Health Informatics

C3-1:

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

Jenny Staab[1]; Donald Bachman[1]; Alan Bauck[1]; Maria Faer[2]; David Gary[2]; Pierre-André La Chance[1]

[1]Kaiser Permanente Northwest; [2]Kaiser 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 efficiences 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 Leader[1]; Amanda Bengier[1]; Jonathan Darer[1]; Azadeh Stark[1]; Victor Vogel[1]

[1]Geisinger 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 Meng[1]; Terhilda Garrido[1]

[1]Kaiser 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 Huang[1]; Ilana Graetz[1]; Richard Brand[2]; John Hsu[3]; Mary Reed[1]

[1]Kaiser Permanente Northern California; [2]University of California, San Francisco; [3]Massachusetts 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