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, PhD[1]; Peter Joski, MSPH[1]; Donald Bachman, MS[2]; Mark Hornbrook, PhD[2]

[1]Center for Health Research Southeast; [2]Center 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, MS[1]; Capp Luckett, BA[1]; Jennifer Ellis, MSPH[1]; Marsha Raebel, PharmD[1]

[1]Kaiser 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 over-ridden 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