JASON MASSEY

919.631.4308
jasonleemassey1991@gmail.com
GitHub: https://github.com/jlmassey1991
Visual Portfolio: https://jasonmassey.crevado.com/

PROFESSIONAL SUMMARY

15 years of experience in data science, statistics, and mathematics. I specialize in the analytic data cycle: performing complex data linkages from many sources, data management, using advanced machine learning methods, conducting analyses and creating professional visualizations. An adaptable team lead and communicator with experience tackling complex public health problems.

Relevant Skills: R programming, SAS, ArcGIS, SQL, Matlab, Python, Macros, PowerBI, Excel, Linear Regression, Logistic Regression, Longitudinal Analysis, Automation, Fixed/Random Effects, Time Series, Imputation, Bootstrapping, Clustering, Reproducibility, QA, Language Learning Models (LLM), GitHub, Databricks, GAMS, Predictive modeling, Relational databases, Agile, Deep learning, Model Training, Validation, Testing, AI, Neural Networks, Hierarchal models/Mixed Models/Multilevel Models, Supervised/Unsupervised Learning, Decision Trees, Random Forests, Causal Inference, Propensity Scores, AWS, Public Speaking, and strong communication

Favorite Packages: Tidyverse, Lubridate, mlr3, Caret, Janitor, Renv, NumPy, Pandas, Matplotlib, TensorFlow, PROC SQL

PROFESSIONAL EXPERIENCE

Senior Mathematical Statistician (Surveillance, Informatics, and Statistics Office)

Centers for Disease Control and Prevention | May 2024-Present

- Lead analyst for a project that measures the effect of diagnostic delays of fungal pathogens on healthcare costs. Performing complex data linkages to capture Marketscan ICD-10 data on symptom and diagnosis dates. Quantile binning probability weighting was used to account for the causal relationship of time. Categorical regressions were conducted for a zero-inflated exposure. Findings estimate daily increases in cost and periods of large jumps in costs as guidance for policy and clinical recommendations. (SQL/R/Python)
- Acted as the main statistical resource to the Mycotic Diseases Branch and others. Including consulting clients, training colleagues, and conducting both collaborative and independent research for manuscripts. (SQL/SAS/R/Python)
- Mentor/guide to multiple fellows. Provided training in areas of statistical analysis, data management, writing manuscripts, and presenting to various audiences. Helped to formulate reproducible code in SAS and R for pulling and managing data, creating analytic files, performing analysis, and producing table shells. Managed progress and professional development.
- Created and led data huddle that holds biweekly meetings to foster upskilling in data science, machine learning, statistics, creating reproducible code, and informatics. Managed projects and teams in ongoing projects. (SQL/SAS/R/Python)
- Used cutting-edge DCIPHER software to rapidly link vast amounts of disparate foodborne, waterborne, and environmental outbreak data from various sources, analyze, and configure operational workflows for public health needs. This includes generating reusable frameworks and leveraging language learning models to increase efficiency for performing tasks related to informatics and analytics.

Biostatistician II (Surveillance Branch)

Centers for Disease Control and Prevention | 2023-2024

- Used the National Healthcare Safety Network to translate multileveled healthcare associated infection surveillance data of more than 65,600 Long Term Care health facilities into meaningful insights and published research. (Python/SQL/R)
- Acted as project lead for the geocoding of LTC facilities including planning, organizing meetings, acquiring data access, validation of geolocations, and conversion of coordinates to census tracts. Linked demographic data such as the social vulnerability index and other social determinant of health metrics.
- Used machine learning package kmlshape to create novel cluster analysis which monitors and examines omicron curve peaks in LTC facilities. This included complex data linkages to the NHSN using bootstrapping to account for instability. Then calculated the bootstrapped adjusted logistic regression to assess the association between high and low peak long term care facilities and outcomes including booster status, insurance type, and socioeconomic demographics.
- Conducted a survival analysis in State Veteran Home facilities which investigated the hazard of being infected by COVID among varying vaccination doses and types. Involved complex data linkages to person-level and facility-level data and creating a nested control group of those who did not receive the bivalent booster. (SAS/SQL/R)
- Implemented automation and modernization of surveillance databases. Merged surveillance data to census, geolocation, CMS, and NHSN data to create analytic datasets for analysis and created interactive maps and dashboards in Power BI. (Power BI/ArcGIS/SQL)
- Oversaw multiple coding training presentations and provides expertise and guidance via tutorials on high-level data science information. Advises colleagues on programming and data science inquiries. Led projects and presentations using decision trees and random forests. (SAS/SQL/R)

Associate Scientist II (Cancer Disparity Team)

American Cancer Society | 2021-2023

- Led a paper in which a hierarchical Bayesian spatial model in R was used to assess the association between healthy food access and life expectancy in United States Census Tracts. This included measuring neighborhood level social determinants of health and assessing public health implications. Findings showed an independent association between life expectancy and access to healthy foods especially when in the presence of reliable public transit. Helped influence public health policy.

- In partnership with the state of Idaho cancer registry developed a data sharing agreement, data dictionary, and preprocessing dataset for streamlining multiple analytic datasets.
- Led a paper in which a multinomial logistic regression was used to assess the association between economic segregation and stage of cancer diagnosis within Idaho census tracts. Principal component analysis was used to measure income inequality extremes and found that every categorical increase in economic segregation resulted in stepwise increase for the odds of late-stage cancer diagnosis among all cancer sites. Research included CDISC standards and oncology knowledge.
- Helped to create a nationwide Cancer Atlas for ACS staff; an interactive Web Map Dashboard with cancer statistics, socioeconomic status, office locations, and congressional districts (R-Shiny)
- Led the monitoring, evaluation, and management of the Cancer Atlas/SEER surveillance systems (SQL)
- Conducted weighted multinomial logistic regression on health disparities among various cancer sites (SAS/SQL/R)
- Used Power BI, Tableau, and ArcGIS to develop dashboards and web mapping tools with cancer statistics

Surveillance Epidemiology Analyst

Oak Ridge Institute | 2020-2021

- Led data management for over 44,000 mothers in BD-Steps database; birth defect case-control. This involved using queries to store and organize data to be used in future analyses. (SAS/SQL/Python)
- Evaluated several public health departments' surveillance systems for federal funding approval
- Led ongoing development and implementation of nationwide surveillance system by strategizing with stakeholders
- Provided consultations, trainings, and workshops with state and local health departments on survey methods, data collection, and data analysis
- Conducted replications, validations, created visualizations for longitudinal analyses, and automated monthly data reports on birth defects data (SAS/SQL/Power BI/Tableau)
- Conducted literature reviews and paper proposals on multiple birth defects risk factors

Biostatistics Research Assistant

Emory University | 2018-2020

- Created a literature review and analysis plan on the burden of influenza from hospital morbidity data in LA county
- Collected and linked large datasets from CDC, NOAA, and Los Angeles Health Department (SQL/SAS)
- Fit time series models and estimate predictions for the burden of influenza ICD 9 and 10 codes (R)
- Performed spline and polynomial methods, stratified by age, adjusted for influenza (primary and secondary diagnoses), time, holidays, facility code, day of week, temperature (min, avg, max), air pollution (CO3, PM25), analyzed effect measurement modifiers, and used a 3-day lag
- Translated reproducible code between SAS and R for courses

Mathematics and Statistics Tutor

North Carolina State University | 2011-2017

- Tutored mathematical and statistical modeling, probability, linear algebra, calculus, differential equations, real analysis

Electronic Data Interchange (EDI) Analyst

Advance Auto Parts | 2016-2017

- Contracted to analyze EDI files that tracked purchase orders and shipping records (Excel/SQL)
- Used internal software to store, manipulate, deduplicate, flag anomalies, and analyze data
- Consulted with clients to fix clerical errors and train in EDI formatting

ACADEMIC EDUCATION

***Emory University Rollins School of Public Health,* (May 2020)**
Master of Public Health, Epidemiology (3.7)

***North Carolina State University*, (May 2013)**
Bachelor of Science in Applied Mathematics (3.2)

VOLUNTEER EXPERIENCE

The Task Force for Global Health
In collaboration with the Gates Foundation, mapped the prevalence of Lymphedema, Filariasis, and Onchocerciasis, or river blindness jointly with the count of infected fly sites that plague the Muheza District, Tanzania. Included advanced techniques in ArcGIS Pro.

Raleigh LGBT Center
Provided public health educational resources regarding HIV and sexual health to men who have sex with men.

Rock Club Rock Climbing Belay Coordinator
As a part of my minor practicum helped to establish and facilitate an instructional rock-climbing belaying clinic in community.

PUBLICATIONS

Jason Massey, Daniel Wiese, Farhad Islami, Ahmedin Jemal, Marjorie McCullough (2023): *"The Association Between Census Tract Healthy Food Accessibility and Life Expectancy in the United States".* Journal of Urban Health. *The Association Between Census Tract Healthy Food Accessibility and Life Expectancy in the United States - PubMed (nih.gov)*

Jeneita Bell et al. (2023): "*Influenza and Up-to-Date COVID-19 Vaccination Coverage Among Health Care Personnel — National Healthcare Safety Network, United States, 2022–23 Influenza Season"*. Centers for Disease Control and Prevention. MMWR. 72(45);1237–1243. Influenza and Up-to-Date COVID-19 Vaccination Coverage Among Health Care Personnel — National Healthcare Safety Network, United States, 2022–23 Influenza Season | MMWR (cdc.gov)

Farhad Islami MD Daniel Wiese, Emily C Marlow, Tyler B Kratzer, **Jason Massey**, Hyuna Sung, Ahmedin Jemal (2023): *Progress in reducing cancer mortality in the United States by congressional district, 1996–2003 to 2012–2020.* Cancer. *https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.34808*

Jason Massey, Daniel Wiese, Farhad Islami, Ahmedin Jemal (expected 2025): *"Index of Concentration of Extremes and Cancer Survival in Idaho Census Tracts"*

Lu Meng*, Jeneita M. Bell*, Sydney Guthrie, Kira Barbre, Larry Mason**, Jason Massey**, Ryan Wiegand, Theresa Rowe, Austin Woods, Iram Qureshi, Hannah Reses, Alfonso Hernandez-Romieu, Matthew J. Stuckey, David Kuhar, Megan C. Lindley, Andrea Benin (expected 2025): *"High versus low SARS-CoV-2 infection peak and resilience to COVID-19 surge among nursing home residents during Omicron variant BA.1 wave in the United States, December 20, 2021-March 20, 2022, National Healthcare Safety Network, United States"*

Lu Meng et al. (expected 2025): *Vaccination and risk of COVID infection in State Veteran Home Facilities*

Kaitlin Benedict, **Jason Massey**, Michelle Fearon Scales, Ian Hennessee, Samantha L. Williams, Mitsuru Toda (expected 2025): *"Impact of delays in diagnosis on healthcare costs associated with blastomycosis, coccidioidomycosis, and histoplasmosis in a commercially insured population"*

REFERENCES

Lu Meng, PhD
Deputy Team Lead
CDC | Division of Healthcare Quality and Promotion
prj8@cdc.gov
404.498.2905

Logan Ray, MPH
Associate Director of Surveillance
CDC | Surveillance, Informatics, and Statistics Office
prj8@cdc.gov
404.498.2905

Lawrence McDonald, MD
Associate Director for Science
CDC | Division of Healthcare Quality and Promotion
ljm3@cdc.gov
404.639.3833

Lyndsay Bottichio, MPH, DrPH
Principal Deputy Director
CDC | Division of Readiness and Response Science
prj8@cdc.gov
404.498.2905

Preeta Kutty, MD, MPH
Medical Epidemiologist
CDC | Division of Healthcare Quality and Promotion
chz6@cdc.gov
404.639.0470

Daniel Wiese, PhD
Principal Scientist (Cancer Disparities)
American Cancer Society | Surveillance
Daniel.wiese@cancer.org
215.459-9483

Theresa Rowe, DO
Unit Team Lead (Long-Term Care)
CDC | Division of Healthcare Quality and Promotion
nnq1@cdc.gov
404.498.2905

Farhad Islami, MD, PhD
Senior Scientific Director (Cancer Disparities)
American Cancer Society | Surveillance
Farhad.islami@cancer.org
678.528-3133