Travis Gerke, ScD

EPIDEMIOLOGIST • DATA SCIENTIST • STATISTICIAN

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

+1 813 745 6448 | travis.gerke@moffitt.org | 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 — *Boston, MA*

Doctor of Science (ScD), Epidemiology — *2014*

Harvard University — *Cambridge, MA*

Master of Arts (AM), Biostatistics — *2011*

University of Florida — *Gainesville, FL*

Bachelor of Science (BS), Statistics — *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 administrative leadership in the data science domain. I have led or collaborated on 14 successful grant applications, many of which helped drive my 70+ peer-reviewed publications. My team science efforts have resulted in over 60 scientific conference presentations, both nationally and internationally. I have also led or co-developed 5 scientific software releases.

H. Lee Moffitt Cancer Center & Research Institute — *Tampa, FL*

Health Informatics Director — *2020–present*

- Serve as the primary liaison between Moffitt data consumers and the Health Informatics team, designing cloud-based data marts and associated informatics tools to 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 and data sharing partnerships

Scientific Director, Collaborative Data Services — *2016–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 — *Gainesville, FL*

Assistant Professor, Department of Epidemiology — *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 — *Boston, MA*

Statistician I / Research Data Analyst, Department of Epidemiology — *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 — *Boston, MA*

Research Data Analyst, Department of Biostatistics — *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 — *Gainesville, FL*

Research Data Analyst, Rehabilitation Outcomes Research Center — *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 undertanding 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 Data Documentation and Management Best Practices Task Force (Member): The goal of this task force was to establish and communicate institutional best practices for preserving and documenting research-generated data
- Moffitt Cancer Epidemiology Research Data Analyst Task Force (Chair): The goal of this task force was to modernize the 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 HRI-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
- Moffitt HRI Molecular Data Training Task Force (Member): This task force facilitates researcher querying of Oracle-based molecular data by designing and implementing a training program for appropriate data science team members
- 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

Selected Mentorship and Teaching Roles

- AACR Integrative Molecular Epidemiology Workshop (annual faculty member, 2017–present): At this workshop delivered to postdocs/early career faculty, I teach sections on effective use of 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 the 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 philosphy on appropriate technology tooling is to use that which most efficiently delivers interpretable value to key stakeholders. This stance guides my regular usage and enthusiam 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 reproduciblity 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 • SAS • git/GitHub • LaTeX •