containing at least one product and one adverse event keyword were collected, de-
identified, and analyzed using a vernacular to MedDRA dictionary. Posts were clas-
sified as resembling an adverse event report (Proto-AEs) or simply discussing a product (Mention). RESULTS: There were a total of 1,410,819 posts categorized as Proto-AEs, 265,838 (19%) from Facebook and 1,144,981 (81%) from Twitter. The top 10 products were ranked in accessions for 946,666 (67%) of the total Proto-AEs in Facebook and Twitter combined. The top 25 accounted for 1,180,040 (84%), the top 50 for 2,125,836 (82%), and the top 100 for 2,245,010 (95%) of the total Proto-AEs. The top 10 products (diphenhydramine, flu vaccine, dextroamphetamine, codeine, morphine, ibuprofen, alprazolam, acetaminophen, oxycodone, and zolpidem) were comprised of six con-
trolled substances, three over-the-counter (OTC) products, and one class of vaccine. Of the top 50 products, controlled substances accounted for 32%, OTC products for 24%, and vaccines for 10%. CONCLUSION: Review of publicly available data over the past two years from two popular social media sites, Facebook and Twitter, offers a high number of potential adverse events (Proto-AEs) for further evaluation. Social listening may be potentially valuable as a supplement to traditional pharmacovigil-
ance practices, particularly for controlled substances, over-the-counter products, and vaccines. These initial findings warrant more research and a closer inspection as to the nature of these posts.

PRM5
HUNTING FOR RANDOMISED CONTROLLED TRIALS (RCTs): A COMPARISON OF SEARCH FILTERS DESIGNED TO IDENTIFY RCTs
Steeves S, Buchanan-Hughes AM, Evans J, Leonard S
Costello Medical Consulting Ltd, Cambridge, UK
OBJECTIVES: There are a number of search filters designed to identify studies with RCTs in the Cochrane Library and PubMed in electronic databases. This study compared three filters for identifying randomised controlled trials (RCTs). METHODS: Searches were conducted on 15thJune 2015 in the Ovid MEDLINE and MEDLINE In-process databases. The filters were: Highly Sensitive Search Strategy for Identifying Randomized Trials in MEDLINE, the SIGN Randomised Controlled Trials MEDLINE filter and the BMJ MEDLINE Randomised Controlled Trial Strategy. Differences were explored using a 323 abstract database. RESULTS: 3,476 abstracts were identified for comparison, a sample of articles returned by all three filters was also reviewed. To estimate the sensitivity of each filter, the detection of 39 publications of RCTs included in a randomly-selected Cochrane Collaboration systematic literature review on cancer was evaluated. RESULTS: 476, 469 and 466 RCTs were identified by the SIGN filter, the BMJ MEDLINE filter and the MEDLINE filter, respectively. The publication missed by the SIGN filter was not the same as for the other filters. A total of 382,523 patients were identified using NLP among the 5 studies. Accuracy among the studies ranged from 95.2% to 100% (95% CI: 95.1%, 100%), with an I2 value of 92.9% (95% CI: 92.9%, 97.7%). CONCLUSIONS: NLP provide a unique opportunity to extract meaningful information from patient-level narrative clinical notes in EMR data sources with high degree of accuracy. This provides additional rich sources of data from narratives.

PRM6
METHODOLOGICAL DIFFICULTIES OF COMPLIANCE ANALYSES BASED ON REAL-WORLD DATA
Andriiska P, Komáromi T, Friggesy R, Salber F
Healthcare Consulting Ltd, Budapest, Hungary
OBJECTIVES: Regarding compliance analysis numerous ratios can be found in literature. The aim of this part of the lecture is to show that there is no golden ratio available as a tool for measuring compliance. The results of our analysis tend to reveal, that choosing an adequate ratio is not sufficient itself, it is essential to know the difficulties and pitfalls of the data management and methodology to the objective assessment of the chosen ratio. The chief aim of our lecture is to demonstrate the pitfalls and factors that may substantially influence the results and the right conclusions, if these factors are modified. METHODS: The analysis is based on prescription refilling’s data of the Hungarian Health Fund in the field of the following indications: diabetes, COPD, oncology. From the ratios available in scientific literature, the FDC (Proportion of Days Covered) was chosen. The following aspects were considered as influencing factors: patient inclusion criteria (index date, time frame, criteria of regular and irregular use), and the use of practical experience, which may substantially influence the results and the right conclusions, if these factors are modified. RESULTS: The analysis is based on prescription refilling’s data of the Hungarian Health Fund in the field of the following indications: diabetes, COPD, oncology. From the ratios available in scientific literature, the FDC (Proportion of Days Covered) was chosen. The following aspects were considered as influencing factors: patient inclusion criteria (index date, time frame, criteria of regular and irregular use), and the use of practical experience, which may substantially influence the results and the right conclusions, if these factors are modified. CONCLUSIONS: Based on the results it may be concluded, that no general best practice can be observed, all settings have both advantages and limitations.

PRM7
UTILITY AND METRICS OF NATURAL LANGUAGE PROCESSING ON IDENTIFYING PATIENTS FOR PHARMACOEPIDEMIOLOGIC STUDIES
1Aetna, Murray, UT, USA, 2VA Salt Lake City Health Care System, Salt Lake City, UT, USA
OBJECTIVES: Electronic medical records (EMR) are increasingly utilized in clini-
care and research, allowing for more efficient availability of rich patient records. However, most use of EMR is limited to coded, structured, and administra-
tive data, while the vast majority of patient information (e.g. disease subtype, severity, medical device usage, etc.) is tied up in narrative clinical notes. The chal-
lenging nature of accessing these valuable data points is improving, however, this has actually been done via timely and costly manual chart review, but as the amount of EHR data increases exponentially, manual chart review becomes impractical and impossible at scale. Advances in Natural Language Processing (NLP) have demon-
strated promising results in combining the capture of additional clinical note information with the efficiency of modern informatics. The objective of this study is to demonstrate the relevancy and utility of NLP to extract health data from EMR to real-world observational studies. METHODS: We conducted a systematic review and meta analysis of performance metrics for five (5) NLP-driven projects involving oncology, inflammation and medical devices, which had similar protocol- and data-driven objectives. We automated and validated the accuracy of NLP algorithms, as well as heterogeneity of accuracy between studies using random effects meta-
analyses (represented by I2 value). RESULTS: A total of 382,523 patients were identified using NLP among the 5 studies. Accuracy among the studies ranged from 95.2% to 100% (95% CI: 95.1%, 100%), with an I2 value of 92.9% (95% CI: 92.9%, 97.7%).