

DISCUSSION/SIGNIFICANCE: Given the implementation of DPP at-scale there is an urgent need to understand the patient and systems-level factors that are associated with referring individuals in the DPP. By detecting characteristics of health systems and patients that warrant special attention, we can improve equitable access to evidence-based diabetes prevention.

56

Using machine learning to predict 30-day readmission and reoperation following resection of supratentorial high-grade gliomas: A national analysis of 9,418 patients.

Abdul Karim Ghaith¹, Marc Ghanem³, Cameron Zamanian^{1,2}, Antonio A. Bon Nieves^{1,2}, Archis Bhandarkar^{1,2}, Karim Nathani^{1,2}, Mohamad Bydon^{1,2}, Alfredo Quiones-Hinojosa⁴

¹Mayo Clinic Neuro-Informatics Laboratory, Mayo Clinic, Rochester, MN, USA ²Department of Neurological Surgery, Mayo Clinic, Rochester, MN, USA ³Gilbert and Rose-Marie Chagoury School of Medicine, Lebanese American University ⁴Department of Neurological Surgery, Mayo Clinic, Jacksonville, FL, USA

OBJECTIVES/GOALS: High-grade gliomas (HGG) are among the rarest, most aggressive tumors in neurosurgical practice. We aimed to identify the clinical predictors for 30-day readmission and reoperation following HGGs surgery using the NSQIP database and seek to create web-based applications predicting each outcome. **METHODS/STUDY POPULATION:** We conducted a retrospective, multicenter cohort analysis of patients who underwent resection of supratentorial HGG between January 1, 2016, and December 31, 2020, using the NSQIP database. Demographics and comorbidities were extracted. The primary outcomes were 30-day unplanned readmission and reoperation. A stratified 80:20 split of the available data was carried out. Supervised machine learning algorithms were trained to predict 30-day outcomes. **RESULTS/ANTICIPATED RESULTS:** A total of 9,418 patients were included in our cohort. The rate of unplanned readmission within 30 days of surgery was 14.9%. Weight, chronic steroid use, pre-operative BUN, and WBC count were associated with a higher risk of readmission. The rate of early unplanned reoperation was 5.47%. Increased weight, higher operative time, and a longer period between hospital admission and the operation were linked to increased risk of early reoperation. Our Random Forest algorithm showed the highest predictive performance for early readmission (AUC = 0.967), while the XG Boost algorithm showed the highest predictive performance for early reoperation (AUC = 0.985). Web-based tools for both outcomes were deployed: <https://glioma-readmission.herokuapp.com/>, <https://glioma-reoperation.herokuapp.com/>. **DISCUSSION/SIGNIFICANCE:** A high fraction of documented early unplanned readmission and reoperation were considered preventable and related to surgery. Machine learning allows better prediction of resected HGG's prognosis based on findings from baseline methods leading to more personalized patient care.

57

Utility of Digital Phenotyping in Big Data to Answer Clinical Questions: Puberty as a Transdisciplinary Science Case Example

David J. Schnabel, Jr., Lorah D. Dorn
Pennsylvania State University

OBJECTIVES/GOALS: A disease-agnostic translational science framework for data mining is proposed for use across disciplines

to: Answer clinical questions, justify future clinical research recruitment, and explore under-represented populations. As a case example, male puberty demonstrates utility of the framework. **METHODS/STUDY POPULATION:** As a case example using the generalizable framework, the following interdisciplinary question was asked: Does early pubertal timing increase the risk of developing type II diabetes (T2d) in boys? A digital phenotype of males < 18 years old was created in the TriNetX Diamond Network utilizing Boolean operator data queries. TriNetX contains patient electronic health record information (ICD-10 diagnoses, anthropometric measures). A case control analysis leveraging patient counts from various digital phenotypes allowed for outcome (T2d) comparison of boys diagnosed with precocious puberty (E30.1, ICD code for early pubertal timing) to those without, while controlling for body mass index (BMI). **RESULTS/ANTICIPATED RESULTS:** Subjects (N=12,996,132) displayed the following digital phenotype: Male, < 18 years old, without ever having a BMI documented >85th percentile. Boys diagnosed with precocious puberty (E30.1) were 6.89 times more likely to develop T2d when aged 14-18 years old than those without (OR 6.89, 95%CI: 5.17-9.19, p. **DISCUSSION/SIGNIFICANCE:** Boys are under-represented in the early pubertal timing literature, justifying future human subjects research on male puberty. This case example demonstrates a broader disease-agnostic framework which can be adapted across disciplines. Opportunities may include public health digital phenotyping.

58

Utilizing VA Data to Define Long COVID and Identify Patients at Risk

Peter L. Elkin^{1,2,4,5}, Skyler Resendez^{1,2}, H. Sebastian Ruiz¹, Wilmon McCray^{1,2}, Steven H. Brown³, Jonathan Nebeker³, Diane Montella³

¹Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo ²Department of Veterans Affairs, VA Western New York Healthcare System and VA Research Service ³Office of Health Informatics, Department of Veterans Affairs ⁴Department of Internal Medicine, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo ⁵Faculty of Engineering, University of Southern Denmark

OBJECTIVES/GOALS: To determine the signs, symptoms, and diagnoses that are significantly upregulated in cases of long COVID while identifying risk factors and demographics that increase one's likelihood of developing long COVID. **METHODS/STUDY POPULATION:** This is a retrospective, big data science study. Data from Veterans Affairs (VA) medical centers across the United States between the start of 2020 and the end of 2022 were utilized. Our cohort consists of 316,782 individuals with positive COVID-19 tests recorded in the VA EHR with a history of ICD10-CM diagnosis codes in the record for case-control comparison. We looked at all new diagnoses that were not present in the six months before COVID diagnosis but were present in the time period from one month after COVID through seven months after. We determined which were significantly enriched and calculated odds ratios for each, organized by long COVID subtypes by medical specialty / affected organ system. Demographic analyses were also performed for long COVID patients and patients without any new long COVID ICD10-CM codes. **RESULTS/ANTICIPATED RESULTS:** This profile shows disorders that are highly upregulated in the post-COVID population and provides strong evidence for a broad definition of long COVID. By breaking this into subtypes by medical specialty, we define cardiac long COVID, neurological long COVID, pulmonary long COVID, and eight others. The long COVID cohort

was older with more comorbidities than their non-long COVID counterparts. We also noted any differences regarding sex, race, ethnicity, severity of acute COVID-19 symptoms, vaccination status, as well as some analysis regarding medications taken. **DISCUSSION/SIGNIFICANCE:** This profile can be utilized to decisively define long COVID as a clinical diagnosis and will lead to consistence in future research. Elucidating an actionable model for long COVID will help clinicians identify those in their care that may be experiencing long COVID, allowing them to be admitted into more intensive monitoring and treatment programs.

59

Where are the viral loads? Searching for additional HIV laboratory results in South Africa's National Health Laboratory Services Database

Keri Geiger, Jason E. Farley
Johns Hopkins University

OBJECTIVES/GOALS: Guidelines suggest people with HIV (PWH) receive routine HIV viral load (VL) testing at least yearly and upon diagnosis with multidrug-resistant tuberculosis (MDR-TB). Many PWH and MDR-TB in South Africa seem to be missing VL results. This study's goal was to find results which may be available from the National Health Laboratory Service (NHLS). **METHODS/STUDY POPULATION:** We abstracted HIV laboratory results, specifically baseline VL and CD4 count and VL at MDR-TB cure, from PWH enrolled in a cluster-randomized clinical trial of nurse case management for PWH and MDR-TB in South Africa who were cured of MDR-TB. For any participant missing one or more of these results, we thoroughly searched the electronic NHLS database using multiple separate searches varying terms including patient name, surname, date of birth, medical record number where available, and South African identification number. Returned results were compared to results abstracted from the parent study and any additional results were entered into the parent study data. **RESULTS/ANTICIPATED RESULTS:** Of 929 PWH cured of MDR-TB, 879 (94.6%) were missing at least one expected VL or CD4 result in the parent study database. Though our search strategy was successful in identifying participants and returning CD4 and VL results, we rarely found additional results that were not already in the parent study database. Following the search and entry of the few additional results retrieved, 116 (12.4%) participants were missing a baseline CD4, 309 (33.3%) missing baseline VL, and 385 (41.4%) missing VL at MDR-TB cure, representing 572 individuals or 61.6% of participants with at least one unavailable result. This high level of unavailability of key laboratory results used to guide MDR-TB and HIV treatment suggests that these tests were either not ordered, not collected, or not completed due to electronic gatekeeping at NHLS. **DISCUSSION/SIGNIFICANCE:** Unknown CD4 count or VL leaves PWH open to including MDR-TB treatment failure and death. Our search strategy found additional results but was time-consuming and cumbersome. Limitations included lack of information on why laboratory results were missing, which limits our ability to make recommendations for better collection and reporting.

60

Derivation and Validation of a Novel Hospital Capability Score for Sepsis

Uchenna Ofoma, Elena Deych, Fei Wan, Marin Kollef, Karen Joynt Maddox
Washington University in Saint Louis

OBJECTIVES/GOALS: Regionalized sepsis care could improve sepsis outcomes. There are no measures of sepsis capability to guide the identification of hospitals that can best serve sepsis patients. We derived Capability-Based (CB) scores from specific hospital characteristics and evaluated their performance as system predictors of mortality among adults with sepsis. **METHODS/STUDY POPULATION:** We used the 2018 State Inpatient Databases to identify 90051 adult sepsis encounters at 157 non-federal NY hospitals (derivation cohort), and 130,249 sepsis encounters at 220 hospitals in FL and MA (validation cohort). We used Principal Component Analysis to analyze to reduce 14 hospital-level resource use characteristics to 3 interpretable, linear data combinations (principal components (PC)). We calculated CB scores for each hospital as a sum of standardized values for each component multiplied by the respective PC loading. We evaluated the correlation of sepsis volume and each CB score with hospital mortality and with outward sepsis transfer proportions. We fitted linear, nested, predictive models to compare the system predictive abilities of CB scores and sepsis volume in relation to hospital mortality. **RESULTS/ANTICIPATED RESULTS:** In the derivation cohort, 83963 (93.2%) patients were non-transferred, of which 20230 (24.1%) died. The mean (range) score was 0 (-3 - +5) with higher scores denoting more capable hospitals. Higher scores were weakly and inversely correlated (spearman's [r]: - 0.28) with outward sepsis transfer proportions. Higher scores had weak but better positive correlation with hospital mortality (r: 0.33), than sepsis volume (r: 0.24). CB scores explained more variation in sepsis mortality ($R^2 = 0.24$, P **DISCUSSION/SIGNIFICANCE:** Capability-based hospital scores account for three times more variation in sepsis mortality than sepsis volume and outperform sepsis volume as a system predictor of mortality. With further refinement and validation, these scores may find utility for improving system-based approaches to sepsis care.

61

Detecting Parkinson's Disease Using Computer Vision

Jacob Simmering, Robert Gerritsen, Nandakumar Narayanan
University of Iowa

OBJECTIVES/GOALS: Can we detect Parkinson's-disease-related motor impairments using computer vision and machine learning? **METHODS/STUDY POPULATION:** A sample of 29 people with Parkinson's disease (PD) and 29 non-Parkinson's disease (non-PD) controls were recruited from the University of Iowa Movement Disorders Clinic. Videos of 3 motor assessment tasks performed using the hands were recorded and hand location information was abstracted using the computer vision program MediaPipe. Measures from the raw data series and FFT were used as features to