NIH Biographical Sketch Common Form

Name: Purkayastha, Soumik

Persistent Identifier (PID) of the Senior/Key Person: https://orcid.org/0000-0002-3619-2804

Position Title: Assistant Professor

Organization and Location:University of Pittsburgh, Pittsburgh, Pennsylvania, United States

PROFESSIONAL PREPARATION

INSTITUTION AND LOCATION	DEGREE	Start Date	Completion Date	FIELD OF STUDY
University of Michigan, Ann Arbor, Michigan, United States	Doctor of Philosophy (PHD)	07/2019	07/2024	Biostatistics
University of Michigan, Ann Arbor, Michigan, United States	Master of Science (MS)	07/2019	04/2021	Biostatistics
Indian Statistical Institute, Kolkata, West Bengal, India	Master of Statistics (MSTAT)	07/2017	06/2019	Statistics (specialization in Biostatistics)
St. Xavier's College (Autonomous), Kolkata, West Bengal, India	Bachelor of Science (BS)	07/2014	06/2017	Statistics

Appointments and Positions

2024 - present	Assistant Professor, University of Pittsburgh, Pittsburgh, Pennsylvania, United States
2026 - present	Secretary, Health Policy in Statistics Section, American Statistical Association, Pittsburgh, Pennsylvania, United States
2024 - present	Research Biostatistician, Center for Healthcare Evaluation, Research, and Promotion, VA Pittsburgh Healthcare System, US Department of Veteran Affairs, , Pittsburgh, Pennsylvania, United States
2023 - 2024	Rackham Predoctoral Fellow, University of Michigan, Rackham Graduate School, Ann Arbor, Michigan, United States
2021 - 2021	AI/ML Intern, Apple Inc, Siri Data, Cupertino, California, United States
2019 - 2023	Graduate Student Research Assistant, University of Michigan, Department of Biostatistics, Ann Arbor, Michigan, United States
2018 - 2018	Data Science Intern, Walmart Labs, Bengaluru, Karnataka, India

Products

Products Closely Related to the Proposed Project

1. Purkayastha S, Song P. Quantification and cross-fitting inference of asymmetric relations under generative exposure mapping models. Statistica Sinica. 2026 January. DOI: 10.5705/ss.202025.0236

2. Purkayastha S. A Framework for Covariate-Adjusted Bivariate Causal Discovery. arXiv: arXiv.2502.10317 [Preprint]. 2025 July . DOI: 10.48550/arXiv.2502.10317

3. Purkayastha S, Song P. fastMI: A fast and consistent copula-based nonparametric estimator of mutual information. Journal of Multivariate Analysis. 2024 May 01; 201:105270. DOI: 10.1016/j.jmva.2023.105270

4. Lamba S, Frank DA, McCoy JL, Purkayastha S, Leder SM, Russell LE, Gordon JH, Procario GT, Moy EM, Hausmann LRM. Health-Related Social Needs Among LGB+ Veterans. JAMA Netw Open. 2025 Oct 1;8(10):e2539986. PubMed Central PMCID: PMC12573039.

5. Russell LE, Frank DA, Purkayastha S, McCoy JL, Leder SM, Gordon JH, Lamba S, Procario GT, Hausmann LR. Racial, Ethnic, and Sex Differences in Social Risks and Social Needs Concordance Among Veterans. JAMA network open. Forthcoming.

Other Significant Products Highlighting Contributions to Science

1. Zhang L, Purkayastha S, Lev-Tov H, Nie R, Kirsner R, Spino C, Song PXK. Determinants of enrolment rate in 397 clinical trials for healing diabetic foot ulcers: a systematic review. BMJ Open. 2025 Jul 6;15(7):e095512. PubMed Central PMCID: PMC12230948.

2. Salvatore M, Purkayastha S, Ganapathi L, Bhattacharyya R, Kundu R, Zimmermann L, Ray D, Hazra A, Kleinsasser M, Solomon S, Subbaraman R, Mukherjee B. Lessons from SARS-CoV-2 in India: A data-driven framework for pandemic resilience. Sci Adv. 2022 Jun 17;8(24):eabp8621. PubMed Central PMCID: PMC9205583.

3. Bhaduri R, Kundu R, Purkayastha S, Kleinsasser M, Beesley LJ, Mukherjee B, Datta J. Extending the susceptible-exposed-infected-removed (SEIR) model to handle the false negative rate and symptom-based administration of COVID-19 diagnostic tests: SEIR-fansy. Stat Med. 2022 Jun 15;41(13):2317-2337. PubMed Central PMCID: PMC9035093.

4. Purkayastha S, Kundu R, Bhaduri R, Barker D, Kleinsasser M, Ray D, Mukherjee B. Estimating the wave 1 and wave 2 infection fatality rates from SARS-CoV-2 in India. BMC Res Notes. 2021 Jul 8;14(1):262. PubMed Central PMCID: PMC8264482.

5. Purkayastha S, Bhattacharyya R, Bhaduri R, Kundu R, Gu X, Salvatore M, Ray D, Mishra S, Mukherjee B. A comparison of five epidemiological models for transmission of SARS-CoV-2 in India. BMC Infect Dis. 2021 Jun 7;21(1):533. PubMed Central PMCID: PMC8181542.

Certification:

I certify that the information provided is current, accurate, and complete. This includes but is not limited to information related to domestic and foreign appointments and positions.

I also certify that, at the time of submission, I am not a party to a malign foreign talent recruitment program.

Misrepresentations and/or omissions may be subject to prosecution and liability pursuant to, but not limited to, 18 U.S.C. §§ 287, 1001, 1031 and 31 U.S.C. §§ 3729-3733 and 3802.

Certified by Purkayastha, Soumik in SciENcv on 2026-02-05 08:31:25

NIH BIOGRAPHICAL SKETCH SUPPLEMENT

Name: Purkayastha, Soumik

Persistent Identifier (PID) of the Senior/Key Person: https://orcid.org/0000-0002-3619-2804

Position Title: Assistant Professor

Organization and Location: University of Pittsburgh, Pittsburgh, Pennsylvania, United States

Personal Statement
eRA Username: SOUMIKP

I have been deeply involved in developing robust statistical frameworks for complex, real-world health data. A majority of my methodological research has focused on leveraging large-scale observational data for mechanistic understanding, specifically in (1) causal discovery and structure learning, (2) spatiotemporal modeling of infectious diseases, and (3) the rigorous analysis of complex survey data. I have published over 15 papers in the last 5 years (2020–2025). Most of these publications appear in high-impact peer-reviewed journals in statistics and health sciences, including Science Advances, Statistica Sinica, and BMJ Open. I have experience in developing accessible statistical tools and turning them into open-source programs, having developed multiple R packages such as fastMI and SEIRfansy. My expertise in analyzing data with complex survey designs was further honed through my work in health services research at the Department of Veterans Affairs (VA). I am also devoted to training students on methodological research; I currently supervise multiple doctoral students whose work focuses on the intersection of machine learning and structure learning. **I am an Early-Stage Investigator by the definition of NIH.**

For the proposed project, I am uniquely positioned to resolve the "chicken-or-egg" dilemma in immunopsychiatry. My recent methodological work, including the development of the fast Fourier transformation based estimation of mutual information (fastMI) and Conditional Generative Exposure Mapping (CGEM) framework, provides the statistical foundation for the Longitudinal Weighted Structure Learning (LoWSL) and Weighted Conditional Mutual Information (WeightMI) methods proposed in this application. These frameworks will allows for the rigorous identification of signaling pathways and mediators on said pathways in the presence of heavy-tailed biological data (e.g., CRP, IL-6) and complex sampling designs, which standard algorithms cannot handle.

In addition to my academic engagements, I serve as Secretary of the Health Policy and Statistics Section of the American Statistical Association. I engage with the academic community through memberships in the American Statistical Association, International Biometric Society, and Institute of Mathematical Statistics.

Honors
2023	Rising Star Award, University of Michigan
2023	Best Paper Award, International Biometric Society, Western North American Region
2020	Richard G. Cornell Fellowship, University of Michigan
2019	Sabyasachi Roy Memorial Gold Medal, Indian Statistical Institute
2023 - 2024	Rackham Predoctoral Fellowship, University of Michigan
2021 - 2024	Rackham Conference Travel Award, University of Michigan

Contributions to Science

1. **Information-Theoretic Frameworks for Structure Learning in Observational Data:** My primary methodological research addresses the limitations of standard structure learning when applied to complex, non-idealized data. I have developed novel statistical frameworks that utilize information theory (specifically Shannon entropy and mutual information) to detect signaling asymmetry and directional influence without relying on restrictive linearity or Gaussian assumptions. This includes the development of the Conditional Generative Exposure Mapping (CGEM) model and the fastMI algorithm for non-parametric mutual information estimation. These methods are critical for the proposed project, as they allow for the disentangling of feedback loops (e.g., depression-inflammation) in the presence of heavy-tailed distributions and high-dimensional confounders.

2. **Spatiotemporal Modeling of Infectious Disease:** During the COVID-19 pandemic, I led the development of compartmental

transmission models to forecast disease trajectory in India. This work involved handling high levels of data uncertainty and missingness to generate counterfactual evidence supporting early public health interventions. This experience in spatiotemporal modeling and handling noisy epidemiological data directly informs my approach to analyzing longitudinal cohort data such as Add Health.

3. **Health Services Research:** In collaboration with clinical investigators at the VA Pittsburgh Healthcare System, I apply advanced biostatistical methods to investigate variation in health and social determinants of health. This research line demonstrates my ability to work within multidisciplinary teams to translate complex data into policy-relevant findings. In my role within VA health services research, I have developed specialized proficiency in the rigorous analysis of survey data characterized by complex sampling designs.

Certification:

I certify that the information provided is current, accurate, and complete. This includes but is not limited to information related to domestic and foreign appointments and positions.

I also certify that, at the time of submission, I am not a party to a malign foreign talent recruitment program.

Misrepresentations and/or omissions may be subject to prosecution and liability pursuant to, but not limited to, 18 U.S.C. §§ 287, 1001, 1031 and 31 U.S.C. §§ 3729-3733 and 3802.

Certified by Purkayastha, Soumik in SciENcv on 2026-02-0508:31:25