

remain at the local sites; there is no multi-site physical database at a centralized data coordinating center. 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, label, description, code values, and value labels. Local site programmers have mapped the data elements from their HMO's 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 wide range of data sources used at HealthPartners Research Foundation to feed information into our local implementation of the VDW datasets. **Results:** The HealthPartners Research Foundation local implementation of the VDW contains detailed medical information on HealthPartners members and patients. These files contain details on 69 million pharmacy dispensings (2000-2011), nearly 58 million unique medical encounters (2000-2011), including 14 million diagnoses, and 20 million procedures. We have some 9 million Vital Signs observations, and 26 million lab results. The VDW Enrollment and Demographic files are derived from several historical and current membership/patient files; the VDW Pharmacy and utilization files are derived from internal HealthPartners systems plus claims files; the VDW tumor data is retrieved from our owned Cancer Registry. **Conclusions:** The VDW at HealthPartners Research Foundation provides an easily employed unified central repository of data from all 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: VDW; HPRF; Virtual Data Warehouse

doi:10.3121/cmr.2012.1100.ps2-55

PS2-56:

Building the Virtual Data Warehouse Using Data from a Healthcare Billing System

Rick Krajenta¹; Lois Lamerato¹; Karen Wells¹

¹Henry Ford Health System

Background: Healthcare utilization data, specifically diagnoses and procedures, can be processed through different administrative systems. Billing data are generated by the care provider while claims data are generated by the payer. These two sources can represent the data content differently and incorporating the data from these systems into the VDW has presented a unique set of challenges. We present options for reconciling these two data sources to create a VDW that includes all patients including those outside of the HMO member realm. **Methods:** Henry Ford Health System (HFHS) captures patients, care providers, procedures, diagnoses and medical supply information through a proprietary system. Outpatient billing information is entered by clinic staff using an optical scanning device and clinic-specific forms. Other additional procedures and/or supplies are entered into the system using a transaction capture application. Ancillary services such as imaging and pathology are imported into the billing system from their proprietary information systems. The main function of this system is to generate bills for services performed. The data elements are standardized to contain required justification for reimbursement from all payer types. Standard codes sets, ICD9, HCPCS and CPT4, are required. The data can also be used to evaluate workload and staffing levels, project future needs and characterize trends in service demands. HFHS uses this data source to build VDW files. **Results:** Major differences between claims and billing data exist. Claims data are based on health plan contract with bundled procedures, coverage exclusions and deductibles affecting content. Billing data are based solely on services provided. The potential for overwhelming amounts of data in billing sources is possible due to the level of detail; however, this same level of detail is a rich source for specific care components that would not otherwise be present in claims records. **Conclusion:** Billing data are less likely to be compromised by contract restrictions, bundling of different procedures under one code or rejected payments. It is often more granular than claims-based data. Billing data is a

reliable source for the VDW utilization data area because it is a more accurate representation of the delivery of health care.

Keywords: Virtual Data Warehouse; Healthcare Billing System

doi:10.3121/cmr.2012.1100.ps2-56

PS2-58:

A Survey of HMORN VDW Tumor Data Sources

Rick Krajenta¹; Dustin Key²; Amy Butani³

¹Henry Ford Health System; ²Group Health Cooperative; ³HealthPartners

Background: There are currently fifteen HMORN sites that have a VDW tumor table. The VDW tumor specifications standardize this data. However, there is variation in each site's cancer registry, and hence the types of data that can be accessed. **Methods:** During 2011 two metadata surveys have been sent out to the VDW implementation group. The participation rate has been over a dozen sites. The surveys ascertain each site's registry source type, (internal local registry, external central registry, such as SEER, NPCR, State, etc), the type of software used at the registry, whether all patients are captured by the registry, whether NAACCR is used to format and populate VDW tumor, the table update frequency, and other factors. **Results:** 1) All responding sites populate their data by means of a cancer registry 2) All but two sites maintain their own facility registry. Just about half the respondents claimed to be a SEER site, 3) Six sites use the NAACCR manual as a dictionary to aid the population of VDW tumor, four sites use the FORDS manual, 4) Planned VDW Tumor file update schedules vary a great deal, from weekly to annually, with four sites each reporting annual and monthly update schedules, 5) There is a diverse collection of software vendors used to collect cancer data at the registries, 6) Geographic central registry coverage area may miss some patients treated for cancer at some sites. **Conclusion:** It is instructive to understand the diversity in tumor data sources. Different central registry requirements dictate which fields are available to the site and may vary in content. While the data is standardized across sites by a common specification, there may still be nuances between sites; understanding the source data will help researchers and programmers navigate these nuances, and help shape requests for new fields or derivations.

Keywords: VDW; Tumor Table; Virtual Data Warehouse

doi:10.3121/cmr.2012.1100.ps2-58

PS2-59:

It Takes a Village to Raise a VDW Data Set: Creating Partnerships to Improve Research Data

Lela McFarland¹; Christine Bredfeldt²

¹Kaiser Permanente, Information Technology; ²Kaiser Permanente Mid-Atlantic, MAPMG

Background: Healthcare data is highly complex, and considerable effort is required to create rich data resources that are reliable, user-friendly and represent valid utilization. Challenges include identifying appropriate sources, interpreting the data in a given source, matching data between sources, and transforming source data to meet desired specifications. These challenges are shared by all healthcare analytic groups, including research, finance, HEDIS reporting, care delivery and membership management. Creating partnerships for data development both within and across institutions could make the process more efficient by pooling specialized knowledge and providing opportunities to share both development strategies and data products. **Objective:** Within KPMAS, our goal was to facilitate development of the Virtual Data Warehouse (VDW) by tapping into existing knowledge about source data and strategies for validating and refining data. Across KP sites, our goal was to reduce the effort required to build and use the VDW by sharing code and analytic infrastructure among sites using highly similar data sources. **Methods:** We used a combined systems and human factors approach to identify opportunities for partnership. Within KPMAS, we invited non-research analysts with similar data needs to participate in discovery activities prior to building the VDW. As an incentive to help us find data solutions, we provided all regional analysts with access to the resulting Regional Data Warehouse. Across KP sites, we initiated and led two workgroups focused on code sharing and sharing potential