Discovering adverse drug events combining spontaneous reports with electronic medical records: a case study of conventional DMARDs and biologics for rheumatoid arthritis

Liwei Wang, MD, PhD[1,2], Majid Rastegar-Mojarad[2], Sijia Liu[2,3], Huaji Zhang[4], Hongfang Liu, PhD[2]
[1] School of Public Health, Jilin University, Changchun, Jilin, China
[2] Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, U.S.
[3] Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, U.S.
[4] Chinese Pharmaceutical Association, Beijing, China

Abstract

The use of multiple data sources has been preferred in the surveillance of adverse drug events due to shortcomings of using only a single source. In this study, we proposed a framework where the ADEs associated with interested drugs are systematically discovered from the FDA's Adverse Event Reporting System (AERS), and then validated through mining unstructured clinical notes from Electronic Medical Records (EMRs). This framework has two features. First, a higher priority was given to clinical practice during signal detection and validation. Second, the normalization by NLP facilitated the interoperation between AERS-DM and the EMR. To demonstrate this methodology, we investigated potential ADEs associated with drugs (class level) for rheumatoid arthritis (RA) patients. The results demonstrated the feasibility and sufficient accuracy of the framework. The framework can serve as the interface between the informatics domain and the medical domain to facilitate ADE discovery.

Introduction

Adverse drug events (ADEs), referring to any undesirable effect of a drug beyond its anticipated therapeutic effects occurring during clinical use[1], are important public health concerns. Although randomized clinical trials (RCTs) are considered a gold standard in identifying pre-marketing safety issues of drugs, there are some existing limitations, primarily within experiments. These limitations can include insufficient patient number, homogeneous population, short trial period and exclusion of patients with comorbid diseases. Therefore, it is well accepted that pre-marketing RCTs may not detect all types of ADEs related to a particular drug in clinical practice.

In post-marketing surveillance for adverse drug events (ADEs), the FDA's Adverse Event Reporting System (AERS) has become an important resource. However, signals from AERS data may contain false positive results, where an association between the drug and ADE is incorrectly identified, as well as false negative results, where a true association or signal is missed. Other data sources have been studied aiming for ADE detection, such as the secondary use of Electronic Medical Records (EMRs) for further validation or comparison of ADEs, which has been paid much attention. EMRs contain rich information in unstructured clinical notes that cannot be overlooked[2]. Recently, Natural language processing (NLP) has been used to extract drug-ADE pairs for signal detection through $\chi 2$ test[3]. The efficacy of mining EMRs for drug-ADE relationship has also been proven[4]. As a demonstration that combining AERS with EHRs can improve the accuracy of ADE signal detection, an approach was proposed to produce a highly selective ranked set of candidate ADEs from both AERS and EMRs based on proportionality analysis[5]. This study could systematically discover ADEs and apply to very general scenarios.

In this study, we proposed a framework where the ADEs associated with interested drugs are discovered from FDA AERS, and then validated through mining unstructured clinical notes where clinical priorities are given in terms of cohort selection and result analysis. To demonstrate the methodology, we investigate potential ADEs associated with drugs (class level) for rheumatoid arthritis (RA) patients.

Background

Rheumatoid arthritis (RA) is the most common type of arthritis in adults in the United States[6]. Conventional disease-modifying anti-rheumatic drugs (DMARDs), including methotrexate (MTX), sulfasalazine, and leflunomide, have been the cornerstone of the treatment of rheumatoid arthritis (RA). Recently biological agents (biologics), for example etanercept, demonstrated major therapeutic advances in treating RA patients[7].

In clinical practice, the safety of medications for RA patients is an important issue. Many studies focus on adverse drug events (ADEs) associated with either DMARDs or biologics, or their combination, through randomized controlled trials (RCTs)[8], clinical trials[9], systematic reviews[10], meta-analysis[11], and chart reviews[12]. Because RCTs or

clinical trials are not able to reveal all potential ADEs due to experimental limits, post-marketing surveillance becomes an important means of evaluating drug safety. The FDA Adverse Event Reporting System has been used for the discovery of ADEs associated with biologics for RA, mainly aiming at specific ADEs, i.e., ischaemic colitis[13],T-cell non-Hodgkin's lymphomas[13], neurological events[14], and pneumocystis[15]. In one study, several data sources were used to compare the magnitude of serious adverse events (SAEs) observed in post-marketing reports of tocilizumab (TCZ), one of the biologics for RA patients [16]. However, ADEs are not systematically discovered. Interested ADEs included only serious hepatic events, gastrointestinal perforation, and cardiovascular events (myocardial infarction and stroke). In this study, we aim to systematically discover ADEs associated with two drug classes (conventional DMARDs and biologics) based on the framework we propose.

Materials and Methods

Figure 1 shows the framework for ADE mining from FDA's AERS and EMRs that includes three steps: preprocessing, signal detection and validation. In preprocessing, NLP was conducted for clinical notes in the EMRs. During the signal detection, interested drugs were first identified from the *AERS data mining set (*AERS-DM), and then data mining algorithms such as reporting odds ratios (ROR) were conducted to generate potential ADE signals. For the EMRs, interested drugs, the cohort on interested drugs, and outcomes were identified, and then the outcomes before drug use were removed. Lastly, the overlap between ADE signals and outcomes from EMR was further investigated to discover potential ADEs. The details are shown below.

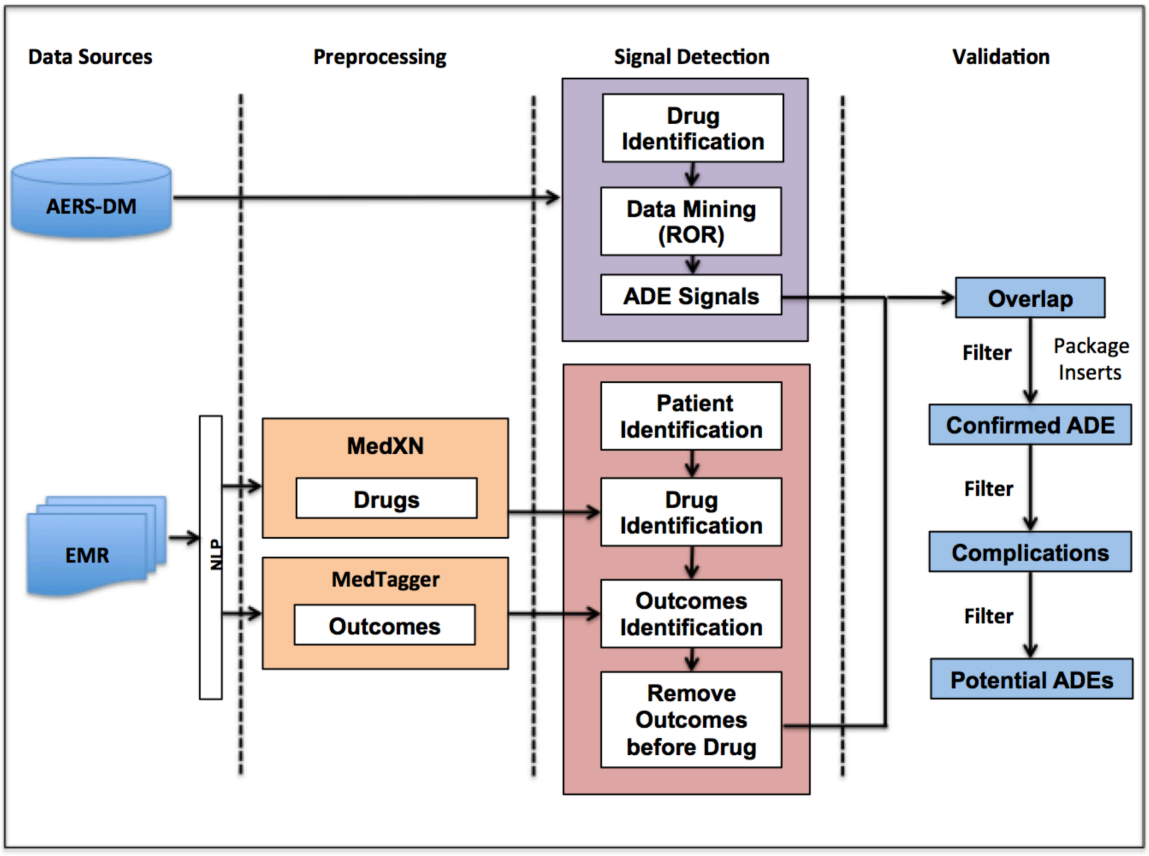

Figure 1. Framework for ADE mining from FDA AERS and EMRs

Data Sources

The FDA's AERS is a database supporting the post-marketing safety surveillance for drug and therapeutic biologic products [17]. However, this database contains redundant data where drugs can also be registered by arbitrary names, including trade names, abbreviations, and even typographical errors. In order to make it convenient for complicated downstream analysis, we previously produced a normalized knowledge-enhanced data mining set based on AERS, i.e., AERS-DM[18]. Three steps were conducted: de-duplication, drug normalization, and data aggregation. First,

redundant reports were removed as suggested by the FDA. This procedure removed multiple reports of the same event. Second, FAERS drug names, along with administration route and dose information, were normalized using a natural language processing (NLP) tool MedEx [19] to RxNorm, a standardized nomenclature for clinical drugs and drug delivery devices [20]. Meanwhile, adverse event terms were mapped to Medical Dictionary for Regulatory Activities (MedDRA)'s preferred term (PT) code and classified into MedDRA System Organ Class (SOC) [21]. Third, adverse events were aggregated according to MedDRA SOC and PT codes, and drugs were aggregated based on National Drug File–Reference Terminology (NDF-RT) classification information through RxNorm [22].

We processed FAERS data from 2004 through 2011 into AERS-DM, which contains 37,029,228 ADE records. In total, 74% of FAERS unique drug names were normalized to 14,489 unique RxNorm concepts, of which 10,221 (71%) were classified in NDF-RT. The datasets of AERS-DM can be downloaded from the website http://informatics.mayo.edu/adepedia/index.php/Download.

EMR clinical notes in our study consist of a cohort of Employee and Community Health (ECH) patients receiving their primary care at Mayo Clinic over a period of 15 years (1998–2013). This cohort include 138,000 patients and covers both inpatient and outpatient settings. Problems (outcomes) in those notes are generally entries which are itemized as either phrases (e.g., Allergic rhinitis/vasomotor rhinitis) or short sentences (e.g, Her asthma appeared to be very mild). In this study, we chose sections related to diagnosis and lab tests for ADE detection.

Preprocessing

To align with the meaningful use requirement, the CORE Problem List Subset was created to better implement Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) in EMRs [23]. The CORE Problem List Subset offers a good coverage of frequently used terms in problem lists [23]. In a previous study [24], we assessed the coverage of SNOMED CT for codifying problem lists in narrative format by extracting itemized entries from clinical notes [25]. In this study, we normalized them to the Unified Medical Language System (UMLS) concepts. We applied the same methodology but kept UMLS concepts that can be mapped to the CORE Problem List Subset codes (the August 2015 version of The CORE Problem List Subset of SNOMED CT was used). Then MedXN was used for the normalization of medications in this cohort to RxNorm codes[26].

Signal detection

DMARDs and biologics are two drug classes for the treatment of RA. DMARDs include methotrexate, leflunomide, hydroxychloroquine and sulfasalazine. Biologics include abatacept, adalimumab, anakinra, certolizumab, etanercept, golimumab, infliximab, rituximab, tocilizumab and tofacitinib.

From AERS-DM, RxNorm codes of these generic ingredients were used to extract records. Drug indications were limited for RA patients. The data mining method reporting odds ratio (ROR) was used to detect associations between drug class DMARDs, biologic use, and ADEs. The calculation of ROR is based on a 2×2 contingency table [27, 28]. The number of reports with drug class and ADE is defined as a. The number of reports with drug class and without ADE is defined as b. The number of reports with drugs other than this drug class and with ADE is defined as c. The number of reports with drugs other drug class and without ADE is defined as d. In this analysis, the lower bound of the 95% confidence interval of the ROR was used[29]. R package PhViD 1.0.6 was used for signal detection[30].

$$ROR = \frac{a \times d}{b \times c}$$

From the normalized data of clinical notes in EMRs, first, synonyms of RA from UMLS were used to identify RA patients, i.e., rheumatoid arthritis or polyarthritis rheumatic. Associated medications, prescription date, and diagnosis date were also extracted. Second, patient cohorts were identified in consideration of clinical priorities based on interested drugs and indications. To study the drug class DMARDs, we identified the cohort of RA patients as those who took any drug in the DMARDs class without drugs of biologics. According to the 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis, conventional DMARDs are usually used for early RA patients, while biologics are often used for moderate or high disease activity, combining with or without DMARDs[31]. To study the drug class biologics, we identified another cohort of RA patients as those who took any drugs within the biologics class, no matter if a drug in the DMARDs class was used in combination. This is also simulating the condition from AERS-DM where data mining of biologics for RA did not consider if DMARSs were used in combination. Therefore, two different cohorts were used for two drug classes. Third, the outcomes of

patients from the two cohorts were identified respectively. Forth, outcomes before the administration of interested drugs were removed to obtain possible ADE signals, i.e., possible consequences of interested drugs.

Validation

After obtaining signals associated with DMARDs and biologics from AERS-DM, MedDRA PT codes were mapped to 2012AB UMLS concepts. The overlapping signals for the two drug classes were further analyzed through mapping PT terms to System Organ Class (SOC) terms. For each drug class, we manually compared the overlapping signals to filter confirmed ADEs from package inserts, and then complications and other confounding factors were filtered to reveal potential ADEs. Some examples were shown using top overlaps of outcomes associated with biologics and DMARDs chosen according to the criteria of ROR more than 2, reporting number in AERS-DM more than 5, and incidence from EMR more than 5%.

Results

In total, there were 497 unique patients with an RA (or synonyms) diagnosis who took only DMARDs, and 365 unique patients with an RA (or synonyms) diagnosis who took biologics no matter if DMARDs were co-administered. Table 1 shows signals from AERS-DM and outcomes from clinical notes. More signals were detected for biologics (152) from both AERS-DM and clinical notes than DMARDs (147).

Table 1. Signal detections from AERS-DM and clinical notes

	Clinical notes			**AERS-DM**		
	No. of patients	No. of outcomes	No. of outcome overlap with AERS-DM (%)	No. of signals	No. mapping to UMLS	No. of signal overlap with clinical notes (%)
DMARDs	497	2,688	**147** (5.5%)	1311	1311	**147** (11.2%)
Biologics	365	2,595	**152** (5.9%)	1450	1448	**152** (10.5%)

The overlapping signals for the two drug classes were further analyzed through mapping PT terms to System Organ Class (SOC) terms. Table 2 shows the number of PT terms (signals) for DMARDs and biologics mapping to SOC. Potential ADEs associated with biologics were involved in more SOCs than those with DMARDs, and the top 6 SOCs were in the same order for potential ADEs associated with both drug classes.

Table 2. Number of PT terms associated with DMARDs and biologics mapping to SOC

System Organ Class (SOC)	**DMARDs**	**Biologics**
Respiratory, thoracic and mediastinal disorders	19	19
Infections and infestations	20	17
Musculoskeletal and connective tissue disorders	16	17
Skin and subcutaneous tissue disorders	15	17
Nervous system disorders	12	15
Surgical and medical procedures	10	13
Injury, poisoning and procedural complications	6	7
Reproductive system and breast disorders	3	7
Blood and lymphatic system disorders	0	6
Gastrointestinal disorders	7	6
Investigations	2	5
Renal and urinary disorders	3	5
Eye disorders	4	3
Hepatobiliary disorders	2	3

Immune system disorders	0	3
Metabolism and nutrition disorders	2	2
Neoplasms benign, malignant and unspecified (incl cysts and polyps)	8	2
Cardiac disorders	1	1
Ear and labyrinth disorders	0	1
General disorders and administration site conditions	2	1
Psychiatric disorders	0	1
Vascular disorders	3	1
Immune system disorders	8	0

For each drug class, we manually compared the overlapping signals with confirmed ADEs from package inserts. Table 3 shows the analysis results. Signals were divided into four categories, the first is confirmed ADEs or signs of ADEs in package inserts such as "vasculitis" for biologics, the second is complications of RA such as "osteoporosis", the third is treatments such as "appendectomy", and the forth is potential ADEs such as "hyperkeratosis" for biologics.

Table 3. Analysis of overlapping signals for each drug class.

	Confirmed ADE	**Complications**	**Treatments**	**Potential ADEs**	**Total**
DMARDs	58 (39.5%)	21 (14.3%)	10(14.7%)	58 (39.5%)	147
Biologics	72 (47.4%)	27 (17.8%)	11 (7.2%)	42 (27.6%)	152

The top potential ADEs associated with biologics and DMARDs were chosen according to the criteria of ROR more than 2, reporting number in AERS-DM more than 5, and incidence from EMR more than 5%. Table 3 and Table 4 show the top potential ADEs for DMARDs and biologics, case number from clinical notes and percentage, report number from AERS-DM, and ROR.

In Table 4, there are 15 signals above the thresholds for DMARDs. There are 7 signals (46%) confirmed as ADEs or signs of ADEs in package inserts, shown in bold and italic. There are 4 signals (27%) identified as complications of RA, shown in italic. Four signals (27%) "Endometrial cancer", "bladder neoplasm", "Sjogren's syndrome", and "Amyotrophic lateral sclerosis" could be possible ADEs following DMARDs that can't be found in package inserts.

Table 4. Top potential ADEs for DMARDs. Bold and italic indicate confirmed ADEs or ADE signs, italic indicates complications of RA, and bold indicates possible ADEs.

Signals	**UMLS codes**	**Case number from clinical notes (%)**	**Report number from AERS-DM**	**ROR**
Endometrial cancer	C0476089	128(25.8%)	30	2.87
Rhinorrhea	C1260880	111(22.3%)	355	2.81
Productive cough	C0239134	100(20.1%)	346	2.83
Bladder neoplasm	C0496930	71(14.3%)	17	2.60
Gastroduodenal ulcer	C0030920	54(10.9%)	7	3.07
Amyotrophic lateral sclerosis	C0002736	44(8.9%)	26	2.71
Sinus congestion	C0152029	40(8.0%)	237	4.87

Sjogren's syndrome	C1527336	36(7.2%)	72	5.54
Respiratory tract congestion	C0242073	36(7.2%)	230	6.59
Bunion	C0006386	34(6.8%)	79	6.41
Sinus headache	C0037195	32(6.4%)	110	3.40
Antinuclear antibody positive	C0151480	32(6.4%)	92	3.01
Metatarsalgia	C0025587	31(6.2%)	7	3.57
Rash pruritic	C0033771	27(5.4%)	330	2.20
Red blood cell sedimentation rate increased	C0151632	26(5.2%)	228	4.07

In Table 5, there are 18 signals above the thresholds for biologics. There are 12 signals (67%) confirmed as ADEs or signs of ADEs in package inserts, shown in bold and italic. There are 2 signals (11%) identified as complications of RA, shown in italic. Two signals (11%) "steroid therapy" and "laparoscopy" could be excluded from ADEs, since they are treatments instead of undesirable effects. The left 2 signals (11%), "Sjogren's syndrome" and "amyotrophic lateral sclerosis" could be possible ADEs following DMARDs that can't be found in package inserts.

Table 5. Top potential ADEs for biologics. Bold and italic indicate confirmed ADEs or ADE signs, italic indicates complications of RA, and bold indicates possible ADEs.

Signals	UMLS codes	Case number from clinical notes (%)	Report number from AERS-DM	ROR
Endometrial cancer	C0476089	101(27.7%)	51	2.49
Rhinorrhea	C1260880	82(22.5%)	887	3.64
Productive cough	C0239134	79(21.6%)	684	2.87
Bladder neoplasm	C0496930	66(18.1%)	28	2.18
Sjogren's syndrome	C1527336	41(11.2%)	106	4.17
Sinus congestion	C0152029	36(9.9%)	516	5.54
Respiratory tract congestion	C0242073	31(8.5%)	574	8.80
Amyotrophic lateral sclerosis	C0002736	30(8.2%)	47	2.50
Sinus headache	C0037195	29(7.9%)	244	3.90
Steroid therapy	C0149783	28(7.7%)	8	3.63
Rash pruritic	C0033771	25(6.8%)	195	2.91
Oral herpes	C0019345	23(6.3%)	235	3.21
Foot operation	C0188413	23(6.3%)	195	10.63
Squamous cell carcinoma	C0007137	21(5.8%)	275	3.28
Wound	C0033119	20(5.5%)	233	2.69
Laparoscopy	C0031150	20(5.5%)	7	6.82
Bunion	C0006386	20(5.5%)	161	6.85
Pneumonia primary atypical	C1412002	19(5.2%)	55	2.01

Discussion

In this study, we demonstrated the framework by exploring potential ADEs associated with drugs for RA patients. ADEs associated with drug class DMARDs and biologics for RA patients were first systematically mined from AERS-DM. Corpuses of RA patients on each drug class were then carefully selected according to the clinical guidelines. Following that, outcomes following drug uses were revealed from unstructured EMRs, and the overlaps between the signals and the outcomes of RA patients on these drugs were further analyzed to identify potential ADEs. RA is a systemic autoimmune disease with the characteristics of chronic inflammation that results in a

destructive polyarthritis. Many complications may occur after RA. Therefore, we fully considered the features of RA to exclude possible complications from overlaps between signals from AERS-DM and outcomes from EMRs.

In view of various regimens used among different institutions, some drugs used in one institution may not be used in another. EMR data from only a single institution, i.e., Mayo Clinic, was used in this study. To avoid omitting information on drugs and indications, our method doesn't aim for screening whole databases as done in the previous study[5]. Instead, demonstrated as a framework interfacing informatics domain and medical domain, it employed more refined strategies based on interested drugs and indications. In the future, we will develop more general methodology once EMR data from multiple institutions, such as Optum lab, can be obtained.

Some adverse events occur after a short time following drug use, from several minutes to several hours. Others occur only after several days, weeks, months or even years of exposure [4]. Therefore, when extracting outcomes, we have not limited the time of outcome occurrence after drug use. This allows detection of late-onset events. However, it may be interesting to observe the difference of time of outcome occurrences in the future.

During the result analysis, we found that potential ADEs such as "Endometrial cancer" and "bladder neoplasm" for conventional DMARDs could also be the natural consequences of RA. Because the disease is a systemic autoimmune disease, patients with RA are at an increased risk for cancer[32]. In the future, we will integrate case-control study design into the framework based on EMR data to further discriminate such potential ADEs from co-morbidities with indications of interested drugs.

Conclusions

We proposed a framework for discovering potential ADEs associated with drugs combining both FDA AERS and EMRs. This framework has two features. First, more priority was given to clinical practice. Second, the normalization by NLP facilitated the interoperation between AERS-DM and EMRs. The results demonstrated the feasibility and sufficient accuracy of the framework. The framework can serve as the interface between the informatics domain and the medical domain to facilitate ADE discovery.

Acknowledgements

This work was made possible by joint funding from National Institute of Health, R01GM102282, R01LM11369, R01LM011829, and R01 LM011934, and National Natural Science Foundation of China (NSFC), No. 81601574. The authors would like to thank Mrs. Katelyn N Cordie for her editing assistance with the manuscript.

References:

1. Pirmohamed M, Breckenridge AM, Kitteringham NR, Park BK. Adverse drug reactions. British Medical Journal 1998;316(7140):1295.
2. Classen DC, Resar R, Griffin F, et al. 'Global trigger tool'shows that adverse events in hospitals may be ten times greater than previously measured. Health affairs 2011;30(4):581-589.
3. Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. Journal of the American Medical Informatics Association 2009;16(3):328-337.
4. Wang X, Chase H, Markatou M, Hripcsak G, Friedman C. Selecting information in electronic health records for knowledge acquisition. Journal of biomedical informatics 2010;43(4):595-601.
5. Harpaz R, Vilar S, DuMouchel W, et al. Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. Journal of the American Medical Informatics Association 2013;20(3):413-419.
6. Helmick CG, Felson DT, Lawrence RC, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part I. Arthritis & Rheumatism 2008;58(1):15-25.
7. Schmitz S, Adams R, Walsh CD, Barry M, FitzGerald O. A mixed treatment comparison of the efficacy of anti-TNF agents in rheumatoid arthritis for methotrexate non-responders demonstrates differences between treatments: a Bayesian approach. Annals of the rheumatic diseases 2011:annrheumdis-2011-200228.
8. Smolen JS, van Vollenhoven R, Kavanaugh A, et al. Certolizumab pegol plus methotrexate 5-year results from the rheumatoid arthritis prevention of structural damage (RAPID) 2 randomized controlled trial and long-term extension in rheumatoid arthritis patients. Arthritis research & therapy 2015;17(1):1.
9. Isaacs JD, Zuckerman A, Krishnaswami S, et al. Changes in serum creatinine in patients with active rheumatoid arthritis treated with tofacitinib: results from clinical trials. Arthritis research & therapy 2014;16(4):1.

10. Roubille C, Haraoui B. Interstitial lung diseases induced or exacerbated by DMARDS and biologic agents in rheumatoid arthritis: a systematic literature review. Seminars in arthritis and rheumatism: Elsevier; 2014: 613-626.
11. Barnabe C, Martin BJ, Ghali WA. Systematic review and meta‐analysis: Anti‐tumor necrosis factor α therapy and cardiovascular events in rheumatoid arthritis. Arthritis care & research 2011;63(4):522-529.
12. Müller RB, von Kempis J, Haile SR, Schiff MH. Effectiveness, tolerability, and safety of subcutaneous methotrexate in early rheumatoid arthritis: a retrospective analysis of real-world data from the St. Gallen cohort. Seminars in arthritis and rheumatism: Elsevier; 2015: 28-34.
13. Salk A, Stobaugh DJ, Deepak P, Ehrenpreis ED. Ischaemic colitis in rheumatoid arthritis patients receiving tumour necrosis factor-α inhibitors: An analysis of reports to the US FDA adverse event reporting system. Drug safety 2013;36(5):329-334.
14. Deepak P, Stobaugh D, Sherid M, Sifuentes H, Ehrenpreis E. Neurological events with tumour necrosis factor alpha inhibitors reported to the Food and Drug Administration Adverse Event Reporting System. Alimentary pharmacology & therapeutics 2013;38(4):388-396.
15. Kaur N, Mahl TC. Pneumocystis jiroveci (carinii) pneumonia after infliximab therapy: a review of 84 cases. Digestive diseases and sciences 2007;52(6):1481-1484.
16. Curtis JR, Perez-Gutthann S, Suissa S, et al. Tocilizumab in rheumatoid arthritis: a case study of safety evaluations of a large postmarketing data set from multiple data sources. Seminars in arthritis and rheumatism: Elsevier; 2015: 381-388.
17. Questions and Answers on FDA's Adverse Event Reporting System (FAERS). [cited 2016 11 Apr]; Available from: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm
18. Wang LW, Jiang GQ, Li DC, Liu HF. Standardizing adverse drug event reporting data. Journal of biomedical semantics 2014 Aug 12;5.
19. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. Journal of the American Medical Informatics Association 2010;17(1):19-24.
20. Nelson SJ, Zeng K, Kilbourne J, Powell T, Moore R. Normalized names for clinical drugs: RxNorm at 6 years. Journal of the American Medical Informatics Association 2011;18(4):441-448.
21. Pearson RK, Hauben M, Goldsmith DI, et al. Influence of the MedDRA® hierarchy on pharmacovigilance data mining results. International journal of medical informatics 2009;78(12):e97-e103.
22. Pathak J, Murphy SP, Willaert BN, et al. Using RxNorm and NDF-RT to classify medication data extracted from electronic health records: experiences from the Rochester Epidemiology Project. AMIA Annual Symposium Proceedings: American Medical Informatics Association; 2011: 1089.
23. Agrawal A, He Z, Perl Y, et al. The readiness of SNOMED problem list concepts for meaningful use of electronic health records. Artificial intelligence in medicine 2013;58(2):73-80.
24. Liu H, Wagholikar K, Wu ST-I. Using SNOMED-CT to encode summary level data–a corpus analysis. AMIA Summits on Translational Science Proceedings 2012;2012:30.
25. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research 2004;32(suppl 1):D267-D270.
26. Sohn S, Clark C, Halgrim SR, Murphy SP, Chute CG, Liu H. MedXN: an open source medication extraction and normalization tool for clinical text. Journal of the American Medical Informatics Association 2014;21(5):858-865.
27. Rothman KJ, Lanes S, Sacks ST. The reporting odds ratio and its advantages over the proportional reporting ratio. Pharmacoepidemiology and drug safety 2004 Aug;13(8):519-523.
28. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiology and drug safety 2001;10(6):483-486.
29. van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiology and drug safety 2002;11(1):3-10.
30. Ahmed I, Poncet A. PhViD: an R package for PharmacoVigilance signal Detection. R package version 1.0. 6.(2013). 2014.
31. Singh JA, Saag KG, Bridges SL, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis & Rheumatology 2016;68(1):1-26.

32. Chen YJ, Chang YT, Wang CB, Wu CY. The risk of cancer in patients with rheumatoid arthritis: a nationwide cohort study in Taiwan. Arthritis & Rheumatism 2011;63(2):352-358.