distributions of “true” physician scores (means of 73, 88, 95% satisfaction and standard deviations of log odds = 0.5, 1.0 and 1.5), we simulated 5,000 physicians and assigned each a true score and rank within these distributions. We then tested various patient sample sizes (N) from 10-100, and repeated 1,000 simulations under each scenario to calculate 95% inner ranges of observed ratings and ranks for each physician. True and observed ranks were compared and examined as a function of underlying distribution and N. **Results:** The precision of an individual physician’s percentile rank increases dramatically with 3 factors: increase in N; increase in variance of true physician scores; and decrease in overall mean physician score. Precision is also greatest for the best and worst physicians and less precise in the mid-range. In the best case scenario tested (mean 73%, log odds SD = 1.5, N = 100), physicians with an observed rank as low as the 68th percentile were likely to be equivalent to physicians with an observed rank above the 90th percentile. In the worst case scenario (95% mean, log odds SD = 0.5, N = 10), physicians with a “true” 90th percentile rank were likely to receive a ranking as low as the 18th percentile. **Conclusions:** If physician rankings based on patient satisfaction scores are used to measure and incentivize quality, rankings should either be based on a measure with maximal variation among physicians, or incentives should be based on ranges that reflect uncertainty rather than absolute rankings.

Keywords: Patient satisfaction; Quality metrics

doi:10.3121/cmr.2014.1250.ps1-34

Vaccine/Infectious Diseases

PS1-36:

Developing Rapid Molecular E.coli Clonal Diagnostics for Community Laboratory Settings

Delia Scholes¹; Kim Riddell²; Jackie Saint-Johnson¹; Peggy Rogers²; Veronika Tchesnokova³; Evgeni Sokurenko³

¹Group Health Research Institute; ²Group Health; ³University of Washington

Background/Aims: E.coli infections of the urinary tract (UTI) are among the most common infections worldwide. U.S. costs of UTI care are >\$1B annually. Recently, there has been rapid worldwide emergence of resistant E.coli, which increases the percentage of serious events and the difficulty of care. One contributing factor may be that usual care typically involves presumptive antibiotic treatment, with antimicrobial sensitivity results available only after 1-2 days. Inappropriate treatment during that interval is suboptimal for the patient and facilitates E.coli resistance. We conducted a 1-year pilot of a potential new diagnostic technology to test for E.coli infection at the clonal level by: 1) optimizing PCR methods to directly and rapidly test urine samples for two E.coli sub-strains (clones) in a community lab setting (Group Health [GH]); 2) using this method to identify two highly-resistant sub-strains in GH samples; and 3) comparing usual care prescribing patterns to potential prescribing based on sub-strain results, focusing on the “drug-bug mismatch”. **Methods:** A research strategy developed at the University of Washington to diagnose E.coli sub-strains was adapted to a commercial PCR platform for two selected sub-strains: ST131 and ST69. We tested 653 urine samples for presence of ST131 and ST69 and reviewed EMRs to collect the clinical diagnoses and antibiotics prescribed at the visit. **Results:** In the GH lab, the PCR-based diagnostic test required about 1.5 hrs. Prevalence (UW-confirmed) was 10% (65/653) for ST131 and 5% (33/653) for ST69. Of 200 EMR reviews to date, these two clones represent 65% (17/26) of fluoroquinolone and 17% (5/29) of trimethoprim/sulfa resistant isolates. Five percent (9/200) of patients were empirically prescribed therapy for which the isolate was resistant. Adjustment of the original prescribed antibiotic based on the E.coli clonal identity would reduce the drug-bug mismatch by 55%. **Conclusions:** These methods delivered accurate, economical results, at the clonal level, faster than usual lab methods. The two E.coli sub-strains are found in a sizable percentage of the isolates that are resistant to the main E.coli therapies. Clonal-based diagnostic strategies can offer opportunities to improve antibiotic prescribing and to impact the continued spread of antimicrobial resistance.

Keywords: Rapid diagnosis; Antibiotic resistance

doi:10.3121/cmr.2014.1250.ps1-36

Virtual Data Warehouse

A4-1:

Validating Electronic Health Record Demographic Data Using Self-reported Data from the Autism Registry Survey

Phillip Crawford¹; Frances Lynch¹; Lisa Croen²; Karen Coleman³; Ashli Owen-Smith⁴; Vincent Yau²; Kathryn Pearson¹

¹Kaiser Permanente Northwest; ²Kaiser Permanente Northern California;

³Kaiser Permanente Southern California; ⁴Kaiser Permanente Southeast

Background/Aims: The use of administrative patient data via the electronic health record (EHR) is very important in research. It’s critical we have a sense of the validity and relative accuracy of key data in this widely available data source. The HMORN Virtual Data Warehouse (VDW) is an example of a large repository of administrative data used in numerous studies. The recent Mental Health Research Network (MHRN) Autism Spectrum Disorder (ASD) Registry Survey was hosted by four HMORN sites inviting Kaiser Permanente members to complete a Web-based questionnaire on behalf of their child identified as having an ASD diagnosis in the EHR. **Methods:** In addition to an extensive battery of questions regarding ASD, the survey collected data on a number of demographic data elements also available in the VDW. These are: age, gender, race/ethnicity (child), income (household) and education (household adult). A total of 1155 adults responded to the ASD Registry Survey. These records were matched with demographic data from the VDW for the same child. **Results:** Preliminary examination of race-ethnicity and gender data shows that there’s a good to excellent level of agreement between the two sources when data are non-missing in both data sources. A total of 992 records had non-missing race/ethnicity data from both data sources. In 81.1% (805/992) of the cases, both data sources agreed (kappa = 0.68, CI = 0.64-0.72). The categories that had the highest level of agreement are White, Black, and Asian, while the Hispanic and multi-racial groupings had a comparatively much lower level of agreement (46.9%). For gender, level of agreement was very high (99.0%, 1118/1129, kappa = 0.97, CI = 0.95-0.99). **Conclusions:** In this study, race in the electronic health record was a very accurate measure for major race categories but was far less accurate when reporting the emergent and important multi-racial category and Hispanic ethnicity. Gender had an excellent level of agreement between the two sources. Age, gender, and race-ethnicity are key covariates for many studies analyzing EHR data. Income and education can serve to illuminate socioeconomic factors very relevant to health care research. Having a sense of the validity and accuracy of these data is crucial to the research process.

Keywords: Data validation; Demographic data

doi:10.3121/cmr.2014.1250.a4-1

A4-2:

ICD-10 CM Transition across Three Research Centers

Suzanne Gillespie¹; Michael Allison¹; Catherine Cleveland¹; Don Bachman¹; Xiuhai Yang¹; Mark Schmidt²; Alan Bauck¹

¹Kaiser Permanente Northwest; ²Kaiser Permanente Hawaii

Background/Aims: The United States currently uses version nine (ICD-9) for diagnosis and procedure coding. Federal regulations require that the US adopt the next version (ICD-10) by 10/1/2014. The new version expands the number of diagnosis codes from about 13,000 to 68,000, providing more granularity to classify diseases. Challenges in moving to ICD-10 include: (a) the crosswalk between versions is a many-to-many match; (b) the relationship between the old and the new codes will sometimes be complex; and (c) it will encompass system-level changes such as expanded field sizes and support for alphanumeric characters. Our challenge is to implement the transition across three research centers at Kaiser Permanente Northwest, Southeast and Hawaii. **Methods:** The three research centers need to account for the new ICD-10 code set and any new data structures introduced by either the EMR or Research Data Warehouse. A smooth transition between coding systems requires working with researchers across the three regions. This includes identifying corresponding ICD-10 codes that correctly identify the research cohort members and remediating the existing programing logic. Kaiser Permanente’s Center for Health Research (CHR) has developed a detailed plan to accomplish these two goals. CHR staff are partnering with the Kaiser

Health Plan ICD-10 implementation team in the development of our project plans. The standard source to be used for mapping at the three Kaiser Research centers is the General Equivalence Mappings based Crosswalk Query Tool, which Kaiser is implementing to support transition mapping needs. **Results:** The CHR ICD-10 remediation team has completed the Framing and Planning phases including identifying programs/tables affected, sharing remediation solutions across the Kaiser regional research teams, and meeting with the Health Plan to coordinate efforts. We have developed a project remediation checklist and will report on the development of validation testing plans and the refinement of vanguard project code mappings. **Conclusions:** In the transition to ICD-10, it is essential to have a well-formed plan and to identify and include all stakeholders in the process. It is also beneficial to share tools and ideas between research centers to improve the overall methods.

Keywords: ICD-10; System-level changes

doi:10.3121/cmr.2014.1250.a4-2

A4-3:

Utilization-based Proxy Enrollment Versus Standard HMORN VDW Enrollment: A Pilot Validation Study

Irina Haller¹; Brian Johnson¹; Karen Riedlinger²; Pinky Barua³; Terese DeFor⁴; Paul Hitzl¹; Roy Pardee⁵; David Tabano⁶

¹Essentia Institute of Rural Health; ²Kaiser Permanente Northwest; ³Geisinger Health System; ⁴HealthPartners; ⁵Group Health Research Institute; ⁶Kaiser Permanente Colorado

Background/Aims: Most HMORN members offer both insurance coverage and health care. Health plan enrollment provides a well-defined population denominator for HMORN-based research. The HMORN common data model combines electronic data, routinely collected in health care delivery or claims processing, into the Virtual Data Warehouse (VDW). Sites with health plans capture insurance enrollment in the VDW enrollment file. Recently HMORN included sites that deliver health care without offering insurance. Since health plan membership is unknown, population denominators must be determined using alternate methods. This study validated a utilization-based proxy enrollment (PE) using standard VDW enrollment (SE) in 5 HMORN sites with Epic EHR and SE files. **Methods:** The utilization-based algorithm defines PE start at first of two non-ancillary health care visits (at least one being a primary care visit) separated by at least 90 days. PE ends at death or at the last qualifying visit if there is no utilization in the following 3-year period. PE files were built at each site by applying utilization-based algorithm to base tables created from Clarity. The PE and SE extracts included study ID, age, gender and the start/end of enrollment periods between 2000 and 2012, based on availability of Clarity data at the sites. The agreement between PE and SE was evaluated using differences in start/end of the first enrollment periods (days). **Results:** The differences between PE and SE starts varied by site with greater variation in children (<20 years) and older adults (64+ years). The differences varied by gender, but the differences between the genders were smaller in the young and the old. The differences were larger for males between ages of ~16 and 64, indicating less utilization by males than by females. The differences between PE and SE ends were generally negative across ages, indicating that PE extended beyond SE. Gender differences between PE and SE starts were similar to differences between PE and SE ends. These empirical results were confirmed by multivariate regression modeling. **Conclusions:** Agreement between PE and SE could be improved using additional parameters, as well as possible adjustment of the time lag for PE end in the utilization-based algorithm.

Keywords: Population denominator; Utilization-based enrollment

doi:10.3121/cmr.2014.1250.a4-3

A4-4:

Do Projects Have Complete Data Capture for Their Study Populations?

Donald Bachman¹; Mark Hornbrook¹; Debra Ritzwoller²; Roy Pardee³; Alan Bauck¹

¹Kaiser Permanente Northwest; ²Kaiser Permanente Colorado; ³Group Health Research Institute

Background/Aims: Investigators often assume that when patients are enrolled in a health plan, they have complete capture of utilization data from their health plan sources. This assumption may not always be true as patients have many incentives to choose care at multiple settings (convenience, price, residence, insurance type, drug coverage), which can give rise to missing data. Complete capture of medical data for population-based research is crucial to our ability to identify populations who have not had particular exposures or outcomes. In case-control designs, where the control group does not have a specified exposure, we measure this condition by absence of data. To address this problem, the VDW enrollment work group created a new enrollment variable, called "Outside_utilization", designed to identify members suspected of incomplete capture of encounters or pharmacy fills. This work reports on a quality assurance analysis of the extent and nature of the data gaps at different sites. **Methods:** For V3, HMORN sites added new VDW variables in their enrollment file including the "Outside_utilization" variable. This variable identifies populations suspected of having incomplete health care utilization capture. Since the reasons for incomplete data capture vary among the sites, the methods for identifying members with incomplete data capture were determined by the local site data managers. The authors distributed a program that computed utilization rates for specific cohorts for "complete" and "incomplete" data capture populations. We also computed rates for those on high deductible plans. We compared differences in these rates by year and site. In addition, we conducted a survey to determine how the incomplete populations were identified for each participating site. **Results:** Some sites clearly identify populations that have incomplete capture of data using the "Outside_utilization" variable. At other sites, the difference in rates is less apparent. The sites that can distinguish patients with incomplete versus complete data have certain common definitions used when designing the "Outside_utilization" variable. **Conclusions:** The "Outside_utilization" variable identifies populations with incomplete data capture at some sites. We recommend that projects use this variable to exclude populations suspected of incomplete data capture when computing population-based utilization rates.

Keywords: Data Population

doi:10.3121/cmr.2014.1250.a4-4

PS1-5:

Converting to ICD-10 Clarity Structures

Jamila Gul¹; Srivardhan Chimmula¹

¹Kaiser Permanente Northern California

Background/Aims: The transition to the ICD-10 code represents a huge undertaking for every Health Care Organization including Analytic teams. In Kaiser Northern California we have effectively engaged KP HealthConnect Application Coordinators, Business Partners, and Regional IT Leads to create and build ICD-10 impact awareness through education and leveraging our partner's vast knowledge to identify critical areas of impact, issues, risks, assumptions and high level business objectives to support a successful transition to the ICD-10. **Methods:** To accommodate ICD-10, Epic has made significant changes to Clarity in their 2010 release including adding new fields, tables and views. These changes are designed to aid in reporting appropriate codes as well as allowing for historical reporting and trending. The new format of the codes will also be part of the changes to Clarity. Changes have been made to three main Epic Master Files: (1) Diagnosis masterfile - maintains a master set of diagnosis records. These diagnosis records can be industry-standard ICD codes used for billing or clinical terms used for clinical encounter documentation, but mapped to industry-standard ICD codes. The diagnosis code changes primarily impact clinical encounter documentation (problem list and a Dx association with encounters and orders), HIM coding, claims for all types of encounters. (2) ICD Procedure masterfile - maintains a master set of ICD hospital procedure records. ICD procedure codes in addition to supporting hospital inpatient billing. (3) Surgical Procedure masterfile - maintains a master set of records for each surgical procedure that can be performed with an operating room facility. It also contains procedure preference records, primarily used for scheduling surgical procedures. We would like to take a real study and switch the codes from ICD9 to ICD10. **Results:** By sharing a real life example and sample codes we are hoping to pass our learning and knowledge to others hoping to help eliminate some of the errors that we made. Also, discuss outstanding challenges and issues that Epic has not resolved. **Conclusions:** The purpose