

Background: Immune thrombocytopenic purpura (ITP) is a rare though well recognized adverse event following measles, mumps, and rubella (MMR) vaccine. Identification of individuals at high risk for these and other vaccine adverse events might inform pre-vaccination screening options for the future. Working in close collaboration with the FDA, we seek to identify genetic differences underlying the susceptibility to ITP following MMR or MMR and varicella (MMRV) vaccination. We will compare the genetic make-up of these cases to ITP occurring after medication exposure and to idiopathic cases of ITP. **Methods:** We plan to recruit 450 persons with ITP. These will include 150 cases with ITP following vaccination from the FDA's Vaccine Adverse Events Reporting System (VAERS) and the Intercontinental Cooperative ITP Study Group (ICIS) and 300 cases (150 with ITP related to prescribed medications, and 150 cases of idiopathic ITP) recruited from ICIS, Kaiser Permanente Georgia, and Kaiser Permanente Southern California. To identify novel candidate genes or pathways, we will perform whole exome resequencing for (at least) six unrelated individuals with a Mendelian form of ITP. Polymorphism frequencies within these candidate genes will be examined and their proportions compared with control data available from the 1000 Genomes Project. The identified polymorphisms will inform the genome-wide association study (GWAS) phase, which will also include a copy number variation (CNV) component. **Results:** To date, we have identified 232 potential cases of ITP following MMR/MMRV vaccination reported to VAERS during 2000-2009. Of these potential cases, 214 have been abstracted and 156 (73%) have been confirmed. Recruitment is set to begin in November 2010. In addition, we have identified six participants from families with congenital thrombocytopenia for exome resequencing. Supplemental funding was obtained to support recruitment efforts and additional sequencing. **Conclusions:** This study involves a unique collaboration between the FDA and the HMORN, leveraging HMORN experience in pharmacogenomics and vaccine adverse event research. The ultimate goal of this project will be to use new resequencing strategies, combined with GWAS and CNV, to identify variants which may enable the prediction of risk, at the individual level, of developing ITP after vaccination.

Keywords: ITP, MMR/MMRV, Vaccine safety

doi:10.3121/cmr.2011.1020.c-c5-05

Virtual Data Warehouse

PS1-33:

CESR Proof-of-Concept Projects: Lessons Learned from the DCC Perspective

Karen Riedlinger, MPH, MT (ASCP); Kaiser Permanente Northwest, Center for Health Research

Background/Aims: Two multi-site proof-of-concept projects were initiated to demonstrate the viability and sustainability of the newly created Kaiser Permanente (KP) Center for Effectiveness and Safety Research (CESR). This presentation highlights lessons learned from the Data Coordinating Center (DCC) perspective as both projects conclude. **Methods:** Each proof-of-concept project, charged with finishing within one year of start-up, assembled teams, including investigators and programmers from each KP site. CESR's DCC supplied an overall project manager as well as consultation services on Institutional Review Board (IRB) issues, Data Use and Sharing Agreements (DUSA) and secure transfer of collected data. Other DCC services included SAS multi-site extraction programs for Virtual Data Warehouse (VDW) data and other non-VDW data. When non-VDW data was requested, the DCC also defined specifications for these tables. Cleaning of, Quality Assurance (QA) of and final combination of data from multiple sites was performed at the DCC. **Results:** The proof-of-concept projects helped to refine CESR processes for future projects. Lessons learned include 1) consult a VDW specialist early in the planning process, including before IRB submission; 2) describe data to be extracted and outlined in IRB applications and DUSAs broadly enough to include related variables to minimize the number of time-consuming modifications (e.g., ask for newborn statistics, not head circumference or *agpar5*); 3) create and use templates for extraction programs, for programs which combine multi-site data and simple statistical processes; 4) date stamp programs and returning

datasets and related output; 5) develop one location to allow investigators and programmers to access project related information and documents; 6) develop documents for tracking programs (including modifications) and datasets flowing between multiple sites and data coordinating center; and 7) invest time to create, update and communicate data flow diagrams (DFD) for each SAS program sent to sites. **Conclusions:** Documenting and analyzing the processes involved in proof-of-concept projects can help inform and contribute to efficiencies in future projects.

Keywords: Project management, Organization

doi:10.3121/cmr.2011.1020.ps1-33

C-B1-04:

Piloting Use of the HMO Research Network Virtual Data Warehouse for Pelvic Inflammatory Disease Surveillance and Research

Delia Scholes, PhD, MS¹; Marsha Raebel²; Onchee Yu, MS¹; Jane Grafton, BA¹; Hillard Weinstock, MD, MPH³; Catherine Satterwhite, MSPH, MPH³

¹Group Health Research Institute; ²Kaiser Permanente Colorado; ³Centers for Disease Control and Prevention

Background/Aims: National surveillance for pelvic inflammatory disease (PID) has been problematic for some time, as treatment and practice patterns have shifted away from inpatient settings and surgical procedures. Research and surveillance may be enhanced by use of automated data now available in many health plans. We evaluated the use of the HMO Research Network (HMORN) Virtual Data Warehouse (VDW) at two sites (GH and KPCO) to compare rates and 9-year trends in PID. As part of these evaluations, we explored the feasibility of using the VDW to obtain related chlamydia testing and infection information. **Methods:** Using standard ICD-9 codes for PID, we identified all cases occurring among GH women aged 15-44 years during 2000-2008. We then modified these programs for use by other HMORN sites by applying the VDW data dictionary standard variable naming conventions. We piloted these programs at KPCO. We calculated annual PID rates per 100,000 person years by 5-year age groups. We also sought information on chlamydia testing and test result within 7 days of the coded PID diagnosis date. The VDW includes relevant CPT codes for chlamydia testing, but microbiology test results are currently unavailable. Therefore, each site accessed their respective lab databases and used agreed-upon text strings to extract test result data. **Results:** PID rates were lower at KPCO than at GH. From 2000-2008, there were no significant PID linear trends at either plan. At GH, the rate was 562/100,000 person-years in 2000 and 526/100,000 in 2008; at KPCO, rates were 532/100,000 in 2000 and 485/100,000 in 2008. PID rates declined significantly only in the 20-24 year-old age group at KPCO. The proportion of cases receiving inpatient treatment was low (8%-16%). At both sites, approximately one-half of PID cases had evidence of chlamydia testing within 7 days. **Conclusions:** Given the national surveillance challenges for PID, the HMORN VDW is a potentially valuable resource that can add to PID surveillance and research by providing standardized data from multiple health plans across the U.S. With some expansion, the VDW could include additional information on characteristics of infection and aspects of care that have not previously been widely available.

Keywords: Surveillance, Automated data, Infectious disease

doi:10.3121/cmr.2011.1020.c-b1-04

PS1-12:

Load I2B2 with VDW Data

David Eastman, BS¹; Dustin Key, BS²; Gene Hart, BS²; Garth Arnold, BS²

¹Kaiser Permanente Southeast; ²Group Health Research Institute

The Virtual Data Warehouse (VDW) and Informatics for Integrating Biology & the Bedside (I2B2) are important components of our population-based health research collaboration efforts. Much hard work is put into cleaning and standardizing a health care institution's data for inclusion in the VDW. The I2B2 tool offers slick drag and drop, point and click query access to health care data. A method for leveraging the VDW for populating a Microsoft SQL Server version of the I2B2 database was developed using a

small set of SAS programs and a T-SQL script. The results at Group Health show the methodology to be an effective means of using the VDW to quickly populate I2B2.

Keywords: Data loading, I2B2, VDW

doi:10.3121/cmr.2011.1020.ps1-12

C-A1-01:

Using Data Transformations, Derived Values, and Cryptographic Functions to Protect PHI in the VDW

Daniel Ng, BSE, MBA; Kaiser Permanente Northern California, Division of Research

Background and Aims: HIPAA policies define values derived from PHI as also being PHI. Consequently, applying algorithmic functions to PHI has been viewed as having little benefit to research data. However, short of full de-identification, the use of transformed PHI may reduce compliance risk and increase security of routine data handling. Our aim is to: 1) present a general framework for evaluating methods of de-identifying/protecting PHI, and 2) evaluate how well selected mathematical functions, including common cryptographic functions, can enhance protection of PHI in the HMORN Virtual Data Warehouse (VDW). **Methods:** The methods used include a review of technical literature/material, from both within and outside traditional research disciplines, followed by analysis and application of findings to the issues addressed here, including: 1) reviewing both the regulatory and practical context for protecting PHI in research data; 2) developing a set of criteria to evaluate the benefits and costs of methods for PHI protection/de-identification; 3) reviewing the basic uses of general cryptography; 4) comparing/contrasting the needs of general cryptography with those of PHI protection in research data; 5) evaluating selected methods of protecting PHI against the proposed criteria. **Results:** The proposed framework for evaluating PHI protection methods includes five criteria: 1) effect on usefulness of data; 2) effect on ease of use or analytical efficiency; 3) net impact on data security; 4) system implementation costs; 5) negative effect on data quality. There are several methods for protecting PHI that can be easily implemented in the VDW, including: 1) the creation of linking variables that eliminate the need to routinely query PHI variables, such as service dates; 2) the MD5 cryptographic hash function, which can be used to obscure any PHI data and is relatively easy to implement in SAS, the standard analysis software platform used in the HMORN. **Conclusions:** Stewards of research data, like the VDW, should adopt the use mathematical functions, including cryptographic hash functions, to transform PHI into derived values. Such methods do not replace the need for full de-identification, but can enhance security and reduce compliance risk during routine data handling.

Keywords: VDW, De-identification, PHI

doi:10.3121/cmr.2011.1020.c-a1-01

C-B1-01:

Psychotherapeutic Prescription Patterns Across Health Plans in the Mental Health Research Network

Christine Stewart, PhD¹; Greg Simon, MD¹; Irina Miroshnik, MS²; Brian Ahmedani, PhD³; Matthew Lakoma, BS²; Dana Larkin, BS³; Karen Niefenfueller, BS⁴; Robin Whitebird, PhD, MSW⁴; Carsie Nyirenda, BS⁵; Arne Beck PhD⁵; Marvin Adams, BS⁶; Robert Davis, MD, MPH⁶; Mark Schmidt, BS⁷; Ameena Ahmed, MD, MPH⁷; Jamila Gul, BS⁸; Phillip Crawford, BS⁹; Frances Lynch, PhD⁹; Julie Liu, BS¹⁰; Karen Coleman, PhD¹⁰

¹Health Research Institute; ²Harvard Pilgrim Health Care Institute; ³Henry Ford Health Systems; ⁴Health Partners Research Foundation; ⁵Kaiser Permanente Colorado; ⁶Kaiser Permanente Georgia; ⁷Kaiser Permanente Hawaii; ⁸Kaiser Permanente Northern California; ⁹Kaiser Permanente Northwest; ¹⁰Kaiser Permanente Southern California

Background: The need for effectiveness and dissemination research is at least as great in mental health as in other areas of health care. The Mental Health Research Network aims to use the VDW for multi-site studies of mental disorder epidemiology, treatment patterns, or treatment effectiveness.

We expect that the existing infrastructure is satisfactory in some areas, while others will need improvement or expansion. We present early results from validation studies of enrollment and pharmacy data on patterns of psychotherapeutic drug use. **Methods:** We identified the relevant denominator populations using existing VDW variables indicating continuous enrollment with any drug coverage in 2009. After developing National Drug Code (NDC) lists for drugs used in the treatment of depression, bipolar disorder, attention-deficit disorder, and other mental health conditions, we calculated rates of use of these drugs across ten health plans in 2009. **Results:** Antidepressant exposure rates vary across sites, although our data include no diagnostic criteria and will include uses for other conditions. Initial results indicate overall antidepressant use rates vary more than two-fold between ten sites, from 62 to 162 exposures per thousand members. In addition, the preferred member of a class (for example SSRIs), and the rate of any pharmacy utilization also differs across sites, although to a lesser degree. **Conclusions:** Data validation is an iterative process where results are compared to what is known and expected and discrepancies are investigated. The considerable variation in overall antidepressant exposure rates at different sites is unexpected, based on reports of the incidence of depression (ref) and warrants further exploration. Possible explanations include use of antidepressants for other conditions, cultural attitudes toward mental health treatment in different populations, or delivery system practices. Alternatively, they may be artifacts of information systems or incorrect interpretations of the underlying data. These possibilities can be explored with a combination of local expertise from each delivery system and further analysis of VDW data.

Keywords: Mental health, Pharmacy

doi:10.3121/cmr.2011.1020.c-b1-01

PS1-38:

Benefits of Best Practice for Multi-Site National Drug Code Identification

Christine Stewart, PhD¹; Gene Hart, MS¹; Chester Pabiniak, MS¹; Brian Ahmedani, PhD²

¹Group Health; ²Henry Ford Health Systems

Background: Prescription drugs are identified in the VDW pharmacy file by National Drug Code (NDC.) Although the NDC constitutes a standard nomenclature, there is a no single, comprehensive resource for codes, either in the HMORN or beyond. Instead, each site maintains an “everNDC” file based on its own internal and external sources. Therefore, to identify NDC codes for drugs of interest for a multi-site study, the recommended process (VDW EverNDC Data Structure, Version 3) involves the collection and compilation of codes from each site. This is a frequent and time-consuming practice. We will present an exploration of the benefits of this approach in the experience of the Mental Health Research Network (MHRN.) **Methods:** We will compare the results of the recommended practice to those from a shortcut limited to RXnorm (RXN), an open-source library from the National Library of Medicine and First DataBank (FDB), a proprietary database licensed by Group Health, for antidepressants, lithium, anticonvulsant mood stabilizers, first and second generation antipsychotics, benzodiazepines, and stimulants. **Results:** In our process of identifying antidepressant NDCs from a list of 32 active ingredient names, RXN and FDB generated a list of 15401 codes. At Group Health (GH), our historical NDC file had 1653 additional codes. However, these additional codes only match 35 fills in our pharmacy file, the most recent being in 2002. For comparison, the full list of antidepressant NDC codes yielded an average of 480,000 fills per year at GH over the period 2000-2009. From nine other sites, we received an additional 1329 codes, increasing the number of codes by about 8%. **Conclusions:** Our initial results indicate that for antidepressants, site-specific NDC codes at GH represent rare and older prescriptions. The RXN/FDB method yielded ~90% of the total NDC codes identified at our site, and >99.99% of the prescriptions identified. We need to confirm this across other sites and other drug classes, but it is possible that a simplified process for defining a list of NDC codes for a drug class would be suitable for many multi-site studies.

Keywords: Pharmacy, National Drug Code, everNDC

doi:10.3121/cmr.2011.1020.ps1-38