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 notes on specific DMARDs and 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 93.9% (95% CI: 91.7%-96.1%) 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.

PRM42 BUILDING A REAL-WORLD NON-SMALL CELL LUNG CANCER COHORT BY LINKING A CLAIMS DATABASE TO ONCOLOGY ELECTRONIC MEDICAL RECORDS

Chen C.C.1, Wade R.L.1, Manley Daumont M.2, Yousif A.3, Gridchyna I.4, Heisel O.4, Moride Y.5, Gregory V.6, 1IMS Health, USA, 2IMS Health, Parsippany, NJ, USA, 3IMS Health, Charlotte, NC, USA, 4Human Services Research Institute, Cambridge, MA, USA, 5IMS Health, Helsinki, Finland, 6IMS Health, Latin America

OBJECTIVES: Lung cancer (LC) leads cancer-related mortality in the U.S.; often described as a cancer of the poor. The treatment of LC, especially for patients living with non-small cell lung cancer (NSCLC), accounts for approximately 85% of all LC. Real-world research in NSCLC necessitates clinical and longitudinal data 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 (PMXT) 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 or use of pemetrexed and without evidence of etoposide use were identified from the PMXT and OEMR databases between 1/1/2006 and 7/31/2013. Patients were linked to either the PMXT or OEMR database and were included if they had a minimum of two visits (within 1 year). Linkage was confirmed by histology or use of pemetrexed and 145 of 152 patients (95.4%) were linked. RESULTS: 85% of the linked patients were greater than 80%. These results indicate that the NLP software can be used to compute 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 93.9% (95% CI: 91.7%-96.1%) 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.

PRM43 REAL WORLD RESEARCH IN LATIN AMERICA: OPPORTUNITIES, SOURCES AND METHODOLOGIES

Chen C.C.1, Wade R.L.1, Manley Daumont M.2, Yousif A.3, Gridchyna I.4, Heisel O.4, Moride Y.5, Gregory V.6, 1IMS Health, USA, 2IMS Health, Parsippany, NJ, USA, 3IMS Health, Charlotte, NC, USA, 4Human Services Research Institute, Cambridge, MA, USA

OBJECTIVES: Real world evidence (RWE) is critical for the assessment of health technologies. This study was conducted to define and compare the governance of RWE 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. Researchers’ key informant teams were identified and accessed for content, quality and bias by two investigators. Data was summarized in major and minor domains for each country. RESULTS: Governance of RWE differed between countries, both from a regulatory and an ethics committee perspectives. BACKGROUND: Real world data are required to be 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 1,800 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 national registries or databases including countries in LA, 245 reported national registries or databases within one country in LA, 599 reported registries or databases from a single or multiple countries within a country. Principal administrative categories included claims, prescription and economic data sources, while principal clinical categories included data sources related to cancer, cardiology, neurology, respiriology, 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 administrative systems, important outcomes 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

Chen C.C.1, Manley Daumont M.2, Yousif A.3, Gridchyna I.4, Heisel O.4, Moride Y.5, Gregory V.6, 1IMS Health, USA, 2IMS Health, Parsippany, NJ, USA, 3IMS Health, Charlotte, NC, USA, 4Human Services Research Institute, Cambridge, MA, USA, 5IMS Health, Helsinki, Finland, 6IMS Health, Latin America

OBJECTIVES: 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 (EVR), would limit comparative research. Although existing traditional and non-traditional data sources on lung cancer and renal cell carcinoma (RCC) patients in Europe, it/determine the utility of each identified data source for data collection and/or research. METHODS: A systematic literature search was conducted using MEDLINE and Embase(01/01/06-20/12/13). MeSH and Emtree terms included real-world evidence (RWE) and Canada. RESULTS: 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 administrative systems, important outcomes 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.

PRM45 UTILIZATION OF THE TRUEN NATIONAL WEIGHTS TO ESTIMATE THE CHRONIC CONDITIONS AND THEIR ASSOCIATED QUALITY MEASURES IN THE UNITED STATES COMMERCIAL EMPLOYER SPONSORED HEALTH INSURED POPULATION

Carrall C, Priest J, Lu CX, Le HV1, 1UMass, Amherst, MA, USA, 2RTI, 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 was used to identify commercial insured employees and the employer population for California, New York, Texas, 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 to potential differences which should be considered when selecting data to answer specific research questions.