successfully used as a resource by Group Health in its i2b2 and SHRINE installations, providing the metadata needed to build the complex querying interface. The programs which accomplish this have been repurposed and we've built a prototype for VDW metadata table structure which encompasses pharmacy, procedure, and diagnosis data. We've built draft standard macros to query the metadata and create NDC lists based on an input class of drugs. The completeness of VDW procedures has been measured via UMLS metadata. Finally, we've identified areas of extreme interest which warrant additional research, such as the potential mapping of ICD-10 and ICD-9 codes via UMLS concept unique identifiers. **Discussion:** The UMLS provides a standard way to distribute and prepare metadata that can be used for VDW research and infrastructure program code sets. Since this resource is freely available, well documented, and supported we recommend its adoption by other sites in their programs.

Keywords: Universal Medical Language System; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.cb6-02

CB6-03:

Data of the Dead – An HMORN-wide Comparison of Death Data Sources

David Eastman[1]; Donald Bachman[2]; Daniel Ng[3]; Wei Tao[3]

[1]Kaiser Permanente Southeast; [2]Kaiser Permanente Northwest; [3]Kaiser Permanente Northern California

Background/Aims: Dates and causes of death are available from a variety of sources, both within HMO organizations (e.g., Clarity Patient table at EPIC sites) and from other entities such as the federal government (e.g., Social Security Administration). These sources differ in the quality and completeness of information provided. For example, HMOs frequently miss out-of-hospital deaths which can account for a substantial percent of adult deaths. Governmental sources of data have more detail such as cause of death, but there is potential error involved in matching those records to the correct HMO members. Reliable death information is crucial for health care research. Misclassifying patients could mistakenly misinform researchers and could lead to erroneous conclusions. In order to increase the level of understanding of available death data sources and improve the selection of death data of adequate quality for inclusion in the VDW and HMORN studies, we will describe the types of death data available, how they compare, and how many sites are using each type. We will also perform validity checking by comparing to enrollment, pharmacy, and medical encounter data (e.g., did a supposedly deceased individual have an amoxicillin prescription after death?). At several sites, we will compile all death dates available from all death data sources and perform an inter-source agreement analysis using % agreement or Kappa statistic if possible and present the results.

Keywords: Death; Data Quality; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.cb6-03

CB6-04:

Standardizing Naming Conventions and Document Templates for Virtual Data Warehouse

Bhushan Mahajan[1]; Dean Kjar[1]; Debbie Godwin[1]; Monica Silcott[1]

[1]Scott & White Health System

Background/Aims: SAS Programmers support their Virtual Data Warehouse (VDW) in many venues. These include collaboration and deliverables for multiple projects both internal and external. Such requests include ad-hoc, feasibility and grants proposal data requests from both internal and external investigators and researchers. Different VDW workgroups and VDW site programmers use different naming conventions and standards; this adds to the confusion and produces an unclean environment. SAS Programmers have to make changes to programs over time as the data requirements continually change and must be maintained and accurate per update. The aim of this paper is to suggest a unified method to ensure a cohesive culture amongst the VDW participants by standardized naming conventions. This paper also focuses on setting guidelines for VDW workgroup members and multi-site study programmers regarding creating documents, complying with standard templates, and ensuring guidelines. **Methods:** We suggest standardization of: 1. SAS Program names 2. SAS Datasets names 3. SAS Library names 4. SAS Variable names 5. SAS Logs and Outputs names 6. Multisite SAS programs header template 7. Directory structures (folder names on Windows) for projects 8. Common template for CRN portal documents (Minutes of Meetings, Work plans, Guidelines, etc) consisting of label, header and footer information 9. Processes for multisite programmers and for SAS programmer working on multisite studies. **Results:** will provide a standardized naming convention process for VDW multi-site programmers to follow before sending out programs. Provided are guidelines for standard naming conventions for VDW SAS programmers at each site, to configure their environment to work on multiple data requests efficiently. Also provided are document templates that assist in making the document more informative and containing revision control. **Discussion:** The common standardized process in conjunction with consistent naming conventions for the HMORN VDW members will benefit current and new programmers collaborating on multi-site studies and multiple data requests. The aim is to establish standardized naming conventions at all HMORN participating sites. This provides a better understanding and allows for a unified culture.

Keywords: SAS; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.cb6-04

PS2-37:

Design Considerations and Population Identification for Implementing a Dental Virtual Data Warehouse: Feasibility, Table Structure and Data Elements

Jay Fuehrer[1]; Aaron Miller[1]; Jimmy Kayastha[1]; Paul Hitz[1]; Amit Acharya

[1]Marshfield Clinic

Background/Aims: The population seeking dental care at the Marshfield Clinic has grown substantially in recent years, partly as a result of increased capacity with the construction of new Marshfield Clinic dental clinics. The influx of additional dental data presents a great opportunity to enhance our dental research within the Marshfield Clinic Research Foundation and facilitate external collaborations. In order to collaborate in multi-site dental research studies, creation of a standardized data structure (analogous to the HMORN Virtual Data Warehouse) will be an enormous asset. **Methods:** An initial table structure and mapping was proposed based on researcher needs and available data elements. Through collaborations with couple of other HMORN sites, common elements between sites were identified and adjustments were made to the proposed table layout. Patients seen at Marshfield Clinic dental facilities were included in the patient population. **Results:** The table structure is currently composed of data comprising dental enrollment, dental medication orders, treatment plans, intervention, prescription and medical histories, tooth surface, and risk assessment tables. The Marshfield Clinic served more than 3,500 dental patients in 2007 – this number has increased to approximately 45,000 seen during 2011; we anticipate that our dental patient population will likely continue to grow, as further dental facilities are added (1 facility in 2003; 7 as of 2011; 2 more are planned for 2012). Of the dental patients seen in 2011, 70.3% (31,880/45,322) of them also had a medical visit in Marshfield Clinic's integrated electronic health record system. **Conclusions:** This dental table structure that will be shared across the sites involved will allow collaboration with these sites in future dental studies. The data elements used herein are common to dental care delivery throughout the country. The substantial overlap between dental and medical care in the Marshfield Clinic system offers great opportunities for Oral-Systemic research.

Keywords: Dental; Informatics; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.ps2-37

PS2-38:

VDW Data Sources: Harvard Pilgrim Health Care

Lingling Li[1]; Irina Miroshnik[1]

[1]Harvard Pilgrim Health Care Institute, Harvard Medical School

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 remain at the local sites; there is no multi-site physical database at a centralized coordinating center. At the core of the VDW are a series of