

Travis Gerke, ScD

EPIDEMIOLOGIST • DATA SCIENTIST • STATISTICIAN

H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, Tampa, FL 33612

☎ +1 706 346 4899 | ✉ travisgerke@gmail.com | 🏠 travisgerke.com | 📷 [travisgerke](#) | 🗣 [tgerke](#) | 📺 [travisgerke](#)

Education

Summary: I am jointly trained as a statistician and epidemiologist. I apply these skills as a data science leader to guide data collection strategies (epidemiology) that will enable optimal insights and robust decision-making from downstream consumption (statistical translation).

Harvard School of Public Health

DOCTOR OF SCIENCE (ScD), EPIDEMIOLOGY

Boston, MA

2014

Harvard University

MASTER OF ARTS (AM), BIostatISTICS

Cambridge, MA

2011

University of Florida

BACHELOR OF SCIENCE (BS), STATISTICS

Gainesville, FL

2009

BACHELOR OF ARTS (BA), MATHEMATICS

2009

Experience and impact

Summary: My career as a quantitative scientist spans not only the traditional principal investigator role, but also includes enterprise leadership in the data science domain. I have led or collaborated on 14 successful grant applications, many of which helped drive my 80+ peer-reviewed publications and 6 scientific software releases. At the enterprise level, I guide implementation of data science platforms that enable rapid and reproducible analytic insights.

H. Lee Moffitt Cancer Center & Research Institute

HEALTH INFORMATICS DIRECTOR

Tampa, FL

2020–present

- Create value for Moffitt data consumers as the agile product owner of cloud-based data marts (AWS/Snowflake) and associated informatics tools (RStudio/Python/Shiny) designed meet the clinical, research, and operational analytics needs of Moffitt's faculty and administration
- Facilitate data workflows across multiple teams around the Center, and provide the foundation for predictive analytics, production pipelines for machine learning/AI applications, and data sharing partnerships

SCIENTIFIC DIRECTOR, COLLABORATIVE DATA SERVICES

2018–present

- Provide scientific and technical leadership for Moffitt's research data provisioning hub. Key initiatives include: (i) transition a traditionally Excel-based data group to modern R frameworks with git versioning; (ii) enhance interfacing with biostatistics/bioinformatics end-users to deliver analytically-ready data

ASSISTANT MEMBER, DEPARTMENT OF CANCER EPIDEMIOLOGY

2016–2020

- Direct academic research lab with emphasis in: (i) developing modernized data science workflows for molecular epidemiology research; (ii) causal inference for decision science and real world evidence studies; (iii) reproducible data science frameworks

University of Florida

ASSISTANT PROFESSOR, DEPARTMENT OF EPIDEMIOLOGY

Gainesville, FL

2014–2016

- Direct research lab with emphasis in: (i) machine learning in studies of gene expression; (ii) methods for high-dimensional data reduction

Harvard School of Public Health

STATISTICIAN I / RESEARCH DATA ANALYST, DEPARTMENT OF EPIDEMIOLOGY

Boston, MA

2011–2014

- As lead statistician in an epidemiology lab, ensure appropriate application of statistical methods and provide accompanying mentorship of students/postdocs as needed; as research data analyst, assist with manuscript and grant development

Dana-Farber Cancer Institute

RESEARCH DATA ANALYST, DEPARTMENT OF BIostatISTICS

Boston, MA

2012–2013

- Lead Python developer in a collaboration with the Google Earth team to estimate public health impacts of global glacier retreat

Department of Veterans Affairs

RESEARCH DATA ANALYST, REHABILITATION OUTCOMES RESEARCH CENTER

Gainesville, FL

2008–2009

- Perform SAS-based econometric health-care delivery analyses

Advisory roles, mentorship, and outreach

Summary: As a result of my deep appreciation for modernized data governance, scientific computing, and reproducible data science frameworks, I provide strategic input for a wide array of relevant organizational initiatives. Successful implementation of such initiatives requires understanding the needs and partnership/mentoring of data science end-users. As an active and visible member of the broader data science community, I enjoy sharing successes/challenges and learning from the experience of others.

SELECTED ADVISORY ROLES

- Moffitt Research Data Analyst Task Force (Chair): The goal of this task force was to modernize the institution's research data analyst job ladder to align with data science job seekers, both with respect to search criteria and salary expectations
- Moffitt High Performance Computing Steering Committee (Member): This committee ensures that services and technology stack offered by the Moffitt compute cluster meet the evolving needs of research end-users
- Moffitt Health and Research Informatics (HRI) Governance Committee (Member): This committee provides guidance and decision-making for data warehouse-related activities, and aims to build trust in the institution's high quality data assets by establishing shared understanding through data availability, quality, and usability
- Flatiron Health Academic Research Oversight Committee (Member): This committee provides guidance on how to best use Flatiron national data to advance academic research and identify opportunities for collaboration across sites
- Flatiron R Package Universe (Core Team Member): This team of developers contributes to R packages which facilitate best practices in feature engineering and analysis of Flatiron real-world data

SELECTED MENTORSHIP AND TEACHING ROLES

- AACR Integrative Molecular Epidemiology Workshop (annual faculty member, 2017–2021): At this workshop delivered to postdocs/early career faculty, I teach sections on effective data management and sharing, big data resources, and predictive modeling/machine learning
- University of Florida (2015–2016): In addition to formally mentoring 8 graduate students on thesis/dissertation committees, I designed and taught graduate courses focused on (i) measurement and data capture in health outcomes research and (ii) introductory causal inference
- Harvard School of Public Health Summer Program in Epidemiology, Introduction to Biostatistics and Statistical Programming (2013): I designed and taught an 8 week course which focused on using the R language to implement common biostatistical methods

SELECTED OUTREACH ACTIVITIES

- Founding Organizer, Tampa Bay R Users Group: Founded in 2017, this data science group has grown to over 500 members, and meets monthly to discuss introductory topics through state of the art in R programming
- Executive Board Member, Florida Interdisciplinary Data Sciences Consortium (2019–present): Provides a forum to advance interdisciplinary data science expertise and training to students, the USF community, and the state; features a biweekly seminar series and short courses/workshops

Technical experience

Summary: My philosophy on appropriate technology tooling is to use that which most efficiently delivers interpretable value to key stakeholders. This stance guides my regular usage and enthusiasm for R/RStudio solutions which feature a full statistical toolkit with deliverables commonly designed for the modern web browser. I also have experience in cloud and/or high-performance computing cluster environments, where key stakeholders may be other data scientists who place high value on efficiency at scale; in such settings, I have deployed Python, SAS, or SQL (often through R database interfaces) as appropriate. In all technical work, I emphasize that reproducibility and readability for future users of code is of paramount importance, with git playing a vital role in this regard.

• R/R Markdown/Shiny • Python • Snowflake • SQL • SAS • git/GitHub • \LaTeX •