OBJECTIVES: Capturing dosing of biologic and non-biologic Disease Modifying Anti-Rheumatic Drugs (DMARDs) using electronic medical records is challenging. Precise estimates of weekly dose using structured pharmacy data alone are difficult since *quantity dispensed* is not standardized for injectable products. Natural Language Processing (NLP) software was developed to extract elements of the prescribing provider's medication instructions (SIGs) to compute the average weekly dose. The objective was to evaluate the accuracy of the NLP software's computation of the average weekly dose by medication and route. **METHODS:** The NLP software computed the average weekly dose for biologic and non-biologic DMARDs and was evaluated against the annotator-derived reference standard. Using an annotation guideline, trained annotators annotated relevant information from SIGs including unit strength, dose per administration event and the schedule of administration events. The NLP software was then trained on 11,937 records. A validated set of the annotator-derived reference standard that contained 140 SIGs per medication and route was used to evaluate the NLP accuracy and compute the 95% Confidence Interval (CI) of accuracy. **RESULTS:** The overall accuracy for injectable biologic and oral and injectable non-biologic DMARDs was 89.1% (95% CI: 87.9%-90.3%). Accuracy was 95.3% (95% CI: 91.9%-98.7%) for oral methotrexate, 84.7% (95% CI: 78.9%-90.5%) for injectable methotrexate, 87.9% (95% CI: 82.5%-93.3%) for sulfasalazine, and 92.9% (95% CI: 88.7%-97.2%) for hydroxychloroquine. For biologics, accuracy was 92.1% (95% CI: 87.6%-96.6%) for etanercept, 98.6% (95% CI: 96.7%-100%) for adalimumab and injectable abatacept 90.0% (95% CI: 85.2%-94.8%). The lower bound of the 95% CI ranged from 79.1%-100% for biologic DMARDs and from 78.9%-91.9% for non-biologic DMARDs. **CONCLUSIONS:** The lower bounds of the 95% CI for most medications were greater than 80%. These results indicate that the NLP software can be used to extract information to calculate the weekly dose of DMARDs from narrative medication schedules.

PRM41

ALL PAYER CLAIMS DATABASES, STATE BASED SOLUTIONS TO LEVELING THE HEALTH DATA INFORMATION PLAYING FIELD

Garfield S[1], Noyes S[2], Armstrong S[1]

[1]*GfK Market Access, Wayland, MA, USA,* [2]*Human Services Research Institute, Cambridge, MA, USA*

OBJECTIVES: Private claims databases have been a rich source of information related to health care delivery, though the fragmented nature of U.S. health insurance market limits their ability to provide a comprehensive view of a given geography. Additionally, claims data are often expensive (>$25,000) - restricting access to those with the economic means to afford it. To provide a more thorough view of health costs and practice at the state level, 11 states have created all payer claims databases (APCDs). An analysis was conducted to review their designs and consider how they will impact health economic analyses and health services research. **METHODS:** A detailed review of 11 APCDs was conducted using a combination of secondary research and key informant interviews. This included reviewing the legislation that mandated and funded their creation, the rules under which they operate, and the data they collect and make available. The multi-state analysis was then informed by a deeper analysis of the data integration opportunities, reporting abilities, and costs for data exchange for Maine's newly revised APCD. **RESULTS:** Across states, APCDs make publicly available a wide array of health data in a manner that ensures privacy and security of personal health information. Each APCD provides varying levels of granularity in health care resource utilization, demographic comparisons, expenditures, and quality/safety data. Ability to customize data pulls is growing, though private claims datasets appear to have greater flexibility. In Maine's APCD, commercial and Medicaid medical, pharmacy, and dental data are available to researchers, with quarterly block releases and opportunities for individual queries. **CONCLUSIONS:** APCDs are providing researchers with an open-access, affordable pathway to understand health care resource utilization and practice across entire populations. As more come online, greater participation in health economic and health services research can be expected.

PRM42

BUILDING A REAL-WORLD NON-SMALL CELL LUNG CANCER COHORT BY LINKING A CLAIMS DATABASE TO ONCOLOGY ELECTRONIC MEDICAL RECORDS

Chen CC[1], Wade RL[2], Khan N[1], Pokras SM[3], Semonelli C[1], Zeng S[1], Marcus M[1]

[1]*IMS Health, Plymouth Meeting, PA, USA,* [2]*IMS Health, Parsippany, NJ, USA,* [3]*IMS Health, Alexandria, VA, USA*

OBJECTIVES: Lung cancer (LC) leads cancer-related mortality in the U.S.; often diagnosed at an advanced stage with poor prognosis. Non-small cell lung cancer (NSCLC) accounts for approximately 85% of all LC. Real-world research in NSCLC necessitates clinical and longitudinal detail typically unavailable from any single retrospective data source. This study describes and benchmarks a comprehensive disease record for NSCLC patients created by linking the PharMetrics Plus (PMTX) database to an oncology electronic medical records (OEMR) database. **METHODS:** Adult NSCLC patients (ICD-9-CM code 162.2 to 162.9) with either confirmed histology, use of pemetrexed, or without evidence of etoposide use were identified from PMTX and OEMR databases between 1/1/2006 and 7/31/2013. Patients were linked using a HIPAA-compliant encryption algorithm. Key patient characteristics from the linked sample were compared with the National Cancer Institute SEER registries NSCLC population (2006-2010, n=152,465). **RESULTS:** 182,425 NSCLC patients were identified (PMTX:156,029; OEMR:26,906), with 510 patients linked to both databases. Among linked patients, 50% were female, 36% were considered non-squamous (confirmed by histology or use of pemetrexed) and 47% were metastatic at diagnosis. Treatment data (surgical resection, radiation, or chemo/biological therapy) were available for 97% of the linked patients; 64% with platinum and 30% with pemetrexed therapy. The SEER population had a higher proportion of patients >=65 years at diagnosis (64% vs. 25%) but a lower portion (21%) metastatic at diagnosis compared to the linked sample. **CONCLUSIONS:** This study demonstrates that databases can be linked to provide a longitudinal, clinically rich view of an NSCLC population. Comparisons between the linked sample and the SEER population highlight potential differences which should be considered when selecting data to answer specific research questions.

PRM43

REAL WORLD RESEARCH IN LATIN AMERICA: OPPORTUNITIES, SOURCES AND BARRIERS

Gregory V[1], Barbeau M[1], Machnicki G[2], Voko Z[3], Heisel O[4], Keown P[4]

[1]*Novartis Pharmaceuticals Canada Inc., Dorval, QC, Canada,* [2]*Novartis Argentina SA, Buenos Aires, Argentina,* [3]*Syreon Research Institute, Budapest, Hungary,* [4]*Syreon Corporation, Vancouver, BC, Canada*

OBJECTIVES: Real world evidence (RWE) is critical for the assessment of health technologies. This study was conducted to define and compare the governance of, and data sources available for, real world research (RWR) in Latin America (LA) and Canada. **METHODS:** Systematic literature review was conducted to examine administrative and clinical data for 10 major countries ranked by population. Research terms were defined by an international team, and reports were reviewed and assessed for content, quality and bias by two investigators. Data was summarized in major and minor domains for each country. **RESULTS:** Governance of RWR differed between countries, both from a regulatory and an ethics committee review perspective. Regulatory review was required for non-interventional studies in 4/10 countries. Ethics review varied importantly in complexity, site (local, central) and duration (which could exceed 1 year). Administrative and clinical search terms returned over 1800 reports from LA, principally from Brazil, Mexico, Argentina and Chile, of which over 700 contained contributory information on data sources for RWE. Of these, 156 addressed international registries or databases including countries in LA, 245 reported national registries or databases within one country in LA, and 308 reported registries or databases from a single or multiple institutions within a country. Principal administrative categories included claims, prescription and economic data sources, while principal clinical categories included data sources relating to cancer, cardiology, neurology, respirology, and diabetes. In contrast, a total of over 2000 reports were obtained for Canada alone, with a similar categorical distribution to that observed in LA. **CONCLUSIONS:** Latin America is a region with diverse health systems, inputs and outcomes. While sources for RWE exist in several larger countries, comprehensive national or regional databases are uncommon compared with a mature public health care system such as Canada. Improvement of database quality and well-designed prospective population studies are critical to enhance the RWE base.

PRM44

EVALUATION OF CURRENT DATA SOURCES IN EUROPE FOR THE CONDUCT OF REAL-WORLD STUDIES ON LUNG AND RENAL CELL CARCINOMA: A SYSTEMATIC LITERATURE REVIEW

Gridchyna I[1], Manley Daumont M[2], Yousif A[1], Hirji I[2], Lees M[2], Salvo F[3], Moride Y[1]

[1]*Faculty of Pharmacy, Université de Montréal, Montreal, QC, Canada,* [2]*Global Health Economics and Outcomes Research – Europe, Bristol-Myers Squibb, Rueil-Malmaison, France,* [3]*Bordeaux University, CHU Bordeaux, Bordeaux, France*

BACKGROUND: Real-world data are required to fully assess the value of treatments in clinical practice. Given the importance of having all available data to make this assessment, while ensuring efficiencies in preventing duplication of data collection, the absence of a central repository of longitudinal data sources on cancer patients in Europe remains a challenge. **OBJECTIVES:** i)Identify and characterize existing traditional and non-traditional data sources on lung cancer and renal cell carcinoma (RCC) patients in Europe; ii)Determine the utility of each identified data source for real-world research. **METHODS:** A systematic literature search was conducted using MEDLINE and Embase(01/01/06-20/12/13). MeSH and Emtree terms included the following concepts: lung cancer (small-cell and non-small-cell), RCC, real-world data (registry, cohort, etc), and European countries. Clinical trials, case reports, case series, literature reviews were excluded. Identified abstracts were screened and reviewed based on relevance to key concepts. For each data source, utility was assessed using the following variables: country, setting, type (e.g. registry, EMR, etc.), clinical and baseline characteristics, tumor, biomarkers, test results, treatments, clinical and patient reported outcomes, resource utilization. **RESULTS:** The search yielded 2,478 abstracts for lung cancer and 698 for RCC. Following screening, 192(7.7%) and 102 (14.6%) were retained respectively. Data sources for lung cancer were equally distributed between general cancer and tumor specific registries. Most originated from the UK (33.3%) or France (13.8%). RCC data sources were primarily tumor specific and local with population size ranging from 56-1800. Data sources originated from Italy (22.5%), France, UK, and Germany (each 16.7%). Gap analyses with respect to the availability of specific data elements and utility for outcomes research will be presented. **CONCLUSIONS:** Data sources on cancer are highly fragmented in Europe, especially for RCC, emphasizing the need for database mapping. Completeness and consistency of the data elements needed to support technology assessments are variable.

PRM45

UTILITZATION OF THE TRUVEN NATIONAL WEIGHTS TO ESTIMATE THE CHRONIC CONDITIONS AND THEIR ASSOCIATED QUALITY MEASURES IN THE UNITED STATES COMMERCIAL EMPLOYER SPONSORED HEALTH INSURED POPULATION

Carroll C, Priest J, Lu CX, Le HV

GlaxoSmithKline, RTP, NC, USA

OBJECTIVES: To establish quality of care benchmark estimates (HEDIS, PQA, or NQF) for Asthma, COPD, and Diabetes in the US Commercial Employer Sponsored Health Insured Population (EIS) for 2011. **METHODS:** The Marketscan Commercial claims database from 2010-2011 was used to identify the EIS population during 2011 and those patients with Asthma, COPD, and Diabetes. Patients needed at least 6 months of continuous coverage during 2011 and coverage during December 2011. Asthma, COPD, or Diabetes must have been recorded during 2011 or in the 6 months prior and was identified using a clinically approved coding algorithm. A weight was assigned