

Ryan Harrison

Principal Biostatistician / Data Scientist
Clinical Data Science | Real-World Evidence | R/Shiny

Professional Summary

Principal biostatistician and data scientist with 15+ years of experience spanning clinical development, real-world evidence, statistical modeling, and R-based analytical application development. Expert in R and Shiny, with extensive experience building reusable analytical applications, packages, and reproducible workflows for clinical and biomedical data. Combines deep statistical expertise with modern software engineering practices, including Git/GitHub, testing, CI/CD, and enterprise deployment using the Posit ecosystem. Experienced in collaborating across biostatistics, clinical development, translational science, data science, and application development teams to translate complex analytical needs into practical, decision-ready tools and insights.

Core Skills

Advanced Analysis & Modeling: Predictive modeling, machine learning, statistical inference, longitudinal analysis, survival analysis, causal inference, propensity score methods, mixed-effects models

Health Data & Claims Analysis: Claims data, registry data, real-world evidence, clinical trial data, patient-reported outcomes, observational studies

Programming & Data: R (expert), Python, SQL, tidyverse, data.table, DuckDB, Arrow, data validation, data integration, reproducible workflows

R Application Development: Shiny, bslib, golem, Rhino, modular application architecture, reactive programming, R package development, ggplot2, plotly, gt, gtsummary

Reporting & Visualization: Quarto, interactive dashboards, data visualization, analytical reporting, stakeholder presentations

Software Engineering & DevOps: Git/GitHub, pull requests and code review, GitHub Actions, CI/CD, renv, testthat, shinytest2, pre-commit, Posit Connect, Posit Workbench

Collaboration & Leadership: Cross-functional collaboration, stakeholder engagement, mentoring, technical leadership, end-to-end project ownership

Recent Professional Experience

Principal Biostatistician – Syneos Health

2026–Present

- * Develop and maintain R/Shiny applications for interactive exploration, visualization, and analysis of clinical trial data.
- * Design modular, reusable application architectures using Shiny, bslib, golem, and Rhino, with emphasis on maintainability, performance, responsive user interfaces, and consistent development patterns.
- * Develop reusable R packages, Shiny modules, and analytical components using modern functional and object-oriented programming practices.
- * Apply modern R software engineering practices including Git/GitHub, pull requests and code review, renv, testthat/shinytest2, pre-commit tooling, and CI/CD with GitHub Actions.
- * Deploy and maintain analytical applications and technical documentation on Posit Connect, including automated deployment workflows.
- * Develop and maintain internal application development guidelines, establishing standards and best practices for R/Shiny architecture, package development, testing, documentation, version control, and deployment.
- * Collaborate with statisticians and application developers to translate clinical and statistical requirements into intuitive interactive analyses and visualizations.
- * Evaluate and integrate emerging technologies and R ecosystem tools to improve application development, analytical workflows, performance, and user experience.
- * Provide technical guidance, training, and consultation on R/Shiny development, application architecture,

Git/GitHub workflows, and development best practices.

- * Support exploratory analyses of clinical trial data using modern R tools for data manipulation, statistical analysis, visualization, and reporting.

Principal Clinical Data Analyst (Contract via Bayside Solutions) – Arcus Biosciences 2025–2026

- * Develop and maintain analytical applications and data pipelines supporting Translational Sciences teams, enabling exploration and analysis of clinical trial and biomarker datasets.

- * Design and deploy reproducible analytical workflows in R to support exploratory data analysis, statistical modeling, visualization, and dashboard-driven review.

- * Implement and maintain scalable, production-grade analysis pipelines and interactive dashboards (R/Shiny) to operationalize insights for scientific and cross-functional stakeholders.

- * Develop and maintain internal R packages and analytical applications for standardized reporting and branded visualizations, aligned with Good Statistical Practice and an emphasis on transparency, validation, and reproducibility.

- * Collaborate with cross-functional partners (Clinical Development, Biometrics, Translational Sciences, Data Engineering) to clarify analytical requirements, prototype solutions, and deliver validated analytical tools and dashboards.

- * Apply large language model (LLM) techniques, including retrieval-augmented generation (RAG), to prototype context-aware tools for unstructured data summarization and knowledge discovery.

- * Serve as technical lead for analytical infrastructure, including administration of Posit Connect and Workbench, CI/CD automation, and version-controlled deployment workflows.

- * Mentor peers and promote best practices in code review, documentation, and reproducible research across the data science team.

Senior Principal Data Scientist - Biogen

2022–2024

- * Led advanced statistical and predictive modeling efforts across neurology programs, applying regression, mixed-effects models, tree-based methods, penalized regression, and longitudinal modeling techniques.

- * Designed, developed, and deployed interactive analytical applications (R Shiny) to support portfolio-level data exploration, model interpretation, and clinical review.

- * Developed simulation and modeling approaches, including non-homogeneous Markov models, to support disease progression analyses in Alzheimer's disease.

- * Applied dimensionality reduction and unsupervised learning techniques (e.g., PCA, clustering) to analyze high-dimensional biomarker, genomic, and proteomic datasets.

- * Applied differential expression, feature importance, and machine learning pipelines to identify proteins associated with Alzheimer's Disease status and progression, reusing and extending these pipelines across 30+ diseases of interest.

- * Built analytical pipelines and applications using UK Biobank data to support large-scale statistical modeling and marginal effects analysis.

- * Built reproducible analysis pipelines in R & Python across AWS, Databricks, and high-performance computing environments.

- * Promoted reproducibility and transparency through version-controlled workflows, interpretable modeling approaches, and collaborative peer review.

- * Collaborated closely with clinicians, biostatisticians, and data science partners to translate complex analyses into decision-ready insights.

- * Contributed to abstracts, manuscripts, and internal reports supporting scientific communication and regulatory-facing analyses.

Lead Biostatistician - CorEvitas, LLC

2016–2022

- * Developed statistical and machine learning models on large registry and claims datasets, focusing on longitudinal trends, survival outcomes, causal inference, and treatment effects to support real-world evidence generation.

- * Authored statistical analysis plans, analysis reports, manuscripts, and abstracts for internal teams, external clients, and peer-reviewed publications.
- * Consulted with pharmaceutical clients to scope, design, and execute analysis projects across inflammatory and autoimmune diseases.
- * Automated recurring analytical workflows, migrated legacy codebases from STATA to R, and improved efficiency, reproducibility, and analytical consistency.
- * Provided technical leadership and direct supervision for biostatistics analysts, mentoring four team members while maintaining hands-on analytical ownership.
- * Contributed to peer-reviewed publications and conference presentations supporting dissemination of real-world clinical research.

Associate Research Scientist/Biostatistician - New England Research Institutes	2010-2016
---	-----------

- * Provided statistical support for clinical trials and survey-based studies using R and SAS.
- * Wrote statistical methods sections and created tables and visualizations for client reports and peer-reviewed publications.
- * Specialized in generalized linear models, mixed-effects models, and missing data imputation techniques.

Education	
M.S. Statistics - University of Massachusetts, Amherst, MA	2010
B.A. Mathematics - Boston University, Boston, MA	2007

Continuing Education & Certifications	
The Complete Python Bootcamp - Udemy	2023
DataOps Fundamentals Certification – DataKitchen	2022
The Complete SQL Bootcamp - Udemy	2022
Machine Learning A-Z: Hands-On Python & R in Data Science - Udemy	2021
Writing in the Sciences - Coursera	2021

Professional Memberships	
American Statistical Association – Boston Chapter ASA	2010-present

Publications	
* Extensive record of peer-reviewed publications derived from clinical trial and observational cohort data, contributing quantitative and statistical expertise to collaborative, multi-disciplinary research teams. * Multiple publications evaluating treatment effectiveness, patient-reported outcomes, disease burden, and clinical quality across chronic conditions. * Regular contributor to conference presentations, abstracts, and posters supporting dissemination of real-world clinical research. * (Full publication list available upon request or via ORCID)	