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 Chimmula¹

¹Kaiser 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 Fortmann¹; Jason Hermosillo²; Daniel Ng³; Tyler Ross⁴; Wei Tao³

¹Kaiser Permanente Northwest; ²Kaiser Permanente Colorado; ³Kaiser Permanente Northern California; ⁴Group 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 Hitz¹; Brian Johnson¹; Jeremy Meier¹; Bruce Wasbotten¹; Irina Haller¹

¹Essentia 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