B4-4:

A Novel Technique for Analysis of Uncontrolled Confounding in Non-experimental Comparative Effectiveness Research

Caroline Thompson¹; Onyebuchi Arah²

¹Palo Alto Medical Foundation Research Institute; ²University of California Los Angeles

Background/Aims: Comparative effectiveness research (CER) investigates the effects of treatments and practices, thus requires causal inference. Routine data such as billing, pharmacy or EHR, while often incomplete on important confounding variables, are the usual sources of information for nonexperimental CER. The lack of randomization introduces important considerations regarding uncontrolled confounding, especially in large datasets, which potentially magnify systematic error. Yet, quantitative bias analysis in CER is not common practice. In this paper we formalize and demonstrate easy-to-implement record-level simulation techniques for analysis of uncontrolled confounding in cancer treatment CER. **Methods:** We use recent advancements from the causal theory and risk analysis literature, specifically directed acyclic graphs (DAGs), and Monte-Carlo simulation techniques to introduce a novel form of record-level missing variable imputation that can be implemented during the core data analysis stage, making bias analysis more accessible using standard statistical packages. Further, our methods take into account varying levels of uncontrolled confounding by research center, or other clustering variable that may predict the level of unknown information in the dataset, and are specifically designed for implementation in large datasets, or data from multiple sources. We demonstrate these methods with two example sensitivity analyses of uncontrolled confounding in cancer treatment CER. **Results:** Our methodology highlights the underlying causal model assumed for the main analysis in CER. Our technique uses the observed data lacking important confounding variables and informed estimates of the unmeasured variables to impute missing variables. The new variables now have a joint distribution with the observed data that would have been the case had they been observed fully under the assumed interrelationships. This technique is intuitively in line with the missing data framework and inference using partially observed distributions. **Conclusions:** Sensitivity analysis for uncontrolled confounding is feasible and indispensable for CER. Unlike existing formula-intensive external adjustment techniques, the new technique can be implemented during core data analysis, is not outcome model specific, is at most semi-parametric and requires no esoteric software. Quantitative uncertainty analysis should be routine practice for CER in large observational data sources. Flexible methods accessible to all researchers should be a priority in this growing area of research.

Keywords: Bias analysis; Comparative effectiveness research

doi:10.3121/cmr.2014.1250.b4-4

PS1-10:

How Can the Same Practice Be Classified as Having 2 and 900 MDs? NAMCS Data Collection in a Changing Ambulatory Care Environment

Meghan Halley¹; Katherine Gillespie¹; Katharine Rendle¹; Harold Luft¹

¹Palo Alto Medical Foundation Research Institute

Background/Aims: Since 1973, the National Ambulatory Medical Care Survey (NAMCS), administered by the National Center for Health Statistics (NCHS) has been widely used in studies of ambulatory care. With the growth in large multispecialty practices – including many members of the HMORN – there is a need to understand how NAMCS data are collected and whether current processes yield accurate and reliable data. NAMCS collects data from physicians about their practices and abstracts a sample of patient visit records. This study reports on the physician component. **Methods:** In collaboration with NCHS, nine physicians were randomly sampled from a multispecialty clinic using standard NAMCS recruitment procedures; eight physicians were eligible and agreed to participate. Using their standard protocols, three Field Representatives (FRs) conducted NAMCS physician interviews while a trained ethnographer (MH, KR) observed and audio-recorded each interview. Transcripts and field notes were analyzed using a grounded theory approach to identify key themes. **Results:** Data have been collected and analyzed. They are currently undergoing standard confidentiality

review by NCHS. However, this process has been delayed due to the government shutdown. We fully anticipate that results will be released in time for presentation at the HMORN conference. **Conclusions:** Though we are precluded from disseminating results at this time, we will provide a full report of our results in our HMORN conference presentation.

Keywords: NAMCS; Survey research methods

doi:10.3121/cmr.2014.1250.ps1-10

PS1-11:

A Comparison of Electronic Medical Records vs. Claims Data for Rheumatoid Arthritis Patients in a Large Healthcare System: An Exploratory Analysis

Daniel Maeng¹; Joseph Boscarino¹; Walter Stewart²; Xiaowei Yan¹; Nancy Steigerwalt¹

¹Geisinger Health System; ²Sutter Health

Background/Aims: Electronic medical records (EMR) and claims data offer two potential data sources for researchers to examine healthcare utilization patterns and cost of care. In particular, combining the clinical and epidemiological variables typically available in EMR with cost information available in the claims data is not only intuitively sensible but also increasingly more feasible with growing standardization of EMR across healthcare delivery systems. **Methods:** In this study, we compare EMR and claims data within a cohort of rheumatoid arthritis patients who received care from Geisinger Health System and had concurrent Geisinger Health Plan (GHP) coverage. We also develop a cost “imputation” method to obtain GHP claims-based cost estimates within EMR even for those who did not have GHP coverage. **Results:** EMR-based estimated means of total cost of care and utilizations tend to substantially underestimate the total cost of care. In particular, EMR substantially understates emergency department (ED) visits [4% (125 of 3,131) in EMR vs. 11.2% (352 of 3,131) in claims], X-rays [4% (125 of 3,131) vs. 22% (689 of 3,131)], and CT scans [5.1% (160 of 3,131) vs. 7.3% (229 of 3,131)]. Use of biologic agents appear to be slightly higher in EMR than in claims [7.2% (226 of 3,131) vs. 6.7% (210 of 3,131)], although the difference is not statistically significant. The correlation between log-transformed EMR-based cost of care and log-transformed claims-based cost of care was modest ($R^2 = 0.81$). **Conclusions:** The findings confirm that there is significant disagreement between EMR and claims data and suggest that each represent a different set of “reality.” The main source of such discrepancies between EMR and claims appears to be missing utilization of certain types of care in EMR. In particular, discrepancies seem greater for the types of services for which patients have more alternative choices in the area. Claims data reflects health plan’s coverage decisions and utilization management, while EMR reflects clinicians’ decisions and practice patterns. Thus, researchers should carefully consider which “reality” they are interested in capturing in their analysis. Lastly, the fact that both EMR and claims are collected for clinical and administrative purposes, not for research purposes, must be emphasized.

Keywords: Cost of care; Electronic medical records

doi:10.3121/cmr.2014.1250.ps1-11

PS1-34:

Sensitivity of Patient-reported Physician Percentile Rankings to Inter-physician Variability and Patient Sample Size

Jove Graham¹; Daniel Horwitz¹

¹Geisinger Health System

Background/Aims: Patient satisfaction is increasingly being recognized as a desirable measure of physician quality and is used for quality-based financial incentives. Patient satisfaction surveys such as the CAHPS, however, typically exhibit ‘ceiling effects’ where most patients report maximal satisfaction, and so physicians are often ranked based on their percentage of maximum-satisfaction responses (“percentile top box scores,” 0-100%) rather than on raw scores. Even so, physicians express concern that low response rates or tight clustering of underlying scores can have unknown effects on rankings and detrimental consequences. This study used simulation to report the effect of inter-physician variability and sample size on survey-based physician rankings. **Methods:** Assuming 9 different underlying

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