and those from people with comorbidities (cancer, chronic lung disease, or congestive heart failure). We estimated the sensitivity, specificity, and positive predictive value (PPV) of ONYX compared to manual review using multivariable logistic regression, accounting for clustering of reports within people. We examined how ONYX's performance varied by age and comorbidity. **Results:** ONYX classified 26% of reports (1,276/5,000; 38% [841/2,200] of true pneumonias and 16% [435/2,800] of non-pneumonias) as "requiring manual review" based on pre-defined criteria. Reports from older people were more likely to require manual review than those from younger people. Among reports that could be classified, ONYX had a sensitivity of 91% (1,242/1,359), specificity of 92% (2,170/2,365), and PPV of 81% (1,242/1,437, modeled based on pneumonia prevalence in the source database). Sensitivity and specificity were similar regardless of comorbidity. **Conclusions:** NLP offers potential for identifying pneumonia outcomes from EMR data. Next steps include 1) further training to decrease the proportion of reports requiring manual review and 2) evaluating the accuracy of ONYX in other health systems.

Keywords: Natural language processing, Pneumonia, Validity
doi:10.3121/cmr.2011.1020.plenary-iii-02

C-C4-02:

Using a Natural Language Processor to Remove All Elements of Personal Health Information (PHI) to De-identify Clinical Annotations for the Specimen Retrieval System (SRS)

Andrew Glass, MD¹; Sheila Taube, PhD²; Andrew McMurry, BA³; Chris Eddy, BS⁴; Pierre-Andre La Chance, BS⁴; Lisa McShane, PhD²; Mei-Yin Polley, PhD²; Isaac Kohane, MD, PhD³

¹Kaiser Permanente Northwest; ²National Cancer Institute; ³Harvard Medical School; ⁴Kaiser Center For Health Research

PHI in HMORN medical records must be systematically removed prior to use and sharing data with co-investigators at other institutions. The problem is particularly acute when one needs to link clinical and follow-up data with pathologic specimens from cancer patients in studies of prognostic and predictive tumor markers. Natural language processing (NLP) can be customized to identify and remove PHI in selected clinical records by substituting nonsense characters. Such a process of deidentification can proceed largely without human intervention and, if successful, can allow efficient linkage of clinical notes with similarly deidentified specimens for laboratory investigators at other institutions. The Shared Pathology Informatics Network (SPIN) was funded by the Cancer Diagnosis Program of the National Cancer Institute to develop a computerized program that would search pathology department text files and reports from several institutions and retrieve, in a database, all records that met search criteria. SPIN was designed to access information from each institution's electronic records without affecting the records and collating data from several sites into a single report. Our current project, the Specimen Retrieval System, requires the identification of large numbers of cancer specimens of specific types from Kaiser Permanente Northwest (KPNW) and, subsequently, other Cancer Research Network sites, retrieval of those specimens and linkage with clinical annotations that describe staging, treatment and outcome. We have used the existing databases of the KPNW Tumor Registry and Department of Pathology to identify cases. The EPIC electronic medical record and several other computerized text files provided the clinical notes for these patients. Using SPIN technology we then processed text files to 'scrub' them of PHI. All records were then manually inspected to assess the completeness of the process, which elements of PHI persisted and which had been successfully 'scrubbed' of PHI. We processed several hundred files from more than 100 patients and will report detailed statistical analysis. This technology has demonstrated significant capability to facilitate searches for pathologic specimens and clinical annotations using conventional reports from pathology departments. It is a valuable tool to remove PHI, deidentify medical records and ease sharing of clinical information with investigators at other sites.

Keywords: Personal Health Information, Deidentification of Clinical Annotations, Cancer Specimen Retrieval

doi:10.3121/cmr.2011.1020.c-c4-02

PS1-09:

Readmission and 30-day Mortality Among CAP Patients in Two Systems of Care

Laurel Copeland, PhD, MPH¹; Debbie Godwin, BS²; Dean Kjar, MS²; Daniel MacCarthy, BS³; Dennis Myers, MD²

¹Scott & White/Veterans Affairs ; ²Scott & White Healthcare; ³South Texas Veterans Health Care System

Objectives: Readmission and mortality among elderly patients admitted to inpatient care are critical outcome measures for healthcare systems. While the Veterans Health Administration (VA) has long made national administrative databases available for health services research, the HMORN Virtual Data Warehouse (VDW) at Scott & White (SWHP) has only been developed in the past year. The current study sought to utilize both data sources while examining outcomes of hospitalization for community-acquired pneumonia (CAP) in two systems of care in the same geographic region. **Methods:** Patients >65 years of age hospitalized with CAP (ICD9 codes 480-483, 485-487) in Central Texas in VA fiscal year 2009 (Oct 2008-Sep 2009) were identified. Dates of care and inpatient death were sought to determine outcomes subsequent to CAP admission. Covariates included sociodemographic measures (such as age, gender) and comorbidity. **Results:** There were 469 SWHP CAP patients and 152 VA CAP patients admitted in the year studied. The SWHP fiscal year differed from the VA fiscal year, potentially complicating the work of the SWHP data team. VA CAP patients averaged 77.2 years (SD 7.6), were 96% male, and experienced an inpatient death rate of 18%. SWHP CAP patients averaged 81.6 years (SD 8.2), and 56% were women. Date of death as an inpatient was not explicitly available in VDW data. Patients in the two systems varied significantly as VA patients were more likely to be male and younger. **Conclusions:** Patients in the federal VA system can be studied in conjunction with private-care patients in the same catchment area using newly available VDW data. The data presented required two teams of data analyst/programmers, one working within the VA system and the other working within the SWHP VDW system. Coordination was facilitated by the VDW Site Director Godwin working together with the jointly appointed VA/SWHP investigator Copeland. Sharing SAS code between VDW and VA programmers helped to lend consistency to the data analytic approaches across systems. Adjusted analyses employing high-dimensional propensity scores could be employed to assess whether inter-system differences in CAP-related outcomes can be validly estimated.

Keywords: Community acquired pneumonia, Healthcare systems, Epidemiology

doi:10.3121/cmr.2011.1020.ps1-09

C-B4-02:

Enhancing the Quality and Efficiency of Obstructive Sleep Apnea Screening Using Health Information Technology: Results of a Geisinger Clinic Pilot Study

Robert Moser, RRT, MHA¹; Elizabeth Boyer, BA¹; Diana Lupinski, RN¹; Jonathan Darer, MD, MPH¹; Tammy Anderer, PhD, CRNP¹; Alexander Villareal, MD¹; Peter Berger, MD¹

¹Geisinger Health Systems

Background: Health information technology (HIT) can improve the screening and delivery of care to patients with chronic illnesses. To date there are no published data describing the use of HIT for screening patients with obstructive sleep apnea (OSA), a prevalent under-diagnosed chronic condition representing a substantial and growing socio-economic burden. Geisinger Clinic is a physician-led, multi-specialty group practice in rural Central Pennsylvania with integrated electronic medical records. **Aim:** We sought to determine whether HIT could be used to identify OSA amongst patients at risk for the disease through the use of a simple questionnaire and a physician approved referral protocol. **Methods:** Using an internet-based portal through which patients can review their Electronic Health Record (EHR), schedule appointments, and receive communication, patients with a body mass index

(BMI) >40 and without a diagnosis or prior evaluation for sleep apnea were invited to complete the “STOP” questionnaire, a self-administered validated screening tool consisting of 4 “yes/no” questions; “yes” answers to 2 questions identify individuals at high risk of having OSA. Patients who screened positive were then contacted to schedule a sleep medicine evaluation. **Results:** Electronic invitations were sent to 2,283 eligible patients to take the questionnaire; 279 (12%) completed and submitted their responses. Based on their answers, 122/279 (44%) were identified as being at high risk for having OSA and recommended to undergo evaluation in the sleep clinic. To date, 104/122 patients (85%) were offered an appointment with a sleep specialist; the remaining 18 (15%) are waiting scheduling. Of those offered a clinic appointment, 41/104 (40%) have undergone a sleep medicine evaluation, 38/104 (36%) are waiting to be scheduled, and 25/104 (24%) cancelled or did not keep their appointment. Of the evaluated patients, 38/41 (93%) were diagnosed with OSA, 2/41 (5%) had negative sleep studies, and 1/41 (2%) are awaiting polysomnographic evaluation. The positive predictive value of the STOP questionnaire among patients who underwent polysomnography was 95%. **Conclusions:** This pilot study suggests that utilizing an internet-based patient portal via the EHR can identify patients at high risk of OSA and facilitate their ultimate evaluation and diagnosis through the administration of a simple questionnaire.

Keywords: Obstructive sleep apnea, Sleep apnea screening
doi:10.3121/cmr.2011.1020.c-b4-02

PS1-20:

Developing an Ontology: Informatics for Integrating Biology & the Bedside (I2B2) and Structuring the Virtual Data Warehouse (VDW) Content Areas

Dustin Key, MAS, BA¹; Gene Hart, MS¹; Garth Arnold, BA¹; David Eastman, BS²

¹Group Health Research Institute; ²Kaiser Permanente Southeast

Background/Aims: Informatics for Integrating Biology & the Bedside (I2B2) offers a dynamic graphical interface for querying VDW data. Part of the I2B2 effort involves designing the drop-down menu of standardized nomenclatures provided by coding schemes such as CPT-4, ICD-9, and RxNorm, by which the user queries the data. This menu of content hierarchies can be complex in terms of the number of nodes and levels of nested folders. In addition, there often isn't an established way to organize these codes into categorization scheme that allows for a clean and organized querying interface. **Methods:** We will demonstrate the use of SAS and SQL to craft these hierarchies and how to leverage the resources provided by the Unified Medical Language System (UMLS) along with resources that may be available at your local site. **Results:** We have been able to develop an ontology that spans many of the VDW content areas including Procedures, Diagnoses, Pharmacy Fills, Vitals, Demographics, Tumor, and Enrollment. **Conclusions:** It is possible to build an ontology that organizes the various coding schemes using SAS and SQL, sourcing public resources such as UMLS. Therefore, our knowledge about building an ontology can be readily transferred across HMORN sites and other consults, empowering additional I2B2 deployments. The development of an ontology for VDW subject areas may have additional uses insofar as organizing standard nomenclatures in a programmatic way is useful.

Keywords: Ontology, I2B2, SAS
doi:10.3121/cmr.2011.1020.ps1-20

PS1-32:

Psychology in Primary Care: An Evaluation of Best Practices

Karon Phillips, PhD, MPH¹; Emily Smith, MSPH¹; Alan Stevens, PhD¹

¹Scott & White Healthcare

Background/Aims: Integrating psychology and mental health professionals into primary care settings has emerged as a means to improve the access to and utilization of mental health services. Three main delivery models of psychology in primary care settings have emerged: (1) referring a patient to

a psychologist/mental health professional located in a facility outside of the primary care physician's (PCP) office; (2) referring a patient to a co-located psychologist/mental health professional who does not directly interface with PCP; (3) integrated and co-located model where the PCP and the psychologist/mental health professional discuss the patient's health. The overall purpose of this project was to compare patient utilization of psychology/mental health services across the Scott & White Health care system. **Methods:** Using electronic medical records and the virtual data warehouse (VDW), mental health clinics/facilities were categorized into one of the three main psychology models. Patients that had depression, anxiety or ADHD DRG codes in their EMR and were aged 18 and older (n=37,310) were included in the analysis. The following additional variables were controlled for: gender, race/ethnicity, and chronic physical health conditions (i.e. arthritis, lung disease, heart disease, diabetes, hypertensive disease, and osteoporosis). ANOVA/ANCOVA analyses were performed to determine the differences across the three models in the length of time between the PPC referral and the first appointment with the psychologist/mental health professional. **Results:** Patients that were seen by psychologist in facilities with integrated co-located models of care experienced a shorter amount of time between their referral and their first appointment with the psychologist/mental health professional, compared to the other two models of care. **Conclusions:** These initial results begin to inform best practices for delivering mental health services within primary care and provide physicians and health care systems with data on issues that facilitate integrated, person-centered care.

Keywords: Psychology, Best practices, Primary care
doi:10.3121/cmr.2011.1020.ps1-32

PS1-28:

The Grid-Enabled Measures (GEM) Database: A Research Tool to Facilitate the Use of Standardized Measures and Data Sharing

Richard Moser, PhD¹; Min Hua, PhD²; Paul Courtney, MS³; Dianne Reeves, RN, MSN¹; Riki Ohira, PhD⁴; Abdul Shaikh, PhD, MHS¹; Bradford Hesse, PhD¹

¹National Cancer Institute; ²Fox Chase Cancer Center; ³SAIC-Frederick Inc/ National Cancer Institute at Frederick; ⁴Booz Allen Hamilton

Background: We are in an era of unprecedented access to data and the potential to move areas of science forward rapidly but currently there is a lack of coordination of research efforts in regards to: 1) Using standardized measures; and 2) sharing harmonized data. The Grid-Enabled Measures (GEM) database has been conceptualized by the National Cancer Institute (NCI) to help accomplish these goals. **Methods:** GEM encourages and enables standardization of measures and data harmonization by supporting a virtual community of researchers who interact with each other using Web 2.0 technologies. Through GEM, researchers are able to upload health-related measures and associated meta-data for use by others and provide comments and ratings about these measures and their associated theoretical constructs. Input from users—including ratings, usage statistics and psychometric properties of the measures—established criteria to prioritize measures to be curated for use on the Cancer Biomedical Informatics Grid (caBIG®). Curation starts with models that are used to represent the measure and its items in a form understandable to both researchers and informaticians. The outcome of the curation process is registration of a measure as a caBIG® common data element so that the variable representing the measure's score can be shared via the grid. **Results:** Models for measure curation have now been developed for five measures in GEM (e.g., Center for Epidemiologic Studies Depression Scale (CESD)). GEM currently has 216 registered users and 110 measures based on 73 constructs and these numbers will change over time with user interaction. Next steps in developing GEM include making datasets—based on GEM-curated measures—available on the grid for use by other systems and users connected to caBIG®. **Conclusions:** GEM represents a critical step in advancing the research enterprise and complements other NIH-funded systems being developed such as Phen-X (consensus measures for Phenotypes and eXposures) which seeks to build consensus measures used in genome-wide association studies. In addition to describing the rationale and development of GEM and how measures and data are curated into caBIG®, this presentation will include a live demonstration of the site