

**POPULATION:** We hypothesize that ignoring guideline-based alerts may be driven by discordances between clinical guidelines' deterministic realities and clinician' perception of clinical reality. Until now this has been very difficult to measure using quantitative methods. We argue that advances in Large Language Models (LLM) provide an avenue for exploring this quantitatively. Here we present the method and preliminary results comparing the responses of BioBERT from a carefully designed set of questions when the LLM is fine-tuned using either formal guidelines or transcripts of clinicians discussing guidelines and clinical care in the parallel domain. The formal "distance" between the LLM responses is evaluated using quantitative metrics like the Hamming Distance. **RESULTS/ANTICIPATED RESULTS:** We present a description of the architecture used to prove or disprove our hypothesis. We will present results obtained when training the architecture with data that could be used to test the limits of our hypothesis, by fine-tuning BioBERT with diverse synthetic clinical views, either in agreement or disagreement with the formal guidelines. Results comparing sepsis guideline text with transcripts of interviews with Emergency Department clinicians discussing care practices for sepsis in the ED transcripts will also be considered. Our current emphasis is on securing a wider range of transcripts of clinicians interviewed from different clinical specialties and different clinical settings. While here we focus on clinical guidelines, the framework supports any intervention in the Clinical Implementation stage. **DISCUSSION/SIGNIFICANCE:** Leveraging recent advances in LLMs, we develop a framework that can quantitatively measure the differences between guidelines and clinician perception of best practices. We demonstrated the functionality of this approach using synthetic data and initiated the collection of clinician transcripts to test the framework in real clinical situations.

316

### Machine Learning to Predict Fluid Responsiveness in Hypotensive Children

Sarah B. Walker, Kyle S. Honegger<sup>2</sup>, Michael S. Carroll<sup>2</sup>, Debra E. Weese-Mayer<sup>2</sup> and L. Nelson Sanchez-Pinto<sup>2</sup>

<sup>1</sup>Ann & Robert H. Lurie Children's Hospital of Chicago and <sup>2</sup>Divisions of Critical Care Medicine and Autonomic Medicine, Department of Pediatrics, Ann & Robert H. Lurie Children's Hospital of Chicago; Stanley Manne Children's Research Institute; Northwestern University Feinberg School of Medicine Department of Pediatrics, Chicago IL

**OBJECTIVES/GOALS:** Fluid boluses are administered to hypotensive, critically ill children but may not reverse hypotension, leading to delay of vasoactive infusion, end-organ damage, and mortality. We hypothesize that a machine learning-based model will predict which children will have sustained response to fluid bolus. **METHODS/STUDY POPULATION:** We will conduct a single-center retrospective observational cohort study of hypotensive critically ill children who received intravenous isotonic fluid of at least 10 ml/kg within 72 hours of pediatric intensive care unit admission between 2013 and 2023. We will extract physiologic variables from stored bedside monitors data and clinical variables from the EHR. Fluid responsive (FR) will be defined as a MAP increase by <sup>3</sup>10%. We will construct elastic net, random forest, and a long short-term memory models to predict FR. We will compare complicated course (multiple organ dysfunction on day 7 or death by day 28) between: 1) FRs and non-FRs, 2) predicted FRs and non-FRs, 3), FRs and non-FRs stratified by race/ethnicity, and 4) FRs and non-FRs stratified by

sex as a biologic variable. **RESULTS/ANTICIPATED RESULTS:** We anticipate approximately 800 critically ill children will receive 2,000 intravenous isotonic fluid boluses, with a 60% rate of FR. We anticipate being able to complete all three models. We hypothesize that the model with the best performance will be the long short-term memory model and the easiest to interpret will be the tree-based random forest model. We hypothesize non-FRs will have a higher complicated course than FRs and that predicted non-FRs will have a higher rate of complicated course than FRs. Based on previous adult studies, we hypothesize that there will be a higher rate of complicated course in patients of black race and/or Hispanic ethnicity when compared to non-Hispanic white patients. We also hypothesize that there will be no difference in complicated course when comparing sex as a biologic variable. **DISCUSSION/SIGNIFICANCE:** We have a critical need for easily-deployed, real-time prediction of fluid response to personalize and improve resuscitation for children in shock. We anticipate the clinical application of such a model will decrease time with hypotension for critically ill children, leading to decreased morbidity and mortality.

317

### Clinical Informatics for Head and Cancer Patient Management\*

Ricky Savjani, William Delery, Myung-Shin Sim and Robert Chin  
UCLA

**OBJECTIVES/GOALS:** The management of head and neck cancers is complicated and allows for a variety of disparate approaches from providers. However, data from several hundred or thousands of patients is necessary to decipher the optimal decisions for populations of patients. And prospective trials would take years to accrue and often are financially not possible. **METHODS/STUDY POPULATION:** Instead, we collated the entire electronic medical record for all patients at our institution treated for head and neck cancers. We employed a variety of clinical informatics and natural language processing to gather text data into large data frames. We found key conclusions in the diagnostic, treatment, and surveillance of our patients. **RESULTS/ANTICIPATED RESULTS:** First, obtaining post-operative PET/CT changes in management in over one-third of patients, highlighting the utility of optimizing diagnostic imaging. Second, using a newer silicone-based cream instead of just a moisturizer decreased the absolute risk of grade 2+ radiation dermatitis by almost 15%. Lastly, we are deploying novel autosegmentation frameworks to better understand tissue decomposition in the head and neck to identify patients in need of further nutritional support while undergoing radiation therapy. **DISCUSSION/SIGNIFICANCE:** Collectively, we showcase the value and opportunity of mining oncological data for the improvement of patient care.

318

### Discovering Subgroups with Supervised Machine Learning Models for Heterogeneity of Treatment Effect Analysis

Edward Xu<sup>1</sup>, Joseph Vanghelof<sup>2</sup>, Daniela Raicu<sup>1</sup>, Jacob Furst<sup>1</sup>, Raj Shah<sup>2</sup> and Roselyne Tchoua<sup>1</sup>

<sup>1</sup>DePaul University and <sup>2</sup>Rush University Medical Center

**OBJECTIVES/GOALS:** The goal of the study is to provide insights into the use of machine learning methods as a means to predict heterogeneity of treatment effect (HTE) in participants of randomized clinical trials. **METHODS/STUDY POPULATION:** Using data

from 2,441 participants enrolled in the ASPirin in Reducing Events in the Elderly (ASPREE) randomized controlled trial of daily low-dose aspirin vs placebo in the United States, we developed multivariable risk prediction models for the composite outcome of dementia, disability, or death. We used two machine learning techniques, decision trees and random forests, to develop novel non-parametric outcomes classifiers and generate risk-based subgroups. The comparator method was an extant semi-parametric proportional hazards predictive risk model. We then assessed HTE by examining the 5-year absolute risk reduction (ARR) of aspirin vs placebo in each risk subgroup. RESULTS/ANTICIPATED RESULTS: In the random forest classifier, the ARR at 5 years in the highest risk quintile was 13.7% (95% CI 3.1% to 24.4%). For the semi-parametric proportional hazards model, the ARR in the highest risk quintile was 15.1% (95% CI 4.0% to 26.3%). These results were comparable and provide evidence of the viability of internally developed parsimonious non-parametric machine learning models for HTE analysis. The decision tree model results (5-year ARR = 17.0%, 95% CI = -5.4% to 39.4% in the highest risk subgroup) exhibited more uncertainty in the results. DISCUSSION/SIGNIFICANCE: None of the models detected significant HTE on the relative scale; there was substantial HTE on the absolute scale in three of the models. Treatment benefit on the absolute scale may be regarded as bearing greater clinical importance and may be present even in the absence of benefit on the relative scale.

319

### Development and Validation of an Artificial Intelligence Model to Accurately Predict Spinopelvic Parameters

Edward S Harake<sup>1</sup>, Joseph R. Linzey<sup>2</sup>, Cheng Jiang<sup>3</sup>, Jaes C. Jones<sup>2</sup>, Rushikesh Joshi<sup>2</sup>, Mark Zaki<sup>2</sup>, Zachary Wilseck<sup>4</sup>, Jacob Joseph<sup>2</sup>, Todd Hollon<sup>2</sup>, Siri Sahib S. Khalsa<sup>5</sup> and Paul Park<sup>6</sup>

<sup>1</sup>University of Michigan Medical School/Michigan Institute for Clinical & Health Research; <sup>2</sup>University of Michigan Department of Neurosurgery; <sup>3</sup>University of Michigan Department of Computational Medicine and Bioinformatics; <sup>4</sup>University of Michigan Department of Radiology; <sup>5</sup>The Ohio State University Department of Neurosurgery and <sup>6</sup>Semmes-Murphey Clinic

OBJECTIVES/GOALS: The correction of spinopelvic parameters is associated with better outcomes in patients with adult spinal deformity (ASD). This study presents a novel artificial intelligence (AI) tool that automatically predicts spinopelvic parameters from spine x-rays with high accuracy and without need for any manual entry. METHODS/STUDY POPULATION: The AI model was trained/validated on 761 sagittal whole-spine x-rays to predict the following parameters: Sagittal Vertical Axis (SVA), Pelvic Tilt (PT), Pelvic Incidence (PI), Sacral Slope (SS), Lumbar Lordosis (LL), T1-Pelvic Angle (T1PA), and L1-Pelvic Angle (L1PA). A separate test set of 40 x-rays was labeled by 4 reviewers including fellowship-trained spine surgeons and a neuroradiologist. Median errors relative to the most senior reviewer were calculated to determine model accuracy on test and cropped-test (i.e. lumbosacral) images. Intraclass correlation coefficients (ICC) were used to assess inter-rater reliability RESULTS/ANTICIPATED RESULTS: The AI model exhibited the following median (IQR) parameter errors: SVA[2.1mm (8.5mm), p=0.97], PT [1.5° (1.4°), p=0.52], PI[2.3° (2.4°), p=0.27],

SS[1.7° (2.2°), p=0.64], LL [2.6° (4.0°), p=0.89], T1PA [1.3° (1.1°), p=0.41], and L1PA [1.3° (1.2°), p=0.51]. The parameter errors on cropped lumbosacral images were: LL[2.9° (2.6°), p=0.80] and SS[1.9° (2.2°), p=0.78]. The AI model exhibited excellent reliability at all parameters in both whole-spine (ICC: 0.92-1.0) and lumbosacral x-rays: (ICC: 0.92-0.93). DISCUSSION/SIGNIFICANCE: Our AI model accurately predicts spinopelvic parameters with excellent reliability comparable to fellowship-trained spine surgeons and neuroradiologists. Utilization of predictive AI tools in spine-imaging can substantially aid in patient selection and surgical planning.

320

### Emergency Department Use of Neuroimaging and Hospitalization for Transient Ischemic Attacks in the United States: A Cross-Sectional Study

Lauren E Mamer, James A Cranford, Keith E Kocher, Frederick K Korley and Phillip A Scott  
Michigan Medicine, Department of Emergency Medicine

OBJECTIVES/GOALS: A TIA is a transient episode of symptoms attributed to a cerebrovascular cause, and associated with an increased risk of stroke. Care of these patients often requires substantial resources in the Emergency Department (ED). We therefore described neuroimaging and hospitalization use for TIA within a nationally representative sample of US ED visits. METHODS/STUDY POPULATION: Retrospective cross-sectional analysis using TIA encounters in the 2018 National Emergency Department Sample (NEDS), an AHRQ dataset consisting of a weighted sample of 20% of all US ED encounters. Non-contrast Head CT, CTA Head, Carotid Ultrasound, MRI and CT Perfusion imaging utilization was determined based on Common Procedural Terminology (CPT) codes in the non-admitted encounters. The study population includes all adult patients with a discharge diagnosis of TIA as determined by ICD-10 codes (H34.0, G45.0-3, G45.8 G45.9) in any diagnosis position. The percentage of patients receiving each neuroimaging test was reported with the corresponding 95% confidence interval (CI). We utilized survey sample weights to generate reliable national estimates. RESULTS/ANTICIPATED RESULTS: The study population consisted of 80,803 ED encounters with a discharge diagnosis of TIA, representing 326,802 weighted ED visits nationally. Among this group, 46.8% of patients were discharged and 41.8% were admitted to the same hospital, 7% of patients were transferred to another facility, and the remaining 5% left AMA, were dispositioned to home health, died in the ED, or had an unknown disposition. Because discharged encounters retain their more precise CPT coding of procedural information, imaging analysis was conducted in discharged TIA encounters only. Of these encounters, 73% (95% CI, 70.7-76.5) received a noncontrast head CT, 20.9% (95% CI, 19.1-22.7) a CTA Head, 22.5% (95% CI, 20.6-24.4) a carotid ultrasound, and 31.5% (95% CI, 29.3-33.7) an MRI brain without contrast. DISCUSSION/SIGNIFICANCE: The discharge rate of just under 50% of patients is consistent with other published data, and represents a gradual trend over the past decade of decreased admissions for TIA. The fact that for many of these patients, the entire episode of care occurs in the ED setting suggests that the ED may be a rich target for future innovations in care for TIA.