Jake William Coldiron

Contact Information

Email: jakewilliamcoldiron.apps@gmail.com
Phone: (202) 878-9282
Website: https://jakewilliamcoldiron.github.io
LinkedIn: https://www.linkedin.com/in/jakewilliamcoldiron/
GitHub: https://github.com/jakewilliamcoldiron

Professional Summary

Statistician, data scientist, and quantitative analyst with expertise in predictive analytics, longitudinal data, and machine learning. Proficient in statistical modeling, complex analysis, and real-world data with hands on experience across data ingestion, transformation, modeling, and reporting using R, Python, SAS, Julia, SQL and more. Experienced in data engineering practices including ETL pipelines, schema design, scalable workflows, and data optimization. Familiar with foundational AI and ML workflows, including data preprocessing, feature engineering, and model evaluation. Applied tools in supervised learning tasks, with a focus on interpretability, cross-validation, and reproducible results in real-world health data contexts. Capable of translating data into actionable insights across clinical, business, and financial domains.

Skills

Statistics: Statistical analysis, descriptive statistics, inferential statistics, Generalized Estimating Equations (GEE), Generalized Linear Models (GLM), mixed-effects models, multilevel modeling, linear regression, logistic regression, Poisson regression, survival analysis, Cox proportional hazards model, longitudinal data analysis, growth curve modeling, repeated measures ANOVA, ANCOVA, MANOVA, causal inference, propensity score matching, inverse probability weighting, multiple imputation, maximum likelihood estimation, time-to-event modeling, Kaplan-Meier estimation, model diagnostics, model validation, goodness-of-fit tests, cross-validation, hypothesis testing, confidence intervals, effect size estimation.

Programming: R, SAS, Python, SQL, Julia, LaTeX, Git, REDCap, NVivo, Covidence, EndNote, Zotero, RMarkdown, RShiny

Data: Data wrangling and cleaning, ETL pipelines, data harmonization and standardization, reproducible research workflows, federated learning implementation, version control (Git), data integration across sources, database management, secure data handling and governance.

Education

Master of Science (MS) in Statistics and Biostatistics, Columbia University
New York, NY, United States of America
August 2023 - May 2025

Bachelor of Science (BS) in Public Health (Honors), Georgia State University
Atlanta, GA, United States of America
August 2018 - December 2022

Selected Work Experience

Statistican, Voices of Atrial Fibrillation Patients
New York, NY, United States of America
December 2023 - August 2025

- Lead end-to-end data science workflows for clinical and real world data, including ingestion, cleaning, transformation, and statistical modeling of structured EHRs, imaging, lab results, and wearable data.
- Apply advanced statistical and machine learning methods including survival analysis, mixed effects models, propensity score matching, latent class modeling, and time to event analysis using R, Python, and SQL to generate insights, stratify risk, and model treatment outcomes.
- Develop reproducible data workflows across multi-institutional datasets using secure data platforms, and oversee team based research execution, regulatory reporting, and clinical dashboard development with a focus on statistical rigor and actionable insight.
- Integrate predictive modeling into stakeholder decision frameworks, ensuring statistical outputs are production-ready and aligned with real world clinical utility.

Clinical Data Scientist, DataUnite
San Francisco, CA, United States of America
July 2024 - January 2025

- Conducted advanced statistical modeling (GLMs, mixed-effects models, survival and time-to-event analysis) to evaluate drug efficacy and safety in RCTs using large-scale, multi-modal healthcare data (EHRs, imaging, pathology).
- Engineered scalable ETL pipelines in Python and SQL to harmonize structured and unstructured clinical data, developed predictive models, and applied cross-validation, AUC-ROC, and calibration metrics.
- Contributed to federated learning and virtual pooling frameworks, enabling decentralized, privacy-preserving machine learning, implemented drift detection, model governance protocols, and communicated results via R Markdown, white papers, and dashboards.
- Designed and validated statistical feature sets for inclusion in decentralized modeling environments, balancing interpretability and predictive performance.

Poll Manager (Operations Analyst Equivalent), Douglas County Board of Elections
Douglasville, GA, United States of America
September 2020 - August 2023

- Built and led high performing teams through structured recruitment, training, and mentorship, fostering accountability, adaptability, and a shared commitment to operational excellence.
- Implemented project management strategies to streamline operations, delegating tasks, monitoring KPIs, resolving on site technical and personnel issues, and coordinating with judges, election officials, and law enforcement to ensure secure, efficient elections.
- Oversaw task delegation and process execution, aligning people, priorities, and systems while managing documentation, troubleshooting, and cross team coordination to ensure smooth, scalable workflows.
- Developed data tracking templates and performance logs to assess operational efficiency, improve team coordination, and guide iterative process improvement.

Researcher, Georgia Health Policy Center
Atlanta, GA, United States of America
February 2020 - January 2021

- Led quantitative analyses using R and SAS, applying descriptive statistics, cross-tabulations, and logic model based evaluations to assess the impact of public health interventions using administrative and registry data.
- Built and maintained data pipelines integrating diverse sources such as policy archives, health registries, and interview data into harmonized formats for mixed methods modeling and reproducible reporting.
- Executed end to end analytic workflows, from data collection to dissemination, translating statistical outputs into stakeholder reports, policy briefs, and presentations to support evidence based decision making.
- Developed automated scripts in R to streamline data cleaning, ensure reproducibility, and accelerate turnaround for health program evaluations.

Research Assistant, Georgia State University
Atlanta, GA, United States of America
August 2019 - December 2022

- Executed advanced statistical analyses using R, SAS, and UCINET applying MANOVA, logistic regression, and social network analysis to investigate behavioral and policy-related health outcomes across longitudinal datasets.
- Led data cleaning, transformation, and exploratory analysis for mixed-methods studies, ensuring analytic rigor and reproducibility across research initiatives involving mental health and justice-involved populations.
- Developed and refined analytic workflows, integrating quantitative and qualitative data, contributing to grants, manuscripts, and scientific presentations grounded in rigorous statistical methodology.

Publications

Coldiron, J. W. et al. (2022). *Individual and Community-Level Factors Related to Contraceptive Access, Family Planning, and Reproductive Health Challenges Among Women in Kumasi, Ghana: A Field Study*. IJTMRPH. doi: 10.21106/ijtmrph.393

Coldiron, J. W. et al. (2021). *Genital Chlamydia Trachomatis Infection: Prevalence, Risk Factors and Adverse Pregnancy and Birth Outcomes in Children and Women in sub-Saharan Africa*. IJMCHA. doi.org/10.21106/ijma.523

Conferences and Lectures

Coldiron, J. W. (2025, May 2). *Building and Validating a Federated Learning Algorithm with Virtual Pooling for Rare Disease Research* [Conference presentation]. Columbia University Biostatistics Symposium, New York, NY.

Coldiron, J. W. (2023, February 16) *Genital Chlamydia Trachomatis Infection* [Conference presentation]. Georgia State University Public Health Conference, Atlanta, GA.

Coldiron, J. W. (2022, November 8) *Understanding Department of Juvenile Justice Stakeholder's Attitudes* [Conference presentation]. American Public Health Association Annual Conference and Expo, Boston, MA.

Coldiron, J. W. (2022, April 9). *Quantitative Analysis of Women's Health Trends in Ghana* [Conference presentation]. Troy MathFest, Troy, AL.

Coldiron, J. W. (2024, April 26). *Geographical Information Systems* [Invited talk]. Columbia University, New York, NY, United States.

Coldiron, J. W. (2022, February 5). *From Panic to Published* [Invited talk]. Georgia State University, Atlanta, GA, United States.

Coldiron, J. W. (2021, June 10). *Lessons Learned from Leveraging Peer-to-Peer Networks in Universities to Fight the Pandemic* [Invited talk]. Frontier Set, Seattle, WA, United States.

Licenses and Exams

Securities Industry Essentials (SIE) Exam, Financial Industry Regulatory Authority (FINRA)
Expected September 2025

Certifications

Advanced Cardiovascular Life Support (ACLS), American Heart Association
Issued September 2022
Credential ID: 225408189332

Basic Life Support for Healthcare Providers (BLS), American Heart Association
Issued June 2022
Credential ID: 225415630202

Human Subjects Research, Collaborative Institutional Training Initiative (CITI)
Issued November 2022
Credential ID: 50411525

Social and Behavioral Research, Collaborative Institutional Training Initiative (CITI)
Issued March 2020
Credential ID: 33581061

Community Engaged Research, Collaborative Institutional Training Initiative (CITI)
Issued March 2020
Credential ID: 33581062

Responsible Conduct of Research, Collaborative Institutional Training Initiative (CITI)
Issued March 2020
Credential ID: 33581057

Leadership

Academic: Biostatistics Department Liason, Academic Affairs Council, Council on Education for Public Health, Dean Search Committee
Professional: American Statistical Association, American Public Health Association, Council of State and Territorial Epidemiologist
Civic: Riverside Park Conservancy, Billion Oyster Project, Fathers Matter ATL, Boy Scouts of America