Jason Massey, MPH

[e-mail]
[Linkedin]
[Visual Portfolio]
[GitHub]

10+ years experience in **data science**, statistics, and causal inference. Expert in the data lifecycle: complex **EMR** linkages, management, machine learning/statistical analysis, and visualizations. An empathetic, cross-functional leader in **healthcare** with experience improving efficiency, performance, **health outcomes**, and presenting to stakeholders.

WORK EXPERIENCE

Senior Mathematical Statistician (Data Science and Health Informatics)

Centers for Disease Control and Prevention | May 2024-Present

- Used propensity weighted truncated gamma model from EHR data prompting increased clinical testing and reduced healthcare costs of insured population
- Acted as the main statistical resource to multiple disease branches. Included consulting clients, pulling EHR data, performing regression, running statistical tests, developing and validating machine learning models
- Mentored a team of junior statisticians in analysis, coding, and machine learning. Supervisor and manager of professional development
- Created/led data group holding biweekly meetings to communicate upskilling in data science, machine learning, statistics, reproducible code, and informatics. Topics include model development, tuning hyperparameters etc.
- Designed data pipelines for disease tracking. Included generating reusable frameworks and leveraging language learning models to increase efficiency in outbreak detection and forecasting disease progression

Healthcare Data Scientist (Value-Based Care and Clinical Effectiveness)

Centers for Disease Control and Prevention | April 2023-May 2024

- Used the NHSN to translate trillions of insurance claims data of more than 65,600 Long Term Care (LTC) facilities into research, improve clinical practices for staff, and value-based care for patients
- Implemented pilot study and developed random forest and logistic regression models in select facilities to reduce catheterization-induced urinary tract infections by 30% and over-prescribed antibiotic usage by 10%
- Used cluster analysis to identify longitudinal COVID curves in LTC facilities and logistic regression to assess the odds of association between high and low-peak LTC facilities in vaccination rate, insurance type, and demographics
- Ran doubly robust IPW time-to-event analysis comparing the hazard of COVID among vaccination dose and type in LTC facilities; Booster adherence had a reduced hazard of infection among minorities with Medicare/medicaid
- Helped in development of R package which pulls and joins unstructured, census, demography, CMS, NHSN, Truveta, insurance claims, HEDIS, stars, and other EMR/ERH notes to create qualitative/quantitative variables for analysis
- Created/led both the social determinants of health and statistical coding workgroups. Presented trainings on methods in data science, coding, machine learning, health equity and study design in health services research

Health Equity Data Scientist (Analytic Performance)

American Cancer Society | June 2021- April 2023

- Used hierarchal Bayesian model to assess the spatial association between healthy food access and life expectancy in U.S. Census Tracts. Findings showed reduced life expectancy in vulnerable neighborhoods informing policy changes
- Data curator for health equity science team; developed a data sharing agreement, data dictionary, and analytic datasets for planning execution of science initiatives; used electronic health records to uncover cancer disparities
- Mixed effects multinomial logistic regression model and principal component analysis found that economically segregated neighborhoods resulted in an increase in the odds of late-stage cancer diagnosis among all cancer sites
- Created dashboard metrics to evaluate app usage on health outcomes. Compared interventions using difference-in-differences financial models. Led to increased app usage by 20%, reduced healthcare costs, and reduced ER visits
- Monitored/evaluated surveillance system cancer dashboards to report cancer statistics and plan population health strategies
- Ran repeated measure analysis with flatiron health data to assess drug therapy effects on tumor size. Used a linear spline model and autogressive covariance structure to measure average treatment effect and individual patient variation

LANGUAGES/ TOOLS

R programming, SQL, ArcGIS, SAS, sparklyr, Python,Tableau, PowerBI, RStan, spark

FAVORITE LIBRARIES

Tidyverse, Lubridate, mlr3, PSWeight, boot, geepack, ggeffects, Caret, Janitor, Renv, NumPy, Pandas, Matplotlib, Scikit-learn

SKILLS/METHODS

Advanced epidemiological methods, Linear Regression, Logistic Regression, Longitudinal Data, NLP, LLM, ICD, CPT, Truveta, Epic, Marketscan, Fixed/Random Effects, Imputation, ETL, big data, Bootstrapping, discontinuity regression ,Clustering, OMOP, GitHub, snowflake, MLOps, Databricks, GAMS, Predictive modeling, Relational databases, Agile, algorithms, data modeling, quasi-experiments, Model Training, Validation, Hierarchal models/Mixed Models, propensity score matching, RWE, Bigquery, spark, Decision Trees, Random Forests, DAGs, AWS, EHR, Propensity Scores, Azure, sagemaker, HEOR, LASSO, Data Mining, statistical modeling, hypothesis testing, data analytics, manuscripts, proposals, public speaking, communication, quality assurance, insurance claims data, Perplexity, Claude, Gemini, ChatGPT, predictive modeling, data cleaning, analytic data lifecycle, data management, data visualization, SDoH, health disparities

Epidemiology Data Scientist (Maternal Child Health Outcomes)
Oak Ridge Institute | August 2020-June 2021

- Led data management for over 44,000 mothers in the BD-Steps database case control study. Used queries to create analytic datasets. Measured impact of maternal/paternal age, and opioid usage for more than 50 birth defects
- Used regression trees to measure demographic and geographic factors related to propensity to seek and quality of care for mothers of children with birth effects. Used electronic medical records to examine risk factors
- Provided consultations, training evaluations with state and local health clinicians on execution of clinical effectiveness, data collection, and evaluation working with real-world evidence (RWE) and administrative claims data
- Conducted replications, validations, analyses, and automated monthly reports on longitudinal birth defects surveillance and observational data
- Collaborated with commercial facing LabCorp and Quest Diagnostics to perform probability weighted statistical genetic analysis and discover relationship between biomarker, genome, and behavioral RWD

Epidemiology and Biostatistics Researcher
Emory University | January 2019 - July 2020

- Wrote literature review, SAP, and thesis paper on the true burden of influenza from hospital data in LA county
- Linked ICD 9/10 code data from large datasets from CDC, NOAA, and Los Angeles Health Department
- Estimated the Burden of Influenza with Quasi-Poisson Time Series Spline Models via primary and secondary diagnoses
- Used pneumonia as a control, CVD as a negative control, and the Delta method to approximate the model variance

EDI Data Analyst
Advance Auto Parts | 2016-2017

- Provided project management to hundreds of clients on shipping and purchasing data to improve efficiency and ROI
- Used various tools to pull, clean, and analyze data. Created reports to consult clients and data team
- Automated Tableau dashboard generation of KPI's for customers. Indicators that were previously inaccessible and now usable for financial steering

Mathematics and Statistics Tutor
North Carolina State University | 2011-2018

- Taught hundreds of clients in courses ranging from school to graduate level mathematics and statistics. Divided complex concepts into accessible parts. Communicated solutions via graphical, numeric, and algebraic approaches.

ACADEMIC EDUCATION

Master of Public Health in Epidemiology
Emory University Rollins School of Public Health | May 2020

Spatial certificate: spatial epidemiology in R, health and GIS, spatial statistics, remote sensing, geography of health care delivery

Bachelor of Science in Mathematics (Actuarial Science)
North Carolina State University | May 2013

Minor in Outdoor Education: adventure pedagogy, emergency preparedness, wilderness first responder, rock climbing I, II

VOLUNTEER EXPERIENCE

Surveyor | Raleigh LGBT Center
Conducted behavioral surveys with men who have sex with men
Belay Instructor | City of Raleigh Parks and Recreation
As a part of my minor practicum I helped to establish and facilitate an instructional rock-climbing belaying clinic in the community
Co-Chair | CDC Built Environment Work Group
Facilitated research on urban/rural infrastructure, walkability, greenspace, heat deserts, sustainability, and public transportation

Publications available on LinkedIn & References available upon request

COURSEWORK

MPH - Causal inference methods, biostatistics I, II, epi models, study design, survival analysis, longitudinal analysis, applied machine learning, time series analysis, Bayesian methods, social epi, confounding, interaction, effect measure modification, selection, bias, information, bias, cohort, case-control, clinical trials (RCT)

B.S. - Honors probability, statistical inference, SAS, differential equations, advanced calculus, linear algebra, stochastic modeling, long/short term actuarial models, Monte Carlo methods, econometrics, signal processing, risk analysis, Fourier analysis