Star Liu

U.S. Citizen   👤 Personal Website   🗋 Latest CV   🗋 Latest Resume   📣 Google Scholar   @ sliu197@jhmi.edu   📱 +470 538 5974

🏛 Education

2023-Present	**Johns Hopkins School of Medicine (JHSOM)**
Degree	PhD in Biomedical Informatics and Data Science
2021-2023	**JHSOM**
Degree	Master of Science in Biomedical Informatics and Data Science Research
Advisor	Harold P. Lehmann
Thesis	*Model Usefulness: Aligning Utility and Prevalence in Use and in Training of Clinical Machine Learning Models*
2017-2021	**Emory University**
Degree	Bachelor of Science in Quantitative Sciences, Biology Concentration
Advisor	Seunghwa Rho
Capstone	*Building a Predictive Model for Fire Incident Risk in Atlanta*

💻 Competencies

Programming	Python, R, SQL, HTML, JavaScript, CSS, Java
Frameworks	Flask, Dash, React JS
Databases	Microsoft SQL Server, CosmosDB, MongoDB
Vocabularies	ICD-9, ICD-10, RxNorm, LOINC
Data Sources	EHR, Claims, Healthcare Cost and Utilization Project, Medical Expenditure Panel Survey
Softwares	GitHub, Overleaf, Microsoft Excel, Slurm, FileZilla, Tableau, QGIS, Google Analytics

📄 Publications and Preprints

Publication - Journal and Conference

1. **Liu S**, Golozar A, Buesgens N, McLeggon JA, Black A, Nagy P. A framework for understanding an open scientific community using automated harvesting of public artifacts. *JAMIA Open*. 2024;7(1):ooae017. Published 2024 Feb 29. doi:10.1093/jamiaopen/ooae017 🗗 PMID : 38425704

2. Hassan S, **Liu S**, Patel SA, Emmert-Fees K, Suvada K, Tandon N, Sosale A, Anjana RM, Mohan V, Chwastiak L, and Ali MK. Association of collaborative care intervention features with depression and metabolic outcomes in the INDEPENDENT study : A mixed methods study. *Prim Care Diabetes*. Published online February 14, 2024. doi:10.1016/j.pcd.2024.02.001 🗗 PMID : 38360505

3. **Liu S**, Wei S, Lehmann HP. Applicability Area : A novel utility-based approach for evaluating predictive models, beyond discrimination. *AMIA Annu Symp Proc*. 2024;2023:494-503. Published 2024 Jan 11. 🗗 PMID : 38222359

4. Benameur K, Gross PJ, **Liu S**, Koneru S. The Mediterranean Soul Food Diet Intervention After Stroke, a Feasibility Study in the Stroke Belt. *American Journal of Lifestyle Medicine*. 2023;0(0). 🗗 Sage Journal Article : 10.1177/15598276231208383

5. **Liu S**, Ding X, Belouali A, Bai H, Raja K, Kharrazi H. Assessing the Racial and Socioeconomic Disparities in Postpartum Depression Using Population-Level Hospital Discharge Data : Longitudinal Retrospective Study. *JMIR Pediatrics Parenting*. 2022;5(4):e38879. Published 2022 Oct 17. 🗗 PMID : 36103575

6. Belouali A, Bai H, Raja K, **Liu S**, Ding X, Kharrazi H. Impact of Social Determinants of Health on Improving the LACE Index for 30-Day Unplanned Readmission Prediction. *JAMIA Open*. 2022;5(2):ooac046. Published 2022 Jun 10. 🗗 PMID : 35702627

7. Sood P, **Liu S**, Lehmann HP, Kharrazi H. Assessing the Effect of Electronic Health Record Data Quality on Identifying Patients with Type 2 Diabetes. *JMIR Medical Informatics*. 🗗 PMID : 38850555

Publication - Under Review

8. **Liu S** and Martin B, Park WY, Northington F, Valdez RC, Makker K, Kuiper J, Nagy PG, Minty E and Aziz KB. Phenotype Development and Evaluation for Neonatal Hypoxic Ischemic Encephalopathy Using Electronic Health Record. *Frontiers in Neurology*. [Under Review]

9. Li ZY, **Liu S**, Ho JC, Narayan KMV, Ali MK, Varghese JS. Translating Subphenotypes of Newly Diagnosed Type 2 Diabetes from Cohort Studies to Electronic Health Records in the United States *Nature Medicine*. 🗗 Preprint

10. Sood P, **Liu S**, Kharrazi H. Evaluating the Effect of Type-2 Diabetes Phenotype Definitions in Identifying Populations of Interest for Clinical Studies and Research. *BMJ Open Diabetes and Care*. [Under Review] 🗗 Link to Manuscript

11. Sood P, **Liu S**, Pandya C, Kharrazi H. Assessing the Impact of Computable Type-2 Diabetes Phenotypes on Predicting Healthcare Utilization using Electronic Health Records and Administrative Claims. *Health Services Research*. [Under Review]

12. Sood P, **Liu S**, Kharrazi H. Assessing the racial bias in Predicting Healthcare Utilization Stemmed from Variations in Diabetes Phenotype Definitions. *JAMIA Open*. [Under Review]

*See latest CV for updates (found in the header).

🎤 Presentations, Posters, and Case Studies

1. **Liu S**, Lehmann HP. Clinical Utility Profiling (CUP) : A practical method for choosing predictive model cutoff ranges based on target deployment. *AMIA Annual Symposium*, 2024. [Podium Presentation] 🔗 Link to Slide and Abstract

2. **Liu S**, Lehmann HP. Identifying and Visualizing the Optimal Cutoff on the Receiver Operating Characteristic (ROC) Curve. *SMDM*. 2024. [Poster] 🔗 Link to Poster

3. **Liu S**, Wei SX, Lehmann HP. Applicability Area : A Novel Utility-based Approach for Evaluating Predictive Models Beyond Discrimination. *AMIA Annual Symposium*, 2023. [Podium Presentation] 🔗 Link to Podium Presentation

4. **Liu S**, Lehmann HP. Assessing the Dimensions of a Dataset Most Sensitive to Cost-Sensitive Loss Functions : A Simulation Study. *SMDM*. 2023. [Poster] 🔗 Link to Poster

5. Sood P, **Liu S**, Kharrazi H. Evaluating the Variations in Healthcare Utilization Predictions Across Common Phenotypes of Type 2 Diabetes Using EHR and Claims Data. *AMIA Annual Symposium*. 2023. [Poster] 🔗 Link to Poster

6. Sood P, **Liu S**, Kharrazi H. Measuring the Effect of EHR Data Quality in Identifying Type-2 Diabetes Population Across Common Phenotype Definitions of Diabetes. *AMIA Annual Symposium*. 2022. [Poster] 🔗 Link to Poster

7. **Liu S**, Golozar A, McLeggon JA, Black A, Nagy P. The OHDSI Community Dashboard : Tracking the Health and Impact of the Open Science Observational Health Data Sciences and Informatics Community. *Observational Health Data Sciences and Informatics (OHDSI) Symposium*. 2022. [Software Demonstration] 🔗 Link to Abstract   🔗 Link to Software

8. **Liu S**, Wall E, Patel SA, Park Y. COVID-19 Health Equity Dashboard - Addressing Vulnerable Populations. *IEEE Visualization for Communication Workshop*. 2020. doi:10.31219/osf.io/2frha. [Case Study] 🔗 Link to Case Study

☰ Research Experience

Present - **January 2023**	**Graduate Researcher | PI : Jim C. Fackler | JHSOM, Baltimore, MD** Estimating the risk of adverse drug events (ADEs) associated with broad-spectrum antibiotics among pediatric patients with sepsis 　› Conduct cohort discovery and describe the risk of ADE across partner OMOP databases 　› Define the risk profiles of ADEs and support broad-spectrum antibiotics administration 　› Predict the risk of ADEs occurrence among pediatric patients with sepsis ADEs ｜ Antibiotics Stewardship ｜ Pediatrics ｜ Sepsis ｜ OMOP ｜ Network Study ｜ Phenotyping
Present - **June 2023**	**Graduate Researcher | PI : Khyzer B. Aziz | JHSOM, Baltimore, MD** Supporting antibiotics stewardship by detecting signals in nSOFA scores (neonatal severity of illness) 　› Engineered features from nSOFA scores and queried Neonatal Intensive Care Unit (NICU) data using SQL 　› Conducted cross-sectional (Random Forest, XgBoost, and SVM) and time-series data modeling (LSTM) to find signals for reducing unnecessary antibiotics 　› Prepare manuscripts for submission Antibiotics Stewardship ｜ NICU ｜ Feature Engineering ｜ Time-series Data ｜ Python ｜ Manuscript
Present - **November 2021**	**Research Assistant | PI : Hadi Kharrazi | Bloomberg School of Public Health, Baltimore, MD** Understanding the effect of data quality on using phenotype definitions to identify cohorts with diabetes 　› ETL 200K+ patient EHR and claims data using SQL and derived descriptive statistics using R 　› Simulated 3 induced data quality issues on ICD, RxNorm, and LOINC for both claims and EHR data 　› Constructed 20+ logistic regression models to assess disparities in data quality by race in claims and EHR 　› Assisted the first author in writing and preparing 4 manuscripts for submission (second author) Data Quality ｜ Phenotyping ｜ EHR ｜ Claims ｜ R ｜ SQL ｜ Logistic Regression ｜ Manuscript
Present - **November 2021**	**Graduate Researcher | Advisor : Harold P. Lehmann | JHSOM, Baltimore, MD** Examining the impact of asymmetric cost in machine learning – clinical utilities analysis 　› Programmed fully factorial simulations with 6 parameters and 6 machine learning models using Python 　› Established a novel utility-based approach to evaluating machine learning model performance 　› Applied the novel utility-based approach to 3 medical datasets and compared against traditional metrics 　› Led the writing of a manuscript accepted by AMIA 2023 Symposium and held a podium presentation Model Evaluation ｜ Decision Analysis ｜ Python ｜ Simulation ｜ Manuscript

May 2023
October 2021
	Graduate Researcher | Advisor : Paul G. Nagy | JHSOM, Baltimore, MD
	OHDSI Community Dashboard : Assessing the health and impact of the OHDSI ecosystem
	(http://ohdsi.azurewebsites.net/)
	    › Co-developed an automated ETL pipeline from PubMed, Google Scholars, YouTube, and EHDEN and mapped texts to standard terminologies using standard APIs
	    › Implemented/debugged the first version of the interactive tool with 3 front-end dashboards each with 3 features using Python, Flask, Dash, and Azure back-end
	    › Established an open-source framework for evaluating open-source scientific communities
	    › Led the writing of a manuscript accepted by JAMIA Open
	OHDSI · APIs · Flask · Dash · HTML · Azure CosmosDB · GitHub · Manuscript

May 2023
January 2021
	Research Assistant | PI : Mohammed K. Ali | Rollins School of Public Health, Atlanta, GA
	The associations of collaborative care intervention components with depressive symptoms and metabolic outcomes in the INDEPENDENT study : A mixed methods study
	    › ETL and analyzed 196 patients' longitudinal data and conducted multiple logistic regressions using R
	    › Wrote the statistical methods section and produced the appropriate figures for the manuscript
	    › Assisted the first author in writing a manuscript accepted by Primary Care Diabetes (second author)
	Longitudinal Clinical Trials Data · Mixed Methods · R · Regression Analysis · Manuscript

September 2022
October 2021
	Graduate Researcher | Advisor : Hadi Kharrazi | JHSOM, Baltimore, MD
	Assessing the Racial and Socioeconomic Disparities in Postpartum Depression Using Population-level Hospital Discharge Data : A Longitudinal Retrospective Study [Published]
	    › Conducted literature reviews on postpartum depression, diagnosis, and racial disparities
	    › Constructed logistic regression, multinomial regression, survival analysis, and sensitivity analysis using R and Healthcare Cost and Utilization Project datasets for Maryland (2016-2019)
	    › Published a manuscript and led its writing, journal submission, revisions, and correspondence
	Manuscript · Maternal Health Disparity · HCUP · R · Regression Analysis

June 2021
May 2020
	Research Assistant | PI : Shivani Patel | Rollins School of Public Health, Atlanta, GA
	COVID-19 Health Equity Dashboard (https://covid19.emory.edu/).
	    › Managed daily ETL process via Python scripts, integrating incident, demographic, vaccine, and SDoH data
	    › Implemented and managed back-end MongoDB database solutions using Python
	    › Implemented/debugged 20+ front-end features using NPM, React JS, React Simple Maps, and HTML
	    › Led the writing as the first author of a case study accepted by the IEEE VisComm 2020
	COVID-19 Interactive Dashboard · Python · MongoDB · JavaScript · Case Study Presentation

May 2020
January 2020
	Undergraduate Researcher | Advisor : Seunghwa Rho | Emory University, Atlanta, GA
	Atlanta Fire and Rescue Department : Predicting Fire Incident Risk ↗ Link to Report ↗ Link to PPT
	    › Pre-processed and aggregated +8 million observations of geospatial data on historical fires
	    › Optimized a random forest model to predict fires with 40% sensitivity and 90% specificity
	    › Delivered an R Shiny interactive dashboard to visualize predictions in use
	    › Received positive feedback from the Atlanta Mayor's Office, City Planning, and the Fire Department
	Geospatial and Longitudinal Data · Predictive Modeling · Random Forest · Interactive Shiny App

June 2020
September 2019
	Volunteer Research Assistant | Rollins School of Public Health, Atlanta, GA
	    › Wrangled Medical Expenditure Panel Survey and Healthcare Cost and Utilization Project data using R
	    › Conducted literature reviews on diabetes, preventable hospitalization, and care utilization
	    › Conducted consistent 30, 90, and 120-day phone interviews with a cohort of 24 patients with stroke
	Diabetes · Survey Data · Panel Data · R · Literature Review · Phone Interviews

💼 Professional Experience

August 2019
May 2019
	Analyst | Supervisor : Michael Trader | RightPatient, Dunwoody, GA
	    › Liaised and delivered 2 million images from public safety agencies for the R&D project on opioid addiction
	    › Developed 4 client leads for the ACO patient-centered care software and relayed to the demo team
	    › Was entrusted with managing 358 client accounts on Zoho CRM for our patient wellness initiative
	Healthcare Technology Startup · R&D · Customer Relations · Sales

👥 Teaching & Mentorship

2025	**Graduate Teaching Assistant	Johns Hopkins School of Medicine, Baltimore, MD**
2023	Clinical Decision Analysis

> Assisted the faculties in teaching decision sciences and utility theory and hosted weekly office hours

TA | Decision Analysis | Treeage | Office Hours

2025	**Graduate Teaching Assistant	Johns Hopkins School of Medicine, Baltimore, MD**
2023	Implementing Fast Healthcare Interoperability Resources (FHIR)

> Assisted the faculties in teaching FHIR and hosted weekly office hours

TA | FHIR | Office Hours

2023	**Graduate Teaching Assistant	Johns Hopkins School of Medicine, Baltimore, MD**
2022	Clinical Data Analysis with Python

> Assisted the faculties in teaching Python and hosted weekly office hours

TA | Python | Office Hours

July 2022	**Bootcamp Consultant	Instructor : Hadi Kharrazi	Bowie State University, Bowie, MD**
July 2022	Public Health Informatics and Technology 4-Week Bootcamp

> Assisted the program faculties in teaching R and Tableau to a class of 15 students
> Hosted daily office hours and coached 4 capstone project groups

Bootcamp | R | Visualization | Capstone projects

May 2021	**Academic Fellow Captain, Academic Fellow	Emory University, Atlanta, GA**
May 2018	Annual International Student Welcome and Mentoring Program

> Coordinated two orientations with 14 colleagues, oversaw 26 fellows, and mentored 24 mentees

Peer Mentorship | Program coordination

✒ Peer Reviews

2021 - Present	Reviewer	JAMIA Open

✺ Certificates

April 2021 - No Expiration Date	Information Privacy Security	JHMI CITI Program	Credential ID : 45798326
April 2021 - No Expiration Date	Health Privacy Issues for Researchers	JSPH CITI Program	Credential ID : 45562106
October 2021 - October 2026	Basic Human Subjects Research	JSPH CITI Program	Credential ID : 45562104
October 2021 - October 2024	Social and Behavioral Research Best Practices for Clinical Research	JSPH CITI Program	Credential ID : 45562105
October 2021 - October 2023	Human Subjects Research - Biomedical Research	JHMI CITI Program	Credential ID : 45733041

🌐 Languages

English	Native
Chinese	Native
Spanish	Beginner