also hope to save other sites from having to spend as much time doing the initial analysis work that KPNC did. **Conclusions:** For KPNC, pulling chemotherapy infusion data was eye-opening. The complexity of the data was significant, particularly linking the data from the initial "planned" treatments to orders, episodes, flowsheets, dispensing, mixing the medications, and finally to what is administered. Often times treatment plans were changed along the way due to chemotherapy reaction or toxicity, and linking the data back to the original plan was therefore complicated. We will review what we learned in this process with other users, so that they may benefit from KPNC's experience.

Keywords: Chemotherapy Data; Infusion Data

doi:10.3121/cmr.2013.1176.a4-3

A4-4:

Visualizing VDW Lab Results Data: Why You Should, and How You Can - Easily!

Roy Pardee[1]

[1]Group Health

Background/Aims: One of the most difficult tasks facing the VDW lab results file implementer is figuring out which records from local lab data should be included in their VDW file, and which should be left out. These decisions frequently hinge on neat points of clinical science that are often outside the expertise of the programmer. We will describe a set of graphics that implementers can use to shed light on these decisions, useful during implementation, and afterwards as indication of data quality and even as documentation. **Methods:** SAS' new Statistical Graphics procedures allow unprecedented control and ease of use in the creation of descriptive graphs and charts. The ODS Graphics Designer utility, paired with the SGDESIGN procedure make it easy to create a single image composed of multiple different graphs, each of which can use its own dataset. These tools allowed us to create something of a "data report card" for each VDW test_type, depicting: (1) number of result records over time; (2) number of result values that are numeric vs. character, stratified by the local lab codes used; 3) distributions of numeric result values, by unit; (4) number of numeric result values, by unit and local lab code; (5) values of character results by local lab code. The report card is produced by a standard VDW program available from the author, which can be run at any VDW lab results-implementing site with access to SAS version 9.2 or greater. **Results:** The graphics produced by this program allow both implementers and end-users to evaluate at a glance, how cohesive the data from various different local codes are, how much data there is, how it waxes and wanes over time, whether the values are of the expected types, and whether units and character values are within valid values. Being able to depict all of this disparate information in a single, compact display allows users to glean insights that, for example, viewing series of larger graphs separately would not afford. **Conclusions:** The author has found this program an extremely useful tool in implementing VDW labs by quickly pinpointing misclassified local codes. Other sites surely will also.

Keywords: Graphics; Laboratory; SAS

doi:10.3121/cmr.2013.1176.a4-4

PS3-10:

Patient of Two Professions: Medicine and Dentistry

Sheryl Kane[1]; Amy Butani[1]; Richard Paskach[1]; Paul Jorgenson[1]; D. Brad Rindal[1]; William Rush[1]

[1]HealthPartners

Background/Aims: The literature in both medicine and dentistry has increasingly focused on conditions relevant to both professions. However, research addressing these conditions is limited by the scope of available data, the size of their samples, and data sets that can be combined. The aim of this project is to create a Dental VDW using data from the three HMORN members that have both medical and dental data (i.e., Kaiser Permanente Northwest, Marshfield, and HealthPartners Institute). **Methods:** A team of investigators and VDW programmers was formed. Initial meetings, aimed at coming up with a common design that would accommodate the dental data from all three sites, were held. Each site has a different electronic dental record system and different ways of coding diagnoses and identifying risk. The project has the further challenge that it was being built in anticipation of future projects and so does not have any project related funding. **Results:** Twelve content areas were identified. Of those, four of the tables have been designed and are having test data loaded into them. In order to simplify these early steps they are designed based on the VDW medical model. The existing 'Demographics' table will accommodate both medical and dental patients. 'Dental Utilization' will hold data common to all the procedures within an encounter and is very similar to the comparable VDW table. Where dental data starts to diverge from medical is the 'Dental Diagnosis and Procedures' table. This is because each dental procedure is related to a specific tooth and surface and can have its own diagnosis code. The 'Provider' table is similar to the medical VDW but with different provider specialties. **Conclusions:** The HMORN is working to create a Dental VDW within an existing Medical VDW that will support efforts to conduct research that crosses the divide that currently exists between medical and dental research. It will be a multi-directional system, allowing medical data within dental studies and dental data within medical studies.

Keywords: VDW; Medical VDW; Dental VDW

doi:10.3121/cmr.2013.1176.ps3-10

PS3-11:

Beyond General Equivalency Mappings (GEMs): Understanding the Implications of ICD 10 in Research

Rick Riordan[1]

[1]Kaiser Permanente Northern California

Background/Aims: With the implementation of ICD 10 CM and ICD 10 PCS less than two years away, there are still unanswered questions as to how research teams will effectively translate or use ICD 10 codes in research. Approximately 84% of the ICD 10 codes have only approximate matches with 10% having multiple matches and only 5% have exact one-to-one matches between ICD 9 and ICD 10. With the number of codes increasing five-fold, this offers additional opportunities and risks when pulling data. **Methods:** Besides looking at the General Equivalency Mappings and other tools that are used to translate ICD 9 codes to ICD 10 codes, we will examine some common research areas where only approximate matches between ICD 9 and ICD 10 exist. We will also discuss how the finer level of detail that ICD 10 gives allows research teams to pinpoint exactly what type of asthma, Crohn's disease, and diabetic retinopathy they wish to study without including some of the other cases that do not meet their research criteria. **Results:** There are significant ambiguities and irregularity in several common areas such as diabetes, mental health, asthma, and gastroenterology due to approximate, multiple, or combination matches. Even in the case of exact matches such as an old myocardial infarction where there is an exact match, the definition of when a myocardial infarction becomes "old" is different. **Conclusions:** ICD 10 offers a finer level of detail and a higher level of specificity, thereby allowing research teams to be more targeted when pulling data. On the other hand, research teams need to exercise caution when using GEMs and other tools to translate ICD 9 codes into ICD 10 codes and vice versa, especially if they are looking at data that overlaps the implementation date of October 1, 2014.

Keywords: ICD 10; Code; Coding

doi:10.3121/cmr.2013.1176.ps3-11

PS3-12:

VDW Dashboards to Assure the Quality and Stability of Monthly Updates: The Heart and Vascular Health (HVH) Study Experience

Dustin Key[1]; Gene Hart[1]; Nicholas Smith[2]; Susan Heckbert[2]; James Floyd[2]; Kerri Wiggins[2]; Bruce Psaty[2]

[1]Group Health; [2]University of Washington

Background/Aims: Since the 1980s, the HVH study, run at Group Health (GH) and University of Washington, has used computerized billing and hospital-discharge data to identify potential cases of myocardial infarction (MI). Recently, HVH adopted the local implementation of the VDW as a replacement for the legacy system. Monthly dashboards that compare the yields and subjects identified by the two systems were enacted to monitor the transition and assure comparable ascertainment over time. **Methods:**

Potential MI cases were identified monthly. Counts of MI diagnoses generated by the VDW were calculated with each system update. These were overlaid with prior VDW updates, as well as those generated by the legacy sources. Cohort eligibility criteria such as age, which make use of other data sources, were factored into the comparisons. Plans were made to compare new data to a running mean by the use of a dashboard. Unexpected and non-trivial divergence in the number of new cases identified by the two approaches will prompt additional investigation. **Results:** The dashboard was implemented in September and October of 2012. Counts for MI diagnostic codes appearing in each month starting in January 2004 were compared across the September and October runs. The two series were nearly coincident between these successive versions of the VDW for events ranging from 2004 through February 2012. The divergence that began in March 2012 and continued through August 2012 represented the expected new case volume. Additional dashboard plots confirmed that the distribution of new cases by event month was not due to missed events or to variance in other data sources such as enrollment, demographics, pharmacy, and opt-outs. **Conclusions:** From the point of view of HVH, the dashboard monitors a transition between case-identification methods and assures comparability over time. From the point of view of GH, the dashboard also serves as a method of monitoring the quality of the VDW data updates and alerting the GH local VDW infrastructure team to warnings about unexpected deviations over time. This process can be expanded to monitor other major elements of the VDW and adopted as a local quality improvement effort.

Keywords: Dashboards; Quality Assurance; VDW

doi:10.3121/cmr.2013.1176.ps3-12

PS3-21:

Extracting Utilization Data from Clarity into VDW Using Oracle and SAS

Srivardhan Chimmula[1]

[1]Kaiser Permanente Northern California

Background/Aims: The purpose of the presentation is to demonstrate how we use SAS and Oracle to load VDW_Utilization, VDW_DX, and VDW_PX tables from Clarity at the Kaiser Permanente Northern California (KPNC) Division of Research (DOR) site. **Methods:** DOR uses the best of Oracle PL/SQL and SAS capabilities in building Extract Transform and Load (ETL) processes. These processes extract patient encounter, diagnosis, and procedure data from Teradata-based Clarity. The data is then transformed to fit HMORN's VDW definitions of the table. This data is then loaded into the Oracle-based VDW table on DOR's research database and then finally a copy of the table is also created as a SAS dataset. **Results:** DOR builds robust and efficient ETL processes that refresh VDW Utilization table on a monthly basis processing millions of records/observations. The ETL processes have the capability to identify daily changes in Clarity and update the VDW tables on a daily basis. **Conclusions:** KPNC DOR combines the best of both Oracle and SAS worlds to build ETL processes that load the data into VDW Utilization tables efficiently.

Keywords: Virtual Data Warehouse; Utilization; ETL

doi:10.3121/cmr.2013.1176.ps3-21

PS3-22:

HMORN VDW Mortality QA: Evaluating Completeness and Consistency of Death and Cause of Death Data

Stephen Fortmann[1]; Jason Hermosillo[2]; Daniel Ng[3]; Tyler Ross[4]; Wei Tao[3]

[1]Kaiser Permanente Northwest; [2]Kaiser Permanente Colorado; [3]Kaiser Permanente Northern California; [4]Group Health

Background/Aims: Patient mortality data is a critical component of health research data, especially where death is a clinical outcome of interest. Mortality data is also important for cohort selection and to avoid communications with families of deceased patients. The HMORN Virtual Data Warehouse (VDW) has a DEATH file and a CAUSE OF DEATH file that contain data on whether a patient is dead or alive, date of death, confidence of death status, and causes of death. VDW mortality data is synthesized from multiple sources, such as state death certificates, federal data, and health plan utilization and membership data. These sources are combined using complex, probabilistic algorithms. There is variation in both source data and the transformation logic used by sites. This poses potential challenges for researchers to derive accurate results from multi-site studies. Our aim is to measure the quality of VDW mortality data across HMORN sites, highlight areas of consistency and variation, report issues for remediation, and identify opportunities to improve data specifications and guidelines. **Methods:** A multi-site QA program was distributed to all HMORN sites by the VDW Mortality Ad Hoc Workgroup. Results were returned to the workgroup, which combined, analyzed, and reported the results. **Results:** 14 sites returned results for DEATH and 12 sites for CAUSE OF DEATH. The DEATH file had good overall data quality: 3 sites had minor issues related to missing values or duplicates. Records at 11 sites were checked for post-death patient activity and all sites scored 93% or better. There was significant variation in the CONFIDENCE and SOURCE variables, indicating potential areas for improvements to the specifications. The CAUSE OF DEATH file had mixed quality: 7 sites had issues related to missing values and disagreement with the DEATH file, but there was good agreement on the most common causes of death. Additional research is needed to understand the sources of variation in the data and to recommend specification improvements. **Conclusions:** The VDW DEATH and CAUSE OF DEATH files provide quality mortality data that satisfy a number of uses for multi-site studies. Improving the quality of VDW mortality data will require sustained effort by all HMORN sites.

Keywords: VDW; Mortality; Death

doi:10.3121/cmr.2013.1176.ps3-22

PS3-23:

VDW Data Source: Essentia Health

Paul Hitz[1]; Brian Johnson[1]; Jeremy Meier[1]; Bruce Wasbotten[1]; Irina Haller[1]

[1]Essentia Health

Background/Aims: The Virtual Data Warehouse (VDW) was created as a mechanism to produce comparable data across sites for purposes of proposing and conducting research. The VDW is not a multi-site physical database at a centralized location, but a distributed 'virtual' database with the data remaining at the local sites. At the core of the VDW are a series of standardized file definitions. Content areas and data elements that are commonly required for research studies are identified, and data dictionaries are created for each of the content areas, specifying a common format for each of the elements - variable name, extended definition, code values. Local site programmers have mapped the data elements from their legacy data systems into this standardized set of variable definitions, names, and codes, as well as onto standardized SAS file formats. This common structure of the VDW files enables a SAS analyst at one site to write one program to extract and/or analyze data at all participating sites. **Methods:** This poster demonstrates the data sources used at Essentia Health (EH) for our local implementation of the VDW. **Results:** EH local implementation of the VDW contains detailed medical information on EH patients. These files contain details on 18.5 million unique medical encounters (2002-2012), 25.9 million diagnoses, and 37.7 million procedures. The VDW Enrollment file, which was created to define patient population engaged with EH, has 390,000 "enrollment" periods for 378,000 unique patients. The Demographics file has 2.4 million records. **Conclusions:** The EH VDW provides an easily employable unified central repository of data from available source files. This resource enables the sharing of compatible data in multi-site studies, and also improves programming efficiency, accuracy, and completeness for local single site studies by expending resources to link these legacy systems only once.

Keywords: Data Sources; Essentia VDW

doi:10.3121/cmr.2013.1176.ps3-23