PS1-22:

Tumor—VDW Table Structure Dictated by National Agency Standards

Rick Krajenta, BS[1]; Lois Lamerato, PhD[1]; Karen Wells, BA[1]

[1]HFHS

Background: North American Association of Central Cancer Registries (NAACCR) was established in 1987 as a collaborative organization for cancer registries, government agencies, professional associations and private groups. NAACCR develops and promotes uniform data standards, provides education and training, certifies population-based registries, processes and publishes data from central registries and promotes the use of cancer surveillance data for research, public health and patient care. Cancer registries in the US include the national central registries, NCI-SEER and CDC-NPCR, individual State registries and hospital-based (care provider-based) registries. The Cancer Research Network (CRN) has adopted NAACCR data standards to define the Virtual Data Warehouse tumor registry table (VDW-TR). However, since the inception of the VDW-TR, there have been many version of NAACCR in effect. **Aims/Methods:** VDW-TR needed to have similar and merge-able data for multi-site projects. Data standards set by NAACCR are optimal for construction of this resource as they are designed to collect tumor data centrally from multiple data sources. The standards establish processes for data exchange and record layout in addition to coordinating input from sponsoring organizations, such as AJCC and NCI. NAACCR is also responsible for incorporating new items of interest as the data used to characterize cancers evolve. We describe how these changes were incorporated into the VDW-TR. **Results:** AJCC Collaborative Stage I (CS-1), applicable to cases diagnosed beginning with January 2004, brought many changes to data and data formats required for staging. These changes were not incorporated by the VDW tumor file in 2004 due to the lack of ownership and oversight. Discrepancies eventually developed between VDW data dictionary and NAACCR causing data value decay. ICDO-2 histology lists were expanded and recoded in ICDO-3, providing additional challenges, along with other rules-based changes in tumor classification. CS-1 also mandated addition of anatomic site specific factors. Many additional changes occurred in 2010 with CS-2. The specifications incorporated in our current VDW-TR address all of these data changes. **Conclusion:** The VDW has to adopt NAACCR changes as they are adapted to remain current with all standards. We are now sensitized to monitor and adjust for significant future changes in NAACCR.

Keywords: Tumor registry, NAACCR, VDW

doi:10.3121/cmr.2011.1020.ps1-22

C-A1-02:

Developing a Structure for Programmatic Quality Assurance Checks on the Virtual Data Warehouse

AlanBauck, BS[1]; Don Bachman, MS[1]; Karen Riedlinger, MPH[1]; Kari Walker, MS[1]; Sabrina Luke, MPH[1]; Jerena Donovan, MS[1]; Ping Chen, MS[1]

[1]Kaiser Permanente Northwest

Background/Aims: The HMORN Virtual Data Warehouse (VDW) has an expanding set of quality assurance data checks developed by its data content work groups. These checks improve the reliability and consistency of the data across sites. When the Center for Effectiveness and Safety Research (CESR) Data Coordinating Center sought to provide additional data checks, it defined a standard methodology for additional validation. The structure and methodology provides a basis to apply data checks consistently across the data content areas. **Methods:** The CESR Data Coordinating Center created a work group of data warehouse experts and analysts that began by compiling an inventory of existing quality assurance data checks. Using the inventory as a reference, the work group selected data content areas with few existing data checks to identify checks that could be documented and programmed more consistently. The work group developed a structure for defining, documenting, and programming new data checks. Finally, the work group tested and refined its documentation and programming structure for new data checks by creating new checks and learning how they fit into the existing structure. **Results:** The work group created a consistent process under which new data checks could be defined, documented, and programmed. This process included a structure to: • Improve the consistency of the types of programmatic data checks. • Document the data checks and processes for the VDW CESR Data Coordinating Center quality assurance programs. • Create a structure in which new data checks could be added that would not conflict with existing data checks. • Reuse programming logic across data content areas. **Conclusions:** This standard methodology enables more consistent implementation and documentation of data checks for the VDW. The methodology also provides information about the outcomes of the data checks which have been processed. Finally, the methodology should allow for quicker development of future data checks.

Keywords: Virtual Data Warehouse, Quality assurance

doi:10.3121/cmr.2011.1020.c-a1-02

C-A1-04:

Challenges Faced as New Members of the HMO Research Network Virtual Data Warehouse

Debbie Godwin, BS[1]; Bhushan Mahajan, MS[1]; Monica Silcott, AAS[1]; Dean Kjar, MS[1]

[1]Scott & White Healthcare

Background/Aims: This presentation will focus on challenges faced by new HMO Research Network (HMORN) sites and their experiences with the Virtual Data Warehouse (VDW) tables for version 3 specifications. The step-by-step procedure developed by Scott and White will be showcased to provide a head start for new incoming sites. Sharing lessons learned during our one year in the HMORN VDW, our objective is to provide best practices regarding resources, time allocation, priorities, schedules, financing, backups, and possible available alternatives. **Methods:** The comprehensive procedure followed by Scott & White to build VDW V3 tables:1) Setting up your SAS Environment • Establishing an architecture and components • Analyzing and forecasting server specifications; 2) Identifying and understanding the multiple data sorces embedded in your organization • Claims data • EMR • Lab • Pharmacy • Other; 3)Detailed variable mapping of sources to VDW specifications; 4)Assessment of source data availability; 5) Getting your site's first feed from the source and following through with quality checks; 6)Methods for building transformation logic for the table; 7) Choosing the methodology to approach table-building. **Results:** We will showcase Scott & White metrics for building a VDW starting with version 3 and best practices used. We improved the validity and completeness of the data through person-to-person interactions with claims personnel who often were disinterested in the needs of research. Having strong advocates who think of research as a part of the institutional mission was helpful. The VDW data team included a sole programmer for the first 6 months along with the Site Data Manager; adding a second programmer improved the rate of progress. **Conclusions:** A structured, step-by-step, approach is recommended for new incoming sites joining the HMORN. This approach will optimize efficiency and effectiveness. The more information and guidance the existing members can provide, the better the achieved results will be. Sharing experiences will guide new members to a more pro-active and efficient start and ensure the success of seamless data-sharing.

Keywords: Virtual data warehouse (VDW), Challenges, New member

doi:10.3121/cmr.2011.1020.c-a1-04

PS1-26:

Data Quality Issues/Checks While Building Virtual Data Warehouse Enrollment, Demographics, Provider and Utilization Tables from Scratch

Bhushan Mahajan, MS[1]; Dean Kjar, MS[1]; Debbie Godwin, BS[1]; Anne Robertson, AGS[1]; Teresa Ponder, AAS[1]

[1]Scott & White Healthcare

Background/Aims: This presentation will focus on data quality issues experienced by new HMO Research Network (HMORN) sites while building

Virtual Data Warehouse (VDW) demographics, enrollment, utilization, and provider tables from scratch. Our goal is to benefit new sites joining the HMORN by providing a heads-up on anomalies to look for and what to expect of the initial data quality check. **Methods:** Some of the quality issues to be discussed:• Choosing unique VDW person identifier for the site (Unique 'Pat_ID' for the HealthPlan members and unique 'MRN' for the hospital inpatients). Discuss issues related to both the patient IDs. • Handling missing MRN's / Pat_ID's, duplicate MRN's/ Pat_ID's, multiple MRN's for Pat_ID and multiple Pat_ID's for MRN• Conflicting gender, race, birth date information for single MRN in both claims and electronic medical record (EMR) hospital data • Rolling up of claims data for inpatient stays • Handling multiple diagnoses per single stay and identifying primary diagnosis• Interpreting claims data (Claim ID, line number, etc.) • How to handle: same MRN, same PROCDATE (procedure date), different provider, different POS (place of service), different PX (procedure code) to determine which is the correct POS, primary provider and PX• Variable ëauthorization ID' to identify unique encounter • Building VDW demographics table by merging claims data and EMR data **Results:** We will thoroughly address and document our progress regarding data quality issues, supporting this presentation with examples and tables. **Conclusions:** New sites need to develop VDW version 3 specification tables from scratch while other sites are in the process of migrating from VDW version 2 to version 3 specifications. As health system use and research dissemination is only as accurate as the initial data management protocol, dedicated quality efforts are essential throughout the process to ensure consistent validity across HMORN sites.

Keywords: Data quality, Data issues, VDW

doi:10.3121/cmr.2011.1020.ps1-26

PS1-21:

Geocoding for VDW CENSUS Files Using SAS Geocoding (an Alternative to MapMarker+ Software)

Dean Kjar, MS[1]; Bhushan Mahajan, MS[1]; John Zeber, PhD, MHA[1]; Debbie Godwin, BS[1]

[1]Scott & White Healthcare

Background/Aims: Geocodes are used to link individual enrollee records to demographic information from the Census Bureau. Currently, many of the HMORN VDW sites are utilizing software such as MapMarker+ to geocode their data. Acquiring third-party software capable of street-level geocoding is an additional expense that new and current HMORN sites may not have to incur. Geocoding at the street level has been added to the SAS/GRAPH module in the third maintenance release of SAS 9.2, enabling users more flexibility to create powerful geographic analyses and data graphic presentation. This presentation will focus on creating the VDW CENSUS table using PROC GEOCODE in SAS/Graph. **Methods:** We will present a practical example of how street-level geocoding can be accomplished with SAS/GRAPH and PROC GEOCODE, utilizing an existing VDW project dataset. We will explain what files are needed and where to find them. Talk attendees will learn to take advantage of a readily available SAS procedure to increase their knowledge and skills of an important new analytical research tool. **Results:** The third maintenance release of SAS 9.2 software allows geocoding at the street level. We will generate the 16 character FIPS codes (Federal information processing standards) using SAS software. This will provide HMORN sites an alternative to MapMarker+ software. **Conclusions:** Street-level geocoding with SAS appears to represent a substantially easier way to match Census Bureau demographic information with individual enrollee records. It utilizes a VDW analyst's current SAS expertise without the need to master stand-alone geocoding software or its specialized knowledge requirements, not to mention the additional time or costs involved with incorporating such external programs, e.g. there is no need to convert the data out of SAS datasets and then convert back to SAS datasets.

Keywords: Geocoding, Census, VDW

doi:10.3121/cmr.2011.1020.ps1-21

C-B1-03:

Building Cancer Research Capacity: Exploring How a Non-SEER Site Can Readily and Efficiently Participate in Cancer Studies

Sharon Rolnick, PhD, MPH[1]; Brian Owens, BA[1]; Amy Butani, BA[1]; John Soler, MPH[2]; Jody Jackson, RN, BSN[1]; Thomas Flottemesch, PhD[1]

[1]HealthPartners; [2]Minnesota Department of Health

Background/Aims: Access to large, inclusive cancer databases such as the Surveillance, Epidemiology and End-Results (SEER) registries enables timely and efficient identification of cancer cases and gathering of tumor data. HealthPartners Research Foundation's (HPRF's) ability to conduct cancer research has been hampered by lack of access to such registries. HPRF developed two capacity-building projects to explore how to quickly and accurately identify cancer cases and related data for research. Identification of a viable electronic data source would allow HPRF to populate the virtual data warehouse (VDW) tumor table and participate in multi-site studies within the CRN. **Methods:** Cheri Rolnick, PhD, MPH conducted a study to assess linking with the population-based cancer registry, the Minnesota Cancer Surveillance System (MCSS), maintained by the Minnesota Department of Health (MDH) to identify cancer cases preparatory to research. HealthPartners (HP) automated claims data was used to identify 1450 patients, representing four cancer cohorts. Cases were forwarded to MDH to identify those that met defined criteria by matching records located within MCSS. In addition to connecting with MDH, we are also examining the capacities of the cancer registry at Regions, HP's owned hospital. Thomas Flottemesch, PhD and Sue Braaten, the lead cancer registrar from Regions, have just received internal funding to examine the utility of registry data for research. The goal of this project is to test the feasibility of populating the VDW tumor tables with the data the hospital sends periodically to MDH, and to develop a formal, automated system of data transfer between Regions and HPRF. **Results:** MDH was able to link 79% (1152 of 1450) of the cases identified by HP. Non-matches were due to misclassification by HP-created algorithms, patients not living in Minnesota, or patients with cancers different from those identified by HP records. Concordance for determining eligibility (stage, date of diagnosis) was high but not 100%. **Conclusions:** Findings from the MCSS study created a process of data exchange between HP and MCSS that can be used in future studies. Results of this examination and the aims and design of the hospital registry study will be presented.

Keywords: Cancer research, Tumor registries, Virtual data warehouse (VDW)

doi:10.3121/cmr.2011.1020.c-b1-03

C-A1-03:

Considerations in the Design and Use of an Oracle-based Virtual Data Warehouse

Christine Bredfeldt, PhD[1]; Lela McFarland, BS, BA[2]

[1]Kaiser Permanente, Mid-Atlantic States; [2]Kaiser Permanente, Information Technology

Background/Aims: The amount of clinical data available for research is growing exponentially. As it grows, increasing the efficiency of both data storage and data access becomes critical. Relational database management systems (rDBMS) such as Oracle are ideal solutions for managing longitudinal clinical data because they support large-scale data storage and highly efficient data retrieval. In addition, they can greatly simplify the management of large data warehouses, including security management and regular data refreshes. However, the HMORN Virtual Data Warehouse (VDW) was originally designed based on SAS datasets, and this design choice has a number of implications for both the design and use of an Oracle-based VDW. From a design standpoint, VDW tables are designed as flat SAS datasets, which do not take full advantage of Oracle indexing capabilities. From a data retrieval standpoint, standard VDW SAS scripts do not take advantage of SAS pass-through SQL capabilities to enable Oracle to perform the processing required to narrow datasets to the population of interest.