

addition to publicly available summary statistics from two prior studies, we generated GWAS data from three electronic health record biobanks (BioVU, eMERGE, and PMBB). In total, we utilized data from 359,378 individuals (4,411 cases and 354,967 controls). Leveraging this large-scale biobank data importantly allows for detection of complex factors contributing to the diverse etiology of PE. Cases across cohorts were defined using PE-specific ICD-9/ICD-10 codes and phecodes. Cohorts included pregnant individuals of self-identified non-Hispanic Black, non-Hispanic White, and East Asian ancestry. RESULTS/ANTICIPATED RESULTS: 2 of 20,204,625 loci achieved genome-wide significance ($p < 5 \times 10^{-8}$) when minor allele frequency was limited to common variants (>0.01). The most significant locus was rs138180605 ($p = 1.77 \times 10^{-8}$), located in an intergenic region between FGFR2 and ATE1, both previously associated with breast cancer. The other significant locus was rs137895377 ($p = 2.33 \times 10^{-8}$), located in an intronic region of PLEKHO1. Another 225 loci achieved suggestive significance ($p < 1 \times 10^{-5}$). 203 loci could be mapped to 109 unique genes, some previously associated with related phenotypes such as hypertension. Next steps will focus on functional analyses, including genetically predicted gene expression incorporating placental tissue, followed by construction of a PE polygenic risk score to demonstrate predictive utility of results. DISCUSSION/SIGNIFICANCE: This work has contributed to the limited body of knowledge surrounding maternal genetic susceptibility to PE by identifying several loci warranting further investigation. Further work will expand on these results to improve understanding of genetic factors and clarify clinical risk of disease.

17

Data Loofah: A web-based app for efficiently identifying erroneous data

Jeffrey R. Fine, Sandra L. Taylor
University of California, Davis

OBJECTIVES/GOALS: The goal was to create and deploy an intuitive, easy-to-use tool that clinical investigators can apply to their data to identify erroneous or inconsistent data entries. Investigators can then correct any errors prior to sharing the data with their statistician for analysis. METHODS/STUDY POPULATION: We developed an interactive shiny app, the Data Loofah, using R Studio that researchers or data analysts can use to examine data. After an investigator uploads data, the app reports which variables are numeric or categorical. Means, standard deviation, median, 25th and 75th quantiles, range and number of missing values are reported for numeric variables. Counts and percentages of categorical variables are summarized. Graphical displays further enhance identification of errors. Access to the Data Loofah is through a secure, university-maintained website with access restricted to university personnel. Supporting materials consisting of instructional step-by-step handouts and videos were developed to assist investigators in the use of the app. RESULTS/ANTICIPATED RESULTS: We will integrate use of the Data Loofah into our Clinical and Translational Science Program's biostatistics consultative practice. Investigators will use the Data Loofah to pre-screen their data prior to sending it to a statistician, identify errors and inconsistencies, and facilitate making necessary corrections. Statisticians will also use the Data

Loofah to review data with investigators prior to starting analyses. Through use of this app, investigators are expected to develop a better understanding of their data specifically and more generally about requirements for preparing data for statistical analysis. Most significantly, regular use of the Data Loofah is expected to result in higher quality data and more efficient use of statistician resources due to reduced effort for data cleaning. DISCUSSION/SIGNIFICANCE: Data cleaning is a time-consuming task and finding data errors can be difficult for data analysts not familiar with clinical variables under study. Further, failure to identify data errors can lead to erroneous results. By facilitating identification of data errors by clinical investigators, the Data Loofah will improve and enhance research output.

19

Development of the Puerto Rico Neoplasm and CNS Tumor Registry (PUNCTURE)

Julio Rosado Philippi¹, Natasha L. Frontera¹, Aixa de Jesus Espinosa¹, Valerie Wojna², David Blas Boria³, Juan L. Perez-Berenguer⁴, Kimberly Hoang⁵, Ali Alawieh⁶, Miguel Mayol Del Valle⁷
¹School of Medicine, University of Puerto Rico-Medical Sciences Campus, ²Neurosurgery Section, University of Puerto Rico-Medical Sciences Campus, ³Department of Neurology, University of Puerto Rico-Medical Sciences Campus ⁴Department of Pathology, University of Puerto Rico-Medical Sciences Campus, ⁵School of Medicine, Emory University, ⁶Department of Neurosurgery, Emory University, ⁷School of Medicine, Department of Neurosurgery MD, University of Puerto Rico-Medical Sciences Campus, Neurosurgery Section

OBJECTIVES/GOALS: To describe and compare clinical data and outcomes for patients with CNS tumors and tumor mimics in Puerto Rico who are undergoing surgical and nonsurgical management. Thus, increasing data from an underrepresented group which can serve as a foundation for investigating determinants of outcomes. METHODS/STUDY POPULATION: This proposal will examine patient charts, radiology and pathology reports, financial data, and treatment details from the electronic medical record of patients receiving surgical and nonsurgical treatment for CNS tumors and mimics in the University of Puerto Rico Medical Sciences Campus and all associated institutions. Data will be analyzed retrospectively between January 1, 2014 and June 30, 2022, and prospectively for ten years until December 31, 2032. Patients with primary and metastatic CNS tumors and tumor mimics in the brain, meninges, ventricles, spinal cord, cranial nerves, orbit, facial sinuses, bony skull, vasculature will be included. The registry will include patients from birth onward. RESULTS/ANTICIPATED RESULTS: We plan to compare different surgical and non-surgical techniques and devices in terms of technical and clinical outcomes after surgical interventions for CNS tumors. We are collaborating with the CNS Tumor Outcome Registry at Emory (CTORE) and plan to continue collaboration with other institutions. Combining our data, we aim to develop predictive models of patient outcomes after surgical and nonsurgical intervention for CNS pathologies using supervised and unsupervised machine learning strategies. DISCUSSION/SIGNIFICANCE: There is a significant lack of literature on CNS intervention outcomes in Puerto Rico. This registry will provide the platform for cost-analysis studies for techniques and clinical protocols applicable to pre-operative, intra-operative, post-operative, and conservative management of patients, in Puerto Rico and beyond.