Samuel Anyaso-Samuel

CONTACT INFORMATION	Biostatistics Branch Division of Cancer Epidemiology & Genetics National Cancer Institute 9609 Medical Center Drive Rockville, MD 20850	Cell: (208) 216-9207 Office: (240) 276-5168 Email: samuel.anyaso-samuel@nih.gov Website: samuelanyaso.github.io

RESEARCH INTERESTS

Microbiome, Statistical genetics, Biological networks, Cancer Epidemiology, Cluster-correlated data analysis, Informative Cluster Size, Univariate and Multivariate time-to-event data.

EDUCATION & TRAINING

Postdoctoral Research Fellow　　　　　　　　　　　　　　　**Sept. 2023 - Present**
Biostatistics Branch,
Division of Cancer Epidemiology & Genetics,
National Cancer Institute, Rockville, MD, USA

University of Florida, Gainesville, FL, USA
Doctor of Philosophy (Ph.D.), Biostatistics　　　　　　　　　　　　**Aug. 2023**

- Thesis Topic: *Advances in cluster-correlated data analysis when cluster size is informative*
- Advisor: Somnath Datta, Ph.D.

Boise State University, Boise, ID, USA
Master of Science (M.S.), Mathematics　　　　　　　　　　　　　　**May 2019**

- Thesis Topic: *Dynamic Sampling Versions of Popular SPC charts for Big Data Analysis*
- Advisor: Partha Mukherjee, Ph.D.
- Area of Study: Statistics

Federal University of Technology, Owerri, Nigeria
Bachelor of Technology (BTech.), Statistics　　　　　　　　　　　　**Dec. 2014**

- Thesis Topic: *Some contributions to the interpretation of Fuzzy Regression Intervals*
- Advisor: Benson Onoghojobi, Ph.D.

PROFESSIONAL EXPERIENCE

Biostatistician　　　　　　　　　　　　　　　　　　　**Aug. 2020 - Aug. 2023**
Brain Rehabilitation Research Center,
NF/SG VHS Malcom Randall VA Medical Center, Gainesville, FL, USA

Graduate Research and Teaching Assistant　　　　　　　**Aug. 2019 - Aug. 2023**
Department of Biostatistics,
University of Florida, Gainesville, FL, USA

Graduate Teaching Assistant　　　　　　　　　　　　　**Aug. 2017 - May 2019**
Department of Mathematics,
Boise State University, Boise, ID, USA

HONORS AND AWARDS

- Selected for Multi Omics NETwork Analysis (MONET) Workshop at University of Colorado Anschutz Medical Campus (June 2025)
- Blackwell-Tapia Conference − Travel Award (Nov. 2024)
- Southern Regional Council on Statistics Summer Research Conference − Travel Award (May 2024).
- 2023 Lifetime Data Science Conference − Student Poster Award (June 2023).
- 2023 Symposium on Data Science & Statistics - Student & Early Career Travel Award (May 2023).

- 2023 Fostering Diversity in Biostatistics (ENAR) - Student Scholarship (Mar. 2023).
- Poster Award (Data Science); 2022 College of Medicine Research Day (Apr. 2022).
- UF Department of Biostatistics PhD Travel Award (Mar. 2022).
- 2020 Intelligent Systems for Molecular Biology - Fellowship Award (July 2020).
- 2019 JSM Diversity Workshop and Mentoring Program - Student Scholarship (July 2019).
- 36th ASA Quality & Productivity Research Conference - Student Scholarship (June 2019).
- 2019 ASA/IMS Spring Research Conference - Student Scholarship (May 2019).
- Best Poster; Workshop on Emerging Data Science Methods for Complex Biomedical and Cyber Data, *Department of Population Health Sciences, Medical College of Georgia* (Mar. 2019).
- Best Poster from College of Engineering; Boise State University service-learning student exhibition, *Boise State University* (Dec. 2018).
- ACM Richard Tapia Celebration of Diversity in Computing Conference Scholarship (Sept. 2018).
- Student Representative, American Statistical Association, *Boise State University* (June 2018).
- Graduate Summer Fellowship, Department of Mathematics, *Boise State University* (May 2018).
- Alfred M. Dufty Jr. Award, *Boise State University* (May 2018).
- Computing Research Association (CRA) Sponsorship for CRA URMD workshop (Mar. 2018).
- Graduate Residential Scholars Program, *Boise State University* (Aug. 2017).

PUBLICATIONS

10. Bather, J., **Anyaso-Samuel, S.**, Chen, Y., Elliott, L., Bennett, A., and Goodman, M. Cluster-weighted modified Poisson regression for estimating risk ratios in longitudinal data with informative cluster sizes. doi.org/10.1101/2025.05.23.25328253 (2025+).

9. **Anyaso-Samuel S.**, and Datta S. Nonparametric estimation of a future entry time distribution given the knowledge of a past state occupation in a progressive multistate model with current status data. doi.org/10.48550/arXiv.2405.05781 (2025+).

8. **Anyaso-Samuel S.**, Datta S., Roos E., and Nevalainen J. Can the unit size predict outcomes? Testing for informativeness in three-level designs. *Statistics in Medicine*, (In Print) (2025).

7. **Anyaso-Samuel S.**, and Datta S. Testing for marginal covariate effect when the subgroup size induced by the covariate is informative. *Statistical Methods in Medical Research*, 33(7):1264-1277, (2024).

6. Sarkar S., **Anyaso-Samuel S.**, Qiu P., and Datta S. Multiblock Partial Least Squares and Rank Aggregation: Applications to Detection of Bacteriophages Associated with Antimicrobial Resistance in the Presence of Potential Confounding Factors. *Statistics in Medicine*, 43(13):2527-2546 (2024).

5. Leinen M., Grandy E., Gebel L., Santana T., Rodriguez A., Singh S., Fernandez M., Dalugdug J., Garcia-Colon E., Lybeshari K., Alexander D., Maura M., Gonzalez M., Almeida C., **Anyaso-Samuel S.**, Datta S., and Schiefer M. Bilateral Subdiaphragmatic Vagal Nerve Stimulation Using a Novel Waveform Decreases Body Weight, Food Consumption, Adiposity, and Activity in Obesity-Prone Rats. *Obesity Surgery*, 34(1):1-14, (2024).

4. **Anyaso-Samuel S.**, Bandyopadhyay D, and Datta S. Pseudo-value regression of clustered current status data with informative cluster or subcluster sizes in a multistate model. *Statistical Methods in Medical Research*, 32(8):1494-1510, (2023).

3. **Anyaso-Samuel S.**, and Datta S. Adjusting for informative cluster size in pseudo-value based regression approaches with clustered time-to-event data. *Statistics in Medicine*, 42(13): 2162-2178 (2023).

2. **Anyaso-Samuel S.**, Sachdeva A., Guha S., and Datta S. Bioinformatics preprocessing of microbiome data with an application to metagenomic forensics. In *Statistical Analysis of Microbiome Data*, (pp. 45-78), Eds: S. Datta and S. Guha, Springer (2021).

1. **Anyaso-Samuel S.**, Sachdeva A., Guha, S., and Datta S. Metagenomic geolocation prediction using an adaptive ensemble classifier. *Frontiers in Genetics, 12*, p.642282 (2021).

OTHER ABSTRACTS

3. Ashby F., **Anyaso-Samuel S.**, Gamlin P., Kabbej N., Andraka N., Mandel R., Riva A., Datta S., Heldermon C. AAV-barcoding for High-throughput Screening of Vector Transduction Efficiency in the CNS of Cynomolgus Macaques Compared to C57BL/6 Mice. *Florida Genetics Symposium*, Gainesville, FL, (Nov. 2022).

2. Kabbej N., Ashby F.J., Riva A., **Anyaso-Samuel S.**, Datta S., Heldermon C.D. Transcriptomic Disparities Between Male and Female Non-Human Primates Related to AAV Transduction Efficiency. *American Society of Gene & Cell Therapy (ASGCT) 25th Annual Meeting*, Washington DC, (May 2022).

1. Ashby F., Kabbej N., Riva A., Rouse C.J., Hawkins K., Andraka N., **Anyaso-Samuel S.**, Gamlin P., Mandel R., Kondratov O., Zolotukhin S., Datta S., Heldermon C. Genetic Barcoding Identifies Similar Transduction Efficiency Rankings within Disease Models of Sanfilippo Syndrome Type-B and Controls. *19th Annual WORLDSymposium*, Orlando, FL, (Feb. 2022).

PRESENTATIONS

Invited, Refereed & Seminar Talks

- Identifying high-dimensional genomic factors associated with biological networks, *Department of Biostatistics at Virginia Commonwealth University* (Aug. 2025).

- Inference for current status observations in a multi-state setting, *Department of Statistical Sciences and Operations Research at Virginia Commonwealth University* (Aug. 2025).

- Early career panel: Transition from graduate school to the workplace, *2025 Diversity Mentoring Workshop at the Joint Statistical Meetings*, NASHVILLE, TN (Aug. 2025).

- Nonparametric estimation of a future state occupation given the knowledge of a past state occupation in a multistate model with current status data, *2025 Lifetime Data Science Conference*, BROOKLYN, NY (May 2025).

- Identifying High-Dimensional Genomic Factor Associations in Biological Networks, *Department of Quantitative Health Sciences at the University of Hawai'i* (Apr. 2025).

- Estimating marginal association in clustered data with informative subgroups induced by a given covariate, *The 7th International Conference on Econometrics and Statistics*, BEIJING, CHINA (July 2024) [virtual presentation].

- Regression analysis for clustered multistate current status data using the pseudo-value approach, *2023 Symposium on Data Science & Statistics*, ST. LOUIS, MO (May 2023).

- Regression analysis of clustered time-to-event data when the cluster size is informative: a pseudo-value approach, *Department of Epidemiology and Biostatistics, West Virginia University* (Mar. 2023).

- Pseudo-value-based regression analysis of clustered multistate time-to-event data when the cluster size is informative, *Biostatistics Branch, National Cancer Institute* (Mar. 2023).

- Regression analysis of clustered time-to-event data when the cluster size is informative, *Division of Computing, Analytics, and Mathematics, University of Missouri*, KANSAS CITY (Feb. 2023).

- Regression analysis of clustered time-to-event data when the cluster size is informative, *UFSTAT Student Seminar Series*, GAINESVILLE, FL (Feb. 2023).

- Bioinformatics Pre-processing of Microbiome Data with an Application to Metagenomics Forensics, *2021 Joint Statistical Meetings*, VIRTUAL CONFERENCE (Aug. 2021).

- Metagenomic Geolocation Prediction Using an Adaptive Ensemble Classifier, *28th Conference on*

Intelligent Systems for Molecular Biology, VIRTUAL CONFERENCE (July 2020).

- Fuzzy Regression Intervals, Graduate Student Seminar, *Department of Mathematics, Boise State University*, BOISE, ID (Dec. 2017).

Contributed Talks
- Identifying high-dimensional genomic factors associated with biological networks, *2025 Joint Statistical Meetings,* NASHVILLE, TN (Aug. 2025).

- Can the unit size predict outcomes? Testing for informativeness in three-level designs, *2025 ENAR Spring meeting,* NEW ORLEANS, LA (Mar. 2025).

- Nonparametric estimation of a future state occupation given the knowledge of a past state occupation in a multistate model with current status data, *2024 Joint Statistical Meetings,* PORTLAND, OR (Aug. 2024).

- Testing for Marginal Covariate Effect when the Subgroup Size Induced by the Covariate is Informative, *2024 ENAR Spring meeting,* BALTIMORE, MD (Mar. 2024).

- Regression analysis of multistate current status data with informative cluster sizes: a pseudo-value approach, *UF PHHP Research Day 2023*, GAINESVILLE, FL (Feb. 2023).

- Adjusting for Informative Cluster Size in Pseudo-Value-Based Regression Approaches with Clustered Time-to-Event Data, *2022 Joint Statistical Meetings*, WASHINGTON DC (Aug. 2022).

- Pseudo-value based regression for clustered time-to-event data when cluster size is informative, *UF PHHP Research Day 2022*, VIRTUAL CONFERENCE (Feb. 2022).

- Bioinformatics Pre-processing of Microbiome Data with an Application to Metagenomics Forensics, *UF PHHP Research Day 2021*, VIRTUAL CONFERENCE (Feb. 2021)

Posters
- Testing for Marginal Covariate Effect When the Subgroup Size Induced by the Covariate is Informative, *Blackwell-Tapia Conference,* PROVIDENCE, RI (Nov. 2024).

- Testing for Marginal Covariate Effect When the Subgroup Size Induced by the Covariate is Informative, *The Southern Regional Council on Statistics 59th Summer Research Conference,* CLEMSON, SC (June 2024).

- Pseudo-Value Regression of Clustered Current Status Data with Informative Cluster or Subcluster Sizes in a Multistate Model, *2023 Lifetime Data Science Conference,* RALEIGH, NC (May 2023).

- Pseudo-Value Regression of Clustered Current Status Data with Informative Cluster or Subcluster Sizes in a Multistate Model, *2023 Annual ASA Florida Chapter Meeting,* GAINESVILLE, FL (Mar. 2023).

- Pseudo-Value Regression of Clustered Current Status Data with Informative Cluster or Subcluster Sizes in a Multistate Model, *2023 ENAR Spring meeting,* NASHVILLE, TN (Mar. 2023).

- Pseudo-value based regression for clustered time-to-event data when cluster size is informative, *2022 International Chinese Statistical Association (ICSA) Applied Statistics Symposium*, GAINESVILLE, FL (June 2022).

- Pseudo-value based regression for clustered time-to-event data when cluster size is informative, *UF College of Medicine Research Day 2022*, GAINESVILLE, FL (Apr. 2022).

- EWMA Control Chart with a Dynamic Sampling Scheme, *2019 Quality and Productivity Research Conference*, WASHINGTION D.C. (June 2019).

- Dynamic Sampling Versions of Popular SPC charts for Big Data Analysis, *2019 IMS/ASA Spring Research Conference*, BLACKSBURG, VA. (May 2019).

- Statistical Process Control Charts for Monitoring Big Data Streams, *Workshop on Emerging Data Science Methods for Complex Biomedical and Cyber Data*, AUGUSTA, GA. (Mar. 2019).

- Using Data Science to help Idaho cities make hiring decisions, *Boise State University Service-learning student exhibition*, BOISE, ID. (Dec. 2018).

- Some contributions to the interpretation of Fuzzy Regression Intervals, *Computing Research Association URMD Workshop*, SAN DIEGO, CA. (Mar. 2018).

GRANTS

3. **U.S. Department of Veterans Affairs IPA**, "Brain Rehabilitation Research Projects". Sept. 2022 - Aug. 2023. Role: Principal Investigator (Mentor: Somnath Datta; VA PIs: D. Clark, R. M. Bauer).

2. **U.S. Department of Veterans Affairs IPA**, "Brain Rehabilitation Research Projects". Sept. 2021 - Aug. 2022. Role: Principal Investigator (Mentor: Somnath Datta; VA PIs: D. Clark, R. M. Bauer).

1. **U.S. Department of Veterans Affairs IPA**, "Brain Rehabilitation Research Projects". Sept. 2020 - Aug. 2021. Role: Principal Investigator (Mentor: Somnath Datta; VA PIs: D. Clark, R. M. Bauer).

SOFTWARE

5. **Anyaso-Samuel S.** and Datta S. crspack R package to conduct inference based on rank-sum statistics for cluster-correlated data with informativeness of the total cluster size, informativeness of a binary covariate distribution or informativeness of a subject-level covariate distribution.

4. **Anyaso-Samuel S.**, Bandyopadhyay D., and Datta S. mspack2. R package for estimating several temporal functions (e.g. state occupation probabilities) for current-status data from of a general multistate model. The code estimates the SOP for the setting where the current-status data is either uncorrelated or cluster-correlated.

3. **Anyaso-Samuel S.** and Datta S. pseudoReg-ICS. R program for estimating the state occupation probability for cluster-correlated data from a multistate model. The program allows for adjusting for informative cluster size.

2. **Anyaso-Samuel S.**, Sachdeva A., Guha S., and Datta S. metagenomic_data_analysis Suite of programs for the bioinformatics pre-processing and downstream analysis of raw sequence metagenomics data.

1. **Anyaso-Samuel S.**, and Mukherjee P. DyAEWMA R package for estimating the average time to signal (ATS) of an adaptive EWMA chart with a dynamic sampling scheme or the average run length (ARL) of the adaptive EWMA chart.

TEACHING EXPERIENCE

Short courses

- **Division of Cancer Epidemiology & Genetics**, *National Cancer Institute*   **Feb. 2025**
  Co-instructor for the course "Multistate models for studying the natural history of cancer" delivered to staff and fellows within the division as part of a lecture series on statistical modeling in cancer epidemiology.

- **Fostering Diversity in Biostatistics**, *ENAR*   **Mar. 2025**
  Co-instructor for the course "Introduction to Statistical Analysis using R" delivered to high

5

school and undergraduate students.

Regular courses

- **Department of Biostatistics**, University of Florida

  Instructor   **Fall 2022**

  - STA 6177 - Applied Survival Analysis.

  Guest lecturer   **Spring 2021, Spring 2022**

  - PHC 7066: Large Sample Theory.
    - Gave lectures on *Modes of convergence* and *Asymptotic normality of the MLE* to PhD students.

  Teaching Assistant   **Fall 2019 - Summer 2023**

  - PHC 6937: Bayesian Biostatistical Methods
  - PHC 6089: Pubic health computing
  - PHC 6937: Fontiers in Biostatistics
  - PHC 6937: Introduction to Applied Biostatistical Computing Using SAS
  - PHC 6052: Introduction to Biostatistical Methods
  - PHC 6937: Data Visualization in the Health Sciences
  - PHC 6092: Introduction to Biostatistical Theory

- **Department of Mathematics**, Boise State University

  Instructor   **Fall 2017, Fall 2018, Spring 2019**

  - MATH 149: Pre-Calculus.
  - MATH 108: Intermediate Algebra.

  Tutor   **Spring 2018**

  - Tutored students enrolled in *Intermediate Algebra*, *College Algebra* and *Pre-calculus* classes.

PROFESSIONAL
DEVELOPMENT

Preparing Future Faculty
Center for Teaching Excellence, University of Florida   **Aug. 2022 - Dec. 2022**

- Competitive and selective semester-long workshop focused on preparing participants for future careers in various academic settings.
- Devoted emphasis on evidence-based teaching, learning practices, expanding mentoring team, and strategies for being a successful faculty member.

SERVICE

Scholarly Journal Refereeing

Referee for   *Biometrics (3), Statistical Methods in Medical Research (3), Statistics in Medicine (3), Journal of Applied Statistics (2), Biometrical Journal (1), Lifetime Data Analysis (1), Sankhya B (1).*

Conference Session Organizer/Chair

- *2025 Joint Statistical Meetings* — Innovations in Biological Network Modeling: Unraveling Omics Data Analysis
- *2025 Lifetime Data Science Conference* — Beyond Right-censoring: Unveiling New Insights with Interval-censored Data Models
- *2025 Eastern North American Region, Spring meeting* — Innovative Strategies for Integrating Epidemiology and Biostatistics in Cervical Cancer Screening
- *2024 Joint Statistical Meetings* — Recent advances in the analysis of multistate survival models

- *2024 Eastern North American Region, Spring meeting* − Advances in Epidemiologic Methods
- *2023 Symposium on Data Science & Statistics* − Methods in Health & Medical Research
- *2023 Eastern North American Region, Spring meeting* − Clustered data methods

Conference & Seminar Service

- Poster Judge for *2025 NIH Graduate Student Research Symposium*
- Planning committee member for *2025 ENAR Fostering Diversity in Biostatistics Workshop*
- Poster Judge for *2025 ENAR Spring Meeting's Student Poster Competition*
- Judge for *2025 LiDS Student Paper Awards*

University Service (University of Florida)

- *Member*, Student recruitment committee; Department of Biostatistics (Nov. 2021 - Apr. 2023).
- *Vice President*, Biostatistics Students' Organization (Sept. 2021 - Sept. 2022).
- *President*, Biostatistics Students' Organization (Sept. 2022 - Apr. 2023).

PROFESSIONAL MEMBERSHIPS

American Statistical Association (ASA)	2018 - Present

- Lifetime Data Analysis Section
- Biometrics Section
- Section on Statistics in Epidemiology

Royal Statistical Society (RSS)	2014 - 2023
Institute of Mathematical Statistics (IMS)	2014 - 2023
International Biometric Society, Eastern North American Region (ENAR)	2022 - 2023
International Chinese Statistical Association (ICSA)	2022 - 2023
International Society for Computational Biology (ISCB)	2020 - 2021
Mathematical Association of America (MAA)	2017 - 2019

COMPUTER SKILLS

General Software

- *Operating systems.* WINDOWS, LINUX and MACOS.
- *Productivity applications.* Advanced skills in WORD, EXCEL, and POWERPOINT.

Computing & Programming

- `Parallel Computing` in selected scripting languages.
- Extensive experience with **R**/RStudio, **C++**, MATLAB, PYTHON/JUPYTER, UNIX, LaTeX.
- Intermediate experience with SAS, STATA, SPSS.
- `Version control:` GitHub user @samuelanyaso

Bioinformatics

- Extensive experience in building pipelines for pre-processing and analysis of large-scale sequencing data.
- Downstream analysis of -omics data.

REFEREES

Available upon request.