combining and sharing data across EMR systems and across institutions. In short, CER requires aggregation and analysis of disparate electronic clinical data sources held by different institutions, each with its own representation of relevant events and concerns for protecting data. **Methods:** The AHRQ-funded CER-HUB project will provide infrastructure for flexible, rapid, and efficient utilization of distributed, heterogeneous, electronic clinical data for CER analyses. The CER- HUB will provide a means for collaborative development and validation of standardized, study-specific processors of distributed clinical data. These processors extend a classification technology (MediClass) that uses both free-text and coded data, enabling comprehensive capture of the clinical events of interest. A central website hosts these tools including a library of processors that can be iteratively refined, validated, and reused by any end-user. **Results:** Dr. Hazlehurst will describe 18 months of progress toward CER-HUB study aims, including progress reports on two demonstration studies (addressing the effectiveness of smoking cessation services and step-up therapies for asthma control) being conducted by a consortium of six geographically-distributed and demographically-diverse health systems. He will demonstrate the processes of using the CER-HUB including (1) policies and methods for creating a new research study, (2) building and validating a study-specific data processor addressing study questions, and (3) generating standardized study data from multiple health care systems while protecting patient confidentiality. **Discussion:** CER-HUB users collaboratively build software applications (specializations of the base MediClass system) that process EMR data within their respective healthcare organizations, creating standardized datasets that can be pooled for analysis. The CER-HUB provides an open platform for researchers, program evaluators, and health system managers to use EMR data to answer diverse healthcare study questions.

Keywords: Natural Language Processing; Comparative Effectiveness Research

doi:10.3121/cmr.2012.1100.cc4-01

CC4-02:

Comparative Effectiveness of Chemotherapy Regimens for Advanced Lung Cancer

Debra Ritzwoller[1]; Lawrence Kushi[2]; Nikki Carroll[1]; Mark Hornbrook[3]; Elizabeth Loggers[4]; Thomas Delate[1]

[1]Kaiser Permanente Colorado; [2]Kaiser Permanente Northern California; [3]Kaiser Permanente Northwest; [4]Group Health Cooperative

Background/Aims: Research has demonstrated that platinum-based chemotherapy may prolong short-term survival and improve symptom control in patients with advanced (stage IIIb-IV) non-small cell lung cancer (NSCLC). Newer chemotherapy agents (e.g., taxanes, antimetabolites, monoclonal antibodies, drugs targeting EGFRs), used as singlet, doublet, or triplet regimens, may improve survival. However, these agents are expensive and may produce side effects requiring hospitalization. Little is known regarding variation in both use (singlet vs. doublet vs. triplet regimens) and outcomes (survival, hospitalizations, cost) in community practices over time for non-aged populations, adjusting for comorbidities. Aims: Using information from the HMO Cancer Research Network's (CRN) Virtual Data Warehouse (VDW), we examined the impact on variation over time in use of first-line chemotherapy regimens (singlet vs. doublet vs. triplet) in stage lIIb-IV NSCLC patients on survival, hospitalizations, and costs. **Methods:** Patients aged less than 21 years with stage IIIb- IV NSCLC diagnosed between 2000-2007 at four CRN sites were included in the analysis. Patients were followed from diagnosis date through 2008 (or death or disenrollment). Patient demographics, comorbidities, chemotherapy treatment data, and mortality were obtained from the CRN VDW. Propensity-adjusted survival and Poisson modeling were employed to examine variation in survival days and hospitalizations. Average wholesale price data were used to examine the relative difference in costs by chemotherapy regimen. **Results:** We identified 3,072 stage lIIb-IV NSCLC patients who received first-line chemotherapy of which 24% were <65 years old at diagnosis. The distribution of first-line therapy changed significantly over time with the introduction of taxane, monoclonal antibody, and antimetabolite agents in the later years of the study period. Those receiving singlet regimens were older and had more comorbidities. Unadjusted survival rates were higher for those receiving triplet therapy, but only against singlet regimens in adjusted models. Those receiving singlet regimens had the greatest number of hospitalizations. The costs of doublet and triplet regimens were significantly higher than the singlet regimens. **Discussion:** There was significant variation over time in chemotherapy regimens used in the CRN. Triplet therapy appeared to be associated with the best outcomes and fewest side effects. However, cost considerations for triplet therapy warrant assessments of their cost-effectiveness.

Keywords: Chemotherapy; Lung Cancer; Comparative Effectiveness Research

doi:10.3121/cmr.2012.1100.cc4-02

CC4-03:

The Use of Administrative Data and Natural Language Processing to Estimate the Incidence of Statin-related Rhabdomyolysis

James Floyd[1]; Susan Heckbert[1]; Noel Weiss[1]; David Carrell[2]; Bruce Psaty[1]

[1]University of Washington; [2]Group Health Cooperative

Background/Aims: Despite the problem of misclassification, many studies of adverse drug reactions (ADRs) rely on administrative data rather than events validated by medical record review. We evaluated the new ICD-9 code for rhabdomyolysis (728.88) and natural language processing (NLP) as methods to identify cases of statin-related rhabdomyolysis, and estimated the incidence with various statins and doses. **Methods:** We conducted a population-based study of statin users in Group Health Cooperative from 2006-2010. Person-years of statin use among all enrollees by statin and dose were estimated using computerized pharmacy records. Trained abstractors reviewed the medical records of selected statin users to identify cases of statin-related rhabdomyolysis and myopathy, defined as muscle injury with a peak creatine kinase level = 10x and 5-10x the upper limit of normal, in the absence of other likely etiologies. **Results:** Review of medical records for 361 statin users with a qualifying administrative code yielded 24 cases of statin-related rhabdomyolysis and 12 of myopathy. The positive predictive value of the rhabdomyolysis ICD-9 code was only 7.5% (22/292). NLP methods identified another 5 cases of rhabdomyolysis and 6 of myopathy. Using validated cases, the rhabdomyolysis incidence rate (IR) per 100,000 person-years of statin use was 13.2 for simvastatin compared with 5.2 for users of other statins (IRR 2.61; 95% CI, 1.03-7.84), and 64.8 for the 80 mg/day dose of simvastatin compared with 5.3 for the 20 mg/day dose (IRR 12.2; 95% CI, 3.6-52.3). In an analysis using only the ICD-9 code for rhabdomyolysis, the IRR for simvastatin compared with other statins was 1.03 (95% CI 0.80-1.34), and for 80 mg/day dose of simvastatin compared with 20 mg/day it was 1.77 (95% CI 1.05-2.88). **Discussion:** Use of the administrative diagnostic code for rhabdomyolysis was highly nonspecific for this ADR and resulted in weaker associations than methods that verified ADRs with medical record review. The use of administrative data alone in surveillance studies of other medication ADRs with multiples causes, such as drug-induced liver injury, may fail to detect actionable and clinically important harms.

Keywords: Rhabdomyolysis; Administrative Data; Comparative Effectiveness Research

doi:10.3121/cmr.2012.1100.cc4-03

CC4-04:

An Application of Doubly Robust Estimation

Brian Johnson[1]; Charles Gessert[1]; Colleen Renier[1]; Adnan Ajmal[1]

[1]Essentia Health

Background/Aims: Observational data often has treatment exposure confounded with baseline covariates. Doubly robust estimation utilizes both a regression model and an additional model for the exposure, often the propensity score, to estimate the causal effect of an exposure on an outcome. **Methods:** In a study designed to evaluate change in hemoglobin (Hb) with use of angiotensin- converting enzyme inhibitors (ACEI) or angiotensin receptor blockers (ARB), in a primary care patient population, we found that receiving ARB vs. ACEI was associated with several of the baseline covariates. These baseline covariates, Hb, treatment start date, chronic kidney disease, congestive heart failure, diabetes mellitus, hypertension, sex, and age, initially thought to be informative in estimating follow-up (F/U) Hb, were chosen a priori. No medical explanation for the differential ordering of