Health Informatics

C3-1:

Creation of Deidentified Test Datasets from the VDW for Code Development Prior to IRB Approval

Jenny Staab¹; Donald Bachman¹; Alan Bauck¹; Maria Faer²; David Gary²; Pierre-André La Chance¹

¹Kaiser Permanente Northwest; ²Kaiser Foundation Research Institute

Background/Aims: The turn-around time of data-only studies could be shortened by about a month if programmers were to start developing extraction and analytic code before official IRB approval has been granted. However, while code development is likely more successful and efficient if the programs can be checked against data, fully identified datasets like the VDW cannot be accessed without the proper authorizations. A solution to this dilemma is the creation of deidentified test datasets. These test datasets are based on the VDW and preserve its structure and much of the richness of the data. Yet, due to sufficient deidentification, their use for code development falls outside of the purview of the Privacy Rule's research provisions and does not qualify as Human Subjects research, so IRB approval is not required. **Methods:** We created VDW based test datasets for 25,000 randomly selected individuals. Only patients up to the age of 90 and within the mid 90% of the age-specific height and weight distribution were eligible. All identifiers, such as MRN, encounter ID, provider ID, and facility code, were replaced with randomly assigned study identifiers, and crosswalks from VDW to study identifier were destroyed immediately after test dataset creation. The same study identifier was used for a given VDW identifier across all tables. Every date was shifted by a specific number of days that varied randomly across individuals, but was consistent for all dates associated with one person. Additional deidentification measures such as random sorts and grouping into larger categories were also applied. **Conclusions:** A combination of methods serves to sufficiently deidentify a group of test datasets that can be used for code development while awaiting IRB approval for the project. The employment of consistent study identifiers across datasets and date shifting by a person-specific constant preserves important relationships across content areas and time. Use of these deidentified test datasets allows for programming to begin as soon as the study population and protocol are sufficiently defined. This, in turn, enhances research efficiencies as data analysis can begin sooner, data-driven decisions can happen earlier, and studies can be completed more quickly.

Keywords: Deidentification; VDW; Compliance

doi:10.3121/cmr.2013.1176.c3-1

C3-3:

Using the Electronic Health Record and the National Cancer Institute's Breast Cancer Risk Assessment (SAS) Macro to Identify Women at Increased Risk for Breast Cancer

Joseph Leader¹; Amanda Bengier¹; Jonathan Darer¹; Azadeh Stark¹; Victor Vogel¹

¹Geisinger Health System

Background/Aims: Women at increased risk for breast cancer (BC) are eligible to take selective estrogen receptor modulators (SERMs) to reduce their risk; Food and Drug Administration (FDA) approval of tamoxifen or raloxifene for BC risk reduction and American Society of Clinical Oncology guidelines for the use of SERMs recommend the two drugs for any woman over the age of 35 years with a 5-year risk of 1.67% or greater, but identifying those women can be both challenging and costly. Fortunately, the National Cancer Institute (NCI) has developed an open source Breast Cancer Risk Assessment Macro (BrCa RAM) that can be run using SAS software. By leveraging the Geisinger Health System (GHS) Electronic Health Record (EHR - EpicCare), the Department of Radiology's software (Centricity RIS-IC), and Department of Pathology's software (CoPath), we were able to calculate 5-year and lifetime risk of developing invasive BC. **Methods:** BrCa RAM calculates risk based on patient age, number of biopsies, if a biopsy ever displayed atypical hyperplasia (Yes/No), age at menarche, age at first live birth, number of first degree relatives with breast cancer, and patient race. We were able to extract and format these elements from EpicCare, RIS-IC, and CoPath. Demographic information (age, race, sex) was obtained from

EpicCare, pathology information (number of biopsies, atypical hyperplasia) was obtained from CoPath, and personal history (number of first degree relatives with breast cancer, age at menarche, and age at first live birth) was obtained from RIS-IC. **Results:** We found 91,692 women between the ages of 35-90 in the RIS-IC database who had ever received a screening mammogram. We identified 9,021 patients with a calculated 5-year breast cancer risk 2% or greater and the mean age was 59.7 years. The numbers of patients by 5-year risk score category were: 2-2.5% (n = 3,551); 2.5-3% (n = 1,946); 3%+ (n = 3,524). **Conclusions:** The BrCa RAM is a powerful tool that enabled GHS to calculate breast cancer risk for our entire population. Using this macro, we were able to identify patients for prophylactic SERM treatment which can potentially prevent or delay a woman's risk of developing BC.

Keywords: Electronic Health Record; Breast Cancer Risk; SAS

doi:10.3121/cmr.2013.1176.c3-3

C3-4:

Primary Care Physician Efficiency in Handling Patient Secure Email and the Impact on Patient Communication and Access

Di Meng¹; Terhilda Garrido¹

¹Kaiser Permanente Program Offices

Background/Aims: All Kaiser Permanente regions have implemented secure email functionality. Regional policies, workflows, and provider best practices have been established regarding secure email encounters with patients. This Phase I study explores effective practices of primary care physicians (PCPs) handling email traffic with patients, provides evidence regarding current secure email workflow and volumes, and estimates physician workloads related to daily secure email encounters with patients. **Methods:** A combination of methods was used: 1. Structured interviews with 27 PCPs with a high volume of secure email encounters with patients; 2. Descriptive and correlation analysis of daily secure email volume and provider response time; 3. Estimation of actual physician workloads related to secure email with patients, based on observed volume and time spent per email; and 4. Approximate randomization to estimate expected workload adjusted by PCP panel size, kp.org registration rate, and number of clinical work days. **Results:** Preliminary results of this Phase I study indicate that, before adjusting for panel size, regional averages for physicians' daily volume of secure email encounters with patients ranged from 1 to 9. Few physicians have high volumes (10-20 per day) of email encounters with patients. Interviews with 27 high-volume physicians revealed two models for managing email traffic: direct physician response and team triage (74%). Some doctors respond quickly to minimize repeat emails/phone follow-ups from patient; others fear that rapid responses will generate higher email volume. First-response times are correlated with email volume in some regions, but average physician response times are not correlated with the volume of email encounters. Preliminary results from approximate randomization provided insight into associations between physician email encounter practices and patient satisfaction with access and communication. **Conclusions:** Practices for handling secure emails with patients varied across physicians and regions. The data did not confirm all high-volume physician perceptions regarding relationships between volume and turnaround times. The majority of physicians was not overwhelmed by the workload related to secure email with patients, and patient satisfaction was associated with higher email volumes.

Keywords: Secure Email; Efficiency; Primary Care Physician

doi:10.3121/cmr.2013.1176.c3-4

C3-5:

Patient Use of a Secure Web Portal and LDL in Patients with Diabetes

Jie Huang¹; Ilana Graetz¹; Richard Brand²; John Hsu³; Mary Reed¹

¹Kaiser Permanente Northern California; ²University of California, San Francisco; ³Massachusetts General Hospital

Background/Aims: Patient use of web portals to interact with their healthcare delivery system and healthcare providers could improve the quality and safety of care. Among patients with diabetes in a large integrated delivery system (IDS), we examined the association between patient use of

the web portals and cholesterol test results. **Methods:** The health system implemented a web-based tool for all patients who registered to use the website in November 2005, allowing members to securely access a personal health record, as well as e-mail their physicians, and view their lab results. In this study, we defined patients as web-portal users when patients emailed their physicians or viewed lab results at the first time. We examined the association of patient web-portal use and low-density lipoprotein cholesterol (LDL) level using linear regression with fixed effect at patient level, adjusting for medical center electronic health record (EHR) implementation, quarter for seasonality, and year for temporal trend. **Results:** The 169,711 patients in the IDS diabetes registry at the start of 2004 were followed through 2009. The number of patients who had used patient web-portal increased dramatically from 16% in 2006 to 35% in 2009. During 2004-2009, a total number of 1,070,856 LDL tests were performed among the study subjects and 18% of the tests were done after patients used the web-portal. Overall patient web-portal use was associated with reduction of LDL value by 0.81 mg/dL (95% CI: 0.64-0.97). Further examination among patients stratified by their baseline LDL (last value in 2003) showed that the largest reduction in LDL was found among those with worst control: on average, LDL level dropped by 0.36 mg/dL (95% CI: 0.14-0.57) among those with baseline LDL <100 mg/dL, 0.90mg/dL (95% CI: 0.63-1.18) among those with baseline LDL 100- <130mg/dL, and 2.04 mg/dL (95% CI: 1.59-2.50) among those with baseline LDL >130 mg/dL. **Conclusions:** Patient use of a web-based portal to review laboratory results or email their clinicians increased substantially between 2006 and 2009. Patient use of web portals was associated with improvement of LDL level, with greater improvement among patients in worse control.

Keywords: Patient Portal; Diabetes; LDL

doi:10.3121/cmr.2013.1176.c3-5

PS2-5:

Prevalence of Migraine in a Diverse Community – Electronic Methods for Migraine Ascertainment in a Large Integrated Health Plan

Alice Pressman¹; Alice Jacobson¹; Andy Avins¹

¹Kaiser Permanente Northern California

Background/Aims: Migraine, a common neurological disorder, is among the top 20 causes of disability worldwide. One-year prevalence of migraine in the US is estimated to range from 8-15%, with women affected at approximately three times the rate of men. Current methods for migraine ascertainment from headache clinics and surveys are costly, affected by clinic-ascertainment bias, and do not work in situations where patient contact is not feasible. With the growing use of electronic medical records (EMR), new methods must be developed for identifying and tracking migraine prevalence over time. We sought to develop an EMR algorithm, to identify migraine, and to characterize its prevalence in Kaiser Permanente Northern California (KPNC). **Methods:** From EMRs of all KPNC members, we collected all outpatient migraine diagnoses (ICD9 code 346.xx) and migraine-specific prescriptions (ergots, triptans, and acetaminophen with caffeine) for 2006-2010. We chart-reviewed a random sample to develop an electronic Migraine Probability Algorithm (score 0-100). We tested the algorithm in a second independent chart review. Using membership data, we calculated prevalences by age, race, and gender. **Results:** We identified 313,174 KPNC members with evidence of migraine - 233,620 women and 79,554 men. The 5-year period-prevalence of migraine among KPNC adults was 17.1% for women and 5.9% for men. Among children, rates did not differ by gender (<2%) until the age of 10, when prevalences were higher - 5.8% for girls and 3.5% for boys. For women, prevalence peaked at ages 25-29. In contrast, males experienced flat prevalence with age (range 5%-6%). Overall, Whites had higher prevalence than Asians, but Blacks did not differ appreciably. **Conclusions:** We used EMR data to capture migraine diagnoses and show prevalence patterns similar to those reported in the literature. Prevalence of diagnosed migraine in KPNC was 2.5-3 times higher in women than men; migraine peaked with age in women, but remained flat for men; and prevalence of migraine among Asian adults was roughly 2/3 that of Whites. These methods for ascertainment of migraine are inexpensive and easy to implement and have applications and implications that extend well to other institutions and debilitating pain conditions.

Keywords: Electronic Medical Records; Migraine; Epidemiology

doi:10.3121/cmr.2013.1176.ps2-5

PS2-6:

Who Opens Alerts to Physicians? (And Who Doesn't?)

Terry Field¹; Hassan Fouayzi²; Shawn Gagne²; Sarah Cutrona¹; Jessica Ogarek²; Devi Sundaresan²; Lawrence Garber²

¹University of Massachusetts Medical School; ²Meyers Primary Care Institute

Background/Aims: With the increasing adoption of electronic medical records (EMR), there are many opportunities to implement systems of information flow, like alerts to providers. Tests of such systems have varied results. **Methods:** Within a multispecialty group practice that uses Epic EMR, we conducted a trial of automated alerts to primary care physicians and their staff when older patients were discharged from hospital to home. We generated alerts for new medications, drug interactions, dose problems, or monitoring needs. Provider staff received alerts to schedule an office visit. Alerts appeared in the recipient's Epic in-basket. Using EMR "digital crumbs", we tracked the length of time before alerts were viewed. We analyzed the impact of physician age, gender, department, and employment status (full-time, part-time). We also analyzed the impact of patient conditions, including age, gender, comorbidity, and number of office visits in the previous year. **Results:** Of 763 alerts to physicians, 616 (81%) were opened within one day. Characteristics associated with timely opening were age <50 (OR 1.7, 95% CI 1.1, 2.6) and full-time employment (OR 2.9, 95% CI 1.6, 5.2). Of 1928 alerts to staff, 1173 (61%) were opened within one day. Physician characteristics associated with staff opening of alerts were gender (staff of male physicians were more likely to open the alerts within one day [OR 1.8, 95% CI 1.4, 2.4]), and being a member of the Family Medicine department (OR 1.9, 95% CI 1.3, 2.6) or sub-specialty department (OR 16.6, 95% CI 2.3, 122.3). Staff of full-time physicians were less likely to open alerts within one day (OR 0.64, 95% CI 0.47, 0.87). Controlling for patient characteristics had no impact on results. **Conclusions:** Special efforts may be required to reach physicians working part-time and older physicians. The characteristics related to staff opening of alerts are specific to this group practice, but the high level of variability across physician types and departments is likely to be an issue in many settings. Design of a system directed at reaching staff quickly may require in-depth assessment of work flow and communication patterns in clinical departments.

Keywords: Health Information Technology; Electronic Medical Record; Alerts

doi:10.3121/cmr.2013.1176.ps2-6

PS2-7:

Automated Ordering and Sample Collection to Leverage Electronic Medical Record Based Genetic Research – The Geisinger MyCode Project

Samantha Avellino¹; Ryan Colonie¹; David Carey¹

¹Geisinger Health System

Background/Aims: The Geisinger MyCode project is a system-wide biobanking project with DNA and serum samples from over 35,000 Geisinger patients. We summarize an automated process for ordering and collection of samples linkable to EMR data for research studies. **Methods:** Eligibility lists for enrolling patients of interest are automatically generated using inclusion criteria and daily clinic schedules. Research staff use eligibility lists to consent patients to participate in MyCode during a regularly scheduled clinic visit. An opt-in consent process is used. After consent, an automatic blood order is placed in the patient's EMR. Electronic orders are executed by a clinical phlebotomist at the time of a clinical blood draw. Samples are stored using a unique study ID, but remain linkable to the patient's EMR data through a data broker process. De-identified clinical data is obtained from Geisinger's enterprise data warehouse to create phenotypes for genetic research. **Results:** The MyCode biobank includes over 35,000 participants and more than 104,000 samples. 85% of patients approached give consent to participate. 90% also agree to be recontacted to participate in future research studies. The standing research blood order allows for collection of serial samples. The retention rate for participation is 99%. More than 24,000 MyCode samples have been used in over 20 research projects. In the past 2 years, externally funded research projects that utilize the MyCode