HOKKAIDO UNIVERSITY

Title	Prodromal signs and symptoms of serious infections with tocilizumab treatment for rheumatoid arthritis : Text mining of the Japanese postmarketing adverse event-reporting database
Author(s)	Atsumi, Tatsuya; Ando, Yoshiaki; Matsuda, Shinichi et al.
Citation	Modern rheumatology, 28(3), 435-443 https://doi.org/10.1080/14397595.2017.1366007
Issue Date	2018
Doc URL	https://hdl.handle.net/2115/71491
Rights(URL)	https://creativecommons.org/licenses/by-nc-nd/4.0/
Type	journal article
File Information	ModRheumatol28_435.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

MODERN RHEUMATOLOGY, 2018
VOL. 28, NO. 3, 435–443
https://doi.org/10.1080/14397595.2017.1366007

Japan College of Rheumatology
MODERN RHEUMATOLOGY
Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS Check for updates

Prodromal signs and symptoms of serious infections with tocilizumab treatment for rheumatoid arthritis: Text mining of the Japanese postmarketing adverse event-reporting database

Tatsuya Atsumi[a], Yoshiaki Ando[b], Shinichi Matsuda[b], Shiho Tomizawa[b], Riwa Tanaka[b], Nobuhiro Takagi[b] and Ayako Nakasone[b]

[a]Division of Rheumatology, Endocrinology and Nephrology, Hokkaido University Graduate School of Medicine, Sapporo, Japan; [b]Chugai Pharmaceutical Co., Ltd., Tokyo, Japan

ABSTRACT

Objective: To search for signs and symptoms before serious infection (SI) occurs in tocilizumab (TCZ)-treated rheumatoid arthritis (RA) patients.

Methods: Individual case safety reports, including structured (age, sex, adverse event [AE]) and unstructured (clinical narratives) data, were analyzed by automated text mining from a Japanese post-marketing AE-reporting database (16 April 2008–10 April 2015) assuming the following: treated in Japan; TCZ RA treatment; ≥1 SI; unable to exclude causality between TCZ and SIs.

Results: The database included 7653 RA patients; 1221 reports met four criteria, encompassing 1591 SIs. Frequent SIs were pneumonia (15.9%), cellulitis (9.9%), and sepsis (5.0%). Reports for 782 patients included SI onset date; 60.7% of patients had signs/symptoms ≤28 days before SI diagnosis, 32.7% had signs/symptoms with date unidentified, 1.7% were asymptomatic, and 4.9% had unknown signs/symptoms. The most frequent signs/symptoms were for skin (swelling and pain) and respiratory (cough and pyrexia) infections. Among 68 patients who had normal laboratory results for C-reactive protein, body temperature, and white blood cell count, 94.1% had signs or symptoms of infection.

Conclusion: This study identified prodromal signs and symptoms of SIs in RA patients receiving TCZ. Data mining clinical narratives from post-marketing AE databases may be beneficial in characterizing SIs.

ARTICLE HISTORY
Received 10 May 2017
Accepted 3 August 2017

KEYWORDS
Adverse events; infections; rheumatoid arthritis; text mining; tocilizumab

Introduction

Current treatment guidelines for rheumatoid arthritis (RA) recommend the use of conventional synthetic disease-modifying antirheumatic drugs (csDMARDs)—methotrexate in the first instance—and biological DMARDs (bDMARDs) with an aim to achieving remission or low disease activity [1–3].

Tocilizumab (TCZ), a recombinant humanized anti-interleukin-6 (IL-6) receptor monoclonal antibody, has been approved for the treatment of patients with RA in more than 100 countries. The efficacy of TCZ, which has a well-established safety profile, as monotherapy or in combination with csDMARDs has been demonstrated in patients with RA who have inadequate responses to csDMARDs. Data from clinical [4–6] and post-marketing surveillance studies [7] of TCZ indicate that infections are the most frequent adverse events (AEs) and serious AEs (SAEs) associated with TCZ therapy. TCZ exerts its therapeutic effects through inhibition of IL-6 signaling, which in turn suppresses the production of proinflammatory acute-phase reactants, including C-reactive protein (CRP) [8,9]. Therefore, there is concern that early signs and symptoms of infection, such as elevated CRP and fever, may not be detected in patients receiving TCZ, potentially allowing the infections to become serious. This highlights the need for identifying signs and symptoms that occur before worsening of an infection to allow for early and adequate safety management.

Electronic medical records and post-marketing AE-reporting databases that contain information on adverse drug reactions (ADRs) are a vast source of valuable patient information [10]. Post-marketing AE-reporting databases are among the fundamental resources in pharmacovigilance [11]. Previous reports, including some in patients treated with TCZ, have focused on structured data elements, such as age, sex, and onset of disease [7,12,13]. However, other recent analyses suggest that clinical narratives may also be useful for obtaining important clinical insights [14–18]. This approach may involve text mining, which enables efficient analysis of unstructured data elements from clinical narratives [15,19]. Indeed, text mining of clinical sources, among them electronic health care data and clinical notes and narratives, has been used in recent large-scale analyses of drug safety and diseases [16,17,20], including those in patients with RA [21].

CONTACT Tatsuya Atsumi at3tat@med.hokudai.ac.jp Division of Rheumatology, Endocrinology and Nephrology, Hokkaido University Graduate School of Medicine, Kita 15, Nishi 7, Kita-ku, Sapporo, 060-8638, Japan

Supplemental data for this article can be accessed here.

© 2017 Japan College of Rheumatology. Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Text mining electronic medical records has been used successfully to obtain insights from clinical narratives, such as enhancing opportunities for the early detection of RA [16,21]. However, this approach may be time consuming and costly because of practical issues requiring collaboration between several hospitals and integration of their data formats [21,22] to acquire enough patient data for analysis. In addition, it is customary to report diagnostic terms (e.g. pneumonia) rather than symptomatic terms (e.g. cough, pain, fever) when identifying AEs, resulting in a large amount of unused textual data regarding the signs and symptoms of AEs.

The analysis reported here used data from the post-marketing AE-reporting database maintained by Chugai Pharmaceutical Co., Ltd. (Tokyo, Japan), which compiles post-marketing AE data from locations throughout Japan and stores them in a consistent format, including unstructured clinical narratives reported with TCZ treatment. A retrospective analysis was performed by text mining data from the post-marketing AE-reporting database to identify symptoms before the development of a serious infection in TCZ-treated patients. To our knowledge, this is the first report of text mining of clinical narratives using a post-marketing AE-reporting database.

Methods

Data source

Data for this analysis were compiled from an AE-reporting database maintained by Chugai Pharmaceutical Co., Ltd., that includes all individual case safety reports (ICSRs) for the manufacturer's authorized products. These ICSRs contain structured (age, sex, concomitant drugs, AE term, laboratory test values) and unstructured (clinical narratives) data from spontaneous reports of ADRs, post-marketing drug use surveillance, and published reports from the literature. All AEs and comorbidities were coded in accordance with the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use guidelines and the Medical Dictionary for Regulatory Activities (MedDRA, McLean, VA) (version 17.1.) [23] by predefined system organ class (SOC) and preferred term (PT). This study was registered at UMIN Clinical Trials Registry (UMIN000024071). The database used in this study is anonymously managed and contains no personal information. Ethics review was completed by the ethical review board of the Japanese Association for the Promotion of State-of-the-Art in Medicine (JAPSAM) on 19 April 2016; no ethical issues related to the study were identified.

Definition of serious infections

The analysis set included TCZ ICSRs, first reported between 16 April 2008 (the date TCZ was first approved for the treatment of patients with RA in Japan) and 10 April 2015, and it met all four of the following conditions: (i) the patient was treated in Japan; (ii) TCZ was administered to treat RA; (iii) the case report contained at least one reported AE of a serious infection following treatment with TCZ; and (iv) a causal relationship between TCZ and the serious infection could not be ruled out. This last condition was defined as follows: included in the 'infections and infestations' SOC in MedDRA [23]; MedDRA judged to be serious by the reporting health care practitioner, the manufacturer, or both; a relationship between TCZ and the serious infection was determined by the reporting health care professional, the manufacturer, or both. The number of patients with serious infections and the proportion of serious infections were determined based on the site where the infection had its onset and on the MedDRA PT. Subsequent analyses were performed in patients for whom the date the infection became serious could be identified from narratives. Infections were considered to be serious on the day the patient was admitted to the hospital or transported to the hospital.

Assessments: signs and symptoms of serious infections

The analysis included investigation of clinical findings and laboratory data that could be signs and symptoms of infections becoming serious.

Signs and symptoms of serious infection were defined as any clinical symptoms that occurred within 28 days before the onset of a serious infection. Using text mining, we extracted possible signs and symptomatic terms from unstructured narratives. These possible signs and symptoms of serious infection for each site at which serious infection had its onset were coded by MedDRA PTs and reviewed to determine whether they were already generally known as signs or symptoms of infection (i.e. previously reported in medical textbooks, such as The Merck Manual of Diagnosis and Therapy [24]) or whether they were newly identified symptoms. If they were newly identified, they were assessed by three of the authors (reviewed by Y.A. and S.T. and confirmed by T.A.). The number of days from the first occurrence of a sign or symptom until an infection became serious was calculated for each site of serious infection and by a sign or symptom; if any part of a date (day, month, or year) was missing, the data were excluded.

Relationships between laboratory data and infections becoming serious were explored based on different thresholds of CRP levels (<1.0 mg/dL/≥ 1.0 mg/dL), white blood cell (WBC) count ($<10{,}000$ μL/$\geq 10{,}000$ μL), and body temperature ($<37.0\,^{\circ}$C/$\geq 37.0\,^{\circ}$C to $<38\,^{\circ}$C/$\geq 38.0\,^{\circ}$C) using laboratory data nearest to the day the infection became serious and measured within 7 days; again, if any part of a date (day, month, or year) was missing, the data were excluded. Also excluded from the analysis as outliers were values outside the body temperature range of 34–42 °C, WBC count within the two-sided 95% distribution range, and CRP level outside the upper bound of the one-sided 95% distribution range.

Statistical analyses

Patient characteristics and data regarding serious infections were summarized using descriptive statistics. Unstructured narratives were analyzed with the use of a text mining

approach to extract signs and symptoms of serious infection (within 28 days of an infection becoming serious) from patient narratives in the database of patients who developed serious infections, and symptoms were converted to structured data (see Supplementary Appendix A1 for the structuring procedure).

Text mining was performed using VextMiner version 9.0 (Vext Inc., Tokyo, Japan), and data processing and analysis were performed using SAS 9.4 (SAS Institute Inc., Cary, NC). Coding was performed to standardize terms in ICSRs using MedDRA version 17.1.

Results

Patient characteristics

A total of 7653 patients with RA were identified from the AE-reporting database, of whom 1221 met the four criteria for serious infection and 782 had known onset dates of serious infections (Figure 1). Patient characteristics were similar among the overall RA population, patients who developed serious infections, and patients with known onset dates of serious infections (Table 1). There were slight differences in age; patients in the total RA population were younger than those who developed serious infections and those who had known onset dates of serious infections (mean age: 60.0 versus 63.4 and 63.2 years, respectively). Glucocorticoid use was reported by approximately one-third (399/1221; 32.7%) of patients who developed serious infections and those with known onset dates of serious infections (300/782; 38.4%) (Table 1). Less than 5% of patients who developed serious infections had previously received bDMARDs (53/1221; 4.3%); the most common previous bDMARDs reported were etanercept [received by 20 patients (1.6%)] and adalimumab [received by 17 patients (1.4%)]. Concomitant csDMARD use was reported by approximately two-thirds (771/1221; 63.1%) of patients who developed serious infections and patients with known onset dates of serious infections (547/782; 69.9%). The most common concomitant csDMARD reported was methotrexate, which was received by 574 patients (55.4%) (Table 1).

Serious infection

In total, 1591 serious infections were reported in 1221 patients. The mean (±standard deviation) number of days from the initiation of TCZ treatment to the onset of serious infection in 750 patients with available data was 318.0 (±423.0) days. Sites for the onset of serious infection were respiratory (623/1591; 39.2%), skin (329/1591; 20.7%), gastrointestinal tract (183/1591; 11.5%), sepsis (137/1591; 8.6%), urinary tract (72/1591; 4.5%), and other (247/1591; 15.5%) (Table 2). Common serious infections were pneumonia (253/1591; 15.9%), cellulitis (158/1591; 9.9%), and sepsis (80/1591; 5.0%).

Signs and symptoms of serious infection

Of the 782 patients for whom the date of confirmed diagnosis of serious infection could be identified, 475 (60.7%) had signs or symptoms within 28 days before the diagnosis of serious infection, 256 patients (32.7%) had signs or symptoms but their date could not be identified, 13 patients (1.7%) had no signs or symptoms, and it was unknown whether 38 patients (4.9%) had signs or symptoms.

The most common signs and symptoms of serious infection were pyrexia, pain, cough, and swelling (Table 3).

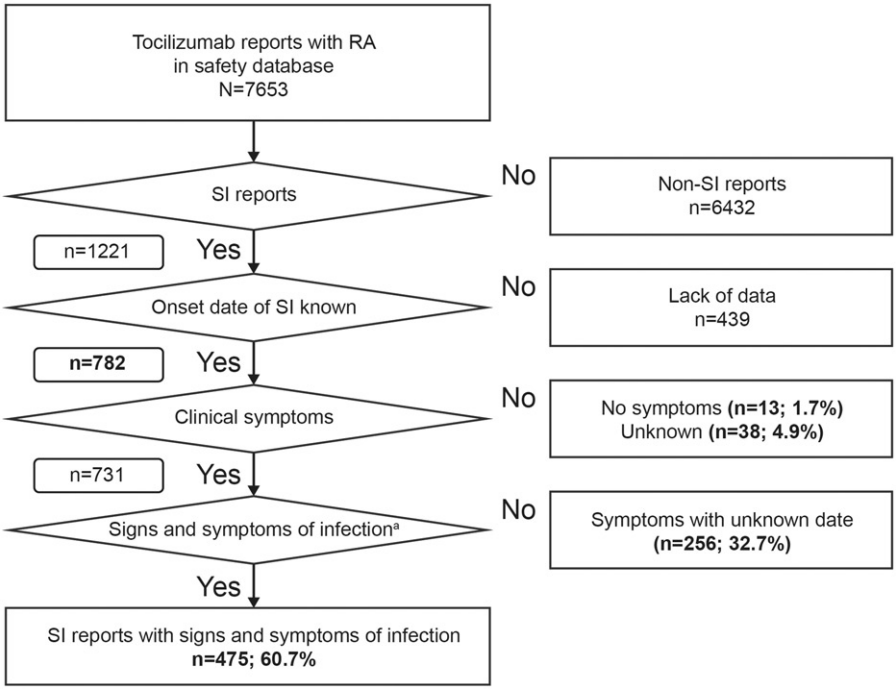

Figure 1. Data disposition used to analyze reports of serious infection. [a]Symptoms within 28 days before the date of onset of serious infection. RA: rheumatoid arthritis; SI: serious infection.

Table 1. Patient characteristics.

	All patients with rheumatoid arthritis, $N = 7653$	All patients who developed serious infections, $N = 1221$	Patients with known onset dates of serious infections, $N = 782$
Sex			
Female	5959 (77.9)	890 (72.9)	583 (74.6)
Male	1417 (18.5)	301 (24.7)	197 (25.2)
Unknown	277 (3.6)	30 (2.5)	2 (0.3)
Age, mean (SD), years	60.0 (12.7)[a]	63.4 (11.7)[b]	63.2 (11.8)[c]
Age category, *n* (%)			
<65 years	3507 (45.8)	434 (35.5)	315 (40.3)
≥65 years	2322 (30.3)	460 (37.7)	336 (43.0)
Unknown	1824 (23.8)	327 (26.8)	131 (16.8)
Weight, mean (SD), kg	53.5 (11.5)[d]	53.1 (11.8)[e]	53.1 (11.9)[f]
BMI, mean (SD)	22.5 (17.2)[g]	22.0 (3.9)[h]	22.0 (3.9)[i]
Number of days from initiating tocilizumab until initial onset of serious infection, mean (SD)	–	318.0 (423.0)[j]	–
Other medications			
Concomitant glucocorticoid use			
Yes	–	399 (32.7)	300 (38.4)
No	–	822 (67.3)	482 (61.6)
Prednisolone-equivalent glucocorticoid dose (mg/day), mean (SD)	–	7.3 (5.1)[k]	7.3 (5.1)[l]
Concomitant use of csDMARD			
Yes	–	771 (63.1)	547 (69.9)
No	–	450 (36.9)	235 (30.1)
Type of concomitant csDMARD			
Methotrexate	–	574 (55.4)	–
Tacrolimus hydrate	–	153 (14.8)	–
Salazosulfapyridine	–	132 (12.7)	–
Bucillamine	–	86 (8.3)	–
Mizoribine	–	37 (3.6)	–
Leflunomide	–	16 (1.5)	–
Cyclosporine	–	8 (0.8)	–
Penicillamine	–	8 (0.8)	–
Iguratimod	–	7 (0.7)	–
Azathioprine	–	6 (0.6)	–
Gold sodium thiomalate	–	5 (0.5)	–
Actarit	–	2 (0.2)	–
Auranofin	–	2 (0.2)	–
Previous use of a biologic			
Yes	–	53 (4.3)	42 (5.4)
No	–	1168 (95.7)	740 (94.6)
Type of previous biologic			
Etanercept	–	20 (1.6)	21 (2.7)
Adalimumab	–	17 (1.4)	18 (2.3)
Infliximab	–	12 (1.0)	10 (1.3)
Abatacept	–	7 (0.6)	6 (0.8)
Golimumab	–	3 (0.2)	3 (0.4)

BMI: body mass index; SD: standard deviation. All data are *n* (%) unless otherwise stated. Data for additional medications are available only for patients who developed serious infections.
[a]$n = 5829$.
[b]$n = 894$.
[c]$n = 651$.
[d]$n = 2246$.
[e]$n = 482$.
[f]$n = 353$.
[g]$n = 1515$.
[h]$n = 380$.
[i]$n = 286$.
[j]$n = 750$.
[k]$n = 378$.
[l]$n = 291$.

Swelling and pain were counted in large numbers regardless of the site of infection, but many other symptoms were specific to the site of serious infection (Supplementary Table S1).

The five most commonly reported signs or symptoms of infection were most often reported in the week preceding the diagnosis of serious infection (Figure 2). The median number of days from the first occurrence of signs or symptoms of serious infection to the onset of a serious infection ranged from 1 to 14 days (Supplementary Table S2). Signs and symptoms of serious infection occurred on the day the infection became serious in 475 (60.7%) of the 782 patients for whom the date of diagnosis of the infection as serious could be identified. The most common signs and symptoms that occurred on the day the infection became serious were pyrexia (138/782; 17.6%), productive

cough (95/782; 12.1%), and dyspnea (77/782; 9.8%) (Supplementary Table S3).

Relationship between laboratory data and serious infection

Among patients whose laboratory test results were available, similar proportions of CRP level measurements were <1.0 mg/dL (43.5%) or >1.0 mg/dL (56.5%), similar proportions of body temperature measurements were <37.0 °C (51.1%) or ≥37 °C (48.9%), and similar proportions of WBC counts were <10,000/μL (57.2%) or ≥10,000/μL (42.8%) when measured within 7 days before infections became serious (Table 4). In reports with known onset dates of serious infection, 8.7% (68/782) of patients had normal laboratory test results (CRP <1.0 mg/dL, body temperature <37.0 °C, and WBC count <10,000/μL). Among these 68 patients with normal laboratory test results, four had no signs or symptoms of serious infection and 64 had signs or symptoms of serious infection (Table 4).

Table 2. Site of onset for most frequently reported[a] serious infections in patients who developed serious infections ($N = 1221$).

Type of serious infection by SOC and PT[a]	Incidence of event, *n* (%) (total number of events = 1591)
Respiratory tract infections	623 (39.2)
Pneumonia	253 (15.9)
Pneumocystis jirovecii pneumonia	59 (3.7)
Nontuberculosis mycobacterial infection	55 (3.5)
Pneumonia bacterial	44 (2.8)
Pneumonia pneumococcal	25 (1.6)
Infectious pleural effusion	24 (1.5)
Skin infection	329 (20.7)
Cellulitis	158 (9.9)
Herpes zoster	56 (3.5)
Subcutaneous abscess	17 (1.1)
Gastrointestinal infection	183 (11.5)
Peritonitis	48 (3.0)
Diverticulitis	43 (2.7)
Urinary tract infection	72 (4.5)
Pyelonephritis	26 (1.6)
Urinary tract infection	26 (1.6)
Sepsis	137 (8.6)
Sepsis	80 (5.0)
Septic shock	36 (2.3)
Other infection	247 (15.5)
Arthritis bacterial	51 (3.2)
Arthritis infective	17 (1.1)

[a]Data for serious infections reported for ≥1% of patients.

Discussion

This retrospective analysis of a Japanese postmarketing AE-reporting safety database successfully used text mining to identify trends for signs and symptoms of serious infections reported within 28 days before the infections became serious.

Demographics and characteristics of patients in the AE-reporting safety database were generally similar between all RA patients and the subset who developed serious infection, except that patients who developed infection tended to be older than the total population, which might be expected given the increased risk for infection in older patients [7]. It is difficult to provide an interpretation of difference in body weight between the patient groups because data were missing for more than half the patients.

The use of text mining methods in our analysis enabled automatic extraction of potential candidate signs and symptoms of serious infection from clinical narratives entered in the post-marketing AE-reporting database. In this analysis, the most common signs and symptoms indicative of an infection becoming serious were pyrexia, pain, cough, and swelling, and the most common signs and symptoms on the day the infection became serious were pyrexia, productive cough, and dyspnea. Pain was identified as a new sign or symptom for respiratory tract infections, skin infections, gastrointestinal infections, sepsis, urinary tract infections (as well as back pain), and other infections. Other newly identified signs or symptoms of infection were swelling with skin infections, sepsis, and other infections, rash with skin infections, and nausea with urinary tract infections.

Analysis of laboratory results for patients who developed serious infections suggested that laboratory test data (CRP level, body temperature, WBC count) were not associated with the development of a serious infection; however, it was considered that these measures alone may not be sufficient to identify infections likely to become serious in patients treated with TCZ. Of note, only four patients with normal laboratory test results did not have any signs or symptoms of infection, and some patients had signs and symptoms of infection despite normal laboratory test results, which meant that most patients in the analysis with normal laboratory test results had some type of sign or symptom of infection. Therefore, to prevent infections from becoming serious during TCZ treatment, it is crucial to acknowledge the signs and symptoms of infection detected in this analysis and to educate patients to contact their health care professionals should they experience any of these signs or symptoms.

Table 3. Signs and symptoms of infection[a] that occurred in patients who developed serious infection and had known onset dates of serious infections ($N = 782$).

Rank	Respiratory tract infection ($n = 623$ events)	Skin infection ($n = 329$ events)	Gastrointestinal infection ($n = 183$ events)	Sepsis ($n = 137$ events)	Urinary tract infection ($n = 72$ events)	Other infection ($n = 247$ events)	All infection ($n = 782$ patients)
1	Cough	Swelling[b]	Abdominal pain	Pyrexia	Vomiting[b]	Pain[b]	Pyrexia
2	Pyrexia	Pain[b]	Pain[b]	Pain[b]	Pyrexia	Swelling[b]	Pain[b]
3	Productive cough	Erythema	Diarrhea	Cough[b]	Pain[b]	Pyrexia	Cough[b]
4	Dyspnea	Rash	Pyrexia	Swelling[b]	Back pain[b]	Erythema[b]	Swelling[b]
5	Pain[b]	Pyrexia	Vomiting	Malaise	Nausea[b]	Arthralgia	Productive cough[b]

[a]MedDRA PT.
[b]Previously unknown as a sign or symptom of infection.

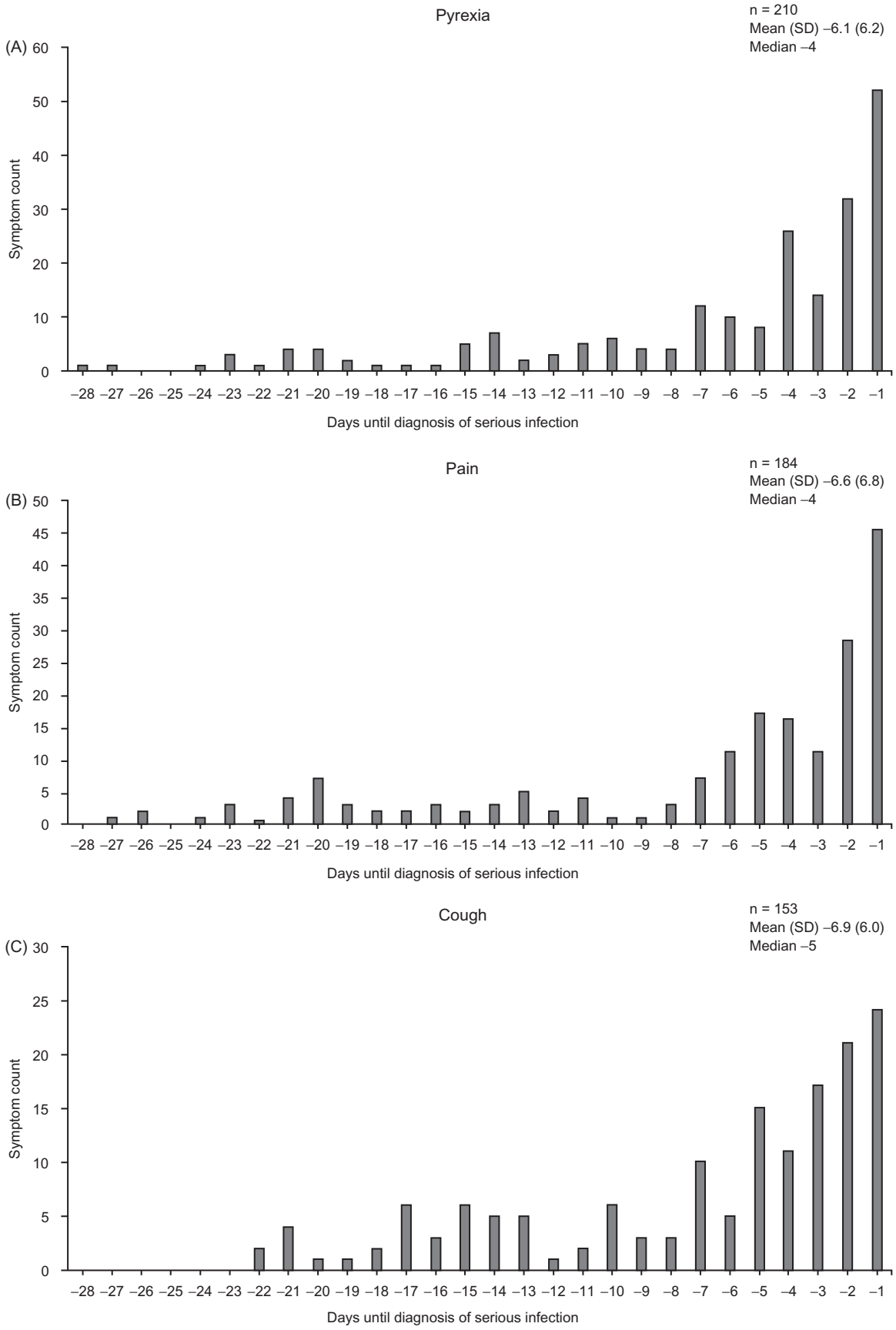

Figure 2. Frequency of the five most common signs or symptoms of infection (A–E) according to time before diagnosis of serious infection. SD: standard deviation.

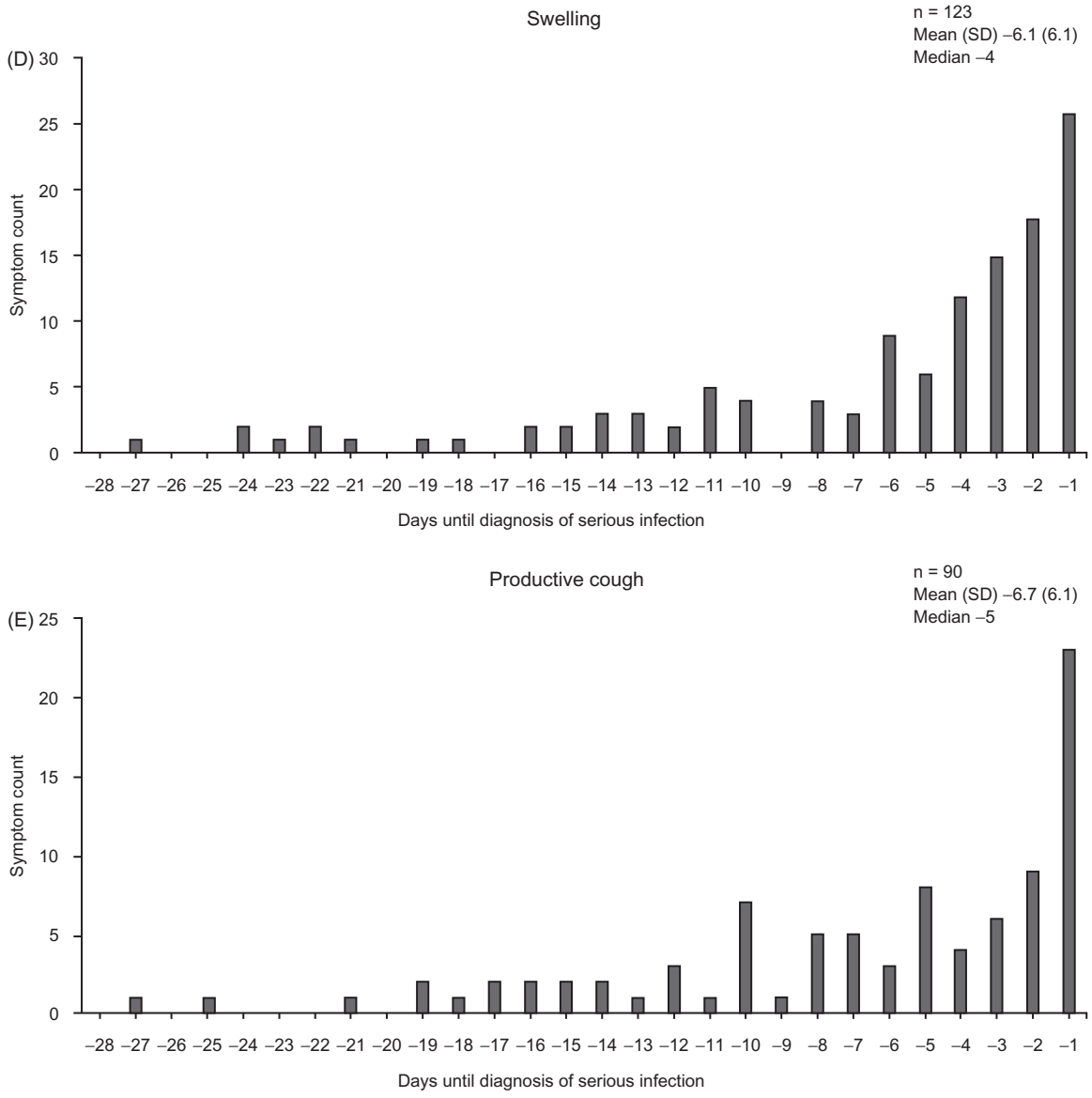

Figure 2. Continued.

Table 4. Laboratory results measured within 7 days of infections becoming serious.

	Result		
Laboratory test	Onset date of SI known, $N = 782$[a]	SI reports with signs or symptoms, $N = 475$[a]	No symptoms and unknown signs or symptoms, $N = 51$[a]
CRP, mean (SD), mg/dL	3.4 (4.4), $N = 619$[b]	NR, $N = 433$[b]	NR, $N = 25$[b]
Incidence <1.0 mg/dL, *n* (%)	269 (43.5)	181 (41.8)	16 (64.0)
Incidence ≥1.0 mg/dL, *n* (%)	350 (56.5)	252 (58.2)	9 (36.0)
Body temperature, mean (SD), °C	37.0 (0.8), $N = 262$[b]	NR, $N = 202$[b]	NR, $N = 7$[b]
Incidence <37.0 °C, *n* (%)	134 (51.1)	107 (53.0)	6 (85.7)
Incidence ≥37.0 °C to <38.0 °C, *n* (%)	82 (31.3)	60 (30.0)	0
Incidence ≥38.0 °C, *n* (%)	46 (17.6)	35 (17.4)	1 (14.3)
WBC count, mean (SD), /μL	9716.4 (4484.9), $N = 635$[b]	NR, $N = 455$[b]	NR, $N = 22$[b]
Incidence <10,000/μL, *n* (%)	363 (57.2)	262 (57.6)	15 (68.2)
Incidence ≥10,000/μL, *n* (%)	272 (42.8)	193 (42.4)	7 (31.8)
Patients with results of all three laboratory tests, *n*	186	140	6
Normal results of laboratory tests (CRP <1.0 mg/dL, body temperature <37.0 °C, and WBC count <10,000/μL), *n*	68	52	4

NR: not reported; SD: standard deviation; SI: serious infection; WBC: white blood cell. [a]*N* number of patients. [b]*N* number of patients with measurements, excluding missing data. All data are number of measurements (%) in the category unless stated otherwise. Not included in the analysis were 272 missing measurements for CRP, 555 missing measurements for temperature, and 262 missing measurements for WBC.

To our knowledge, this is the first published report of text mining data from clinical narratives of a post-marketing AE-reporting database. Previous reports have used clinical notes and narratives, or other electronic health care data, to facilitate the early identification of drug–drug interaction signals or to determine AE co-occurrence [16,17]. For example, in patients with RA, annotation analysis of unstructured clinical notes and subsequent mining of the resultant annotations has been successfully used to calculate the risk for myocardial infarction in those treated with rofecoxib [14]. This highlights the applicability of data mining approaches to evaluate potentially SAEs associated with drugs already on the market.

This study includes several limitations. First, the potential effect of underreporting AEs, which is an inherent limitation of post-marketing AE-reporting databases [25], resulted in some misclassification of information for AEs and previous or concomitant use of drugs. In addition, the infection course and symptoms might not have been adequately reported (e.g. the treating physician might have considered symptoms to be related to infection but might not have reported them in the narrative if they were not severe; therefore, absence of a symptom in the report does not necessarily preclude its occurrence). This lack of adequate reporting has the potential to introduce a large bias. However, it might be expected that there would be less underreporting of SAEs than of non-SAEs because this analysis focused on SAEs, and physician motivation on AE reporting is expected to be higher for SAEs [26]. The fact that manufacturers carefully queried reporting physicians during follow-up for SAEs could also have helped decrease the effects of underreporting. Second, the lack of comparison with patients who developed infections that did not become serious means that the signs and symptoms identified are indicative only of the development of serious infection, and a causal relationship cannot be claimed. Third, there is a need for technical improvements in text mining approaches. For example, because we considered only fragmented words rather word combinations, it was difficult to distinguish several pain-related words (e.g. abdominal pain, back pain, chest pain, pain in throat), and these words were considered as pain. As a result, frequently appearing words to describe symptoms such as pain tended to be easily identified in our analysis. Considering the use of additional text mining techniques, such as named entity recognition, may improve the quality of the results. Finally, in some cases (e.g. reports from the literature), crucial information such as the patient's medical history was not available unless the clinical notes were described in detail. Thus, insights obtained from this study should be interpreted as new hypotheses rather than confirmed evidence; validation studies are necessary to confirm our findings. Although AE-reporting data do have several limitations, the advantage of a relatively quick analysis and the opportunity for a safety database text mining approach represent an important opportunity for obtaining further safety insights and generating new hypotheses.

In clinical practice, patients are likely to experience multiple signs and symptoms of infection simultaneously; therefore, the time to onset of serious infection and the sequence of signs and symptoms may aid in early detection of infections before they become serious. This study successfully identified signs and symptoms of serious infection by site in patients with RA who developed infections that became serious after the administration of TCZ. The findings highlight that retrospective analyses of post-marketing AE databases using text mining of clinical narratives may be beneficial for characterizing serious infections associated with the treatment of RA.

Acknowledgements

The authors thank all the investigators and the study team at Chugai Pharmaceutical Co., Ltd., especially Yukiko Hayashi and Yoshiyuki Moriguchi, for their contributions. Professional writing and editorial assistance was provided by Liselle Bovell, PhD, and Sara Duggan, PhD, of ApotheCom on behalf of F. Hoffmann La-Roche Ltd.

Conflict of interest

T.A. has received personal fees from Chugai Pharmaceutical Co. Ltd., during the conduct of the study as well as grants, personal fees, and other from Astellas, Mitsubishi Tanabe, Chugai Pharmaceutical Co. Ltd., and Eisai; personal fees and other from Takeda; personal fees from Pfizer, AbbVie, and UCB Japan; grants from Sanofi, Alexion, Bristol-Myers, and Janssen; and other from Daiichi Sankyo and Bayer outside the submitted work. Y.A., S.M., S.T., R.T., N.T., and A.N. report personal fees from Chugai Pharmaceutical Co. Ltd., during the conduct of the study.

Funding

This study was funded by Chugai Pharmaceutical Co., Ltd.

References

1. Smolen JS, Landewe R, Breedveld FC, Buch M, Burmester G, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. Ann Rheum Dis. 2014;73:492–509.
2. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Care Res. 2016;68:1–26.
3. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Ann Rheum Dis. 2016;75:3–15.
4. Nishimoto N, Miyasaka N, Yamamoto K, Kawai S, Takeuchi T, Azuma J. Long-term safety and efficacy of tocilizumab, an anti-interleukin-6 receptor monoclonal antibody, in monotherapy, in patients with rheumatoid arthritis (the STREAM study): evidence of safety and efficacy in a 5-year extension study. Ann Rheum Dis. 2008;68:1580–4.
5. Genovese MC, Rubbert-Roth A, Smolen JS, Kremer J, Khraishi M, Gómez-Reino J, et al. Longterm safety and efficacy of tocilizumab in patients with rheumatoid arthritis: a cumulative analysis of up to 4.6 years of exposure. J Rheumatol. 2013;40:768–80.
6. Kivitz A, Olech E, Borofsky M, Zazueta BM, Navarro-Sarabia F, Radominski SC, et al. Subcutaneous tocilizumab versus placebo in combination with disease modifying antirheumatic drugs in patients with rheumatoid arthritis. Arthritis Care Res. 2014;66:1653–61.

7. Koike T, Harigai M, Inokuma S, Ishiguro N, Ryu J, Takeuchi T, et al. Effectiveness and safety of tocilizumab: postmarketing surveillance of 7901 patients with rheumatoid arthritis in Japan. J Rheumatol. 2014;41:15–23.
8. Mihara M, Kasutani K, Okazaki M, Nakamura A, Kawai S, Sugimoto M, et al. Tocilizumab inhibits signal transduction mediated by both mIL-6R and sIL-6R, but not by the receptors of other members of IL-6 cytokine family. Int Immunopharmacol. 2005;5:1731–40.
9. Nishimoto N, Terao K, Mima T, Nakahara H, Takagi N, Kakehi T. Mechanisms and pathologic significances in increase in serum interleukin-6 (IL-6) and soluble IL-6 receptor after administration of an anti-IL-6 receptor antibody, tocilizumab, in patients with rheumatoid arthritis and Castleman disease. Blood. 2008;112:3959–64.
10. Harpaz R, Dumouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther. 2012;91:1010–21.
11. Nomura K, Takahashi K, Hinomura Y, Kawaguchi G, Matsushita Y, Marui H, et al. Effect of database profile variation on drug safety assessment: an analysis of spontaneous adverse event reports of Japanese cases. Drug Des Devel Ther. 2015;9:3031–41.
12. Dore DD, Seeger JD, Arnold CK. Use of a claims-based active drug safety surveillance system to assess the risk of acute pancreatitis with exenatide or sitagliptin compared to metformin or glyburide. Curr Med Res Opin. 2009;25:1019–27.
13. Yamamoto K, Goto H, Hirao K, Nakajima A, Origasa H, Tanaka K, et al. Longterm safety of tocilizumab: results from 3 years of followup postmarketing surveillance of 5573 patients with rheumatoid arthritis in Japan. J Rheumatol. 2015;42:1368–75.
14. LePendu P, Iyer SV, Fairon C, Shah NH. Annotation analysis for testing drug safety signals using unstructured clinical notes. J Biomed Semantics. 2012;3(Suppl. 1):S5.
15. Harpaz R, Vilar S, Dumouchel W, Salmasian H, Haerian K, Shah NH, et al. Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. J Am Med Inform Assoc. 2013;20:413–19.
16. Iyer SV, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. J Am Med Inform Assoc. 2014;21:353–62.
17. Roitmann E, Eriksson R, Brunak S. Patient stratification and identification of adverse event correlations in the space of 1190 drug related adverse events. Front Physiol. 2014;5:332.
18. Imai T, Aramaki E, Kajino M, Miyo K, Onogi Y, Ohe K. Finding malignant findings from radiological reports using medical attributes and syntactic information. Stud Health Technol Inform. 2007;129:540–4.
19. Gonzalez GH, Tahsin T, Goodale BC, Greene AC, Greene CS. Recent advances and emerging applications in text and data mining for biomedical discovery. Brief Bioinform. 2016;17:33–42.
20. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of adverse drug events from free text electronic patient records and information in a large mental health case register. PLoS One. 2015;10:e0134208.
21. Chin CY, Weng MY, Lin TC, Cheng SY, Yang YH, Tseng VS. Mining disease risk patterns from nationwide clinical databases for the assessment of early rheumatoid arthritis risk. PLoS One. 2015;10:e0122508.
22. Bruland P, McGilchrist M, Zapletal E, Acosta D, Proeve J, Askin S, et al. Common data elements for secondary use of electronic health record data for clinical trial execution and serious adverse event reporting. BMC Med Res Methodol. 2016;16:159.
23. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf. 1999;20:109–17.
24. Porter RS, Kaplan JL, eds. The Merck manual of diagnosis and therapy. 19th ed. Kenilworth (NJ): Merck Publishing; 2011.
25. Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. Drug Saf. 2006;29:385–96.
26. Matsuda S, Aoki K, Kawamata T, Kimotsuki T, Kobayashi T, Kuriki H, et al. Bias in spontaneous reporting of adverse drug reactions in Japan. PLoS One. 2015;10:e0126413.