**Yiyuan Wu**

dretheo1114@gmail.com | (732) 570-3246|website: https://yiw1990.github.io/yiw1990/index.html

**EDUCATION**

**Rutgers University**, School of Arts and Science, New York, NY (GPA 3.8) 2015-2016

• **Master of Science**, Department of Statistics, Applied Biostatistics and Data Science

**University of Kentucky**, College of Arts and Sciences, Lexington, KY (GPA 3.9) 2009-2012

• **Bachelor of Science**, Major: Mathematics

**SKILLS**

**Data**: EDA, wrangling, manipulation, reporting, analysis, management, and visualization

**R** (RStudio, R Markdown): tidyverse, ggplot2, shiny, janitor, dlpyr, knitr, lubridate, sqldf, MASS, survival, here, glmnet, glm, htmlwidget

**SAS** (Base, Enterprise, Studio): Linear and logistic regression, ANOVA, survival analysis, GEE and MLM models, propensity score matching, macro language

**Relational Databases and Languages**: Microsoft Access, SQL (MySQL, PROC SQL)

**Research interest:** Deep learning, causal mediation analysis, genomics, mobile health data

**EXPERIENCES**

**Weill Cornell Medicine**, New York City, NY Feb 2019- Present

**Research Biostatistician II | Research Biostatistician I**, Oct2 2017- Present

* Provided consultation service for investigators with their long-term or short-term projects
* Helped with study design, data wrangling, preparation and grant writing support
* Visualize data and provide descriptive reports
* Utilized statistical analysis skills with various projects such as generalized linear mixed model, GEE, joint modeling, Machine learning algorithm such as random forest, gradient boosting machine
* Applied cross validation techniques to validate the performance of a predictive model
* Often used lasso and elastic net for feature selection
* Reviewed, modified statistical section written into manuscript
* Organized workflow by writing reproducible R functions and version control through git and GitHub
* Worked on a long- term project that involves analysis using Medicare claim data (big dataset)

**Montefiore Medical Center**, Bronx, NY, Aug 2016-Oct 2017

**Data Analyst**

* Developed health score alert system in the emergency department using random forest algorithm and EHR data with patients’ diagnosis and other clinical information
* Used logistic regression model to look at 30-day readmission risk for patients with COPD got discharged from hospital stay in Montefiore
* Wrote SAS macro to automatize reports that provides table 1 and other descriptive measures
* Performed statistical analysis such as survival analysis, propensity score matching, logistic regression and modeling (GLM, Cox model, GEE and Mixed Model) for various studies using SAS and R
* Worked with informatics team to implement a hospital score based alert system using random tree forest algorithm
* Created, updated and managed clinical databases as well as extracted data from electronic medical database based on study criteria to build study cohort

**Farm Bureau Financial Services**, Des Moines, IA,May 2016-July2016

**Predictive Modeling Intern**

* Wrote in-depth SQL queries to extract Iowa claim data from the past 8 years
* Visualization in geospatial softare for multiple perils in Iowa State for rate filing of 2017
* Used R to build a GLMM to predict next year’s premium rate using claims data collected in the previous 8 years
* Used 10-fold cross validation to validate the performance of model

**PUBLICATIONS**

* Danielle S. Jackson, MD MPH (1), Samprit Banerjee, PhD (1, 3),Jo Anne Sirey, PhD (1),Cristina Pollari, MPH (1),Richard Novitch, MD (2),Alexandra Chalfin, BS (1), **Y. Wu, MS** (3) ,George S. Alexopoulos, MD (1). Two Interventions for Patients with Major Depression and Severe COPD: Impact on Quality of Life.
* Mikhail A Dziadzko1,2, Paul J Novotny3, Jeff Sloan3, Ognjen Gajic4, Vitaly Herasevich 1, Parsa Mirhaji 5, **Y. Wu, MS5** and Michelle Ng Gong6. Multicenter derivation and validation of an early warning score for acute respiratory failure or death in the hospital.
* Khushabu Kasabwala a, Neal Patel a , Eliza Cricco-Lizza a , Adrian A. Shimpi b , Stanley Weng a , Rose M. Buchmann b , Samaneh Motanagh c , **Y. Wud**, Samprit Banerjee d , Francesca Khani a,c , Daniel J. Margolis e , Brian D. Robinson a,c , Jim C. Hu. The Learning Curve for Magnetic Resonance Imaging/Ultrasound Fusion-guided Prostate Biopsy.
* Miriam Samstein, MD, PhD, **Y. Wu, MS**, Linda M. Gerber, MS, Erika Abramson, MD, MS. Pediatric Chief Resident Perspective on High Value Cost Conscious Care Training During Residency: A National Survey.
* Yongkang Zhang, Zachary Grinspan, Eloise O'Donnell, **Y. Wu**, Dhruv Khullar, Rainu Kaushal. Using Computable Phenotypes to Identify High-Need, High-Cost Patients.
* Lisa R. Koenig, BA,1 Russell Rosenblatt, MD,2 Rahil M. Patel, BS,3 **Y. Wu, MS**,4 Anton Orlin, MD,1 Robison V. Paul Chan, MD,1 Szilard Kiss, MD,1 Donald J. D’Amico, MD,1 Sonal Kumar, MD,2 Mrinali P. Gupta, MD1. Hepatitis C does not modulate prevalence or severity of diabetic retinopathy.
* A.A. Hope1, **Y. Wu1**, J.-T. Chen1, D. Kaufman2, D. Talmor3, D. J. Kor4, O. Gajic4, M. N. Gong1.The Association Between Pre-Hospital Vulnerability, The Development of ARDS And Mortality Among At-Risk Adults: Results From The Lips-A Randomized Clinical Trial.
* Ellen M. Whyte MD​, Patricia Marino, PhD​, Samprit Banerjee, Ph.D.​ **Y. Wu,** MS. Sustaining Remission of Psychotic Depression: ​The STOP-PD II Study -- Focus on older adults American Society of Clinical Psychopharmacology of Symposium presentation.