rates increased. mended by HCUP, and reweighting by adjustment cells performed similarly in mean data sets. In general, biases in magnitude and MSE of mean adjusted costs increased as data sets. To evaluate the quality of an estimator, we assessed bias, variance, and data set with non-missing CCRs as a benchmark complete data set. We picked one using four techniques for handling missing CCRs.

PRM35

EMPIRICAL INVESTIGATION OF TECHNIQUES FOR HANDLING MISSING COST-TO-CHARGE RATIO (CCR) IN NATIONWIDE INPATIENT SAMPLE (NIS) Yaqoob A, 1Thompson A, 1Jones R, 1Henderson D 2
1Department of Internal Medicine, University of Utah, Salt Lake City, UT, USA; 2Salt Lake City Health Care System, Salt Lake City, UT, USA

OBJECTIVES: Researchers use the Healthcare Cost and Utilization Project (HCUP) Cost-to-Charge Ratio (CCR) to estimate total charges in the Nationwide Inpatient Sample (NIS) into accurate costs. We investigated the impact on cost estimates using four techniques for handling missing CCRs. METHODS: This is a cross-sectional study using nationally-representative hospitalization data in the US in 2011. We investigated four techniques including: 1) complete case analysis, 2a) using a weighting technique recommended by HCUP, 2b) reweighting by adjustment cells, and 3) using hotdeck imputation by adjustment cells. We used five disease codes: T20 (diabetes), K70 (hypertension), M32 (malignant neoplastic disease of liver), N18 (chronic liver disease and intrahepatic cirrhosis), and N30 (alcohol-related disorders). We created a disease cohort data set with non-missing CCRs as a benchmark complete data set. We picked one state at a time and set their CCRs to missing to create a data set with missing CCR data. We compared the CCR-adjusted cost estimates (mean and total) from both data sets. To evaluate the quality of an estimator, we assessed bias, variance, and the mean squared error (MSE) of an estimator. RESULTS: The range of missing rates was 1% to 29% of sample one to 32% of the other data sets, a one-to-one matching might be needed. A systematic approach is needed to overcome the limitations and complexities. METHODS: Nineteen Charlson Comorbidity Index (CCI) medical conditions plus 21 others were chosen for mapping. CCI's 2013 GEMs were reviewed when no GEMs existed in total CCRs. Pro-prietary mapping systems were utilized. When necessary, expertise in coding and clinical knowledge was leveraged to improve the mapping. RESULTS: The 164 unique ICD-9-CM codes required to fully define the 40 medical conditions, mapped to 3742 unique ICD-10-CM codes. A total of 6355 code-to-code mappings were obtained. 1135 (17.9%) had a one-to-one match, 3636 (57.2%) had a many-to-one match (ICD-10-CM less granular), 425 (6.7%) had a one-to-many match (ICD-10-CM more granular), and 1156 (17.6%) had a more complex many-to-many code map. CONCLUSIONS: GEMs provides a strict and limited mapping of many conditions necessitating additional work to define these conditions. When ICD-10-CM codes are in effect, validation studies are needed to verify the accuracy of the proposed mappings, particularly for commonly used conditions such as CCI. The CCI's reliability and specificity, particularly, must be maintained with the new ICD-10-CM codes in order to continue to be useful as an important research tool.

PRM36

DEVELOPMENT OF A SPATIALLY-ENABLED PUBLIC-USE DATABASE FOR END-STAGE RENAL DISEASE POLICY STUDIES Stephens M, 1 Maione-Dowing B, 1 Brotherton S A, 1 Gitlin M D 2
1Prima Healthcare Analytics, Weymouth, MA, USA; 2InnPeritus, Geneva, Switzerland

OBJECTIVES: While patients with end-stage renal disease (ESRD) represent only 1 percent of Medicare spending, this proportion of 1 percent accounts for 8 percent of all Medicare spending. The disproportionate cost of treating ESRD has resulted in a high volume of research focused on ESRD prevention or management. Some important findings are as follows: 1) The use of biologic DMARDs administered at Veterans Health Administration (VHA) facilities are well-documented in the electronic medical record but data are not consistently entered into the pharmacy dispensing or nurse administration structured data sources. Although CPT codes can be used to identify many infusion events but inconsistent coding does not allow estimation of the administered dose. To address this, we developed Natural Language Processing (NLP) software to identify potential infusion notes. We used the NLP software to extract documented dosage information, and standardize results. METHODS: Trained reviewers compared the NLP extractions to source documents and judged whether the software correctly extracted and standardized data. The software contains a display window allowing reviewers to directly assess the NLP Extraction. NLP was run on all notes from Medicare (8,635) and SelectCare (6,128) CDM version. Different numbers of coding were performed for each data source to identify biologic DMARD infusion data when coding is inconsistent.

PRM37

BUILDING A BRIDGE: ICD-9-CM TO ICD-10-CM MAPPING CHALLENGES AND SOLUTIONS Searle J, 1 Yong Y F, 2 Zarotovs K V 3
1Opuntia, Eden Prairie, MN, USA; 2Zsbekas Pharmaceuticals International Inc, Deerfield, IL, USA

OBJECTIVES: The mandated implementation of ICD-10-CM (International Classification of Diseases, Tenth Revision, Clinical Modification) codes in 2014 changes the way health care organizations determine the medical conditions of patients. This creates a need for clinical and coding experts to work with the ICD-10-CM code sets to ensure that the correct diagnosis is recorded. The magnitude of the challenge is due to the number of conditions that are mapped from ICD-9-CM to ICD-10-CM. As a result, the mapping process has become a critical activity that must be completed accurately and efficiently. Despite the existence of tools that can assist in mapping, a systematic approach is needed to overcome the limitations and complexities. METHODS: Antirheumatic Drugs (DMARDs) administered at Veterans Health Administration (VHA) facilities are well-documented in the electronic medical record. Although CPT codes can be used to identify many infusion events but inconsistent coding does not allow estimation of the administered dose. To address this, we developed Natural Language Processing (NLP) software to identify potential infusion notes. We used the NLP software to extract documented dosage information, and standardize results. METHODS: Trained reviewers compared the NLP extractions to source documents and judged whether the software correctly extracted and standardized data. The software contains a display window allowing reviewers to directly assess the NLP Extraction. NLP was run on all notes from Medicare (8,635) and SelectCare (6,128) CDM version. Different numbers of coding were performed for each data source to identify biologic DMARD infusion data when coding is inconsistent.

PRM38

USE OF A COMMON DATA MODEL TO FACILITATE RAPID ANALYTICS SUPPORTING HEALTH OUTCOMES RESEARCH Kim H, 1 Joo S, 1 Anstatt D, 1 Morrison J, 1 Reisinger S, 1 Germscheid L, 2 Murray R 2
1Bristol-Myers Squibb Company, Pennington, NJ, USA; 2UBC, Harrisburg, PA

OBJECTIVES: Our research using observational data is complex due to massive size and disparate formats. To consistently and efficiently apply a research methodology across disparate databases, facilitating consistent and efficient application of research methods. The objective of this study was to rapidly and efficiently estimate how well disparate sources reflect “known” population characteristics, utilizing standardized patient selection, analyses and visualization software on data that has been transformed into a CDM. METHODS: Using CEWorks® software, AF patients treated with warfarin or NOA were selected from multiple Administrative Claims and EHR databases previously transformed into a CDM. Rates of selected disease states were calculated using a CEWorks analysis module, and then compared to results published in a recent study. Results across disparate databases were imported into a visualization tool for further comparison among data sources. RESULTS: Preliminary results indicate that rates of selected disease state across disparate databases are similar to those published in a previous study. Total number and percentage reported in the study for NOAC 10,789 - 35%) and warfarin [19,964 - 65%] patients matched comparably at 8,093 [29%] NOAC patients and 20,133 [65%] warfarin patients in CDM version. Different data sources also show similar prevalence rates on selected disease state, although such similarities show some gaps introduced by region and other demographic variables, covering 100% of current Medicare spending.