Haoyun Hong (Elizabeth)

haoyunhong507@gmail.com | 510-631-5698
132 NE 145th Street, Unit 703, Shoreline, WA, 98155

EDUCATION

University of Washington, Seattle — **Seattle, WA**
PhD. Biomedical and Health Informatics. — Sep. 2025 – present

University of California, Berkeley — **Berkeley, CA**
B.A., Data Science. — Aug. 2018 - May. 2022
Coursework: Structure and Interpretation of Computer Programs, Discrete Mathematics and Probability Theory, Principles and Techniques of Data Science, Data, Inference and Decisions, Data Mining and Analytics, Natural Language Processing

RESEARCH EXPERIENCE

American Heart Association, Data Science and Evaluation Team — **Dallas, TX/ Remote**
Data Science Intern — Jan. 2022 – Aug 2022
Data Science Fellow/Statistical Analyst — Aug. 2022 – May 2025

- **Advisors**: Juan (Wendy) Zhao, Jennifer Hall

Race-Based differences in STEMI process metrics and mortality from 2015 through 2021.

- Conducted comprehensive data cleaning, processing and analysis on over 150,000 patient records in R on a cloud platform to examine racial disparities in process metrics and mortality for STEMI care, leveraging the Get with the Guideline – Coronary Artery Disease (GWTG-CAD) national registry.
- Performed chi-square tests, Kruskal-Wallis tests, and Dunn test with Bonferroni correction to assess statistical significance of between-group differences; applied Generalized Estimating Equations (GEE) models to estimate odds ratio with for comorbidities and socioeconomic status.
- Found significant racial disparities in treatment goals among men and women that persisted post-adjustment; identified higher adjusted in-hospital mortality for Hispanic women, with no significant differences observed in other groups.
- Created tables and figures for the final manuscript published in *Circulation* (IF:37.8)

Other cross-sectional studies using GWTG-CAD national registry.

- Collaborated with external researchers to refine study objectives and develop research plans for 7 studies with following areas of focus:
  - *Trend in STEMI care stratified by transportation mode from 2018-2021 (JAMA, IF: 120.7)*
  - *STEMI care throughout waves of COVID-19 pandemic*
  - *STEMI care trends among patients who have non-system reason for delay **
  - *Site-level variability in STEMI care (JAMA Cardiology, IF: 14.7)**
  - *Social determinants of health differences in the prescription of P2Y12 inhibitors at discharge**
  - *Association between social determinants of health factors and cardiac rehabilitation referral rate**
  - *Association between system failure and STEMI care outcomes**
- Led and conducted data and statistical analysis for 5 research projects, using multivariate regression to assess trends and between group disparities. (*denotes projects with both leading and contributing effort)
- Most manuscripts are currently in preparation or under review for publication with findings aimed at informing clinical guidelines and identifying key areas for improving equitable STEMI care.

De-biasing clinical care algorithms in cardiovascular care (DECCA)

- Analyzed over 60,000 longitudinal patient records from an electronic health record (EHR) system that utilizes OMOP common data model (CDM) linked with third-party social determinants of health (SDOH) data, utilizing SQL, Python and R with Apache Spark
- Examined the disparities in the development of long-term outcomes (CVD and non-CVD conditions) between renters and homeowners based on risk predictions from Pooled Cohort Equation (PCE) and PREVENT
- Project was funded by American Heart Association as part of Doris Duke Foundation's initiative of Racial Equity in Clinical Equations to contribute novel rigorous evidence to the scientific landscape of race in clinical care algorithms to identify and validate best practices and solutions.
- Abstract accepted for presentation at AMIA 2025 informatics summit

Utilizing machine learning (ML) models to improve data quality in GWTG-CAD national registry

- Developed ML based models (Random Forest, XGBoost, LightGBM, etc) to impute critical covariates for CAD mortality prediction models leveraging nested cross validation with a calibration layer.
- Evaluated model performance in sex, race/ethnicity subgroups through false negative rate (FNR) and false positive rate (FPR) and analyzed feature importance using SHapley Additive exPlanations (SHAP) values
- Abstract published and presented in American Heart Association Scientific Sessions 2023

ETL pipeline development & registry maintenance

- Contributed to the redesign and development of ETL pipeline for 5 GWTG national registries using Python including extracting information from data documentation, clean and process registry data, anonymize patient ID and generate documentation for cleaned data.
- Developed and maintained 8 tutorial notebooks for Research Goes Red (RGR) registry survey data summarizing data distributions in Python and R, ensuring up to date content with each new data harvest.
- Harmonized and maintained data updates for AHA Tobacco Center for Regulatory Science (ATRAC) 2.0 registry, employed unsupervised machine learning (K-means clustering, agglomerative clustering) to analyze volatile organic compound biomarkers data from non-smokers, e-cigarette and cigarette users.
- Developed automated updates on survey data for Chapter 13: Sleep of 2025 *Heart Disease and Stroke Statistics: A Report of US and Global Data From the American Heart Association*.

PUBLICATIONS

Journal Articles

- Asishana Osho, Marcelo F. Fernandes, Remy Poudel, James de Lemos, **Haoyun Hong**, Juan Zhao, Shen Li, Kathie Thomas, Daniel S. Kikuchi, Jessica Zegre Hemsey, Nasrien Ibrahim, Nilay S. Shah, Lori Hollowell, Jacqueline Tamis-Holland, Christopher B. Granger, Mauricio Cohen, Timothy Henry, Alice K. Jacobs, James G. Jollis, Clyde W. Yancy, Abhinav Goyal. **Race-Based Differences in ST-Segment–Elevation Myocardial Infarction Process Metrics and Mortality From 2015 Through 2021: An Analysis of 178 062 Patients From the American Heart Association Get With The Guidelines–Coronary Artery Disease Registry**. *Circulation*, 2023

- Shiavax J. Rao, Yaa A. Kwapong, Ellen Boakye, Pratheek Mallya, Juan Zhao, William Akel, **Haoyun Hong**, Shen Li, Chigolum P. Oyeka, Faith Elise Metlock, Pamela Ouyang, Roger S. Blumenthal, Khurram Nasir, Abha Khandelwal MD MS8 , Claire Kinzy, Laxmi S. Mehta, Veronique L. Roger, Jennifer L. Hall , Garima Sharma. **Reproductive Experiences and Cardiovascular Disease Care in Pregnancy-Capable and Postmenopausal Individuals: Insights From the American Heart Association Research Goes Red Registry**. *Current Problems in Cardiology*, 2023.

- Remy Poudel, Shen Li, **Haoyun Hong**, Juan Zhao, Shweta Srivastava, Rose Marie Robertson, Jennifer L. Hall, Sanjay Srivastava, Naomi M. Hamburg, Aruni Bhatnagar, Rachel J.Keith**. Catecholamine levels with use of electronic and combustible cigarettes**. *Tobacco Induced Disease*. 2024

- Chandler Beon, Lanjing Wang, Vihaan Manchanda, **Haoyun Hong**, Holly Picotte, Kathie Thomas, Jennifer L. Hall, Juan Zhao, Xue Feng. **Empowering Research With the American Heart Association Get With The Guidelines Registries Through Integration of a Database and Research Tools**. *Circulation: cardiovascular quality and outcomes*. 2024

- Katherine M. Conners, Maryam Hashemian, Claire Kinzy, Jennifer L. Hall, Christine Herr, Reeti Sharma, Pratheek Mallya, Juan Zhao, Nasrien E. Ibrahim, Joseph J. Shearer, **Haoyun Hong**, Véronique L. Roger. **Awareness of Heart Failure, Blood Pressure Management and Self-Efficacy: The Research Goes Red for Women Registry**. *Women's Health*. 2024

- Yasser M. Sammour; Safi U. Khan, Sachin S. Goel, Wissam A. Jaber, Grant W. Reed, Alexander C. Fanaroff, **Haoyun Hong**, Remy Poudel, Jingyuan Wu, Kathie Thomas, Zhao Ni, Abhinav Goyal, Ajay J. Kirtane, Robert W. Yeh, W. Schuyler Jones, Neal S. Kleiman. **Institutional Variability in Processes of Care and Outcomes Among Patients with ST-elevation Myocardial Infarction in the United States: Contemporary Insights from the American Heart Association Get With The Guidelines® – Coronary Artery Disease (GWTG-CAD) Registry.** *JAMA Cardiology*. 2025

- Juan Zhao, **Haoyun Hong**, Joseph Zhai, Remy Poudel, Sanjay Srivastava, Andrew C. Stokes, Pawel Konrad Lorkiewicz, Elizabeth Henderson, Tian Jiang, Rose Marie Robertson, Aruni Bhatnagar, Jennifer L. Hall, Naomi Hamburg, Rachel J. Keith. **Clustering Analysis of Volatile Organic Compound Biomarkers: to Identify Tobacco Exposure and Association with Cardiovascular Health Outcomes using an Observation Study Cohort.** *Tobacco Induced Disease*. 2025

- Chiadi Ndumele, Ankeet Bhatt, Lynne Braun, Steven Chen, Seth Martin, Michael Mullen, Nishant Shah, Stephen Sigal, Tracy Yu-Ping Wang, Rebecca Alicki, Chandler Beon, **Haoyun Hong**, Sara O'Kane, Katherine Overton, Kathie Thomas, Howard Haft. **The American Heart Association National Integrated ASCVD Initiative: An Implementation Initiative to Improve Lipid Management Among Patients with ASCVD**. *Circulation: cardiovascular quality and outcomes*. 2025

- Tetz Cheng-Che Lee, James Jollis, **Haoyun Hong**, Tian jiang, Juan Zhao, Timothy D. Henry, Alice K. Jacobs, Abhinav Goyal, Christopher B. Granger, Kathie Thomas, Jacqueline E Tamis-Holland. **Characteristic and Outcomes of Patients with ST-segment Elevation Myocardial Infarction who have Non-System Reasons for Delay in Treatment: A Report from the AHA GWTG CAD Registry.** (submitted) [https://doi.org/10.1101/2024.11.07.24316939]

- Hend Mansoor, Remy Poudel, **Haoyun Hong**, Shen Li, Kathie Thomas, Stephen Sigal, Jacqueline Tamis-Holland, Abhinav Goyal, Islam Y. Elgendy. **Sex, ethnic and social determinants of health differences in high-potency P2Y12 inhibitors prescription among patients with acute coronary syndrome: An analysis of the American Heart Association Get With The Guidelines®–Coronary Artery Disease Registry.** (submitted)

- **Haoyun Hong**, Remy Poudel, Asishana Osho, Kathie Thomas, Shen Li, Abhinav Goyal, Jennifer L. Hall, Juan Zhao. **Machine Learning-Based Models for Imputing Missing Values of Key Predictors of Important Clinical Outcomes in a Large-Scale Registry of Acute Myocardial Infarction**. (In Revision)

Abstracts

- **Haoyun Hong**, Remy Poudel, Asishana Osho, Kathie Thomas, Shen Li, Abhinav Goyal, Jennifer L. Hall, Juan Zhao. **Machine Learning Based Models for Imputing Heart Failure on First Medical Contact From the Get With The Guidelines®- Coronary Artery Disease (GWTG-CAD) Registry.** American Heart Association Scientific Sessions 2023.

- Remy Poudel, Shen Li, Rose Marie O Robertson, Jessica L Fetterman, Sanjay Srivastava, **Haoyun Hong**, Juan Zhao, Naomi M Hamburg, Aruni Bhatnagar, Rachel J Keith. **Association of Urinary Catecholamines and Their Metabolites With the Use of Electronic and Combustible Cigarettes**. American Heart Association Scientific Sessions 2023.

- Dan L. Musat, Mina K.Chung, Kathie Thomas, Juan Zhao, Shen Li, **Haoyun Hong**, Jonathan P. Piccini, Suneet Mittal. **Dofetilide Initiation in the USA – Is everybody getting with the guidelines? Insights from Get With the Guidelines Registry**. Heart Rhythm Society 2024.

- Hend Mansoor, Remy Poudel, **Haoyun Hong**, Shen Li, Kathie Thomas, Stephen Sigal, Jacqueline Tamis-Holland, Abhinav Goyal, Islam Y. Elgendy. **Sex, ethnic and social determinants of health differences in high-potency P2Y12 inhibitors prescription among patients with acute coronary syndrome: An analysis of the American Heart Association Get With The Guidelines®–Coronary Artery Disease Registry.** American Heart Association Scientific Sessions 2024.

- **Haoyun Hong**, Zihang Gao, Sadiya S. Khan, Yingying Sang, Haoyuan Wang, Chuan Hong, Michael J. Pencina, Jennifer, L. Hall, Juan Zhao. **Examine disparities in development of CVD and non-CVD conditions between renters and homeowners.** American Medical Informatics Association 2025 Informatics Summit

Presentations

- Juan Zhao, Haoyun Hong. **Harnessing machine learning models on GWTG-CAD Registry to improve clinical prediction**. American Heart Association Scientific Sessions 2024.

- **Haoyun Hong**, Remy Poudel, Asishana Osho, Kathie Thomas, Shen Li, Abhinav Goyal, Jennifer L. Hall, Juan Zhao. **Machine Learning Based Models for Imputing Heart Failure on First Medical Contact From the Get With The Guidelines®- Coronary Artery Disease (GWTG-CAD) Registry.** American Heart Association Scientific Sessions 2023

Reviews

- American Medical Informatics Association 2025 Clinical Informatics Conference
- American Medical Informatics Association 2025 Annual Symposium
- Society for Women Engineers FY25 Emerging First-Year Scholarship Reviewer

MISCELLANEOUS

Other Professional Experience

- Data analyst intern for Bank of China, Shanghai — May 2021 – Aug 2021
- Analyst intern for Plug and Play, China — Jan 2021 – Feb 2021
- Data Science intern for Fosun Pharma, Digital Technology and Innovation, Shanghai — Jun 2020 – Aug 2020

Teaching

- Tutor for Data C8: Foundations of data science — Jan 2021 – May 2021
- Lab assistant for CS 61A: Structure and interpretation of computer programs — Aug 2019 – Dec 2019

Volunteer & Membership

- Participant of McKinsey Leadership Essentials Program — Nov 2024 – Feb 2025
- Ambassador for Women in Data Science (WiDS) — Jan 2024 – Present
- Membership for Society of Women Engineers (SWE) — Oct 2024 – Present
- Student volunteer for Cal Hacks 5.0 — Nov 2018
- Student volunteer for Tech Crunch Hangzhou — Jul 2018

Awards & Honors

- Excellent in performance review (top 1%) at American Heart Association in FY22-23 and FY23-24

TECHNICAL SKILLS & LANGUAGES

- **Technical Skills**: Python (NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn), R (dplyr, tidyverse, tidyr, ggplot2, survey), SQL, Apache Spark, AWS, Git, Agile tools (Jira, Confluence)
- **Languages**: English (TOEFL:113/120 , GRE: 339/340 (Verbal: 169/170, Quantitative: 170/170, writing: 4.5/6.0), Chinese (native)