[3] *Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, United States;* [4] *University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States;* [5] *Kaiser Permanente Colorado, Denver, CO, United States*

Background: Studies that evaluate UGIB in electronic healthcare data typically rely on inpatient diagnostic codes for outcome identification. Use of hemoglobin (HGB) lab test results might increase detection of UGIB that do not lead to hospitalization.

Objectives: To evaluate whether use of HGB test results increases UGIB identification using non-steroidal anti-inflammatory drugs (NSAIDs) as a test case.

Methods: From the Mini-Sentinel distributed database, we identified patients ≥18 years old who initiated prescription NSAIDs in 3 Data Partners between January 2008-April 2013. Availability of HGB test results was examined before and after NSAID initiation. Numbers of events and cumulative incidences within 30 days after NSAID initiation were calculated for 4 mutually exclusive outcome definitions: (1) inpatient UGIB diagnosis (standard claims-based definition without lab test results); (2) non-inpatient UGIB diagnosis AND ≥3 g/dL decrease in HGB; (3) ≥3 g/dL HGB decrease alone without UGIB diagnosis in any clinical setting; (4) non-inpatient UGIB diagnosis, without ≥3 g/dL HGB decrease. In secondary analyses, we reviewed all coded diagnoses in patients with outcome 3 to scan for codes indicative of potential UGIB and assessed distributions of specific UGIB diagnoses in patients with outcomes 1, 2, and 4.

Results: We identified 2,289,772 NSAID initiators; 45% had ≥1 HGB result available within 365 days before or 30 days after NSAID initiation. Only 7% had results before and after. Of 7,637 potential outcomes identified from all 4 definitions, outcome 1 accounted for 22%, outcome 2 for 1%, outcome 3 for 34%, and outcome 4 for 43%. Potential cases identified by outcome 3 were mostly associated with codes for non-UGIB or other non-hemorrhagic conditions. Outcomes 1, 2, and 4 were associated with similar distributions of specific UGIB codes.

Conclusions: Using HGB result values in combination with UGIB diagnoses identified few additional potential UGIB cases and with unknown specificity. The use of HGB result values alone did not improve identification of potential UGIB events. The use of non-inpatient diagnostic codes may increase UGIB outcome detection, but would require validation.

263. Cancer Recording In Patients With Type 2 Diabetes In Primary Care And Hospital Admission Data

Rachael Williams[1,2], Arlene M. Gallagher[1,2], Tjeerd van Staa[2], Tarek Hammad[3], Bert Leufkens[2] and Frank de Vries[2]

[1] *Clinical Practice Research Datalink, London;* [2] *Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht;* [3] *Office of Surveillance and Epidemiology, Food and Drug Administration, Washington*

Background: Electronic health records are increasingly used to investigate associations between antidiabetic therapy and cancer. Misclassification can impact results, especially if differential between comparators.

Objectives: Estimate cancer misclassification when using primary care or hospital data alone.

Methods: Adults aged ≥40 years with an insulin or oral antidiabetic prescription in Clinical Practice Research Datalink (CPRD) primary care data at least a year after start of data collection, and no record of type 1 diabetes, were included. Patients were matched by year of birth (stepwise within 5 years), sex and GP practice to up to 1 non-diabetic patient. The cohort was restricted to those eligible for Hospital Episode Statistics (HES) linkage with follow-up during the study period (04/01/97-12/31/06). Follow-up started at the maximum of the registration date with the practice, practice up-to-standard date (a CPRD quality metric), and start of study period. Follow-up ended at the minimum of when the patient left the practice, the date CPRD last collected data from the practice, and end of study period. Cancer was identified in CPRD via Read codes and in HES via ICD-10 codes. For each cancer case in CPRD, analysis evaluated whether there was a corresponding record in HES coded with same, different or unspecified site. Analysis was repeated for cancers identified in HES.

Results: 53,585 diabetic patients were matched to 47,435 non-diabetic patients. 83% of cancer cases in CPRD had a corresponding record in HES (78% with the same type). Misclassification varied by cancer site, ranging from 3% (stomach cancer) to 57% (non-melanoma skin cancer). 83% of cancer cases in HES had a

DOI: 10.1002/pds

corresponding record in CPRD, with all misclassification rates <20%.

Conclusions: A good level of concordance and low level of misclassification of cancer exist between CPRD primary care data and HES. The value of linking these data for establishing cancer outcomes lies more in the complimentary variables held than in reducing misclassification.

264. Authenticity Validation of Lung Cancer Records from a Standardized Inpatient EMR Database in China and Comparison of Different Methods

Meng Shu[2], Wenhua Liang[1], Yongjing Zhang[2] and Jianxing He[1]

[1] *The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China;* [2] *Epidemiology, Janssen Research and Development, Shanghai, China*

Background: Information from pathology reports and discharge diagnoses in Electronic Medical Records (EMR) are frequently used to identify lung cancer patients. Their validity has not been fully assessed in China.

Objectives: To assess the accuracy of the identification of lung cancer inpatients using standardized EMR data from our hospital (The first affiliated hospital of Guangzhou medical university, China).

Methods: The de-identified EMR data from the hospital was transformed into a common data model. Patients with an inpatient lung cancer diagnosis, pathology diagnosis, or discharge diagnosis of lung cancer between January 1, 2012 and September 30, 2015 were identified. Both ICD-10 code and description in free text were used. A random sample of 300 lung cancer inpatients was selected, and their medical records were reviewed and adjudicated by an experienced physician who was blinded to the diagnosis. Positive predictive values (PPV) were calculated as the proportion of lung cancer inpatients in the EMR database that were correctly identified based on the physician assessment.

Results: There were 6862 possible lung cancer inpatients identified per discharge diagnosis, of whom 2703 patients did not have a pathology diagnosis (group i). There were 5274 cases identified according to pathology reports, of whom 1115 did not have a corresponding discharge diagnosis (group ii); 4159 cases were categorized into group iii (patients with both pathology and discharge diagnoses). The PPV for group i was 85% (95% CI, 77.6%–92.4%), which was lower than the 90% (95% CI, 84.1%–95.9%) for group ii. Patients with both lung cancer pathology and discharge diagnoses (group iii) yielded the highest PPV, 99% (95% CI, 97.1%–100.0%). There were more male and older patients in group i, which may be due to a higher number of pre-existing lung cancer patients who did not have pathology records within the predefined period.

Conclusions: In this study, the PPV is highest when cases are identified using both lung cancer pathology and discharge diagnoses. In China, free text analysis alone is not sufficient to identify true cases; combined method is encouraged.

265. Ascertaining Pulmonary Hypertension and Other Conditions in Claims Databases

David Vizcaya[1], Alexander Michel[1], Lin Li[2], Stefanie Breitenstein[3], Gemzel Hernandez[4], Montse Soriano-Gabarro[1] and Susan Jick[2]

[1] *Global Epidemiology, Bayer Pharma AG, Berlin, Germany;* [2] *Boston Collaborative Drug Surveillance Program, Boston University School of Public Health, Lexington, MA, United States;* [3] *Global Clinical Development, Bayer Pharma AG, Wuppertal, Germany;* [4] *Global Clinical Development, Bayer Pharma AG, Whippany, NJ, United States*

Background: Epidemiological studies using claims databases often rely on diagnostic codes for case identification. For certain diseases, this may result in misclassification.

Methods: In the context of a study to estimate the incidence and prevalence of pulmonary arterial hypertension in a US pediatric population, 2010-2013 using MarketScan claims data we identified all potential cases of pulmonary hypertension (PH). Initially, a case was defined as a patient with ≥2 claims for PH that were at least 1 day apart, or ≥1 claim for PH plus ≥1 prescriptions for PH treatment (definition A). Later, we redefined the inclusion criteria to require ≥1 prescriptions for PH treatment for all cases (definition B). We evaluated the impact of each definition as the ratio of both incidence and prevalence of PH using each definition. We searched the literature on other

DOI: 10.1002/pds