network. Although we supplement with laboratory-based diagnosis, using diagnosis codes as labels is problematic as numerous reports suggest low sensitivity of codes for AKI. Future work includes calibration analysis, incremental updating (“online learning”), and a representation learning-based (“deep learning”) extension of the model.

Genetic determinants of recovery after mild traumatic brain injury: Can study samples be linked to DNA biobanks?
Jessica Dennis, Scott Zuckerman, Aaron Yengo-Kahn, Nancy Cox and Gary Solomon
The Vanderbilt Sports Concussion Center. VUMC, Nashville, TN, USA

OBJECTIVES/SPECIFIC AIMS: To develop an algorithm that identifies post-concussion syndrome (PCS) cases and controls from among patients with mild traumatic brain injury (mTBI) in a large academic biobank. METHODS/STUDY POPULATION: The Vanderbilt University Medical Center’s (VUMC) electronic medical record (EMR) research database includes longitudinal medical record data on 2.5 million people. DNA and genotype data were also available for >225,000 of these individuals. Our algorithm used a combination of billing codes and natural language processing to apply inclusion and exclusion criteria. We defined PCS cases as those with a PCS billing code (ICD-9 306.02 or ICD-10 F07.81) and/or symptoms of PCS within 1–6 months of qualifying mTBI. We will compare the positive predictive value of our algorithm to that of 2 simpler case selection schemes: (1) 1 instance of the PCS billing code anywhere in the medical record; and (2) 2 or more instances of the PCS billing code anywhere in the medical record. RESULTS/ANTICIPATED RESULTS: An mTBI was diagnosed in 28,720 patients regularly attending VUMC, and 528 of these patients were classified as PCS cases by our algorithm. The characteristics of our EMR sample reflected known risk factors for PCS. Our cases were more likely than controls to be female (49.4% vs. 38.4%), to have sustained a previous TBI (31.0% vs. 12.0%) and to have comorbid mood disorders. Our PCS cases were also more likely to have a lower age of 42.4% vs. 33.6%) and to have a sports-related keyword associated with the mTBI (44.1% vs. 25.2%), emphasizing the relevance of sports to young athletes. Nonetheless, the number of PCS cases identified by our algorithm was small, and within the VUMC EMR, there were 5039 patients with 1 PCS billing code, and 2457 patients with 2 or more PCS billing codes anywhere in their EMR. Our next step is to calculate the positive predictive values of each selection scheme by manually reviewing the EMR of a selection of cases. Ultimately, we will implement the selection scheme that maximizes both positive predictive value and sample size, and in future work, we will genotype the selected patients to better understand the genetic architecture of PCS. DISCUSSION/SIGNIFICANCE OF IMPACT: EMR and biobanks are the future of human health research, and we asked whether complex algorithms or simple billing codes were best for studying the genetics of recovery after mTBI within the VUMC EMR. Our results are relevant to other studies of brain injury phenotypes within biobanks, including recovery from moderate or severe TBI, recovery from stroke, or the occurrence of delirium after routine surgery, and will help transform biobanks into fruitful research tools.

The design of a patient-centered personal health record with patients as co-designers
Arlene Chung, Haiwei Chen, Grace Shin, Ketan Mane and Hye-Chung Kum

OBJECTIVES/SPECIFIC AIMS: The promise and potential of connected personal health records (PHRs) has not come to fruition. This may be, in part, due to statistical (false discoveries, p-hacking, overuse/misuse of p-values, low power, poor experimental design) and computational (data, code and software management) issues. These require understanding the processes and workflows practiced by an organization, and the development and use of metrics to quantify reproducibility. METHODS/STUDY POPULATION: Within the Foundation of Discovery – Population Health Research, Center for Clinical and Translational Science, University of Utah, we are undertaking a project to quantify reproducibility. OBJECTIVES/SPECIFIC AIMS: Key factors causing irreproducibility of research datasets include those related to appropriate study design methodologies and statistical analysis. In modern statistical practice irreproducibility can arise due to statistical (false discoveries, p-hacking, overuse/misuse of p-values, low power, poor experimental design) and computational (data, code and software management) issues. These require understanding the processes and workflows practiced by an organization, and the development and use of metrics to quantify reproducibility. METHODS/STUDY POPULATION: Within the Foundation of Discovery – Population Health Research, Center for Clinical and Translational Science, University of Utah, we are undertaking a project to quantify reproducibility.
streamline the study design and statistical analysis workflows and processes. As a first step we met with key stakeholders to understand the current practices by eliciting example statistical projects, and then developed process information models for different types of statistical needs using Lucidchart. We then reviewed these with the Foundation’s leadership and the Standards Committee to come up with ideal workflows and model, and defined key measurement points (such as those around study design, analysis plan, final report, requirements for quality checks, and double coding) for assessing reproducibility. As the next step, we are using our finding to embed analytic and infrastructural approaches within the statisticians’ workflows. This will include data and code dissemination platforms such as Box, Bitbucket, and GitHub, documentation platforms such as Confluence, and workflow tracking platforms such as Jira. These tools will simplify and automate the capture of communications as a statistician work through a project. Data-intensive process will use process and workflow management platforms such as Activiti, Pegasus, and Taverna. RESULTS/ANTICIPATED RESULTS: These strategies for sharing and publishing study protocols, data, code, and results across the spectrum, active collaboration with the research team, automation of key steps, along with decision support. DISCUSSION/SIGNIFICANCE OF IMPACT: This analysis of statistical methods and process and computational methods to automate them ensure quality of statistical methods and reproducibility of research.

**2476**

**Identifying strangulated small bowel obstruction with machine learning**

Samuel David Zetumer and Hobart Harris

OBJECTIVES/SPECIFIC AIMS: Historically, logistic regression algorithms (LRAs) have failed to differentiate strangulated small bowel obstructions (SBOs) from nonstrangulated SBOs. Our hypothesis is that a machine learning algorithm (MLA) can differentiate strangulated from simple SBOs better than an LRA can. METHODS/STUDY POPULATION: We used records of patients presenting with acute SBO and managed with exploratory laparotomy to test and train algorithms. We compared MLA to LRA via area under the receiver operating characteristic curve (AUROC) and cut-off points maximizing sensitivity and specificity. RESULTS/ANTICIPATED RESULTS: With 192 patient records, the AUROC of the MLA was 0.85. At the sensitivity cutoff, the MLA had 100% sensitivity and 55% specificity. At the specificity cutoff, the MLA had 45% sensitivity and 100% specificity. We anticipate improvements as more records are incorporated, and that LRA will underperform MLA across all measures. DISCUSSION/SIGNIFICANCE OF IMPACT: Our MLA represents a significant improvement over past LRAs, and may provide decision assistance to surgeons managing SBO. If this MLA maintains its high sensitivity, it may be used in the future to prevent unnecessary surgeries.

**2492**

**Leveraging CTSA informatics capacity to expand global health engagement and research capacity in Latin America and the Pacific**

Timothy De Ver Dye, Thomas Fogg, Margaret Demment, José Pérez-Ramos, Scott McIntosh, Deborah Ossip, Angela Sy, Carmen Velez Vega, Karen Peters and Haq Nawaz

University of Rochester Medical Center, Rochester, NY, USA

OBJECTIVES/SPECIFIC AIMS: The objective of this partnership was to create a global network of clinical and public health researchers and communities conducting technology-assisted research in noncommunicable disease. METHODS/STUDY POPULATION: The University of Rochester’s Clinical and Translational Science Institute (CTSI) has successfully leveraged the informatics capacity into an emerging network of organizations that focus on technology and health in settings outside of the mainland United States. The CTSI’s coordinated with another NIH-funded infrastructure program [the RCMI coordinated with another NIH-funded infrastructure program [the RCMI core] and the University of Puerto Rico and the University of Hawaii, both of which serve as hubs for common research interests in technology and health throughout the Caribbean and the Pacific. This network was formalized as the CDC’s Coordinating Center for its Global and Territorial Health Research Network (the “Global Network”), with additional US partners (Yale, University of Illinois at Chicago, University of North Carolina Chapel Hill, and the University of South Florida) within a wider scope of the CDC’s Prevention Research Centers (PRC) program. RESULTS/ANTICIPATED RESULTS: Through combining 2 main NIH-funded research infrastructure networks (CTSA and RTRN), with a large CDC-funded PRC, the University of Rochester’s Informatics Core was successful in establishing a new productive global health network throughout Latin America and the Caribbean, and in the Pacific, garnering additional research support from NIH Fogarty and other programs. The resulting network not only supports locally-important research in technology and health on compelling health issues (eg, diabetes, Zika, participation in research), but also facilitates community engagement through local partnerships and the cores of the involved networks. In addition, much of the information and communications technology (ICT)-related research and learnings from the Global Network activity is immediately applicable to populations in the United States, served by the various collaborative networks. In total, while new, the Global Network supports a wide range of projects and engagements throughout the world that expand local informatics capacity and use of technology in the research process and to address global health problems, further enhancing the CTSI’s informatics core, to serve the needs of its own constituency and promote research engagement with technology within this population. Local research collaborative projects reinforce the utility of the network and its resources, evidenced by tools, publications, partnerships, and conference presentations that have arisen. Lessons to date from this Global Network collaboration include: specific global research projects provide opportunities for partnership building and meaningful collaboration, team science is of central importance in distributing the work of the network, synergy is multidirectional with expertise and need flowing in all directions, and project team members in all locales learned and contributed substantially in ways that carried into their other responsibilities. DISCUSSION/SIGNIFICANCE OF IMPACT: The overall partnership has created a unique opportunity for South-South collaboration, for adaptation of projects among locales, and has helped boost reputational value for all partners involved. Implications for other CTSI awardees include: global collaboration can serve core research and technical needs for the CTSI itself and its local partners, CTSI status can be leveraged to access resources to support local research, and collaboration in other federally-funded research networks helps expand the insight, scope, and potential for new research.

**2498**

**Individual patient outcome predictions using supervised learning methods**

Abiel Roche-Lima, Patricia Ordoñez, Nelson Schwarz, Adnel Figueroa-Jiménez and Leonardo A. García-Lebrón

University of Puerto Rico-Medical Sciences Campus, San Juan, Puerto Rico

OBJECTIVES/SPECIFIC AIMS: To learn the edit distance costs of a symbolic univariate time series representation through a stochastic finite-state transducer to predict patient outcomes in intensive care units. METHODS/STUDY POPULATION: High frequency data of patient outcome predictions using supervised learning methods.

**2505**

**Understanding quality of life transitions for women: Assessing the impact of EPIC decision support tools to address untreated menopausal symptoms on women's quality of life and provider workflow**

Margaret Demment, Ivelisse Rivera, Morgan Pratte, Miriam Weber, Chris Morley and Tim Dye

University of Rochester Medical Center, Rochester, NY, USA

OBJECTIVES/SPECIFIC AIMS: The goal of this study is to assess how quality of life scores change in menopausal women before and after implementation of this aid. In addition, we are also interested in 2 process evaluation objectives: (1) determine if MyChart, the patient portal, is an effective way for this patient population to provide insight their quality of life to their providers and (2) to evaluate providers use of and reactions to the decision support tool. METHODS/STUDY POPULATION: This project is a collaboration between University of Rochester Medical Center and S.U.N.Y. Upstate Medical