MICHAEL NSIAH-NIMO, M.Sc.

RESEARCH SCIENTIST | MACHINE LEARNING AND STATISTICAL MODELING

CONTACT

- Lincoln, Nebraska
- michael20129648@gmail.com
- Linkedin

EDUCATION

MSc Statistics
Focus : *Machine Learning, Biostatistics, Clinical Trials, Causal Inference, Causal Mediation*

University of Texas at El Paso

2015 - 2017

BSc Actuarial Science

Kwame Nkrumah University of Science and Technology

2010 - 2014

SKILLS & TOOLS

Database Management and Standards:
- Proficient in electronic data capture tools: REDCap and QuestionPro
- Ensured data integrity and compliance with clinical trial standards (CDISC, ICH guidelines) and regulatory authority guidance (FDA, EMA, GCP)
- Experience with large-scale data management using SQL, Hadoop, Spark, and R Shiny

Clinical Trials and Observational Research:
- Expertise in clinical trial planning, execution, reporting, and documentation
- Skilled in protocol development, **SAP creation, and CRO deliverables**
- Experience producing **Tables, Listings, and Figures (TLFs)** for clinical trials
- Ensured ethical and regulatory compliance in clinical research

Statistical Methodologies and Analytical Techniques:
- **Experimental Design** and **Study Design** (including Sample Size Estimation), **Multivariate Analysis, Missing Data Imputation, and Hypothesis Testing, Parametric and Non-Parametric Statistics, Causal Inference, Causal Mediation Analysis Multilevel Modeling, Survival Analysis, and Longitudinal Data Analysis, Machine Learning and Deep Learning** (Autoencoders, Gradient Boosting, Random Forest, Decision Trees), **Bayesian Models** (LASSO, Horseshoe Prior), **Penalized Regression** (LASSO, SCAD, MCP), **Data Mining and Visualization**

Programming Languages:
- **Python** (7 years), **R** (10 years), **SAS** (5 years), **Stata, SQL**
- Developed reusable **Python classes and SAS macros** for data cleaning, validation, and analysis

Libraries and Frameworks:
- Python: **statsmodels, NumPy, Pandas, Plotly, Scikit-Learn, Matplotlib, Seaborn, TensorFlow**
- R: **Tidyverse** (ggplot2, dplyr, tibble, readr, tidyr, broom, modelr), R Shiny, glm, glme, lmer
- SAS: **PROC MEANS, PROC FREQ, PROC MIXED, PROC SQL, PROC GLM, PROC FACTOR, PROC GENMOD, PROC GLMSELECT, SAS Macros**
- Tools: **Jupyter Notebooks, Visual Studio Code IDE, Git (Versioning), Google Cloud, BigQuery**

SUMMARY

Results-driven research scientist with **7+ years** of experience in applied research, machine learning, and statistical modeling, and **deep expertise in both theoretical and applied statistics** with a focus on developing diagnostic and predictive models for complex health outcomes. Consulted on **50+ biomedical and clinical projects, securing over $1M in national grants**. Expert in **deep learning, gradient boosting, penalized regression, and Bayesian methods** for analyzing **large healthcare datasets (clinical, molecular, omics, and sociobehavioral).** Proven success in uncovering novel biomarkers for chronic inflammation and metabolic diseases, integrating social determinants into predictive models, and guiding clinical strategies. Strong background in **designing and conducting large-scale experiments, developing statistical and machine learning pipelines, and translating complex data into actionable insights**. Versatile in advanced statistical analyses and study designs, with a proven ability to navigate shifting priorities and deliver results under pressure.

PROFESSIONAL WORK EXPERIENCE

Staff Research Statistician || Biomedical Research

Border Biomedical Research Center
Integrative Analytic Unit   JULY 2020 - PRESENT

Expert Level Skills: Programming & Development (Python, R, SAS), Statistical Tools (General Linear Modeling, Non Parametric Regressions, Machine Learning, Mixed-Effects Modeling, Survival Analysis), Study Design & Optimization (Clinical Research, Observational Study, Sample Size Estimation and Power Analysis). Data Management, Compliance & Reporting: Regulatory Adherence (FDA, NIH, NIHMD Guidelines, IRB, HIPAA), Statistical Analysis Plans (SAPs), Dynamic Dashboards.

- **Developed a biological age metric** from immune cytokine signatures using a range of machine learning and statistical models, including**:**
  - **Gradient boosting, deep learning with autoencoders, random forests, lasso, ridge, robust regression, Bayesian lasso, and Bayesian models with horseshoe priors.**
  - Identified complex **nonlinear patterns in cytokine data to model biological aging** and predict chronic inflammation risk.
  - Employed **feature selection techniques to reduce model complexity** and improve interpretability, identifying novel biomarkers linked to chronic inflammation rather than acute responses.
- **Validated predictive performance** by comparing test error across different models using:
  - **RMSE, trimmed RMSE, and MAD**.
  - Boosting yielded the best prediction accuracy (11.8 – 14.22).
  - Conducted 20 cross-validation splits with a 75% training and 25% test set, averaging performance metrics to assess model reliability.
  - Average predicted values were calculated from 20 repeated training/testing cycles, with performance measured by mean squared error (MSE) and mean absolute deviation (MAD)
- **Selected novel cytokines** through **machine learning feature engineering and selection and biological pathway analysis** and biological significance testing, uncovering new biomarkers for chronic inflammation and metabolic dysfunction.
- **Developed molecular diagnostics models** using **machine learning and statistical approaches** to identify immune, physiological, and neurohumoral markers contributing to cardiometabolic health disparities.
  - **Adjusted for ethnicity and social determinants of health to isolate** biologically meaningful predictors of disease risk.
  - Built predictive models for clinical trials using **penalized regression (LASSO, SCAD, MCP), decision trees, bagging, and gradient boosting to quantify clinical relevance.**
- **Designed and executed a two - arm (A/B test) randomized clinical trial** evaluating dietary intervention effects on systemic chronic inflammation and metabolic health outcomes.
  - Measured changes in inflammatory cytokines, BMI, and fasting glucose at baseline and post-intervention using **mixed-effects models and logistic regression.**
  - **Modeled the interaction of immune and stress responses using causal mediation analysis, identifying norepinephrine as a key mediator of effects** on metabolic health. Results suggest that stress pathways may causally regulate inflammation-related metabolic dysfunction, informing potential therapeutic targets.
- Oversee data analysis and **validation processes, producing publication-ready reports and dynamic dashboards** to facilitate informed decision-making among stakeholders, while **reviewing study protocols,** Case Report Forms (CRFs), and Data Validation Specifications (DVS) to ensure accuracy in statistical representation and comprehensive data collection
- **Collaborated with a multidisciplinary team** of eight (statisticians, clinicians, bioinformaticians, and faculty) on **hybrid applied research combining machine learning prediction first approaches and statistical inference** to study cardiometabolic diseases, inflammatory aging, and cancer, **securing $1 million in funding.** Developed the Statistical Analysis Plan (SAP) to ensure methodological rigor and regulatory compliance, while coordinating with stakeholders to align objectives, timelines, and analysis strategies..

MICHAEL NSIAH-NIMO, M.Sc.

RESEARCH SCIENTIST | MACHINE LEARNING AND STATISTICAL MODELING

Behavioral Skills

- Ability to convey complex machine learning and statistical concepts to non-technical stakeholders, fostering understanding and collaboration.
- Proven track record of seamless collaboration across interdisciplinary teams, contributing to project success.
- Adept in identifying challenges and implementing innovative solutions ensuring the integrity of the analyses.
- Experience in providing guidance and mentorship to junior team members fostering a collaborative work environment.
- Flexibility with changing priorities, strong attention to detail, ability to work well under pressure and take on unfamiliar tasks

PUBLICATIONS

- **[In Review]** *Exploring Socio-Behavioral Correlates of Metabolic and Inflammatory Risk in Hispanics Living Along the U.S./Mexico Border: A Pilot Study Concomitantly Collecting Survey Data, Blood and Hair Samples, and Physical Measures, Xu, Cai, Frietze, Gabriel, Leung, Ming-Ying,* **Nsiah-Nimo, Michael,** *Sanyal, Nilotpal, Begum, Khodeza Robles, Elisa Flores, Diana P , Mandal, Abhijit, Liang, Panfeng*

In Preparation

- *Biological Age Index of Chronic Inflammation as a Predictor of Cardiometabolic Phenotypes: Differential Associations with General, Central Obesity and Diabetes in Mexican Origin Hispanics, Robles, Elisa, Leung, Ming-Ying,* **Nsiah-Nimo, Michael,** *Sanyal, Nilotpal, Xu, Cai, Frietze, Gabriel, Begum, Khodeza , Flores, Diana P , Mandal, Abhijit, Liang, Panfeng*

- *Associations Between Social Support, Perceived Stress, and Metabolic Health Outcomes: A Cross-Sectional Analysis of General Obesity, Central Obesity, and Diabetes in the Hispanic Health Hub. Robles, Elisa, Leung, Ming-Ying,* **Nsiah-Nimo, Michael,** *Sanyal, Nilotpal, Xu, Cai, Frietze, Gabriel, Begum, Khodeza , Flores, Diana P , Mandal, Abhijit Liang, Panfeng*

CERTIFICATIONS

- **Probability Theory, Generalized Linear Models and Inferential Statistics (Duke University)** | May 2018 – September 2019
- **Supervised Machine Learning**: Regression and Classification | October 2022
- **Clinical Data Analysis with SAS** (Oreilly)

- ***Relevant Graduate Project Work***
Efficacy of Treatment Sessions in a Medical Emergency Department - Utilized Interaction Plots, Hasse Diagrams and Linear Mixed Models to analyze and conduct hypothesis tests to reveal patterns of illicit drug use, such as cocaine and cannabis, among individuals, gender and age

PROFESSIONAL CONFERENCES

- Microsoft AI Research Forum, 2024
- University of Hawaii Bioinformatics Data Science Workshops, 2023
- NIH NIGMS Cloud Based Biomedical Research, 2023
- Better Data for More Equitable Research : Research America Alliance, 2023
- RCMI Seminars at Meharry Medical College, 2023
- College of Science Stakeholders' on Health Disparities and Cancer Research, 2016 , 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024
- JSM Conference 2017, Diversity Scholar

Statistical Data Analyst || Health Research and Analytics

College of Science Research Enterprise,
University of Texas at El Paso   JULY 2018 - JULY 2020

Expert Level Skills: *Data Cleaning & Manipulation (SAS, R), Statistical Tools (GLM, Bayesian Analysis), Study Design (Cross-Sectional Research).* **Data Management, Compliance & Reporting:** *Outlier Detection, Clean Dataset Creation, Compliance with NIH/NIMHD Guidelines.*

- Provide timely and scientifically sound statistical expertise to clinical development projects, actively participating in department and **cross-functional meetings to optimize study designs, endpoints, and analysis strategies that enhance research outcomes**
- Provide analytical support and **generate detailed reports, tables, figures and listings for key projects**, particularly those with UTEP Border Biomedical Research Center, All of Us, and NIH U54 achieving significant research milestones
- **Created reusable Python scripts and SAS Macros**, to streamline data preprocessing, and data validation reducing processing time by 50% and enhanced data integrity.
- Developed and reviewed **SAP, protocols, sample size calculations, randomizations, and blinding techniques, ensuring rigorous and reliable research**
- Analyzed **clinical and survey data** using **longitudinal analysis, mixed effects modeling, logistic regression, and survival analysis, improving model accuracy and predictive power**
- Created and maintained primary databases using tools such as REDCAP and QUESTIONPRO
- Developed **social determinants of health trends for prevalent issues like obesity, diabetes, cancer, and metabolic syndrome** in the Borderplex region, **informing targeted public health interventions**

Lecturer/Research Statistician

Alcohol Research Group,
University of Texas at El Paso   AUGUST 2015 - MAY 2018

Expert Level Skills: *Statistical Methods (DOE, Regression Techniques), Data Management (Summary Tables, Report Writing), Teaching & Mentorship.*
Data Management, Compliance & Reporting: *Compliance with Academic Standards, Publication-Ready Manuscripts.*

- **Analyzed treatment efficacy in emergency department settings** by **employing linear mixed models and interaction plots, uncovering significant patterns of illicit drug use, such as cocaine and cannabis, across** different age groups and genders, to inform tailored intervention strategies.
- **Applied Hasse Diagrams and advanced statistical techniques** to conduct hypothesis testing, providing actionable insights on substance use behaviors, supporting clinical teams in optimizing treatment sessions for Hispanic health initiatives.
- Collaborated on statistical research projects, enhancing methodologies for data interpretation and **contributing to manuscripts reviewed for publication in peer-reviewed journals**.
- **Applied Design of Experiments (DOE) techniques to clinical data analysis,** increasing experimental accuracy and reliability by 20%, resulting in more consistent and valid research outcomes.
- Instructed over 150 students in advanced statistical methods, achieving a 90% pass rate **while mentoring students on real-world statistical applications.**

Professional Contributions

- National Institutes of Health (NIH) U54 projects for the College's Research Core, focusing on health disparities and clinical research outcomes.
- ***Grant Awards: NIH Project No: 5U54MD007592 -27, SUB IDS: 8320, 8321, 8317, NIH Project No: 5U54MD007592 -29 & 30, SUB IDS: 8320, 8321, 8317, NIH Project No: 1C06OD032074-01 & 3U54CA272167-02S1***

Federal and Foundational Organizations Grant Projects

Contributed extensively to over 50 federal and foundational grant projects. Listed are a few of some of the major grants projects:

- National Institute of Health (NIH) U54 projects for the College's Research Core, Clinical Research, and Socio-behavioral Research.
- NIH All of Us Data for Health Disparities Research and Covid-19 U54 Supplements.
- NIH C06 Grants for the Borderplex Biomedical Research Core.
- NIMHD Grants for Health Disparities, RCMI Grants for Hispanic Health Disparities.
- Robert Wood Johnson Foundation Hispanic Health Disparities Grant.
- Bill Gates Foundation Health Equity Grant.
- Arnold Ventures Health Services Inequity and Disparities Grant.
- Paso Del Norte Health Research Grants.

- ***Work Site: Willing to work remotely***