monitoring of this vaccine's safety in early fall 2009. **Methods:** The VSD Project conducts near real-time monitoring of potential adverse events following vaccination for all new vaccines. Each week, each site updates dynamic data files, and the coordinating center analyzes all data cumulatively. Repeated analysis of the same data requires adjustment for multiple testing. To address this issue, VSD uses sequential statistical methods. Evaluation of H1N1 vaccine safety poses special challenges, including confounding by indication, seasonality, and new priority groups including pregnant women. We are using several analytic methods, including: (1) Self-controlled case-series, which compares exposed vs. unexposed time windows within the same individual; (2) Poisson, which compares the outcomes during the current season with expected rates based on historical data for seasonal influenza vaccine; and (3) Difference-in-difference, which compares the relative risk in exposed vs. unexposed time windows for H1N1 vaccine during the current season with the relative risk for seasonal influenza vaccine during cumulative prior seasons. **Results:** Pre-specified outcomes of interest include: Guillain-Barre syndrome, neuropathies, seizures, encephalitis, Bell's palsy, ataxia, anaphylaxis, spontaneous abortion, pre-eclampsia, stroke, myocarditis and wheezing. Historical rates of these outcomes were calculated for time windows after seasonal influenza vaccination in prior seasons (2000/01-2008/09). The sequential analysis methods above will be used, and relative and absolute risks will be estimated for each outcome. For outcomes found associated with vaccination in preliminary analyses, additional evaluation will be conducted using temporal scan analyses, logistic regression, and when appropriate, medical record review.

Keywords: H1N1 safety, Influenza A vaccine monitoring, Vaccine Safety Datalink Project

doi:10.3121/cmr.2010.943.c-b5-02

C-B5-04:

An Algorithm for Classifying Potential Cases of Amyotrophic Lateral Sclerosis from Electronic Health Records Using Decision Tree Analysis

Kathleen Albers, MPH¹; Stuart Hoffman, DO²; Judith Readon, MS²; Mark Hornbrook, PhD³; Alison Naleway, PhD³; Joanna Bulkley, PhD³; Susan Andrade, PhD⁴; Daniel Newman, MD⁵; David Nerenz, PhD⁵; Margaret Gunter, PhD⁶; James Burmester, PhD⁷; Priscilla Velentgas, PhD⁸; Daniel Strickland, PhD⁹; Stephen Van Den Eeden, PhD¹⁰

¹Kaiser Permanente; ²Geisinger Medical Center; ³Kaiser Permanente Northwest; ⁴Meyers Primary Care Institute; ⁵Henry Ford Health System; ⁶LCF Research; ⁷Marshfield Clinic Research Foundation; ⁸Harvard Pilgrim Health Care; ⁹Kaiser Permanente Southern California; ¹⁰Kaiser Permanente Northern California

Background/Aims: The CDC is developing a national amyotrophic lateral sclerosis (ALS) registry that would be based on electronically available medical records and claims. The HMORN contracted with the CDC to help examine the feasibility of doing this in locations (e.g., HMOs) where complete longitudinal records are more likely to be available. Any final system used by the CDC will require an algorithm to classify putative cases. Our aim was to develop an automated algorithm to classify putative cases of ALS and other motor neuron diseases using electronic health records.

Methods: We identified potential cases of ALS by review of inpatient, outpatient, other provider services, mortality and pharmacy data for any evidence of ALS or, more generally, motor neuron disease (MND) from 2001 to 2005 at Kaiser Permanente, Northern California. The records of each individual were reviewed to classify into five categories: probable ALS, not ALS, other MND (not ALS), insufficient information, or unclear. Classification And Regression Tree (CART) analyses were used to create an algorithm based on a comprehensive set of variables that may predict the likelihood of having true ALS (or, conversely, ruling them out). **Results:** Multiple algorithms were developed with key variables being number times an ALS code appeared in the record, physician-type making diagnosis, alternative diagnoses, mortality records, and length of follow-up. In our best model, we were able to correctly classify 99.6% of those originally determined to be probable ALS. However, there remain a substantial number of individuals classified by review or algorithm in the unclear, insufficient, or other MND categories. In addition, with this analytic approach, the "costs"

of misclassification can be modified, and we will present alternative algorithms with a discussion of the benefits and disadvantages of each. Validation analysis results conducted in a second HMO will be presented. **Conclusions:** An algorithm was developed that appropriately classifies the vast majority of probable ALS cases. However, given the lack of detail in some records and the difficulty in diagnosing some clinical cases, there will be some misclassification when based solely on electronic medical records.

Keywords: Amyotrophic Lateral Sclerosis, Electronic health records, ALS Classification

doi:10.3121/cmr.2010.943.c-b5-04

PS1-15:

A Method for Discovering Variant Spellings of Terms of Interest in Clinical Text

David Carrell, PhD¹; David Baldwin, BS¹

¹Group Health

Background and Aims: Mining information from unstructured clinical text requires knowledge of synonyms and variant spellings corresponding to the term of interest. Lexicons such as the Unified Medical Language System identify recognized synonyms but are less helpful in identifying variant spellings or locality-specific language. Natural language processing algorithms employing dictionary look-ups frequently miss terms because of variant spellings, accounting for as much as 25% of algorithm failures. This problem is not easily resolved deductively or through domain expertise. We describe an inductive, empirically-driven method that identifies synonyms/variant spellings of a concept of interest. **Methods:** The conceptual basis for this method is the principle that terms with similar meanings tend to be found in similar textual contexts. Accordingly, sets of words found with high frequency in the context of two synonyms will overlap with high probability. Call the word of interest the target term; call the high-frequency words around it context terms; call words most likely to be synonyms/variants synonym candidates. A four-step process can be used to identify variants. Step 1: Identify the target term's highest frequency context terms. Step 2: Identify terms found in high frequency near each of the context terms. Step 3: Eliminate Step 1 & 2 terms appearing more often outside the context than in it. Step 4: Manually review remaining synonym candidates to identify true synonyms/variants. Using SAS we illustrate the method for target term HER2/NEU in Group Health chart notes for 2008, representing 200,913,281 lines of clinical text. **Results:** Four of the top 20 context terms near HER2/NEU had ratios of in_context_occurrence/out_of_context_occurrence > 1: OVEREXPRESSION, ONCOGENE, OVEREXPRESSED, C-ERBB-2. These context terms occurred 1196, 617, 540, and 170 times, respectively, and produced lists of 205, 119, 143, and 54 of their own context terms, respectively (excluding words with frequency < 2). Manual review yielded a final set of 23 synonyms and spelling variants, including HER-2, HER/2, HER-2-NEU, HER2/NUE (note spelling: "NUE"), FISH, IHC, and C-ERBB-2 among others. **Conclusion:** The method describe provides an easily implemented approach to identifying synonyms/variants of terms of interest that can be used to enhance strategies for identifying concepts in clinical text.

Keywords: Identifying variant spellings, clinical text, medical terminology
doi:10.3121/cmr.2010.943.ps1-15

PS1-16:

Disenrollment Rates Among HMORN Sites by Age, Gender, and Medical Condition

Jennifer Ellis, MBA, MSPH¹; Marsha Raebel, PharmD^{1,2}; Addie Clark, BA¹; Susan Andrade, ScD³; Robert Davis, MD, MPH⁴

¹Kaiser Permanente Colorado; ²University of Colorado; ³Fallon/Meyers; ⁴Kaiser Permanente Georgia

Background: It is often desirable to include enrollment and disenrollment (or, conversely, retention) information in grant applications. Overall disenrollment has been evaluated as part of Virtual Data Warehouse quality

assurance work, but disenrollment by member characteristics has been incompletely characterized. This information is important in longitudinal outcomes research. **Aim:** To calculate member disenrollment across HMORN sites by age, gender, and selected disease states. **Methods:** Baseline membership cohorts were identified for 8 consecutive years (2000-2007) at 9 HMORN sites. Members were included in a baseline cohort year if they were enrolled for at least 305 days in that year. Disenrollment, defined as having fewer than 305 days of membership in a year, was assessed at 1, 3, and 5 years for each cohort. Members who died before the end of each assessment period were censored. Comorbidity was assigned based on ICD9 diagnosis code(s) for diabetes, asthma, heart failure, hypertension, hyperlipidemia or depression over a three-year period ending with the end of the cohort baseline year. SAS code was prepared at one site for distribution to other sites to run against VDW tables. **Results:** While overall disenrollment varied across the participating sites, within-site disenrollment changed little from year to year. In general, 1-year disenrollment rates varied from 7% to > 22%; 5-year disenrollment rates ranged from < 30% to > 60%. Using one site as an example, peak 1-year disenrollment was seen among the 20-29 year old age group; lowest rates were seen among those age 60 years and older. Similar patterns were seen at 5 year follow-up, although differences between highest and lowest rates were smaller. Disenrollment rates were slightly higher among males. Disenrollment among members with any one of the measured diseases was lower than among members without that condition. **Conclusions:** Disenrollment (and retention) rates vary by age and disease state and less so by gender. While overall disenrollment is useful for many purposes, rates by member characteristics can be essential while planning studies of specific populations.

Keywords: HMORN, Participant characteristics, Enrollment studies
doi:10.3121/cmr.2010.943.ps1-16

PS1-17:

Herpes Zoster in Children

Sheila Weinmann, PhD, MPH¹; Colleen Chun, MD¹; Karen Riedlinger, MPH, ASCP¹; Michelle Roberts, MSN, FNP¹; Mary Rix, RD¹; Mona Marin-Nelson, MD²; Scott Schmid, PhD²

¹Kaiser Permanente Northwest; ²Centers for Disease Control and Prevention

Background and Aims: Vaccine-strain herpes zoster (HZ) is a potential adverse outcome of varicella vaccination. Data are needed on the incidence of and risk factors for vaccine-strain HZ among children <18 years of age. The Centers for Disease Control and Prevention (CDC) funded a study at Kaiser Permanente Northwest (KPNW) to determine the number and proportion of pediatric HZ cases caused by varicella vaccine, assess positive predictive value (PPV) of physician diagnosis of HZ, and compute incidence of vaccine and wild-type HZ in the KPNW population. **Methods:** We used electronic data to identify all pediatric medical office visits with a HZ diagnosis. During the visit, the provider collected a lesion specimen and completed a questionnaire. The study nurse called to interview the parent and, if no specimen was previously collected, schedule a home visit to obtain it. A follow-up questionnaire on severity and duration of HZ was returned four weeks after rash onset. Specimens were sent to the National Varicella-Zoster Virus (VZV) Laboratory at the CDC for PCR analysis to confirm presence of the virus and classify it as wild-type or vaccine-type. **Results:** From 5/2005 to 3/2009, we enrolled 308 subjects in the study. Two-hundred eighty-one (96%) had adequate lesion specimens for PCR analysis. We found VZV in 75% of adequate specimens; 69% were wild-type, while 9% were vaccine-type and 1% (2) were vaccine:wild-type recombination. Among vaccinated subjects, 69% had wild-type HZ. The PPV for provider diagnosis of 'definite HZ' was 83%: 85% for unvaccinated subjects and 76% for vaccinated subjects. Incidence rates for vaccine-strain and wild-type VZV in plan children age 0-17 will be presented. **Conclusions:** Vaccine-strain VZV was found in a minority of HZ cases in this age group. We observed recombination of vaccine-strain and wild-type viruses in two subjects. Provider diagnosis of HZ was good overall.

Keywords: Herpes zoster, Diagnosis of herpes zoster, Varicella vaccination
doi:10.3121/cmr.2010.943.ps1-17

PS1-22:

Premature Birth, Initial Hospital Length of Stay, and Costs

Mary Ann Blosky, MS, MHA¹; Zhengmin Qian, MD, PhD¹; G. Craig Wood, MS¹; James Betoni, DO¹; Libby Black, PharmD²; Andrea Wary, BS, RN, MS¹; Harry Mateer, MD¹; Walter Stewart, PhD, MPH¹

¹Geisinger Health System; ²GlaxoSmithKline

Background: Preterm birth disparities with their associated geographic, socioeconomic status and race/ethnicity factors are among the most widely recognized and least understood phenomena in the study of reproductive health. Previous studies have focused on total cost; few studies have disaggregated total charges into their components for both the preterm infants and their mothers or compared such costs to full-term births. **Aim:** To quantify and disaggregate initial hospital charges among preterm infants and their mothers compared to full-term controls. **Methods:** The study populations were comprised of mother-infant dyads that included 746 preterm cases and 3372 controls. Data were obtained from the Geisinger outpatient electronic health record, Geisinger Medical Center billing records, and the inpatient pharmacy billing database in the Geisinger Health System. Measures were derived for hospital charges, lengths of stay for infants and mothers, and the mean delivery hospitalization charges for the subset of infants and mothers who had the most frequent diagnoses, procedures, and drug classes. **Results:** The mean hospitalization charges (\$103,000 versus \$7,000) and the length of stay (16.1 days versus 2.3 days) were substantially higher for preterm infants versus controls. Extreme preterm infants (< 26 weeks) accounted for 1.3% of the total preterm births and approximately 20% of the total initial hospital charges and near term infants (33-36 weeks) accounted for approximately 14% of total births and 30% of the total charges. The most frequent inpatient medical conditions were five to 17 times as frequent among preterm infants as among the controls. The most frequent procedures and drug classes were also generally more prevalent among preterm infants. Preterm infants had generally higher charges related to inpatient diagnoses, procedures and medication utilization. The mean hospitalization charges (\$23,000 versus \$15,000) and length of stay (i.e., 5.1 days versus 2.3 days) was also higher among the preterm mothers than controls. Medication utilization was compared between the two groups and preterm mothers had higher charges related to inpatient diagnoses, procedures and medication utilization compared to controls. **Conclusions:** The most extreme premature (4.3%) of preterm births accounted for 47% of the total preterm hospital charges. Overall costs increased as gestational ages decreased.

Keywords: Preterm birth disparities, Preterm hospitalization costs, Reproductive health

doi:10.3121/cmr.2010.943.ps1-22

PS1-23:

The Environmental Scan: A Valuable Community Health Research Tool

Virginia Scobba, MLS, MA; Group Health Cooperative

Background and Aims: The environmental scan (ES) focuses on the identification, collection, and translation of external information that may potentially influence an organization's decision making process. An ES allows an organization to address social, economic, political, and technical issues that may not be readily apparent and enables a comparative evaluation of the organization's internal strengths and weaknesses. Through understanding external elements and internal processes, the organization can respond in a more auspicious and productive manner. Health researchers are employing ESs to build on established knowledge with the goal of improving the design and effectiveness of community health programs. **Methods:** There is no one correct methodology for conducting an ES, but typically, ESs are characterized by several features. Initially, a broad spectrum of issues is examined, involving social, political, technological and other trends. Second, information may be gathered from a variety of sources, including literature reviews, surveys, interviews, focus groups, and site visits. Third, critical analysis and dissemination of collected data is promoted by the recruitment of experts from both inside and outside the fields being scanned. ESs can utilize a wide sweep of information sources, ranging from formal, established data, such as