

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 'authorization 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, MS¹; Bhushan Mahajan, MS¹; John Zeber, PhD, MHA¹; Debbie Godwin, BS¹

¹Scott & 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, MPH¹; Brian Owens, BA¹; Amy Butani, BA¹; John Soler, MPH²; Jody Jackson, RN, BSN¹; Thomas Flottesch, PhD¹

¹HealthPartners; ²Minnesota 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 Flottesch, 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, PhD¹; Lela McFarland, BS, BA²

¹Kaiser Permanente, Mid-Atlantic States; ²Kaiser 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.

Methods: Beginning in 2009, the research department at Kaiser Permanente in the Mid-Atlantic States (KPMA) has developed an Oracle-based VDW according to the HMORN v3 specifications. In order to take advantage of the strengths of relational databases, KPMA introduced an interface layer to the VDW data, using views to provide access to standardized VDW variables. In addition, KPMA has developed SAS programs that provide access to SQL pass-through processing for first-pass data extraction into SAS VDW datasets for processing by standard VDW scripts. **Results:** We discuss both the design and performance considerations specific to the KPMA Oracle-based VDW. We benchmarked performance of the Oracle-based VDW using both standard VDW scripts and an initial pre-processing layer to evaluate speed and accuracy of data return. **Conclusions:** Adapting the VDW for deployment in an Oracle environment required minor changes to the underlying structure of the data. Further modifications of the underlying data structure would lead to performance enhancements. Maximally efficient data access for standard VDW scripts requires an extra step that involves restricting the data to the population of interest at the data server level prior to standard processing.

Keywords: Infrastructure, Database, Performance

doi:10.3121/cmr.2011.1020.c-a1-03

PS1-35:

Papanicolaou Tests and Data at Group Health

Tyler Ross, MA¹; Joe Webster, BA¹

¹Group Health Research Institute

Background: The Cancer Screening Effectiveness and Research in Community-based Healthcare (SEARCH) project aims to create infrastructure to support conducting comparative effectiveness research in cancer screening. As a proof of principle, we have created a multi-site VDW-like PAP data structure and populated it at the Group Health Research Institute (GHRI). **Methods:** In this poster, we provide the PAP table structure that has been defined by SEARCH and the extract, transform, load (ETL) process employed at GHRI to populate the table. We report basic descriptives on the loaded data including the rates of PAP testing over time by population groups. **Results:** Group Health enrollees whose primary care clinics are owned and operated by Group Health (referred commonly as the Group Practice Division (GPD)) appear to have complete data capture of PAP tests and results. The rate of testing among the population seems to have declined in the most recent years, likely due to changes in cervical screening guidelines. In primary care clinics not owned and operated by Group Health (referred commonly as the Network), the number of PAP results found is considerably lower. This low rate of PAP resulting is due to incomplete data capture. **Conclusion:** Our findings show that research projects like the SEARCH cervical proof of principle that use PAP results as a key component of their study must define their population of interest based on the individual person's primary care clinic and its relationship with Group Health to ensure complete data capture. This data structure, while created under the SEARCH project, has been designed for future use in other studies.

Keywords: PAP test, VDW

doi:10.3121/cmr.2011.1020.ps1-35

C-B1-02:

VDW Patient Medical Record Report -Detailed Information on Patient Utilization

Donald Bachman, MS; Center for Health Research, Kaiser Permanente Northwest

Background/Aims: Project teams often require itemized patient information for their research and often engage a medical record technician to help them understand the details of a patient's utilization. The authors suggest that project teams can either supplant or augment their abstracting efforts by using this person-level VDW (Virtual Data Warehouse) patient medical record report. This report is an inexpensive, quick, and well-organized method to produce a patient medical record using VDW data. Another important benefit of this report is that it also has additional information (such

as claims) not readily found in the electronic medical record. **Methods:** Users of the VDW patient medical record report have access to a wide array of content: diagnosis; DRG and procedure codes and descriptions; outpatient dispenses described by generic name; lab test results; BMI; weight; height; blood pressure; provider department; inpatient admitting source; and discharge status. Users can also opt to include cancer site from the VDW tumor registry. More variables can be added over time. This report includes all encounter types (e.g., inpatient, outpatient, emergency, long-term care, telephone, e-mail) found in the VDW files. In addition, the header includes gender and birth/death year, and users specify a time period. Each data value adds an additional line. It is designed for local review as the report contains PHI. **Results:** This report is easy to create. Because the layout follows chronological order, the reviewer can quickly understand the patient's utilization over a specified time period. Analysts, investigators, and providers have used this report. Programmer analysts use this information to guide their coding. Providers and investigators like it because they can quickly review patient utilization. We also often give this report to abstractors as a supplement to the traditional electronic medical record. Caveats: The report only contains information from the VDW, so it is only as complete as a site's VDW files. The electronic medical record contains much information that is not in the VDW patient medical record report. **Conclusions:** The VDW patient medical record report is an efficient way for research teams to view and understand patient-level utilization found in the VDW.

Keywords: VDW, Patient medical record report

doi:10.3121/cmr.2011.1020.c-b1-02