

## EDUCATION


<b>2023-Present</b>	<b>Johns Hopkins School of Medicine (JHSOM)</b>
Degree	PhD in Biomedical Informatics and Data Science   Mentor : Harold P. Lehmann
Dissertation	<i>Aligning clinical predictive models with clinical decision making via elicitation, calibration, and threshold selection</i>
<b>2021-2023</b>	<b>JHSOM</b>
Degree	Master of Science in Biomedical Informatics and Data Science Research   Mentor : Harold P. Lehmann
Thesis	<i>Model Usefulness : Aligning Utility and Prevalence in Use and in Training of Clinical Machine Learning Models</i>
<b>2017-2021</b>	<b>Emory University</b>
Degree	Bachelor of Science in Quantitative Sciences, Biology Concentration   Mentor : Seunghwa Rho
Capstone	<i>Building a Predictive Model for Fire Incident Risk in Atlanta</i>

## COMPETENCIES


<b>Programming &amp; ML</b>	Python (PyTorch, pandas, numpy, scikit-learn) • R • SQL • Deep Learning • LangChain • Azure
<b>Health Economics</b>	Markov Models, Decision Trees, TreeAge
<b>Databases</b>	Microsoft SQL Server, CosmosDB, MongoDB
<b>Vocabularies</b>	ICD-9, ICD-10, RxNorm, LOINC, SNOMED-CT, CPT
<b>Data Sources</b>	EHR, Claims, <a href="#">Healthcare Cost and Utilization Project</a> , <a href="#">Medical Expenditure Panel Survey</a>
<b>Softwares</b>	GitHub, Overleaf, Microsoft Excel, Slurm, FileZilla, Tableau, QGIS, Google Analytics

## PUBLICATIONS AND PREPRINTS

### Published - Journal and Conference

- Sood PD, **Liu S**, Kalyani RR, Kharrazi H. Comparing computable type 2 diabetes phenotype definitions in identifying populations of interest for clinical research. *BMJ Health Care Inform.* 2025.  PMID : 41125310
- Barrett R, Lawler B, Liu S, Park WY, Davoodi M, Martin B, Kalyanam SM, Makker K, Kuiper JR, Aziz KB. Transforming neonatal care through informatics : A review of artificial intelligence, data, and implementation considerations. *Semin Perinatol.* 2025.  PMID : 40973564
- Sood PD, **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. *Healthcare (Basel).* 2025.  PMID : 41008424
- Sood PD, **Liu S**, et al. Measuring the Impact of Data Quality and Computable Phenotypes on Potential Racial Disparities in Predicting Healthcare Utilization Among Type 2 Diabetes Populations. *Racial and Ethnic Health Disparities.* 2025.  PMID : 40425977
- Al Gharaibeh FN, **Liu S**, Wynn JL, Aziz KB. The utility of neonatal sequential organ failure assessment in mortality risk in all neonates with suspected late-onset infection. *J Perinatol.* 2025.  PMID : 40251303
- Sood PD, **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.* 2024.  PMID : 39189917
- 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.  PMID : 38425704
- 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.* 2024.  PMID : 38360505
- Liu S**, Wei S, Lehmann HP. Applicability Area : A novel utility-based approach for evaluating predictive models, beyond discrimination. *AMIA Annu Symp Proc.* 2024.  PMID : 38222359
- Benamer 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.  Sage Journal Article : 10.1177/15598276231208383
- 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.  PMID : 36103575
- 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.  PMID : 35702627

### Under Review

- 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

## In Preparation

14. **Liu S**, Park WY, Al Gharaibeh FN, Wynn JL, Aziz KB. Antibiotics Stewardship - Characterizing antibiotics administration patterns among neonates < 33 weeks and 1500g with negative blood cultures relative to organ dysfunction. *JAMA Pediatrics*.
15. **Liu S**, Ding X, Lehmann HP. Bootstrapping Bezier Curves to Quantify the Uncertainty Associated with the Decision-Analytic Optimal Point on the Receiver Operating Characteristic Curve. *Medical Decision Making*.
16. **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*.

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

## PRESENTATIONS, POSTERS, AND CASE STUDIES

1. **Liu S**, Park WY, Al Gharaibeh FN, Wynn JL, Aziz KB. Antibiotics Stewardship - Characterizing antibiotics administration patterns among neonates < 33 weeks and 1500g with negative blood cultures relative to organ dysfunction. *Neonatal Research Conference - UT San Antonio*. 2025. [Oral Presentation] [Link to Slide](#)
2. **Liu S**, Barrett RB, Zollo-Venecek K, Riesser B, Martin B. Advancing Electronic Clinical Quality Measure (eCQM) Interoperability : Model Context Protocol (MCP)-Orchestrated CQL-to-OMOP Translation. *Observational Health Data Sciences and Informatics (OHDSI) Symposium*. 2025. [Software Demonstration] [Link to Demo](#)
3. **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](#)
4. **Liu S**, Lehmann HP. Identifying and Visualizing the Optimal Cutoff on the Receiver Operating Characteristic (ROC) Curve. *Society for Medical Decision Making (SMDM)*. 2024. [Poster] [Link to Poster](#)
5. **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](#)
6. **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](#)
7. 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](#)
8. 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](#)
9. **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](#)
10. **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

### Johns Hopkins University School of Medicine

Baltimore, MD

Present - Jan. 2025	<b>Independent Research - Self-initiated with PhD colleague</b> Clinical quality measures automation : LLM-powered mapping from CQL to OMOP SQL <ul style="list-style-type: none"><li>&gt; Reduced manual mapping time from days to minutes via LLM-powered pipeline using LangChain to automate translation of clinical quality measures from Clinical Quality Language to executable SQL on OMOP databases</li><li>&gt; Engineered Model Context Protocol (MCP) integration to provide LLMs with structured access to OMOP vocabularies and VSAC value sets, improving mapping accuracy through enhanced context retrieval</li><li>&gt; Evaluated system performance against gold-standard clinician reviews; preparing manuscript for JAMIA</li></ul> <div>PythonLangChainLLMsMCPOMOP CDMSQLHealthcare Standards</div>
Present - Jan. 2024	<b>Graduate Research   PI : Jim C. Fackler   Anesthesia and Critical Care Medicine</b> Estimating the risk of adverse drug events (ADEs) associated with broad-spectrum antibiotics among septic patients <ul style="list-style-type: none"><li>&gt; Conduct cohort discovery and characterize the risk of ADE across partner OMOP databases</li><li>&gt; Define the risk profiles of ADEs and support broad-spectrum antibiotics administration</li><li>&gt; Predict the risk of ADEs occurrence among pediatric patients with sepsis using statistical and NLP methods</li></ul> <div>ADEsAntibiotics StewardshipPediatricsSepsisOMOPNetwork StudyPhenotyping</div>

Present - Jun. 2023	<b>Graduate Researcher   PI : Khyzer B. Aziz   NICU Precision Medicine Center of Excellence</b> Neonatal sepsis and antibiotics stewardship : Analyzing treatment patterns and outcomes <ul style="list-style-type: none"> <li>&gt; Queried 90K+ neonates across 9 years of multi-million row NICU EHR data using SQL to analyze neonatal outcomes and antibiotic treatment patterns relative to organ dysfunction</li> <li>&gt; Engineered 20+ clinical features from temporal data to train ML algorithms (Random Forest, XGBoost, Clustering, LSTM), revealing that poor predictive performance signals inconsistent stewardship practices</li> <li>&gt; Leading manuscript for JAMA Pediatrics analyzing antibiotic use patterns; co-authoring multi-center studies to inform sepsis treatment guidelines across multi-center study sites</li> </ul> Python SQL Large-scale EHR data Feature Engineering Statistical Methods Predictive Modeling LSTM
Present - Nov. 2021	<b>Graduate Researcher   Mentor : Harold P. Lehmann   Decision Sciences Lab</b> Clinical predictive model evaluation in the context of use : Aligning ML models with clinical decision-making <ul style="list-style-type: none"> <li>&gt; Developed novel ML evaluation framework (Applicability Area) published at AMIA 2023, improving clinical utility assessment by incorporating harm-benefit tradeoffs, enabling better model selection for healthcare deployment</li> <li>&gt; Deployed interactive Azure dashboard for real-time sensitivity analysis across clinical scenarios to optimize ML model prediction thresholds based on context-specific cost-benefit ratios <a href="#">Link to iCUDA dashboard</a></li> <li>&gt; Built 1000-stratum Markov microsimulation evaluating suicide risk prediction strategies across 100K patients with probabilistic sensitivity analysis</li> <li>&gt; Leading manuscript establishing method for identifying optimal ML model operating points using decision analysis; conducted 25 physician interviews to quantify clinical harm-benefit tradeoffs</li> </ul> Python R Statistical Modeling Markov Models Decision Analysis Azure Interactive Dashboards Qualitative Methods
Oct. 2025 - Nov. 2021	<b>Research Assistant   PI : Hadi Kharrazi   Center for Population Health IT</b> Understanding the effect of data quality on using phenotype definitions to identify diabetes cohorts <ul style="list-style-type: none"> <li>&gt; Processed 200K+ patient EHR and claims data; simulated 3 data quality scenarios across medical code systems (ICD, RxNorm, LOINC) to evaluate 4 diabetes phenotypes</li> <li>&gt; Built 20+ logistic regression models in R, demonstrating how data quality and phenotype choice impact downstream racial disparities in healthcare utilization predictions</li> <li>&gt; Co-authored 4 manuscripts informing equitable ML model development for population health</li> </ul> SQL R Large-scale EHR and Claims data Logistic Regression Statistical Modeling Health Equity
May 2023 Oct. 2021	<b>Graduate Researcher   Advisor : Paul G. Nagy   JHSOM</b> OHDSI Community Dashboard : Assessing the health and impact of the OHDSI ecosystem ( <a href="http://ohdsi.azurewebsites.net/">http://ohdsi.azurewebsites.net/</a> ) <ul style="list-style-type: none"> <li>&gt; Co-developed an automated ETL pipeline from PubMed, Google Scholar, YouTube, and EHDEN and mapped texts to standard terminologies using standard APIs</li> <li>&gt; 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</li> <li>&gt; Established an open-source framework for evaluating open-source scientific communities</li> <li>&gt; Led the writing of a manuscript accepted by JAMIA Open</li> </ul> OHDSI APIs Flask Dash HTML Azure CosmosDB GitHub Manuscript
Sep. 2022 Oct. 2021	<b>Graduate Researcher   Advisor : Hadi Kharrazi   JHSOM</b> Assessing the racial and socioeconomic disparities in postpartum depression using population-level hospital discharge data : A longitudinal retrospective study [Published] <ul style="list-style-type: none"> <li>&gt; Conducted literature reviews on postpartum depression, diagnosis, and racial disparities</li> <li>&gt; Constructed logistic regression, multinomial regression, survival analysis, and sensitivity analysis using R and Healthcare Cost and Utilization Project datasets for Maryland (2016-2019)</li> <li>&gt; Published a manuscript and led its writing, journal submission, revisions, and correspondence</li> </ul> Manuscript Maternal Health Disparity HCUP R Regression Analysis

## Emory University and Rollins School of Public Health

Atlanta, GA

May 2023 Jan. 2021	<b>Research Assistant   PI : Mohammed K. Ali   Rollins School of Public Health</b> The associations of collaborative care intervention components with depressive symptoms and metabolic outcomes in the INDEPENDENT study : A mixed methods study <ul style="list-style-type: none"> <li>&gt; ETL and analyzed 196 patients' longitudinal data and conducted multiple logistic regressions using R</li> <li>&gt; Wrote the statistical methods section and produced the appropriate figures for the manuscript</li> <li>&gt; Assisted the first author in writing a manuscript accepted by Primary Care Diabetes (second author)</li> </ul> Longitudinal Clinical Trials Data Mixed Methods R Regression Analysis Manuscript
-----------------------	--

Jun. 2021 May 2020	<b>Research Assistant   PI : Shivani Patel   Rollins School of Public Health</b> COVID-19 Health Equity Dashboard ( <a href="https://covid19.emory.edu/">https://covid19.emory.edu/</a> ) <ul style="list-style-type: none"> <li>› Managed daily ETL process via Python scripts, integrating incident, demographic, vaccine, and SDoH data</li> <li>› Implemented and managed back-end MongoDB database solutions using Python</li> <li>› Implemented/debugged 20+ front-end features using NPM, React JS, React Simple Maps, and HTML</li> <li>› Led the writing as the first author of a case study accepted by the IEEE VisComm 2020</li> </ul> COVID-19 Interactive Dashboard Python MongoDB JavaScript Case Study Presentation
May 2020 Jan. 2020	<b>Undergraduate Researcher   Mentor : Seunghwa Rho   Emory University</b> Atlanta Fire and Rescue Department : Predicting fire incident risk <a href="#">Link to Report</a> <a href="#">Link to PPT</a> <ul style="list-style-type: none"> <li>› Pre-processed and aggregated +8 million observations of geospatial data on historical fires</li> <li>› Optimized a random forest model to predict fires with 40% sensitivity and 90% specificity</li> <li>› Delivered an R Shiny interactive dashboard to visualize predictions in use</li> <li>› Received positive feedback from the Atlanta Mayor's Office, City Planning, and the Fire Department</li> </ul> Geospatial and Longitudinal Data Predictive Modeling Random Forest Interactive Shiny App
Jun. 2020 Sep. 2019	<b>Volunteer Research Assistant   Rollins School of Public Health</b> Diabetes care utilization and health equity research <ul style="list-style-type: none"> <li>› Wrangled Medical Expenditure Panel Survey and Healthcare Cost and Utilization Project data using R</li> <li>› Conducted literature reviews on diabetes, preventable hospitalization, and care utilization</li> <li>› Conducted consistent 30, 90, and 120-day phone interviews with a cohort of 24 patients with stroke</li> </ul> Diabetes Survey Data Panel Data R Literature Review Phone Interviews

## PROFESSIONAL EXPERIENCE

### RightPatient

Dunwoody, GA

Aug. 2019 May 2019	<b>Analyst   Supervisor : Michael Trader</b> <ul style="list-style-type: none"> <li>› Liaised and delivered 2 million images from public safety agencies for the R&amp;D project on opioid addiction</li> <li>› Developed 4 client leads for the ACO patient-centered care software and relayed to the demo team</li> <li>› Was entrusted with managing 358 client accounts on Zoho CRM for our patient wellness initiative</li> </ul> Healthcare Technology Startup R&D Customer Relations Sales
-----------------------	---

## TEACHING & MENTORSHIP

2025 2023	<b>Graduate Teaching Assistant   Johns Hopkins School of Medicine</b> Clinical Decision Analysis <ul style="list-style-type: none"> <li>› Assisted the faculties in teaching decision sciences and utility theory and hosted weekly office hours</li> </ul> TA Decision Analysis Treeage Office Hours
2025 2023	<b>Graduate Teaching Assistant   Johns Hopkins School of Medicine</b> Implementing Fast Healthcare Interoperability Resources (FHIR) <ul style="list-style-type: none"> <li>› Assisted the faculties in teaching FHIR and hosted weekly office hours</li> </ul> TA FHIR Office Hours
2023 2022	<b>Graduate Teaching Assistant   Johns Hopkins School of Medicine</b> Clinical Data Analysis with Python <ul style="list-style-type: none"> <li>› Assisted the faculties in teaching Python and hosted weekly office hours</li> </ul> TA Python Office Hours
Jul. 2022 Jul. 2022	<b>Bootcamp Consultant   Instructor : Hadi Kharrazi   Bowie State University</b> Public Health Informatics and Technology 4-Week Bootcamp <ul style="list-style-type: none"> <li>› Assisted the program faculties in teaching R and Tableau to a class of 15 students</li> <li>› Hosted daily office hours and coached 4 capstone project groups</li> </ul> Bootcamp R Visualization Capstone projects
May 2021 May 2018	<b>Academic Fellow Captain, Academic Fellow   Emory University</b> Annual International Student Welcome and Mentoring Program <ul style="list-style-type: none"> <li>› Coordinated two orientations with 14 colleagues, oversaw 26 fellows, and mentored 24 mentees</li> </ul> Peer Mentorship Program coordination

## PEER REVIEWS

2021 - Present Reviewer | JAMIA Open

## CERTIFICATES

---

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

## LANGUAGES

---

English	Native
Chinese	Native
Spanish	Beginner