EHR to effectively identify large numbers of patients interested in and eligible to enroll in a research study. The ability to more accurately identify patients who also satisfy the more “minor” inclusion/exclusion criteria would improve the efficiency and automation of this approach.

Keywords: Computer assisted recruitment, Electronic health record, Clinical trials

doi:10.3121/cmr.2010.943.ps1-26

PS2-01:

A System for Computer-Assisted Rapid Case Ascertainment of Breast, Lung and Colorectal Cancers

Roy Pardee, JD, MA; Group Health

Background and Aims: An intervention trial requiring recruitment of new cancer cases prompted this investigation into text classifier programs to identify pathology reports describing malignancies. Text classifiers are among the simplest Natural Language Processing methods, and are particularly useful for filtering large streams of documents. Our classifiers allowed study staff to concentrate their time on the reports most likely to result in recruitable subjects. The proposed presentation will describe our use of text classifiers as part of an efficient computer-assisted recruitment pipeline. **Methods:** We gathered corpora of pathology reports for breast, lung and colorectal tissue, each with a ‘gold standard’ assessment of whether they discussed a malignancy. We then randomly divided each into 75% training and 25% evaluation subsamples, trained the classifier on the training subsample and compared its categorizations to the gold standard in the evaluation subsample. We performed this training/evaluation cycle repeatedly, generating a distribution of predictive value statistics, thereby getting insight into the classifiers’ sensitivity to the particular random division of corpus reports. More importantly, it allowed us to experiment with tweaks to the classifiers and to the basic text processing that preceded them, and quickly evaluate whether they contributed to the predictive power of the classifier. **Results:** Several enhancements contributed to the accuracy of the system. In particular, taking multi-word phrases in addition to individual words as features increased the classifier’s predictive power. To our surprise, ‘stemming’ individual words (that is, reducing them to their root forms) did not have a perceptible effect. Once the classifiers were optimized, we were able to cover approximately 1600 path reports/week in approximately 30 minutes of a Research Specialist’s time. As an added benefit, when the RS found that a report was flagged in error, the classifier was trained on that report, thereby improving its future performance. **Conclusions:** Text classifiers are an effective tool for optimizing the use of staff time in rapidly ascertaining cancer cases for recruitment. Furthermore, because the classifiers’ training (including the corrections made during the study) is easily reduced to a file on disk, it becomes an independent asset, useful for future studies needing to do rapid ascertainment of cancer.

Keywords: Natural Language Processing methods, Text classifier programs, Rapid ascertainment of cancer cases

doi:10.3121/cmr.2010.943.ps2-01

PS2-05:

Does Socioeconomic Status Moderate the Association of Race with Morbidities During Pregnancy Episodes?

Douglas Roblin, PhD¹; Peter Joski, MSPH¹; Donald Bachman, MS²; Mark Hornbrook, PhD²

¹Center for Health Research Southeast; ²Center for Health Research Northwest

Background and Aims: African American adults generally have higher prevalence of morbidities than whites in the United States. Living in lower socioeconomic status (SES) neighborhoods increases the likelihood of poor health. The extent to which lower SES contributes to a higher rate of pregnancy complications for African Americans requires investigation. We explored racial differences in likelihood of maternal morbidities and whether these differences were moderated by neighborhood SES. **Methods:** Pregnancies were defined from linked birth certificates and hospital records.

We studied 9,726 pregnancies for non-Hispanic African American women and 10,272 for non-Hispanic white women enrolled with Kaiser Permanente Georgia (KPGA) from 2000-2006. Morbidities during pregnancy were ascertained from all outpatient and inpatient diagnoses and procedures classified by a previously validated maternal morbidity grouper. SES was measured from a factor analysis of 7 US Census measures associated with tract or block group of residence. Factor scores were divided into quartiles. Likelihood of selected morbidities was estimated using logistic regression adjusted for race, SES quartile, and mother’s age. Parameter estimates were adjusted for repeated observations (pregnancies) per woman during the study period. **Results:** 14% of pregnancies were characterized by anemia, 13% by hypertension, and 7% by mental disorders. African American women were more likely than white women to live in lowest SES quartile neighborhoods (34% vs. 16%; $p < 0.05$). Adjusted for age and SES, pregnancies characterized by anemia and hypertension were significantly more likely among African Americans than whites (OR = 2.24 and 1.30, respectively). Mental disorders were significantly less likely among African Americans (OR = 0.50). Adjusted for age and race, low SES neighborhoods had a significantly greater moderating effect among whites than among African Americans. Among whites, the adjusted ORs for the lowest SES quartile (vs. highest) were 1.21 (95% CI: 1.40, 1.01) for hypertension, 1.97 (95% CI: 2.40, 1.61) for anemia, and 1.27 (95% CI: 1.57, 1.03) for mental disorders. **Conclusions:** Neighborhood SES had a significant moderating effect on selected morbidities during live birth pregnancy episodes in this HMO. Compared to African American women, living in low SES neighborhoods significantly increased the risk of major morbidities during pregnancy episodes among white women.

Keywords: Maternal morbidities, Minority health disparities, Socioeconomic status and pregnancy

doi:10.3121/cmr.2010.943.ps2-05

PS2-06:

Feasibility of Determining Medication Adherence from Electronic Health Record Medication Orders

Nikki Carroll, MS¹; Capp Luckett, BA¹; Jennifer Ellis, MSPH¹; Marsha Raebel, PharmD¹

¹Kaiser Permanente Colorado

Background and Aims: Medication adherence is essential to optimizing outcomes for many chronic diseases. Typically, calculation of adherence is based on dispensings identified from pharmacy claims databases. However, this procedure excludes prescription orders that are never dispensed and potentially results in biased adherence estimates. To explore extracting newly-ordered medications from the Electronic Health Record (EHR), we conducted a project to: 1) Determine the process needed to identify ordered medications from the EHR, 2) Apply the process to identify a study cohort, and 3) Validate the process with medical record review. **Methods:** The study team prepared an inclusive list of study drugs, identified EHR tables and fields that contained medication order data, explored methods that facilitated inclusion of only the final medication order (when applicable), and explored optimal sequencing of inclusion criteria. These four steps were then used to identify a patient cohort newly-prescribed an antihyperlipidemic agent, and record review was used iteratively to validate each process step. **Results:** Multiple challenges were encountered and solutions identified. First, drug orders in the EHR were based on an unfamiliar drug categorization scheme. Second, Clarity tables, fields and linkages required exploration to discern useful data elements. For example, the “stop reason” field was not useful in identifying cancelled orders and the “order date” was occasionally overridden by the most recent refill of that order. Other factors complicating identifying new orders included human error, reversed orders, sending orders to incorrect dispensing locations, and medication reconciliation after nursing home discharge. We also found that by changing the order in which we applied the selection criteria resulted in a “cleaner” data set. For example, applying diagnosis criteria earlier in the process allowed us to remove prevalent users and their refill history. **Conclusion:** Identifying medications ordered in the EHR, particularly medications ordered but not dispensed, was a process fraught with multiple challenges. We believe, however, that we have

developed an iterative process that yields a reliable medication order algorithm. This is an important first step in achieving less biased estimates of medication adherence in chronic disease populations.

Keywords: Medication adherence, Electronic Health Record, Medication algorithm orders

doi:10.3121/cmr.2010.943.ps2-06

PS2-09:

Relation of Practice Level Blood Pressure Management to Patient Cardiovascular Events: The Blood Pressure Disease Management Index

Thomas Vogt, MD, MPH¹; Andrew Williams, PhD¹; Adrienne Feldstein, MD¹; Chuhe Chen, PhD¹; Mark Schmidt, BS¹; Richard Meenan, PhD¹

¹Kaiser Permanente

Background and Aims: Despite the proliferation of care guidelines, widespread practice variation is common. Longitudinal electronic health records (EHR) provide an opportunity to observe how care variations are related to care outcomes. This study uses blood pressure management to illustrate the uses of the Disease Management Index (DMI) to relate care quality to cardiovascular outcomes. **Methods:** The Blood Pressure Disease Management Index (BP-DMI) incorporates the level of control of BP and the time out of control into a single index scored from 0 to 100. The practice level DMI for a given year is the average DMI for all diagnosed hypertensive patients in the practice. We observed practice level BP-DMI scores for 199 to 374 primary care physician (PCP) practices for the years 1998-2007 (Oregon) and 2005-2007 (Hawaii) and related those scores to annual incidence rates of myocardial infarction (MI), stroke, and congestive heart failure (CHF) hospital days. **Results:** Systolic BP (SBP) control was much worse than diastolic BP (DBP) control, particularly before 2005, but improved steadily from that point on. Practice variations in BP control were substantial. Propensity scores were used to adjust for the substantial differences in patient risk across practices. On average, clinicians treating elevated DBP had patients who were younger and healthier than those treating elevated SBP. Propensity adjusted incidence rates of MI, stroke, and CHF hospital days were strongly associated with both systolic (S) and diastolic (D) BP-DMI scores ($p < .0001$ to $< .02$). The estimated effect of increasing the SBP and DBP DMI scores from the 50th percentile to the 60th percentile of clinical practices for MI and stroke incidence, and CHF hospitalization was:

SBP: -7.2%, -4.6%, -8.4%, respectively
DBP: -17.6%, -20.1%, -20.0%, respectively.

Conclusions: Practice level BP-DMI scores over a ten-year period are strongly related to the incidence of cardiovascular events including hospital days for CHF. The DMI is an effective tool for relating practice patterns to morbidity risk and health care utilization.

Keywords: Blood pressure management, Disease Management Index, Health outcomes

doi:10.3121/cmr.2010.943.ps2-09

PS2-10:

The Potential Use of Area-Based Population Characteristics for Targeted Selection and Balancing of Recruitment into Health Literacy Interventions

Douglas Roblin, PhD¹; Josephine Calvi, MPH¹; Brandi Robinson, MPH¹; Kathleen Mazor, EdD²

¹Kaiser Permanente Georgia; ²University of Massachusetts Medical School

Background and Aims: Individual-level socioeconomic are rarely available for enrollees of health care systems in the US. Research to selectively target participation from, or to balance recruitment across, specific socioeconomic subgroups must rely on other measures. We used of area-based measures from US Census tract records to target participant recruitment by level of formal education and race into a health literacy study. Our objectives were to obtain balanced proportions of participants: 1) with a high school (HS)

education or less, and 2) who were African American (AA) or white. **Methods:** Kaiser Permanente Georgia (KPG) is one of 3 sites participating in an NCI-funded study to develop and evaluate a computer-based assessment of health literacy. At KPG, the goal is to recruit 300 participants from enrollees 25-74 years of age who were geocoded to their US Census tract using residential addresses (N=185,150). Census tracts were linked to US Census SF-3 data to obtain percent AA residents and percent of adults with a HS education or less in each tract. Approximately 400 participants were randomly selected from each of 9 strata defined by high, moderate, and low tertiles of percent AA residents and, within each tertile of AA, into tertiles of percent of adults with a HS education or less. **Results:** 189 assessments have been completed among 1,947 invited to date. 58.7% (N=111) of participants were AA; 34.9% (N=66) were white. Self-reported race and formal education of participants paralleled the percent AA residents and percent adults with a HS education or less in the US Census tracts. By race tertiles, 85%, 57%, and 16% of participants were AA compared to 86%, 24%, and 4% of US Census tract residents. By education tertiles, 33%, 19%, and 10% of participants had a HS education or less compared to 56%, 32%, and 22% of US Census tract adults. **Conclusions:** Initial results indicate this sampling strategy is achieving study goals of recruitment of proportional numbers of AA and white participants and, within race, adequate numbers with a HS education or less. Area-based measures of race and education may facilitate targeted recruitment in the absence of individual-level data.

Keywords: Computer-based assessments, Health literacy interventions, Targeted recruitment

doi:10.3121/cmr.2010.943.ps2-10

PS3-03:

Creation and Validation of the Electronic Medication Possession Ratio (eMPR)

Joseph Leader, BA¹; G. Craig Wood, MS¹; Nirav Shah, MD, MPH¹; JB Jones, PhD, MBA¹

¹Geisinger Health System

Background and Aims: Medication adherence studies usually rely on a narrow aspect of a patient's medication history. Studies using administrative claims data such as insurance pharmacy claims cannot compute adherence for patients who don't fill their medication orders or who use spousal benefits. Studies using medication orders from the electronic health record (EHR) assume that all patients fill their medication orders. Linking EHR medication orders to pharmacy claims can create a more robust picture of a patient's medication history. The aim of this study was to create an algorithm for linking EHR orders to pharmacy claims to create a new metric of medication adherence – the eMPR (electronic Medication Possession Ratio). **Methods:** Patients included in this study had a Geisinger primary care clinic provider while enrolled with the Geisinger Health Plan. Overlapping information across the EHR medication order database and the pharmacy claims database were used to identify likely matches between a prescription and subsequent fill. The overlapping data includes medical record number, order and fill dates, days supply, and National Drug Code (including medication name and dose). To evaluate the usefulness of the algorithm, the percent of claims that matched to an order was calculated. **Results:** The linking algorithm is currently being tested as part of ongoing research to be completed by the time of the HMORN Annual Meeting (results pending). Several challenges identified include: 1) Developing a method of linking EHR orders to pharmacy claims; 2) Creating a link for medication orders and multiple refills, as each generates unique pharmacy claims; and 3) Linking generic forms of the prescribed medication to name-brand prescriptions written. Initial results indicate a high degree of matching which fluctuates based on assumptions provided to the algorithm. **Conclusions:** Medication orders and pharmacy claims can be used to create a link between a patient's order and their pharmacy claim or "fill." The addition of medication orders to the HMORN Virtual Data Warehouse may be valuable for improving the quality of medication adherence research within the HMORN, and lay the foundation for an eMPR which can be widely used in many settings.

Keywords: Medication adherence, Computer-based medication orders, Electronic Medication Possession Ratio

doi:10.3121/cmr.2010.943.ps3-03