PS1-10:

Laboratory Content Area Development Within the Virtual Data Warehouse

Marsha A Raebel, PharmD, *Kaiser Permanente Colorado Institute for Health Research*; Gwyn Saylor, BA, *KPCO Institute for Health Research*; Richard J Krajenta, BS, *Henry Ford Health Systems*; Jeremy McCauley, MBA, *Marshfield Clinic Research Foundation*; Paul Hitz, BS, *Marshfield Clinic Research Foundation*; Terese A DeFor, MS, *HealthPartners Research Foundation*; Susan Andrade, ScD, *Meyers Primary Care Institute*

Background: HMORN leadership charged several Workgroups to check and describe Virtual Data Warehouse (VDW) data areas, provide observations on data quality across sites, and recommend needed actions. Availability of laboratory (lab) results data in the VDW in a standardized format is desirable, but across HMORN sites, lab content is either incompletely developed or not built at all. Further, there is little consistency across sites with regard to the approach to adding lab content. The work of the VDW Lab Content Area Workgroup (Lab Workgroup) therefore focused on developing a standardized approach to incorporating lab content. We describe development and implementation efforts to create and maintain a standardized lab result content area across HMORN sites. **Methods:** The Lab Workgroup identified priorities: 1) Establishing standardized naming conventions for variables; 2) Determining test results to be initially incorporated; 3) Encouraging site-specific exploration and development of lab information system content as well as lab department contacts; 4) Developing a lab test reference document; 5) Serving as content area experts; and 6) Conducting content quality checks. A list of possible tests to incorporate was circulated to HMORN sites. The Lab Workgroup wrote code to generate descriptive data and data checks across sites for glycated hemoglobin (HGBA1C) and serum potassium result content. A lab test metadata table was posted for site data managers to complete. **Results:** Incorporation of chemistry tests that were most often requested for research was identified as an initial priority. Microbiology and tumor markers content were other high priority areas. Standardized lab variable naming conventions were determined for the list of initial tests. Results of data checks for HGBA1C and potassium concentration results will focus on describing the extent and quality of these files at sites, as well as identify potential issues to be addressed. Sites were surveyed about the barriers, facilitators, and priority of adding lab content to VDW. Results of all this work as well as an excerpt from the lab test reference document will be presented. **Conclusions:** The availability of lab test results information in the VDW in a standardized, extractable format across HMORN sites is a technological advance in data availability and sharing. The work described here will provide important initial insights into the lab content area potential of the HMORN VDW.

PS1-16:

Quality Assurance in the VDW TUMOR File

Richard J Krajenta, BS, *Henry Ford Health System*; Lois E Lamerato, PhD, *Henry Ford Health System*; Karen Wells, BA, *Henry Ford Health System*

Background: The VDW TUMOR database is constructed using nationally standardized variable definitions. The North American Association of Central Cancer Registries (NAACCR) is a professional agency for establishing data standards. The database is NAACCR compatible with respect to variable definitions and record layout. Many of the CRN sites are also Surveillance, Epidemiology, and End Results (SEER) Program sites. SEER collects data from specific geographic regions representing 26% of US population and is demographically diverse. Access to SEER data provides a mature data source and familiar variable list for use in cancer research. Non SEER site can access comparable data using either hospital based or State based data collection systems. **Methods:** The TUMOR content quality team was formed to examine data quality to report back to the VDW operations committee. Frequency of values in selected fields can be used to indicate data quality. The team has decided to test for non standard values, number of unknowns (stage, anatomic site others), sex specific cross tabs (i.e., males with female GYN cancer, females with prostate cancer) and histology frequencies by year to detect outliers. In addition to identifying areas for quality improvement, the TUMOR content team will evaluate the need for additional references to be appended to the VDW data area and suggest future directions to improve the utility of the TUMOR database. The team will also review the need for additional data processing tools (i.e. macros) to facilitate ease of content use. **Results:** Data

from 9 sites have tumor data posted to the CRN web counter. This represents 725,000 tumors. Over 77,000 un-staged cancers (10.6%) were identified using the CRN web based cancer counter. The next step is to determine the proportion for which un-staged is an appropriate designation by examining the tumor behavior and other clinical parameters. This process will be used to examine additional variables and provide the TUMOR quality team with guidance on which specific content areas require attention. This work will allow the TUMOR quality team to identify areas for data improvement without distributing many work-plans to each participating sites. **Conclusion:** Examining and documenting the findings of the TUMOR content team will provide a mechanism to establish best practices for use of the TUMOR data. Expanded knowledge of current status will facilitate future additions to the content.

PS1-19:

Method for Using the Electronic Health Record to Create Population Denominators

G Craig Wood, MS, *Geisinger Center for Health Research*; Joseph B Leader, BA, *Geisinger Center for Health Research*; Walter F Stewart, PhD, MPH, *Geisinger Center for Health Research*

Background/Aims: The HMORN includes research centers that are part of integrated delivery systems where the practice and insurance entities are independent. Rules for defining a population denominator at these centers require a consideration of the practice population that is not a member of the insurance entity, as the notion of enrollment and disenrollment is not relevant. We describe methods for defining a population denominator for primary care patients using data from only the electronic health record (EHR), including consideration criteria for dates that are analogous to membership enrollment and disenrollment. The aim of this study was to create a method for calculating population denominators from the EHR based on utilization activity. **Methods:** We modeled data on the pattern of utilization over time to determine start and termination dates for time under observation. Survival analysis was used to estimate the probability of cohort with drawl after various lengths of utilization inactivity. Optimal cutoffs were derived that balanced false positives with false negatives. Since the length of time required for a patient to become inactive is likely dependent upon age and gender (i.e. a typical gap in utilization will tend to be longer for young males as compared to older males), the cutoffs for defining termination dates varied based on patient demographic. **Results:** For each unique patient, the number of encounters were counted for each 6-month time period (i.e. January 1 – June 31, and July 1 – December 31). To define cohort entry, the 6-month period in which the first office visit encounter occurred was used to define a starting date. Enrollment was considered active until the patient failed to have any encounters with a primary care clinic for an age/gender specific cutoff of number of consecutive 6-month periods, at which point the end date will be retroactively assigned to the end of the 6-month period in which the final encounter occurred. **Conclusions:** EHR utilization can be used to define population denominators. Validation of the proposed method needs to be conducted by comparing results to insurance enrollment spans. This application is limited to clinical areas where there is evidence of relatively complete capture.

PS1-20:

VDW Operational Committee: Current Activities and Future Directions

Jeffrey Brown, PhD, *Department of Ambulatory Care and Prevention, Harvard Pilgrim Health Care*; Sarah McDonald, BA, *Group Health Center for Health Studies*; Kristen Moore, MPH, *Department of Ambulatory Care and Prevention, Harvard Pilgrim Health Care*; Gwyn Saylor, BA, *Institute for Health Research, Kaiser Permanente Colorado*; Gene Hart, MS, *Group Health Center for Health Studies*; Mark Hornbrook, PhD, *Center for Health Research, Kaiser Permanente Northwest*; David Magid, MD, MPH, *Institute for Health Research, Kaiser Permanente Colorado*; Alan Go, MD, *Division of Research, Kaiser Permanente Northern California*

Background/Aims: The Virtual Data Warehouse (VDW) is the HMO Research Network (HMORN) approach for facilitating multisite research while protecting the privacy of members and proprietary corporate information. A valid and sustainable VDW is critical to the success of the HMORN, and is needed for HMORN's inclusion in many of the most important public health initiatives

planned for the next few years. **Methods:** In November 2007 the HMORN Governing Board approved the VDW 5-Year Strategic Plan, including creation of a VDW Operational Committee (VOC), reporting to the Assets Stewardship Committee (ASC) and providing coordinated oversight of the development, maintenance, and enhancement of the VDW. The seven members of the VOC are investigators and analysts representing HMORN's major consortia projects: CERT, CVRN and CRN. **Results:** In 2008, the VOC accomplished two major short term goals: 1) Creation of a comprehensive data quality checking system, implemented by seven content area expert Working Groups (WGs) consisting of 23 investigators and 23 programmer/analysts. The content areas of the WGs are enrollment and demographics, pharmacy, utilization, tumor, vital signs, laboratory, and census. The WGs assessed data availability and completeness for their content area and reported findings to the VOC and ASC, including recommendations for changes and enhancements. WG reports will form the basis for ongoing VDW documentation that will be used to standardize HMORN VDW descriptions for grants and proposals; 2) Development of a budget, staffing plan and priorities for 2009. The 2009 priorities include development of VOC standard operating procedures for creating new data content areas and changing current data areas/definitions, documenting policies for use of the VDW, promoting use of the VDW, and creation of additional WGs for death (new VDW content area), informatics, implementation, and data documentation. The documentation WG will oversee issues such as data dictionaries, data checking documentation, programming resources, a clinical concepts library, user guides, and incorporating updated VDW information in the HMORN Collaboration Toolkit. Looking forward, the VDW 5-year Strategic Plan includes streamlining IRB approval for multisite research and enhancing HMORN informatics capabilities. **Conclusions:** The newly created VOC has made substantial progress in helping make the HMORN VDW an even more valuable and useful resource for the network.

PS1-22:

The VDW Vital Signs File: Strengths, Issues, and Recommendations for the Future

Amy L Butani, BA, *HealthPartners Research Foundation*; Nancy Sherwood, PhD, *HealthPartners Research Foundation*; Kenneth Adams, PhD, *HealthPartners Research Foundation*; G Craig Wood, MS, *Geisinger Center for Health Research*; Yury Vilk, PhD, *Harvard Medical School*

Background/Aims: The Virtual Data Warehouse (VDW) Vital Signs file is an important data source for clinical information that is often used in VDW based research. Vital signs traditionally include pulse, diastolic blood pressure, systolic blood pressure, respiratory rate, and body temperature. This file contains multiple records of these data for each patient as well as related data such as measure date, height, weight, body mass index, and position for blood pressure reading. Despite this importance, it has not thus far been subjected to broad, systematic scrutiny for the extent and quality of the data at implementing sites. That is soon to be remedied by the subject effort of the proposed poster. We have devised an analysis plan and are writing code to generate copious amounts of descriptive data. These data and our group's cross-site analysis thereof will be documented in a report to the HMORN. The report will serve as a basis for establishing the current state of the Vital Signs files at implementing sites, and for making recommendations for future improvements. It should also prove quite useful as a reference for future grant writing efforts. **Methods:** Data set subject area committees were formed as part of the new HMORN governance of the VDW, composed of programmers and investigators from each site. Each committee has been charged with stewarding its particular dataset specification, and assuring the quality of the implemented files at the various HMORN sites. The Vital Signs subcommittee has prepared specifications for a set of descriptive tables, and a metadata survey inquiring about how the sites have interpreted the current Vitals file specification. Subcommittee members are in the process of writing a VDW program that will be run at each site to produce the desired descriptive tables. **Results:** Our report to the Operations Committee will focus on describing the extent and quality of the Vital Signs file data at the various sites, pointing out any inconsistencies in implementations and making recommendations for changes in the specifications. The proposed poster will highlight our findings. **Conclusions:** The report will evaluate the extent and quality of vital signs across HMORN sites and identify current ambiguities in the specification that will need to be remedied.

PS1-24:

Proof of Principle Demonstrations of a Distributed Research Network: Findings and Lessons Learned

Jeffrey S Brown, PhD, *Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care*; John H Holmes, PhD, *University of Pennsylvania School of Medicine*; Ross Lazarus, MBBS, MPH, *Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care*; Robert Langer, MD, MPH, *Geisinger Center for Health Research*; David Magid, MD, MPH, *Institute for Health Research, Kaiser Permanente Colorado*; Andrew Nelson, MPH, *HealthPartners Research Foundation*; Joe Selby, MD, MPH, *Division of Research, Kaiser Permanente Northern California*; Edward Wagner, MD, MPH, *Group Health Center for Health Studies*; Richard Platt, MD, MS, *Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care*

Researchers, policy makers and others commonly use electronic data routinely collected during the delivery of, or payment for, medical care to study the "real world" effectiveness, comparative effectiveness, safety, and costs of medical interventions. However, even very large individual healthcare data resources are often not big enough to adequately conduct post-marketing evidence studies. To improve the ability to use multiple distributed data resources, like the HMORN's Virtual Data Warehouse, the Developing Evidence to Inform Decisions about Effectiveness (DEcIDE) centers at the HMORN Center for Education and Research on Therapeutics (CERT) and the University of Pennsylvania are developing a design for a scalable distributed research network. Funded by AHRQ's Effective Health Care Program, the project intends to develop the framework for a distributed research network that will help close the knowledge gap regarding the "real world" effectiveness, comparative effectiveness, and safety of medical technologies. The network will have these important attributes: distributed architecture, strong local control of data uses, and federated querying. Constructing the network is a challenge, even among sites with fully operational local virtual data warehouse (VDW) installations, because of differences in computing environments and information systems, the need for responsible stewardship of clinical records, concerns related to data privacy, software development obstacles, the use of proprietary data, and the need to develop effective governance mechanisms. To investigate some of these issues, the project includes implementation of a proof of concept prototype involving these HMORN sites: Kaiser Permanente Northern California, Kaiser Permanente Colorado, Harvard Pilgrim Health Care, Group Health Cooperative, and Geisinger Health System. The goals of the proof of concept work are to distribute a query to separately stored data resources, execute the query remotely, and present aggregated results to an authorized system user. The proof of concept will be designed with a single network portal that manages the network software, including security, authorization, authentication, messaging, and query aggregation. We will present our findings from the proof of concept, including information on specific architecture and design used, the success of the proof of concept, and lessons learned from implementation.

PS1-26:

The VDW Demographic and Enrollment Files: Strengths, Issues, and Recommendations for the Future

Mark C Hornbrook, PhD, *Center for Health Research, Kaiser Permanente Northwest*; Roy E Pardee, JD, MA, *Center for Health Studies, Group Health*; Paul Hitz, BS, *Marshfield Clinic Research Foundation*; Mark Schmidt, BA, *Center for Health Research, Kaiser Permanente Hawaii*

Background/Aims: The VDW Demographic and Enrollment data files contain basic demographic data on all HMO members. These data include details of membership (start and end dates, whether drug coverage was included, type of insurance [Medicare, Medicaid, etc]). Demographic data include birthdate, gender, race/ethnicity. Demographic and enrollment data give the information needed to enumerate and begin describing a potential or actual study population. These data can be combined to give interesting and informative views into the different populations at the participating HMOs. This poster will describe the variations in data validity and completeness between the various HMOs, including age distributions, Medicare/Medicaid enrollment, variations in availability of race/ethnicity data, and recommendations for future improvements. **Methods:** The VDW Demographic/Enrollment QA committee was formed as part of the new HMORN governance of the VDW. This committee is charged with analyzing the quality of the demographic and enrollment VDW files at the various