best practice maneuvers, this model will not require manipulation of the patient, have less rigid criteria for reliable interpretation, and not require as specific of a technical skill set to interpret. Furthermore, it will include many common categories of resuscitative therapies (eg, vasopressors, inotropes, fluids) and will allow effects of a combination of interventions to be predicted while making no assumptions of independence between said interventions. This study will also contribute a novel process of sequence prediction using RNNs by incorporating an element of context on top of the sequential data in every training step. An RNN learning the sequence of hemodynamic data comprising a patient’s hemodynamic state would, alone, fit into the realm of sequence prediction. Our innovation is the addition of treatment information with each temporal division of the hemodynamic data. The result is an RNN that combines the task of sequence prediction with sequence translation, the 2 major use cases for RNN learning algorithms.

Immune stress biomarkers correlate to violence and internalization of violence in African American young adults
Latifa Jackson, Max Shestov, Forough Saadatmand and Joseph Wright
Georgetown - Howard Universities, Washington, DC, USA

OBJECTIVES/SPECIFIC AIMS: Allostatic load, the chronic stress-induced wear and tear on the body, has a cumulative deleterious effect in individuals over their lifetime. Recent studies have suggested that socio-economic status, psychological determinants, and biomedical health cumulatively contribute to allostatic load in young adults. Although these finding individually suggest that African American children may be particularly susceptible to the effects of allostatic load due to racially-based discrimination and economic instability, few studies have shown the effect of exposure to violence on the allostatic load carried by young African Americans. METHODS/STUDY POPULATION: The Biological and Social Correlates of Drug Use in African American Emerging Adults (BADU) data set is composed of young African Americans (n = 557 individuals) living in the Washington, DC area, collected from 2010 to 2012. Study participants were sought equally between males and females (n = 283, n = 274, respectively). This data set provides a rich source of information on the behavioral, mental, and physical health of African American young adults (18–25 year olds) living in the Washington, DC area. Analysis of 6 biomedical markers were measured in BADU study participants: C-reactive protein, cortisol, Epstein-Barr virus IgM, IgE, IgA, and IgG, known to be markers of immune stress and allostatic load. Naïve Bayes was used to identify participant responses that were correlated to elevated stress biomarker levels. RESULTS/ANTICIPATED RESULTS: Violence was most closely correlated to elevated EBV-VM IgM and IgE levels. Elevated IgE levels correlated to increased experience of familial violence and sexual abuse; familial drug abuse and depression; violence and chronicity of intervention. Violence was positively correlated to reported emotional state (R = 0.072) and perceived individual discrimination (R = 0.059). DISCUSSION/SIGNIFICANCE OF IMPACT: Allostatic load appears to be high in individuals who self-report exposure to violence. Both perceived mental health and violence were correlated to elevated stress biomarkers. When Epstein-Barr virus viral capsid antigen IgM was compared with violence features characterized in this dataset, we found that internalization of environmental stressors were most strongly correlated to elevated allostatic load markers. This work suggests that internalization of experienced violence may be an important as the actual violence experience.

A machine learning pipeline to predict acute kidney injury (AKI) in patients without AKI in their most recent hospitalization
Samuel Weisenthal, Samuel J. Weisenthal, Caroline Quill, Jiebo Luo, Henry Kautz, Samir Farooq and Martin Zand

OBJECTIVES/SPECIFIC AIMS: Our objective was to develop and evaluate a machine learning pipeline that uses electronic health record (EHR) data to predict acute kidney injury (AKI) during rehospitalization for patients who did not have an AKI episode in their most recent hospitalization. METHODS/STUDY POPULATION: The protocol under which this study was funded was exempt status by our institutional review board. The fully deidentified data set, containing all adult hospital admissions during a 2-year period, is a combination of administrative, laboratory, and pharmaceutical information. The administrative data set includes International Classification of Diseases, 9th Revision (ICD-9) diagnosis and procedure codes, Current Procedural Terminology, 4th Edition (CPT-4) procedure codes, diagnosis-related grouping (DRG) codes, locations visited in the hospital, discharge disposition, insurance, marital status, gender, ethnicity, and length of stay. The laboratory data set contains measured bicarbonate, chloride, calcium, albumin, phosphorus, glucose, glomerular filtration rate, creatinine, urea nitrogen, albumin, total protein, liver function enzymes, and hemoglobin. The pharmacy data set includes, for each medication, a description, pharmacologic class and subclass, and therapeutic class. Data preprocessing was performed using Python library Pandas (McKinney, 2011). Top 1000 binary representation was kept. In total, 508 features and 51004 unique experiments were stored in IPython (Pérez, 2007) notebooks for easy viewing and results reproduction. RESULTS/ANTICIPATED RESULTS: There were 107,036 adult patients that accounted for 199,545 admissions during a 2-year window. Per admission, there were at most 54 ICD-9 diagnoses, 38 ICD-9 procedures, 314 CPT-4 procedures, and 25 hospital locations visited. The admissions were 55% female, the average age was 46 ± standard deviation 20, and average length of stay was 2.5 ± 8.0 days. We excluded 2360 admissions that involved an AKI event that directly followed an admission with an AKI event and 4310 admissions that did not involve an AKI event but directly followed an admission with an AKI event. In total, there were 4561 (5.3%) positive samples (AKI during rehospitalization without AKI in the previous stay) generated by 3699 unique patients and 81,458 negative samples (non-AKI during rehospitalization without AKI in the previous stay) generated by 31,831 unique patients. When using any AKI event as a positive sample (regardless of whether or not AKI was in the most recent stay), the prevalence was 7.3% (6921 positive samples generated by 4395 unique patients and 85,588 negative samples generated by 33,287 unique patients). Best results were achieved with a code precision of 3 digits for which we had a total of 4556 features per patient. Fitted hyper-parameters corresponding to each classifier were logistic regression with 11 penalty C = 2 × 10⁻⁵, logistic regression with 12 penalty C = 1 × 10⁻⁴, random forest number of estimators at 100, maximum depth as 3, minimum samples per leaf as 50, minimum samples per split as 10, and entropy as the splitting criterion; and multilayer perceptron I2 regularization parameter α as 15, architecture as 1 hidden layer with 5 units, and learning rate as 0.001. Five-fold stratified cross validation on the development set yielded AUROC for logistic regression with 11 penalty C = 0.810 ± 0.006, logistic regression with 12 penalty C = 0.796 ± 0.007, random forest 0.828 ± 0.007, and multilayer perceptron 0.841 ± 0.005. In an identical experiment for which an AKI event was considered a positive sample regardless of AKI presence in the most recent stay, we had 4592 features per sample with the same code precision. Five-fold stratified cross validation on the development set identified with identical settings for the hyperparameters yielded AUROC for logistic regression with 11 penalty C = 0.805 ± 0.004, logistic regression with 12 penalty C = 0.819 ± 0.006, random forest 0.833 ± 0.004, and multilayer perceptron 0.853 ± 0.006. DISCUSSION/SIGNIFICANCE OF IMPACT: Our objective was to investigate the feasibility of using machine learning methods on EHR data to provide a personalized risk assessment for “unexpected” AKI in rehospitalized patients. Preliminary model discrimination was good, indicating that this approach could aid clinicians to recognize AKI risk in unsuspicous patients. The authors recognize several limitations. Since our data set corresponds to a time-window sample, patients with high frequency of hospital utilization are likely over-represented. Similarly, our data set contains records from only 1 hospital.
Genetic determinants of recovery after mild traumatic brain injury: Can study samples be identified from electronic medical records 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) electroni

ic 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-9 310.2 or ICD-10 F07.81) and/or symptoms of PCS within 1

6 months of a 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 be older than controls of age (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 PCS 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.