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, MD[1]; Sheila Taube, PhD[2]; Andrew McMurry, BA[3]; Chris Eddy, BS[4]; Pierre-Andre La Chance, BS[4]; Lisa McShane, PhD[2]; Mei-Yin Polley, PhD[2]; Isaac Kohane, MD, PhD[3]

[1]Kaiser Permanente Northwest; [2]National Cancer Institute; [3]Harvard Medical School; [4]Kaiser 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, MPH[1]; Debbie Godwin, BS[2]; Dean Kjar, MS[2]; Daniel MacCarthy, BS[3]; Dennis Myers, MD[2]

[1]Scott & White/Veterans Affairs ; [2]Scott & White Healthcare; [3]South 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, MHA[1]; Elizabeth Boyer, BA[1]; Diana Lupinski, RN[1]; Jonathan Darer, MD, MPH[1]; Tammy Anderer, PhD, CRNP[1]; Alexander Villareal, MD[1]; Peter Berger, MD[1]

[1]Geisinger 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