correlation coefficients showed significant positive relationship between factors like obesity and diabetes, median age and access to health care, and negative relationship between obesity and foreign born. Income, healthcare access, and white population were found to be significantly different SVIs from ANOVA. DISCUSSION/SIGNIFICANCE OF IMPACT: This research study found that some SDoH affect diabetes and obesity in the same direction. The association is positive for median age and negative for income, SVI, percentage of white population, and foreign born. The associations were found between actionable and non-actionable factors like percentage of white population with access to health care.

378

Leveraging large language models to communicate translational science benefits at Weill Cornell Medicine Clinical and Translational Science Center

Michael Bales[1], EA Wood[2], Sigaras A[3], E Campion Sholle[4], TR Jr[2] and J Imperato-McGinley[2]
[1]Weill Cornell Medicine; [2]Weill Cornell Clinical and Translational Science Center, New York, NY, USA; [3]AI-XR Lab, Weill Cornell Medicine, New York, NY, USA; Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA; Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA and [4]Department of Information Technologies & Services, Weill Cornell Medical College, New York, New York; Division of Health Informatics, Department of Population Health Sciences, Weill Cornell Medical College, New York, New York

OBJECTIVES/GOALS: This Weill Cornell Clinical and Translational Science Collaborative (CTSC) project evaluates whether large language models (LLMs) can generate accurate summaries of translational science benefits using the Translational Science Benefits Model (TSBM) framework, aiming to identify optimal LLMs and prompting strategies via expert review. METHODS/STUDY POPULATION: We are using prompt engineering to train multiple LLMs to generate one-page impact profiles based on the TSBM framework. LLMs will be selected via benchmarks, focusing on models excelling in information extraction. Leading LLMs (e.g., Llama 3.2, ChatGPT 4.0, Gemini 1.5 Pro, and Claude) and other high-performing models will be considered. Initial work has utilized Gemini 1.5 Pro. Models use data from CTSC-supported projects in WebCAMP, our local instantiation of a translational research activity tracking system used by >20 CTSA hubs, and manuscripts from the Overton database cited in policy documents. Human experts will evaluate the quality and accuracy of LLM-generated profiles. RESULTS/ANTICIPATED RESULTS: Preliminary results using Gemini 1.5 Pro indicate that LLMs can generate coherent and informative impact profiles encompassing diverse areas within the TSBM. Face validity appears satisfactory, suggesting the outputs align with expectations. We anticipate that further exploration with other LLMs and expert validation will reveal strengths and weaknesses of the LLM approach, including the potential for naccuracies ("hallucinations"), informing further refinement of models and prompting strategies. Analysis of manuscripts cited in policy will provide valuable insights into communicating policy-relevant benefits effectively, and benchmark comparisons will identify optimal LLMs for this use case. DISCUSSION/SIGNIFICANCE OF IMPACT: This project demonstrates LLMs' potential for streamlining and enhancing impact reporting in translational science, enabling broader dissemination of research outcomes and promoting better understanding among stakeholders. Future work will integrate LLM-based reporting into research infrastructure.

379

Using the Delphi method to strategize about health AI

Whitney Welsh and Shelley Rusincovitch
Duke University

OBJECTIVES/GOALS: Our goal was to determine whether a consensus exists around 1) what the main barriers to innovation in Health artificial intelligence (AI) are 2) where there are gaps in education and training in Health AI and 3) where in their workflows organizations should implement AI to see the most immediate impact on productivity. METHODS/STUDY POPULATION: We employed a three-round Delphi method survey to stakeholders with health and/or engineering expertise. The first round was open-ended to generate responses to the three research questions. The second round asked participants to rank the responses and provide feedback as to their reasoning. The third round provided aggregated results and feedback and asked participants to re-rank the responses. Participants were attendees at a conference that brought people with health and/or engineering backgrounds together to discuss innovation in Health AI. 55 people in total participated across the three rounds. RESULTS/ANTICIPATED RESULTS: Consensus emerged on all three questions: lack of trust was seen as the single greatest barrier to innovation, experience with implementation as the greatest gap in training, and automating health documentation as the point of most immediate impact. Consensus also emerged as to which of the 10–15 responses to each question were top priorities, which were somewhat significant, and which were not that important. Some of the rankings (such as implementation) seemed to reflect hot topics of discussion at the conference, but others (such as documentation) only emerged as significant in the surveys. DISCUSSION/SIGNIFICANCE OF IMPACT: We successfully employed the Delphi method to discover what stakeholders think about three important questions in Health AI. Interestingly, although we polled experts from both health and engineering backgrounds, their answers converged on all three questions.

380

National trends in interventional clinical trial participation by race, gender, and age: Insights from EHR data on over 130 million patients

Sarah Fry[1], Sarah E. Fry[2], Pauline Terebuh[2], Pamela B. Davis[2], Lara Jehi[3], Yasir Tarabichi[4] and David C. Kaelber[4]
[1]Case Western Reserve University School of Medicine; [2]Case Western Reserve University School of Medicine, Cleveland, OH; [3]Cleveland Clinic Foundation, Cleveland, OH and [4]Metrohealth Medical Center Cleveland, OH

OBJECTIVES/GOALS: To investigate interventional clinical trial participation overall and by race, gender, and age. METHODS/STUDY POPULATION: We used Epic Cosmos, an aggregated, de-identified EHR platform including over 270 million patients, to examine overall clinical trial participation and the race, gender,

https://doi.org/10.1017/cts.2024.1003

and age composition of participants versus non-participants. Patients ≥5 years old with known race and gender and at least one healthcare encounter between 2021 and 2024 were included. Interventional trial enrollment was identified by a "research flag" indicating current or past participation in an interventional study within an Epic system contributing data to Cosmos. Race was categorized as American Indian, Asian, Black, Native Hawaiian, or White. Age-adjusted relative representation (RR) ratios were used to compare participation, with RR >1 indicating over-representation and RR RESULTS/ANTICIPATED RESULTS: Of 130,455,189 patients meeting eligibility criteria, 0.52% (673,425) of patients were active or inactive in an interventional clinical trial. Results are shown in the figure below. The poorest representation was from Asian and NH/PI persons. Representation was most similar to the patient population for whites and AI/AN persons. Black males participated less and women, more than predicted by patient composition. Older patients participated more frequently than younger (age, mean (SD), y, 53 (22) vs. 46 (23); p DISCUSSION/SIGNIFICANCE OF IMPACT: This is the first study we know of describing interventional trial participation in the USA across millions and millions of patients. Further research is needed to clarify whether these differences are due to the nature of the studies themselves (e.g., OB/GYN trials including only women, etc.) versus disparities in recruitment or otherwise.

381

Host-bacterial immune responses to ventilator-associated pneumonia in COVID-19 patients

Cecilia Chung[1], Yaa Kyerematen[2], Kendrew Wong[2], Miao Chang[2], Rajbir Singh[2], Colin Mccormick[2], Anna Czachor[2], Clea Barnett[2], Yonghua Li[2], Tsay Juh-Chieh[3], Leopoldo N. Segal[2] and Benjamin G. Wu[3]

[1]New York University Clinical and Translational Science Institute (NYU Grossman School of Medicine); [2]New York University Grossman School of Medicine – NYUGSoM and [3]New York Harbor Healthcare System

OBJECTIVES/GOALS: Ventilator-associated pneumonia (VAP) is an infection caused by bacteria, viruses, or fungi during mechanical ventilation. We analyzed a cohort of COVID-19 patients admitted to the intensive care unit with respiratory failure with different VAP outcomes. We hypothesize that the multiomics data can help predict VAP development within this cohort. METHODS/STUDY POPULATION: We recruited participants from a cohort on a NYU IRB protocol (i22–00616), who had COVID19 respiratory failure, admitted to ICU, and required invasive mechanical ventilation (n = 245). We collected and analyzed research specimens (bronchoalveolar lavage [BAL, n = 213], tracheal aspirates [n = 246], background [n = 18]) and clinical cultures (sputum and BAL) for 245 participants. A panel of experts adjudicated VAP within the cohort, resulting in 92 VAP diagnoses. We annotated metatranscriptome (Illumina NovaSeq) using a Kraken/Bracken database, and KEGG for functional annotation of transcriptome data (Illumina HiSeq). We used edgeR (v.4.0.16) to analyze differential expression of metatranscriptome and transcriptome data. RESULTS/ANTICIPATED RESULTS: We diagnosed VAP in n = 92 (38%) participants. We found significant differences in days of overall hospital stay (p DISCUSSION/SIGNIFICANCE OF IMPACT: VAP is a serious complication of mechanical ventilation, and oral commensals alter the lung microbiome and host immunity. We identified a transcriptome-metatranscriptome signature that identifies those at VAP risk. VAP was associate with both pro- and anti-inflammatory gene expression resulting in increased risk for lower airway infection.

382

Online graphical interface for bulk to single-cell transcriptomics

Manoj Kandpal and Hong Hur

Rockefeller University

OBJECTIVES/GOALS: We aim to develop an intuitive interface to understand the possible relationships between data from different RNA-Seq technologies. It can help novice users and educators to understand, analyze, explain, and visualize such datasets from diverse platforms, all without the need for additional software installations or strong programming expertise. METHODS/STUDY POPULATION: An online interactive interface is developed, integrating robust algorithms for three distinct types of analyses: DESeq2 for bulk RNA-seq, CIBERSORTx for deconvolution, and Seurat for single-cell analysis, with plans to include more algorithms. It allows a demo mode for training using the sample datasets and option for tailored analysis using user's partially processed data. The interface provides capability to process bulk RNA-seq data from raw counts or a differential gene list. Further, deconvolution analysis for bulk RNA-seq data can be done using raw counts and single-cell data analysis can be performed using processed sequence reads, organized into three key files: barcodes, matrix, and features. Users also have an option to download the results as a zipped file, for samples from human and mouse studies. RESULTS/ANTICIPATED RESULTS: Users with an active internet connection can access the interface from any major web browser. They can adjust parameters – such as genome type, cutoff thresholds, and batch effect correction – according to their specific needs. Bulk RNA-seq results are presented in the form of volcano plots, heatmaps, clusters, gene expression across samples, DEGs, and enrichment plots from KEGG and GO analyses. Deconvolution analysis can be performed using either the "LM22" signature matrix (for human leukocyte cell types) or Derm22 (for skin-specific cell types). The single-cell workflow provides results including quality control metrics, UMAP clustering, gene expression plots/tables, and cluster annotation using CellTypist. Comprehensive details on methods and tutorials are available in the GitHub repository. DISCUSSION/SIGNIFICANCE OF IMPACT: Although multiple workflows are available to process bulk and single cell RNA-Seq data along with deconvolution methods to bridge the gap between the two, this is the first online interface to provide the capability to explore and analyze data from all three approaches in one place, without requiring strong computational expertise or resources.

383

A CTS team approach to leveraging EHR data for predicting necrotizing enterocolitis in NICU*

Keliy Fordham and Yao Hou

University of Florida CTSI

OBJECTIVES/GOALS: This research aims to harness electronic health records (EHR) combined with machine learning (ML) to predict necrotizing enterocolitis (NEC) in preterm infants using data up to their first 14 days of life. We aim to provide interpretable results for clinical decisions that can reduce infant mortality rates and complications from NEC. METHODS/STUDY POPULATION: Through a retrospective cohort study using data from the

https://doi.org/10.1017/cts.2024.1003