

(pass/warning/fail) reports, and sites sharing ETL (Extract, Transform, Load) code have considerably improved the data quality of the VDW utilization files.

Keywords: Utilization; VDW; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.ps2-51

PS2-52:

VDW Data Sources: Mid-Atlantic Permanente Research Institute

Christine Bredfeldt¹; Lela McFarland²

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

Background/Aims: The Virtual Data Warehouse (VDW) was created as a mechanism for producing comparable data across sites for purposes of proposing and conducting research. It is “virtual” in the sense that the data remain at the local sites rather than 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 range of data sources used at Kaiser Permanente in the Mid-Atlantic States (KPMAS) to feed information into our local implementation of the VDW datasets. **Results:** The KPMAS local implementation of the VDW contains detailed medical information on KPMAS members. These files contain details on 33 million pharmacy dispensings (2004-2011), nearly 27 million unique medical encounters (2005-2011), including 0.5 million hospitalizations, 19 million ambulatory visits, 80 million diagnoses, and 46 million procedures. The data includes 19 million Vital Signs observations, and 40 million lab results. The VDW Enrollment and Demographic files are derived from several historical and current membership files; the VDW Utilization and Pharmacy files are derived from national Kaiser Permanente systems, augmented with data from the KPMAS electronic health record and claims systems; the VDW tumor data is derived from MD, VA and DC state registries. **Conclusions:** The KPMAS VDW provides a centralized, tested repository of data from all available sources. This resource enables data sharing for multi-site studies, and also improves programming efficiency, accuracy, and completeness for KPMAS studies by providing an integrated regional data warehouse.

Keywords: Data sources; Development; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.ps2-52

PS2-53:

VDW Data Sources: Kaiser Permanente Colorado

Heather Tavel¹; Anju Gupta¹; David Tabano¹

¹Kaiser Permanente Colorado

Background: The Virtual Data Warehouse (VDW) is a mechanism for producing comparable data across sites for purposes of proposing and conducting research. It is “virtual” in the sense that the data remain at the local sites; there is no multi-site physical database at a centralized data coordinating center. The core of the VDW is a series of standardized file definitions for content areas and data elements that are commonly required for research studies. Data dictionaries are created for each of the content areas, specifying a common format for each of the files and their respective data elements. Local site programmers map data elements from their HMO’s data systems into file structures with standardized variable attributes. This common structure of the VDW 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 Kaiser Permanente Colorado (KPCO) to feed information into our local VDW datasets. **Results:** The KPCO local implementation of the VDW contains detailed medical information on over 1.5 million enrolled KPCO

members. This includes 50 million pharmacy dispensings (1992-2011), 44 million unique medical encounters, 80 million diagnoses, and 119 million procedures (1998-2011). Vital signs and lab results for over 75 different lab test types have also been loaded into the VDW format. Enrollment and Demographic files are derived from historical and current membership files; Utilization, Vitals and Lab files are derived from legacy and current EMR and claims systems; the VDW tumor data is sourced from Metriq, a cancer registry internal to KPCO; Death data is sourced from Common Membership, Clarity, Metriq, and State Vital Statistics (since 1998). **Conclusions:** The VDW at KPCO 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; Data Sources; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.ps2-53

PS2-54:

The Flow of Tumor Data: A Technical Review of Tumor Data Production, Submission and Use

Jack Richter¹; Lela McFarland¹; Chu-Ling Yu²

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

Background: The Virtual Data Warehouse tumor registry enables important multi-site research in cancer prevention, treatment and outcomes, as well as health communications and quality. Ideally, data are collected from hospitals where the tumors are biopsied, staged, and reported to the state. These data can be difficult to obtain for HMOs such as Kaiser Permanente in the Mid-Atlantic States (KPMAS) that do not own their own hospitals. In such cases, the HMO must request tumor data from the state. This process is particularly complex at KPMAS, which represents three independent jurisdictions (Maryland, Virginia and the District of Columbia). Our end goal is to develop a KPMAS tumor registry that integrates data from electronic health record (EHR) systems and state tumor registries. A key intermediate objective was to develop a comprehensive understanding of tumor registry development at more advanced HMORN sites in order to efficiently build the KPMAS tumor registry. **Methods:** We used a multi-pronged approach to develop the technical structure for the KPMAS tumor registry. First, we surveyed other HMORN sites that have previously successfully developed tumor registries. Second, we reached out to existing tumor registry efforts within KPMAS to 1) reduce duplication, 2) capitalize on existing work in this area, and 3) document the workflow: from data capture through state reporting and integration into our electronic health record. Third, we developed relationships with key tumor registry managers in all three KPMAS jurisdictions to identify the processes required for obtaining tumor data corresponding to our members. During this process, we requested both a data dictionary and a set of dummy data from each of the external jurisdictions. **Results:** Using the data dictionaries and dummy data provided by the external tumor registries, we developed a technical solution for housing and managing tumor related information from multiple sources. Key considerations involved integrating data from external sources with data from the EHR and removing duplicate data that results from integrating data from three external agencies that cover a highly mobile geographic area. **Conclusions:** Integrating tumor data from multiple sources involves both technical challenges and requires that a complete understanding of the data meaning.

Keywords: Tumor registry; Technology; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.ps2-54

PS2-55:

VDW Data Sources: HealthPartners Research Foundation

Amy Butani¹; Lucas Ovans¹

¹HealthPartners

Background: The Virtual Data Warehouse (VDW) was created as a mechanism for producing comparable data across sites for purposes of proposing and conducting research. It is “virtual” in the sense that the data