Methods: We designated 63,681 pelvic ultrasound exams of 44,704 women performed during 1997-2006 as cases or controls. Cases included 214 exams of 188 women who received a pathologically confirmed ovarian cancer diagnosis within 1 year; controls include 63,467 exams of 44,516 women who did not. We divided all exams for cases and 10,000 randomly selected exams from controls into split-half development and validation samples (having 107 cases and 5000 controls each). The full text of development sample reports was concept-coded using caTIES. Guided by domain expertise and trial-and-error we developed an algorithm to identify cases and classified each report according to it. Expert review of the 107 case reports and random samples of 300 false positive and 300 true negative controls is ongoing, as is algorithm modification. The validation sample remains unused. **Results:** A simple algorithm employing concepts 'mass,' 'simple,' 'hemorrhagic,' and 'resolution' achieved sensitivity of 63% and specificity of 93%. Expert review indicates the algorithm may be improved by 1) focusing exclusively on concepts in the report's impression section, 2) incorporating additional concepts to identify true positives and exclude false positives, 3) developing custom NLP rules to associate organ systems and concepts referring to them, and 4) excluding concepts expressed in future and past tenses. **Conclusions:** Progress to date is encouraging for developing schemes useful for ordinal classification of clinical risk of disease. Opportunities for applying NLP to clinical text for medical research are numerous and caTIES appears to be a promising tool, especially if customized to perform domain-specific NLP tasks.

Abstract PS2-29

Incident Asthma Surveillance Using Electronic Medical Record (EMR) Data – Methods from the PASS and PASS2 Studies

Suzanne Gillespie, MS, *Kaiser Permanente Center for Health Research*; MaryAnn McBurnie, PhD, *Kaiser Permanente Center for Health Research*; Victor Breen, MD, *Kaiser Permanente Center for Health Research*; Brad Crane, MS, *Kaiser Permanente Center for Health Research*; Brian Hazlehurst, PhD, *Kaiser Permanente Center for Health Research*; David B. Callahan, MD, *Centers for Disease Control and Prevention*; William M. Vollmer, PhD, *Kaiser Permanente Center for Health Research*

Background: Availability of longitudinal data from the electronic medical record (EMR) provides opportunities to monitor trends in the incidence and treatment of chronic diseases such as asthma. The Centers for Disease Control (CDC)-funded PASS studies developed, evaluated and implemented an algorithm for using the EMR to identify incident cases of asthma in the Kaiser Permanente Northern California (KPNW) region over a 5-year period. **Methods:** The original PASS study was carried out in two phases. In phase I candidate algorithms based on medication dispensings and asthma diagnoses from doctor visits were developed and piloted. A range of options for medication dispensing criteria, health plan eligibility, and length of the surveillance period were considered. The resulting algorithm was arrived at through a collaborative decision process involving the CDC and investigators from two field centers. In phase II the algorithm was validated by recruiting a subset of 219 randomly-selected patients identified by the algorithm as having incident asthma and collecting clinical and self-report measures from them, including pulmonary function tests. Two KPNW pulmonologists reviewed all information relating to each of these cases to arrive at a 'gold standard' asthma classification for these patients. **Results:** The final PASS asthma algorithm requires the following criteria be met: either two dispensings of asthma medication, or two or more visits at which asthma was noted as a diagnosis. The 'at-risk' population is identified by excluding those who meet these criteria in the 4 years prior to the year of interest and those who do not meet membership eligibility. Individuals in the resulting 'at-risk' population who meet the asthma criteria in the year of interest are identified as incident for asthma. The validation phase resulted in KPNW physicians rating 85% of the 219 charts reviewed as probable/possible asthma and 15% as unlikely. This represents a predictive value of 86%. This algorithm has subsequently been implemented in the 5-year PASS2 study which aims to, among other things, carry out asthma surveillance to determine the incidence and prevalence of physician-diagnosed asthma and develop and maintain a research database of incident asthma cases. The next step is to evaluate the scalability of this algorithm in other EMR-based systems. **Conclusion:** The PASS studies have demonstrated that it is feasible to develop and implement an EMR-based algorithm for asthma surveillance. Similar methods can be employed for other chronic diseases.

Abstract PS2-30

Care Coordination and Health Information Technology: Information Availability and Timeliness Across Care Transitions

Ilana P. Graetz, BA, *Kaiser Permanente Division of Research*; Mary Reed, DrPh, *Kaiser Permanente Division of Research*; Jie Huang, PhD, *Kaiser Permanente Division of Research*; Thomas Rundall, PhD, *University of California Berkeley*; John Hsu, MD, MBA, MSCE, *Kaiser Permanente Division of Research*

Background: Patients who require care by multiple clinicians are at risk for problems during these transitions. Use of an electronic health record (EHR) could provide clinicians a mechanism to coordinate activities and information, thereby improving care across transitions. We examined the impact of having EHR on the availability and timeliness of clinical information within an integrated delivery system. **Population:** All adult primary care clinicians (n=396 PCs) working in a large prepaid, integrated delivery system who completed the surveys in both 2005 and 2006. Overall, 57.3% were female, 48.9% were white, 83.1% were physicians, mean age was 45 years, and the mean panel size was 1540 patients. In 2005 only 2.0% of respondents had finished implementing the EHR at the time of their survey versus 39.4% in 2006. **Methods:** Using survey data collected in both 2005 and 2006, we examined clinician reports of the availability and timeliness of relevant clinical information when multiple clinicians are involved with their patients' care. We collapsed the five-point responses into two categories, e.g., high availability when the information was 'always' or 'usually' available. We defined EHR status as pre-, during, or post-implementation using the implementation dates for the medical center. Using multivariate logistic regression, adjusting for clustering by clinician, medical center, year, clinician gender, race, panel size, job title and status, we examined the association between EHR status and both clinical information availability and timeliness. **Results:** In year-one, 42.7% of PCs reported having high levels of clinical information availability, compared with 64.9% in year-two; 39.3% and 62.4% reported having high levels of timely information in years one and two, respectively. Overall, 35.3% and 57.8% reported having high levels of both information availability and timeliness in years one and two, respectively. After adjustment, clinicians working in medical centers that had completed the EHR implementation were significantly more likely to report having high levels of both clinical information availability and timeliness (OR=2.88; 95% CI, 1.46-5.69). **Conclusion:** EHRs could be an important tool for improving the timeliness and availability of clinical information. These improvements could facilitate the coordination of care between clinicians.

Abstract PS2-31

Virtual Data Warehouse 'High Utilizers'

Gene Hart, MS, *Group Health/Center for Health Studies*; Jennifer L. Ellis, MSPH, *Kaiser Permanente Colorado/Institute for Health Research*; Don Bachman, MS, *Kaiser Permanente Northwest/Center for Health Research*; Gwyn Saylor, MS, *Kaiser Permanente Colorado/Institute for Health Research*

Background: The virtual data warehouse (VDW) was created as a mechanism for streamlining the process of proposing and conducting multi-site research by eliminating the need for custom data extraction programming at each site. At the core of the VDW are a series of standard file definitions. Content areas and data elements commonly required for research are identified, and data dictionaries created for each one. The data dictionaries specify a common format for each element—variable name, label, extended definitions, code values, and value labels. Local site programmers then developed programs to aggregate and transform their local legacy data into the standard VDW files. The data at the sites is thus made uniform, making it possible for multiple sites to run the same program against their local VDW data, and produce comparable results. The VDW is 'virtual' in the sense that the actual data remain at the local sites. It is not a centralized database held at a coordinating center and accessible only to a privileged few. Rather, it is a mechanism for sharing and reusing SAS code, and accelerating the research process. **Methods:** This poster presents data from 3 HMOs that shows how the VDW is used at these sites. **Results:** Programmers at Group Health, Kaiser Northwest, and Kaiser Colorado all use the VDW for more than multi-site projects. Using both automated file