

SAMANTHA WENJIN ZHOU

HEALTHCARE DATA SCIENTIST | CLINICAL RESEARCHER | EHR ANALYTICS

samantha.zhou@alumni.emory.edu | LinkedIn

Healthcare data scientist and clinical researcher experienced in translating clinician-led questions into reproducible EHR and research workflows. Hands-on expertise in Epic Cosmos, SQL, R, Python, REDCap, cohort construction, data quality, and statistical modeling across million-scale clinical datasets. Combines healthcare analytics, study design, IRB operations, preclinical research, and clear scientific communication to support evidence-based decisions.

1.26M+

EHR encounters screened

236K+

patients in analytic cohort

98%

reduction in manual analysis time

CORE EXPERTISE

DATA & PLATFORMS

SQL, R, Python, Excel, REDCap, Epic Cosmos, DeepLabCut

ANALYTICS

Cohort design, data QC, logistic regression, sensitivity analysis, post hoc testing, visualization

CLINICAL & RESEARCH

Outcomes research, IRB, NIH R01 operations, clinician collaboration, grant review, scientific writing

RESEARCH, DATA & CLINICAL EXPERIENCE

Emory Healthcare | Atlanta, GA

Dec 2025-Present

Data Scientist (Research Collaboration) - *Complex Benign Gynecologic Surgery, Dr. Heit's Team*

- Built a reproducible Epic Cosmos SQL/R pipeline that screened 1,259,836 minimally invasive hysterectomy encounters and produced a 236,799-encounter analytic cohort representing 236,242 patients and 19,776 providers.
- Designed cohort logic and data-quality controls for fibroid-only outcomes research, including diagnosis and procedure definitions, denominator audits, age/BMI validation, a 730-day Charlson comorbidity lookback, and missing-data checks.
- Conducted bivariate and multivariable logistic regression, adjusted odds ratios with 95% confidence intervals, Bonferroni-adjusted comparisons, and sensitivity analyses for hospital admission and postoperative complications by surgeon specialty and annual volume.
- Translated clinician questions into decision-ready feasibility models and tables; shortened feasibility timelines by approximately 50% and contributed study design, cohort strategy, and preliminary analyses to collaborative grant proposals as a Co-Investigator.

School of Medicine, Emory University | Atlanta, GA

Nov 2025-May 2026

Research Specialist - *Endometriosis & Pain Group, McAllister Lab*

- Led a preclinical NIH R01 Aim 3a dose-finding study in a deep endometriosis model, coordinating experimental workflows and defining dosing parameters for downstream therapeutic testing.
- Built a Python/DeepLabCut pose-estimation and automated-scoring pipeline that reduced manual behavioral analysis from approximately 8 hours to 10 minutes (about 98%) while improving reproducibility.
- Developed objective behavioral metrics and an end-to-end structured data workflow from pose-estimation outputs to support scalable, reproducible phenotyping.
- Created onboarding manuals and checklists that reduced independent training from months to approximately 1 month.

School of Medicine, Emory University | Atlanta, GA

Feb 2026-May 2026

Functional Clinical Research Coordinator - *NIH R01 Clinical Research Program*

- Architected REDCap data systems with automated eligibility logic and specimen-linked quality-control workflows for an NIH R01 study.
- Harmonized surgical findings, menstrual staging, patient-reported pain, and biospecimen metadata into structured, IRB-compliant datasets.
- Led IRB submissions and amendments to expand cohort eligibility and strengthen protocol-data alignment.

EDUCATION & HONORS

Emory University Atlanta, GA	Class of 2025
<i>Bachelor's Degree - Double Major in Biology and English</i> Highest Honors; Dean's List	
McGill University Montreal, Quebec, Canada	Transferred
<i>Advanced honors STEM curriculum</i> Science Department Honors Program	

ADDITIONAL RESEARCH & SCHOLARLY EXPERIENCE

English Department, Emory University Atlanta, GA	Fall 2024-May 2025
<i>Honors Thesis - Highest Honors</i>	
• Authored an 80-page interdisciplinary thesis integrating literary analysis, sociology, and philosophy to examine how secularization reshaped the pastoral elegy from Greek to Romantic literature. • Presented and defended the thesis before faculty committees; awarded highest honors and published in the Emory Theses & Dissertations Repository.	
Biology Department, Sun Yat-Sen University Guangdong, China	Sep 2019-Jan 2022
<i>Research Assistant</i>	
• Supported research on engineering beta-glucosidase to improve glucose tolerance and thermal stability; prepared culture media, collected experimental data, and troubleshoot protocols. • Organized results in Excel tables, formulas, and figures for quantitative analysis and scientific presentation.	

PEER REVIEW & PROFESSIONAL SERVICE

Patient-Centered Outcomes Research Institute (PCORI)	2026
<i>Reviewer - AI/ML & Data Research Networks</i>	
• Invited to evaluate patient-centered comparative effectiveness proposals involving artificial intelligence, machine learning, and large-scale electronic health record data networks. • Provided structured critiques and scores on scientific rigor, methodological validity, and translational impact. Contributed to merit review discussions.	
Emory University Postdoctoral Association (EU-PDA)	2026
<i>Grant Reviewer - Annual Conference Travel Awards</i>	
• Reviewed competitive applications supporting early-career researchers in pharmaceutical development and contributed recommendations based on scientific merit, innovation, and relevance.	

SELECTED PUBLICATION

Zhou, Samantha. *Secularization in Pastoral Elegy: From Sacred Nature to Liberal Selves.* Honors thesis, Department of English, Emory University, 2025. [Emory Repository](#)

CERTIFICATIONS & TRAINING

Epic/Cosmos Data Scientist & Architect	Johns Hopkins (Coursera): Clinical Trials Operations; Biostatistics
CITI: Human Subjects, Clinical Trials, Animal Research (RCR)	AALAS Animal Biosafety & Transgenics; Emory EHSO Rodent Weaning

TECHNICAL & RESEARCH SKILLS

PROGRAMMING & DATA SQL, R, Python, Java, Excel, REDCap, DeepLabCut	HEALTHCARE DATA Epic Cosmos, EHR cohort construction, data quality control, structured datasets, reproducible pipelines
STATISTICAL METHODS Descriptive statistics, logistic regression, adjusted odds ratios, sensitivity analysis, Bonferroni post hoc comparisons, visualization	CLINICAL RESEARCH Study design, IRB submissions and amendments, cohort definition, outcomes research, retrospective studies, patient-reported outcomes
LABORATORY METHODS IACUC, rodent behavioral assays, dosing and injections, surgery, perfusion, histology, PCR, cell culture, immunofluorescence	COMMUNICATION & REVIEW Clinician collaboration, scientific writing, grant proposal support, peer review, presentations, multidisciplinary teamwork