "Is Today's U.S. Air Force Fit? It Depends on How Fitness is Measured." RAND Corporation. RRA552-1.

Matthews, M., C. S. Sims, S. Robson, M. Walsh, S. Rennane, and **J. Snoke** (2022). "Physical Fitness Standards to Support Readiness and Deployability: An Examination of Current Department of the Air Force Policies and Culture." RAND Corporation. RRA552-2.

Robson, S. M., M. C. Lytell, M. Walsh, K. C. Hall, K. M. Keller, V. Kilambi, **J. Snoke**, J. Welburn, P. Roberts, O. Hall (2022). "U.S. Air Force Enlisted Classification and Reclassification: Potential Improvements using Machine Learning and Optimization Models." RAND Corporation. RRA284-1.

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.

Matthews, M., A. R. Morral, T. L. Schell, M. Cefalu, **J. Snoke**, and R. J. Briggs, A. Calkins, L. Cottrell, S. O. Meadows, and R. L. Collins (2021). "Characteristics of Sexual Assault and Sexual Harassment in the U.S. Army: Implications for Prevention Efforts." RAND Corporation. RBA1385-1.

Matthews, M., A. R. Morral, T. L. Schell, M. Cefalu, **J. Snoke**, and R. J. Briggs (2021). "Sexual Assault and Sexual Harassment in the U.S. Army: Where Cases are Highest and Why." RAND Corporation. RBA1013-1.

EDUCATION, HEALTH, & JUSTICE

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.

Doan, S., M. Fernandez, D. M. Grant, J. H. Kaufman, C. M. Setodji, **J. Snoke**, M. Strawn, and C. J. Young (2021). "American Instructional Resources Surveys: 2021 Technical Documentation and Survey Results." RAND Corporation. RRA134-10.

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.

Kaufman, J. H., M. K. Diliberti, G. P. Hunter, D. M. Grant, L. S. Hamilton, H. L. Schwartz, C. M. Setodji, **J. Snoke**, and C. J. Young (2020). "COVID-19 and the State of K-12 Schools: Results and Technical Documentation from the Fall 2020 American Educator Panels COVID-19 Surveys." RAND Corporation. RRA168-5.

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.

FELLOWSHIPS & AWARDS

2023 Privacy Papers for Policymakers Award — 02.2024
Future of Privacy Forum

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 Statsitics, 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 08.2009 - 05.2013
Wheaton College

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

Conference Program Committees

Privacy and Policy Conference 09.2024
Symposium on Data Science and Statistics (SDSS) 06.2020
ACM-IMS Foundations of Data Science (FODS) 10.2020

Academic Journal Reviewer

Annals of Applied Statistics
Biostatistics
Harvard Data Science Review
IEEE Transactions on Information Forensics & Security
Journal of Machine Learning
Journal of Official Statistics
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
Statistica Sinica
Statistical Science
Statistics and Computing
Science Advances
The American Statistician

Invited Conference Sessions Organized

Joint Statistical Meetings 08.2023
Statistically Significant: Equity Concerns in Algorithmic Bias, Privacy, and Survey Representation

SELECTED PRESENTATIONS

Short Courses

Symposium on Data Science and Statistics 06.2022
Statistical Data Privacy Techniques for Sharing Sensitive Data

Invited Talks

American Statistical Association Statistical Consulting Section 09.2023
Navigating Privacy Concerns in the Practice of Statistics: A Necessary Part of the Consulting Toolkit

UCLA Synthetic Data Workshop 04.2023
Applied Methods in Synthetic Data for Structured Surveys

NISS-IOF: Advancing Demographic Equity with Privacy Preserving Methodologies 01.2023
Considering Equity in the Practice of Statistical Data Privacy

Joint Statistical Meetings 08.2022
Disclosing Economists' Perspectives on Privacy

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
Categorical Data Analysis 08.2017 - 12.2017
Applied Data Mining and Statistical Learning 08.2017 - 12.2017

Elementary Statistics 08.2013 - 12.2013

Wheaton College
Teaching Assistant
Intermediate Macroeconomics 08.2011 - 05.2012
Calculus II 01.2011 - 05.2011

COMPUTING

Software
Proficient in R, Python, Git, Excel Familiarity with SAS, STATA, Minitab, Java, SQL

Development
Contributor to R package *synthpop*

Author of RAND Algorithmic Equity Tool:
`https://github.com/RANDCorporation/algorithmic-equity-tool/tree/main`