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