


# MICHAEL NSIAH-NIMO, M.Sc.

RESEARCH AND DATA ANALYTICS PROFESSIONAL  
BIOSTATISTICS | ADVANCED ANALYTICS

 [My Website](#)

## CONTACT

 Lincoln, Nebraska

 michael20129648@gmail.com

 [Linkedin](#)

## EDUCATION

MSc Statistics

Focus: Machine Learning, Causal Inference, Causal Mediation Analysis Predictive Analytics, Survival Analysis, Bayesian Statistics, Study Design

University of Texas at El Paso

2015 - 2017

BSc Actuarial Science

Kwame Nkrumah University  
of Science and Technology

2010 - 2014

## CORE COMPETENCIES

- **Data Analytics & Predictive Modeling** (SQL, SAS, R, Python)
- **Healthcare Claims Analysis** (Medical, Pharmacy, Enrollment Data)
- **ETL Development, Data Integrity, and Quality Control**
- **Machine Learning for Healthcare Outcomes** (Random Forests, Boosting, Deep Learning)
- **Statistical Modeling** (GLM, Bayesian Methods, Logistic Regression, Survival Analysis)
- **Client Engagement & Cross-functional Communication**
- **Health Information Exchange (PHIX) Database Experience** with Diagnoses and treatment plans Allergies , Medications , Immunizations , Lab results Radiology reports , Social Determinants of Health Data. Other clinical notes and reports
- **CMS VRDC Experience** and Familiarity with medicare, medicaid, dual eligibles data, NIH AllofUS Research Workbench with Regulatory Compliance (FDA, NIH, IRB)
- **Clinical Trials and Observational Research Analytics**
- **Dynamic Dashboard Development** (Tableau, Power BI, R Shiny, Python Dash)

## TECHNICAL SKILLS

- **Programming Languages:** R, Python, SAS, SQL
- **Analytics & Modeling Tools:** PROC MIXED, PROC LOGISTIC, Scikit-learn, TensorFlow, R Tidyverse, Seaborn, ggplot2
- **Data Management:** REDCap, SQL, Hadoop, Spark, AWS
- **Statistical Techniques:** GLM, Mixed Models, Bayesian Analysis, Causal Inference, Deep Learning (Autoencoders, Boosting)
- **Reporting Tools:** Dynamic Dashboards (R Shiny, Dash), ODS Graphics (SAS), Tableau
- **Other:** Git Version Control, Google Cloud Platform (GCP) BigQuery

## SUMMARY

Dynamic, results-driven research statistician with over 7 years of experience designing and implementing statistical models, ETL processes, and machine learning approaches for large healthcare datasets. Expertise spans healthcare claims, enrollment, pharmacy data, and regulatory-compliant clinical research, securing national grants and contributing to over 50 federally funded projects (NIH, NIMHD, Robert Wood Johnson Foundation). Adept at navigating complex datasets (SQL, SAS, R, Python), producing analytical deliverables, creating Statistical Analysis Plans (SAPs), and developing dynamic dashboards to support healthcare decision-making. Proven track record of leadership, cross-functional collaboration, and delivering actionable insights aligned with client timelines and expectations

## WORK EXPERIENCE

### Staff Research Statistician

Border Biomedical Research Center, Integrative Analytic Unit  
University of Texas at El Paso

JULY 2020 - PRESENT

- Work closely with a multidisciplinary team of statisticians, principal investigators, clinicians, bioinformaticians, and faculty on clinical research projects focused on cardiometabolic diseases, inflammatory aging, and cancer, contributing to securing over \$1 million in federal funding.
- Designed and implemented advanced statistical models leveraging large-scale healthcare databases (medical claims, EHR, omics, sociobehavioral factors) to predict chronic diseases and improve precision in disease management strategies.
- Developed a biological aging metric derived from nonlinear patterns in protein signatures using nonparametric regressions and machine learning techniques (e.g., random forests, gradient boosting) to serve as an indicator of chronic inflammation and cardiometabolic diseases.
- Performed data mining and exploratory analyses across biological domains (physiological, immunological, metabolomic) and non-biological domains (social determinants of health from surveys, health records, healthcare claims and costs) to identify the most informative features and categorical factor levels to enhance statistical modeling, pattern recognition in healthcare utilization and healthcare costs and to improve healthcare quality , disease prediction, and estimation
- Led independent research using the NIH All of Us longitudinal repository to investigate healthspan markers and social determinants of health, conducting comparative analyses between Hispanic and non-Hispanic White populations to generate hypotheses for primary data collection.
- Implemented a clinical trial examining the impact of a dietary intervention on systemic chronic inflammation and cardiometabolic outcomes, evaluating changes in inflammatory cytokines, fasting glucose, BMI, and waist circumference at baseline and four weeks post-intervention.
- Led Extract, Transform, Load (ETL) processes, ensuring intake, transformation, validation, and quality control of complex clinical datasets, enabling efficient database integration and reliable analyses.
- Drafted Statistical Analysis Plans (SAPs), developed study protocols, and conducted sample size and power analyses to ensure methodologically rigorous study designs aligned with NIH, FDA, and clinical trial standards.
- Conducted randomization, blinding procedures, and statistical analyses for observational and experimental studies submitted for FDA and NIH/NIMHD grants, ensuring regulatory compliance.
- Built dynamic reporting and visualization systems using R Shiny and SAS Macros, improving the turnaround time for project deliverables by 50%, while enhancing data transparency and stakeholder engagement.
- Performed comprehensive statistical analyses using SAS PROCs (PROC MEANS, PROC FREQ, PROC MIXED, PROC GLM, PROC REG, PROC LOGISTIC, ODS GRAPHICS) and R functions (tidyverse, dplyr, randomForest, bagging, rstatix, PCA), combining multiple datasets to produce accurate, validated results.
- Applied advanced statistical methodologies, including mixed-effect modeling, logistic regression, causal mediation, Bayesian LASSO, and penalized regressions, to develop diagnostic and therapeutic strategies addressing Hispanic cancer disparities.
- Conducted rigorous quality control reviews of statistical analyses, ensuring high standards of accuracy and reproducibility; validated analysis datasets, generated statistical output products (tables, listings, figures), and designed dashboards supporting critical decision-making.
- Collaborated with data management and clinical operations teams to develop and validate study-specific data capture systems, improving data quality and ensuring adherence to regulatory standards (FDA, NIH, HIPAA).
- Produced publication-ready analytical reports and dynamic dashboards to support informed stakeholder decision-making, while reviewing study protocols, Case Report Forms (CRFs), and Data Validation Specifications (DVS) for completeness and accuracy.

# MICHAEL NSIAH-NIMO, M.Sc.

RESEARCH AND DATA ANALYTICS PROFESSIONAL  
BIOSTATISTICS | ADVANCED ANALYTICS

## Behavioral Skills

- Adept at using simple language without watering down the key ideas of the research and focus on the purpose and results of the study to capture audience comprehension and relatability
- 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

- **[Publication]** *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*, by Gabriel A. Fietze, Cai Xu, Bibiana Mancera, Elisa Robles, Escajeda, Alyssa A. Martinez, Michelle Gil, Diana P. Flores, Khodeza Begum, Panfeng Liang, Abhijit Mandal, **Michael Nsiah-Nimo**, Nilotpal Sanyal, Ming-Ying Leung, Michael J. Kenney and Robert A. Kirken

## In Review

- *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, Fietze, 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, Fietze, Gabriel, Begum, Khodeza, Flores, Diana P, Mandal, Abhijit Liang, Panfeng

## CERTIFICATIONS

- **Decision Intelligence : LinkedIn**
- **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)**

## 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
- JSM Conference 2017, Diversity Scholar
- Cardwell Foundation Seminar, 2016

## Statistical Data Analyst | Health Research and Analytics

College of Science Research Enterprise,  
University of Texas at El Paso

JULY 2018 - JULY 2020

- Conducted **healthcare analytics projects focusing on social determinants of health, metabolic syndromes, and clinical intervention outcomes.**
- Led **analyses involving health claims and survey data using logistic regression, multilevel modeling, and machine learning methods.**
- Developed **reusable SAS Macros and Python scripts for efficient data management and statistical reporting**, improving project efficiency and deliverable quality.

## Lecturer/Research Statistician

Alcohol Research Group, University of Texas at El Paso

AUGUST 2015 - JULY 2018

- **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 with Dr. Amy Wagler on multiple statistical research projects, **leading to advancements in statistical methodologies that improved data interpretation and contributed to manuscripts reviewed for publications.**
- **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 mathematics, achieving a 90% pass rate while **mentoring students on end-of- semester projects**, enhancing their understanding and application of statistical techniques in real-world scenarios

## BUSINESS AND HEALTH PROJECTS

### Compressing Feature Space For Classification Using PCA

- Applied Principal Component Analysis (PCA) **to reduce the dimensionality of 50 cytokines to capture variation in estimated biological ages among Hispanic Origin Individuals leading to an increase in model performance**

### Enhancing Targeting Accuracy Using ML

- Built a model that would accurately predict the customers that would sign up for a delivery club. **This allowed for a much more targeted approach when running the next iteration of the campaign**

### Predicting Customer Loyalty Using ML

- Built a predictive model using Random Forest in Python to estimate customer loyalty scores for a company's data agency **leading to a 30% increase in the ability to contact customers with promotional material**

### "You Are What You Eat" Customer Segmentation

- Used k-means clustering on grocery transaction data to split out customers into distinct "shopper types" **to accurately target customers with relevant content & promotions**

### Fruit Classification Using A Convolutional Neural Network

- Built & optimized a Convolutional Neural Network to classify images of fruits, with **the goal of helping a grocery retailer enhance & scale their sorting & delivery processes.**

### Quantifying Sales Uplift With Causal Impact Analysis

- Analyzed customer retail practices to quantify the increase in sales attributed to customers joining the company's delivery club compared to what they would have spent without the club. **This analysis revealed a significant 41.1% uplift in sales among club members, indicating the club's positive impact on customer spending**

## Professional Contributions to

## Federal and Non - Federal Grants Projects

- **NIH Project No : 5U54MD007592 -27, SUB IDS : 8320, 8321, 8317**
- **NIH Project No : 5U54MD007592 -29, SUB IDS : 8320, 8321, 8317**
- **NIH Project No: 1C06OD032074-01 & 3U54CA272167-02S1**

- National Institute of Health (NIH) U54's for the College's Research Core :
- NIH All of Us Data for Preliminary Data for Health Disparities Research
- NIH U54 Clinical Research
- NIH U54 Socio behavioral Research
- NIH Covid 19 U54 Supplements NIH C06 Grants for the Borderplex Biomedical Research Core NIMHD Grants for Health Disparities, RCMI Grants for Hispanic Health Disparities Paso Del Norte Health Research Grants