Validation of Claims-based Definitions for Identifying Rheumatoid Arthritis: A Hospital-based Cross-sectional Study Conducted in 64 Hospitals in Japan

Kiyoshi Kubota (kubota@dsrujp.org)
NPO Drug Safety Research Unit Japan
Masaki Yoshizawa
Department of Rheumatology, Shonan Kamakura General Hospital
Satoru Takahashi
Tokushukai General Incorporated Association Tokyo Headquarters
Yoshiaki Fujimura
Tokushukai Information System, Inc.
Hiroko Nomura
Tokushukai General Incorporated Association Osaka Headquarters, Osaka-shi, Osaka
Hitoshi Kohsaka
Rheumatology Center, Chiba-Nishi General Hospital, Matsudo, Chiba

Research Article

Keywords: Rheumatoid arthritis, validation study, sensitivity, specificity, positive predictive value, negative predictive value, bias analysis

Posted Date: December 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-128099/v1

Version of Record: A version of this preprint was published at BMC Musculoskeletal Disorders on April 22nd, 2021. See the published version at https://doi.org/10.1186/s12891-021-04259-9.

Abstract

Background

An administrative database covering a whole population such as the national database in Japan may be used to estimate the nationwide prevalence of diseases including rheumatoid arthritis (RA) when a well-validated definition of the disease is available. In Japan, the record-linkage between the administrative database and medical charts in hospitals is strictly prohibited. A “hospital-based” validation study is one of few possible validation studies where claims kept inside the study hospital are rearranged into the database structure.

Methods

Random samples of 19,734 patients were selected from about 1.6 million patients who received medical care between February 2018 and January 2019 in one of the 64 hospitals of the Tokushukai Medical Group. We excluded patients whose observation period was less than 365 days and identified 334 patients who met the definition of “possible cases of RA” whose medical chart was then evaluated by two rheumatologists independently. In a sensitivity analysis, we evaluated bias due to misclassifying some patients with RA who did not meet the definition of “possible cases of RA” as a patient with no RA.

Results

The kappa coefficient between two rheumatologists was 0.80. The prevalence of RA in the study population was estimated as 0.56%. We found [condition code of RA] and ([any disease-modifying antirheumatic drug] or [oral corticosteroid with no systemic autoimmune diseases (other than RA) and no polymyalgia rheumatica]) had a relatively high sensitivity (around 73%) and a high positive predictive value (around 80%). In a sensitivity analysis, we found that when some patients with RA who did not meet the definition of “possible cases of RA” were misclassified as a patient with no RA, then this would lead to underestimation of the prevalence of the definition-positive patients and the adjusted prevalence.

Conclusions

We recommend using the claims-based definition of RA (found in the current validation study in the 64 hospitals) to estimate the prevalence of RA in Japan. We also recommend estimating the adjusted prevalence using the quantitative bias analysis method because the prevalence of the disease in the “hospital-based” validation study is different from that in the administrative database.

Trial registration: The current study is not a clinical trial and thus is not subject to trial registration.

Background

Administrative databases have been used to estimate the prevalence of several diseases including rheumatoid arthritis (RA) [1–4]. To estimate the disease prevalence using the administrative database,

the database should cover the whole population, and a well-validated definition of the disease of interest (which can accurately identify the patients with the disease) should be used.

In Japan, the prevalence of RA was estimated to be 1.7% in a study conducted in Wakayama Prefecture in 1996 [5]. In a recent study using data from the Comprehensive Survey of Living Conditions, the prevalence of RA was estimated as 0.75% [6]. In a study using the claims database covering 1 million subjects, the prevalence of RA was estimated as 0.6 to 1.0% in the Japanese population aged ≥ 16 to < 75 years [7]. Nakajima et al. used data between April 2017 and March 2018 from the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB Japan) and reported that the prevalence of patients with RA was between 0.46% and 0.88% when seven different definitions were used [8]. They recommended “Definition 3” which was “patients ≥ 16 years old with 1 International Classification of Diseases, 10th revision (ICD-10) code of RA, and prescribed any disease-modifying antirheumatic drugs (DMARDs) for at least 2 of 12 months”. However, as acknowledged by the authors, seven definitions of RA used in the study have not been validated and “Definition 3” excludes patients with RA treated by an oral corticosteroid only but not by a DMARD.

Database studies are relatively new to clinical studies in Japan and there have been only a few validation studies of definitions for database studies to identify patients with breast cancer [9], acute myocardial infarction [10], and other conditions [11]. In North America and Europe, the validation study has been often conducted by the chart review of patients selected from the administrative database [12–15]. However, the record-linkage is strictly prohibited in studies using the administrative database and therefore, subjects selected from the administrative database cannot be linked to the medical charts in hospitals in Japan.

A “hospital-based” validation study is one of few possible options where the claims kept inside the study hospital are rearranged into the database structure, and where claims-based definitions are evaluated by the chart review of patients in the hospital [10]. The representativeness of the “hospital-based” validation study is therefore questionable because the population in the validation study is, in general, different from the population in a future study where the validated definition is used. To improve the representativeness of the validation study, the study may be conducted in a variety of hospitals so that the population in the validation study is more representative of the population covered by the administrative database.

For chronic conditions like RA, another problem exists related to the Japanese health care system in that patients can select the hospital or clinic according to their own preference in general and one patient often receives medical care for two or more conditions in different hospitals (or clinics) during the same calendar period. This is because most hospitals see new patients without a reference letter though some hospitals charge an additional fee for a new patient without a reference letter. It is only a few hospitals that do not see any patients without a reference letter [16]. If a patient is currently receiving care for the disease of interest (e.g., RA) in a different hospital, the records of some medical care (e.g., drug treatment) for such patients may be available in claims of the different hospital but not in claims of the

study hospital. Therefore, such patients may be excluded from the study population. Similarly, some patients may visit multiple hospitals and have the medical care in the study hospital just once or for a short period only. Those patients may be also removed from the study population because the claims-based definition for a chronic condition often requires the information collected during a specific length of time (e.g., “three condition codes over 2 years”) [15, 17, 18].

We conducted a “hospital-based” validation study in the 64 hospitals located in various areas of Japan and evaluated claims-based definitions to identify patients with RA while addressing problems inherent to the “hospital-based” validation study of a chronic condition in the Japanese health care system.

Methods

We used electronically available claims and clinical data from the 64 hospitals of the Tokushukai Medical Group routinely collected by the Tokushukai Information System (TIS) Inc. The study was conducted in accordance with the Declaration of Helsinki and approved by the Tokushukai Group Ethics Committee [19] where obtaining the informed consent from study subjects was waived for the current study but the Committee indicated that the conduct of the study be announced through the internet [20]. Currently, Japanese hospitals may be classified as those following the diagnosis procedure combination/per-diem payment system (DPC/PDPS, simply abbreviated as DPC) [21] and non-DPC hospitals. The 64 study hospitals of the Tokushukai Medical Group are located in 23 of the 47 prefectures in Japan; they include 47 DPC hospitals and 17 non-DPC hospitals (see Table S1 in Additional File 1). The electronically available clinical data include records of electronic medical charts, drugs used in inpatient and outpatient care, laboratory data, and medical procedures such as surgical operation and rehabilitation, and the reference letter (in PDF format). However, X-ray radiographs or other images (e.g., computed tomography or magnetic resonance imaging) are not routinely collected form the 64 study hospitals nor readily accessible through the network. During 1 year of the study period between February 1, 2018 and January 31, 2019, 1,590,669 patients used outpatient care (1,575,464 patients) or inpatient care (222,131 patients) in one of the 64 hospitals.

A total of 13,224 of 1,590,669 patients had a condition code of RA in at least one of 12 monthly claims issued during the study period. We selected two mutually exclusive sets (Set A and Set B) of random samples of 19,734 patients each (about 1.2% of 1,590,669 patients) so that set A and set B included about 160 patients (about 1.2% of 13,224 patients) with a condition code of RA in at least in 1 monthly claim. We expected that from random samples including 160 patients with a condition code of RA, at least 100 cases of ‘definite’ RA would be identified by the chart review allowing the estimation of the sensitivity of RA within ± 0.1 [22]. Of the two sets of 19,734 patients each, one set (Set A) was used in a pilot study conducted prior to the main study. In a pilot study, the electronic medical chart of 20 patients who met the definition of “possible cases of RA” (Table 1) in some of the 64 hospitals were reviewed by one rheumatologist (MY) through the network with the help of a technical engineer of the TIS.

The main purposes of the pilot study were to confirm the practicability of the chart review through the network for patients in hospitals located in various areas in Japan and to know whether a PDF survey form used in the study could effectively collect the relevant information. Using a PDF survey form, age and sex of a possible RA patient, information needed to classify patients according to the criteria by the American College of Rheumatology (ACR) Board of Directors and the European League Against Rheumatism (EULAR) in 2010 (ACR/EULAR classification criteria) [23], and the final judgement made by the rheumatologist about whether the patient had RA were recorded. Information on whether the patient was receiving care for RA in a different hospital was also collected by the survey form, provided that the patient was judged to have RA.

After we confirmed that the chart review through the network was practical and the survey form was able to collect the information effectively in the pilot study, we started the main study using 19,734 random samples of Set B. We first excluded 6,712 patients whose observation period was less than 365 days out of 19,734 patients in Set B. As the research fund was limited, we did not review the medical charts of all of the remaining 13,022 patients whose observation period was 365 days or longer but rather selected 334 patients who met the definition of “possible cases of RA” (Table 1) to classify 13,022 patients as patients having RA or patients having no RA. We assumed that the remaining 12,688 patients who did not meet the definition of “possible cases of RA” had no RA. As the medical chart of 1 of the 334 possible RA patients was unavailable, the chart review was conducted for 333 possible RA patients by two rheumatologists (MY and HK) independently through the network. The agreement of the judgement by two rheumatologists was assessed by Cohen’s kappa coefficient. Finally, two rheumatologists resolved a disagreement where the electronic medical chart could be reviewed through the network when necessary. The disagreement was resolved through discussion to obtain a final judgement on whether the patient had RA, ACR/EULAR classification score, and whether the care for RA was given in a different hospital other than the study hospital.

Although two rheumatologists classified 333 patients into three categories—(1) having RA, (2) having no RA, or (3) RA suspected—they were informed that a “RA suspected” patient would be re-classified as a patient having no RA in the final analysis and thus they requested that patients be classified into either (1) having RA or (2) having no RA whenever possible. We evaluated 26 claims-based definitions specified by combining some of 3 inclusion criteria and 1 exclusion criterion in Table 2 using the reference standard where patients were classified as having RA or not having RA according to the final decision agreed between two rheumatologists.

For each of the 26 claims-based definition of RA, patients were classified into true positive (TP), false negative (FN), false positive (FP), and true negative (TN) cases, and four key measures of diagnostic accuracy were estimated. Of the four key measures, the sensitivity (SE) was estimated as SE = TP/(TP + FN), the specificity (SP) was estimated as SP = TN/(TN + FP), the positive predictive value (PPV) was estimated as PPV = TP/(TP + FP), and the negative predictive value (NPV) was estimated as NPV = TN/(TN + FN). The total number of TN was estimated as the number of TN cases in 333 patients for

whom the medical chart was reviewed plus 12,688 patients who were not selected as “possible cases of RA”.

In the primary analysis we excluded 1 patient whose medical chart was unavailable as well as 39 patients receiving care for RA in a different hospital from the study population. Therefore, the size of the study population was corrected to 12,982 (Fig. 1). In a sensitivity analysis, we included 39 patients receiving care for RA in a different hospital to the study population. In the sensitivity analysis, including the 39 patients with RA, the size of the study population was 13,021 after excluding 1 patient whose medical chart was unavailable.

We also conducted another sensitivity analysis to estimate bias due to misclassification of some patients with RA as a “patient with no RA” because they did not meet the definition of “possible cases of RA” (Table 1). The sensitivity analysis was conducted in conjunction with the quantitative bias analysis [24] assuming that the best claims-based definition found in the current validation study would be used in the future study. It was assumed that the prevalence of RA in the future study would be different from the prevalence in the current validation study, but the sensitivity and specificity were the same as those in the validation study.

To calculate the adjusted number of patients with RA in the future study, we used the equation B1=(B1*-(1-SP*)N)/(SE*+SP*-1), where B1 was the adjusted number of patients with RA, B1* was the number of the definition-positive patients, N was the population size in the future study where the validated definition would be used, and SE* and SP* were the sensitivity and specificity estimated in the current validation study, respectively [25]. If all patients with RA were included in 334 “possible cases of RA” in the current validation study, B1 should be the true number of patients with RA in the future study provided that the sensitivity and specificity were the same between the validation and future studies, but if some patients with RA were not included in “possible cases of RA”, B1 will be biased.

We assumed two scenarios for the future study where the true prevalence of RA was 0.56% and 1.0%, respectively. We examined the effect of the misclassification by comparing the estimated prevalence of the definition-positive patients (B1*/N) and the adjusted prevalence (B1/N) with the true prevalence in the future study (see Appendix 3 of Additional File 1 for the details).

Results

Table 3 shows the number and proportion of patients stratified by sex and age (5-year interval) in the original population of 1,590,669 patients subdivided by Non-DPC hospitals, DPC hospitals in East Japan, and DPC hospitals in West Japan. Age-sex distribution was roughly the same between those three groups of hospitals.

In 19,734 random samples of Set A and Set B, 164 and 169 patients, respectively, were found to have a condition code of RA at least in 1 monthly claim. After excluding 6,712 patients whose observation period was less than 365 days in the main study using Set B, 133 patients had a condition code of RA in the

remaining 13,022 patients. Of 133 patients, 36 patients had a condition code of RA in an inpatient claim, but 97 patients had a condition code in one or more outpatient claims only. Including those 133 patients, a total of 334 patients were selected as “possible RA patients” where 143 patients were selected because they met only component 5 or 6 in Table 1 (i.e., they had the RA-related phrase in free-text in the medical chart but did not meet any other components).

After excluding 1 patient whose medical chart was unavailable from 334 “possible cases of RA”, 333 patients were evaluated by two rheumatologists independently, and Cohen’s kappa coefficient used to assess the agreement of the judgement of having RA was 0.78 when patients were classified into three categories. One rheumatologist classified 11 patients as “RA suspected” while another rheumatologist classified 22 patients as “RA suspected”. When “RA suspected” cases were reclassified as patients having no RA, Cohen’s kappa coefficient was 0.80. After the discrepancy between the two rheumatologists was resolved by discussion, 333 possible RA patients were classified into 112 having RA, 216 having no RA, and 5 having “RA suspected”. In the final analysis, 5 “RA suspected” patients were reclassified as having no RA. Although a total of 112 patients were judged to have RA, 39 received care for RA in a different hospital and thus were removed from the study population in the primary analysis. Among the remaining 73 patients with RA, the ACR/EULAR score was estimated as ≥ 6 according to the final agreement for 26 patients (36%), while in many of the remaining 47 patients with RA, the available information was insufficient to estimate the ACR/EULAR score. The prevalence of 73 patients with RA was 0.56% in the final study population of 12,982 subjects. The mean age [SD] was 73.1 [11.9] years in 62 female patients with RA, and 76.6 [9.3] years old in 11 male patients with RA. In the sensitivity analysis, 39 patients receiving RA care in a different hospital were included in the study population and the prevalence was 0.85% (112/13,021) (Table S2, Additional File 1).

Table 4 shows the number of patients of TP, FN, FP, and TN cases as well as the estimated SE, SP, PPV, and NPV. Table 4 also shows the prevalence of definition-positive subjects. The definitions of [condition code of RA in 1 or 2 monthly claims] and [any DMARDs in 1 or 2 monthly claims] during the study period of 1 year (Definitions 3–6) had a high PPV (> 85%) but a low SE (< 60%). The definitions of [condition code of RA] and [oral corticosteroid] (Definitions 7–10) had a low PPV (< 60%) and a low SE (< 40%). The definition of [condition code of RA] and ([DMARD] or [oral corticosteroid with no systemic autoimmune diseases (other than RA) and no polymyalgia rheumatica]) (Definitions 19–26) had a relatively high SE (around 70%) and a high PPV (around 80%).

For Definition 21 (or 22) where SE = 73% and PPV = 80.3%, the mean age [SD] for 53 TP, 20 FN, and 13 FP cases was 73.5 [11.5], 74.1 [12.1], and 69.8 [18.5] years old, respectively. In 10 of 20 FN cases, identified as “possible cases of RA” because they met component 5 or 6 only (Table 1), did not have a condition code of RA in any monthly claim. In 13 FP cases, in addition to a condition code of RA, 4 had a condition code of systemic autoimmune diseases (3 had suspected systemic lupus erythematodes and 1 had polymyositis). Those 4 FP cases were definition-positive (for Definition 21) because they had a condition code of RA and a DMARD even if they had a condition code of systemic autoimmune diseases.

Table S2 (Additional File 1) shows those estimates in the sensitivity study where 39 patients receiving the care for RA in a different hospital are included in the study population. In Table S2, SE was 8.2 to 20.2% lower, PPV was 0.7 to 5.5% higher, and the prevalence of the definition-positive subjects was 0.01 to 0.12% higher than when those 39 were excluded (Table 4).

Figure 2 shows the results of another sensitivity analysis to estimate the effect of misclassification of patients having RA as a patient having no RA because they did not meet the definition of “possible cases of RA” (Table 1). We estimated the effect in the future study using the validated definition under 2 scenarios where the true prevalence of RA in the future study were 0.56% (as in the current validation study) or 1.0%. We assumed that in the validation study, only a fraction (F) of patients with RA were included in 333 “possible cases of RA” (after excluding 1 patient whose medical chart was unavailable from the study population) and the remaining patients with RA (of which the proportion was 1-F) were misclassified as a patient with no RA. Definition 21 in Table 4 was used to estimate the prevalence of the definition-positive subjects and the prevalence adjusted by SE* and SP* which were the sensitivity and specificity in the validation study, respectively (see Appendix 3 in Additional File 1 for the details).

When F = 1 (i.e., all patients with RA were included in “possible cases of RA”), the prevalence of the definition-positive subjects (P*) was 0.51% and the adjusted prevalence (P_{adj}) was 0.56% when the true prevalence was 0.56%. Similarly, when F = 1, the definition-positive subjects (P*) was 0.83% while the adjusted prevalence (P_{adj}) was 1.0% when the true prevalence was 1.0%. When F < 1, both P* and P_{adj} were underestimated. When F < 1, the estimated values of P* and P_{adj} were about F times the corresponding values when F = 1.

Discussion

In the current validation study of claims-based definitions for identifying RA, we randomly selected 19,734 patients from 1,590,669 patients who had outpatient or inpatient care in one of 64 hospitals located in 23 of the 47 prefectures in Japan. After excluding 6,712 patients who were observed for less than 365 days and 39 patients receiving the care for RA in a different hospital, 73 patients had RA and the prevalence of RA was estimated as 0.56% in this population. Though the main objective of the current study is to evaluate the validity of claims-based definition of RA rather than to estimate the prevalence itself, this prevalence (0.56%) may be viewed to be similar to that in the previous studies [5–8]. In the current validation study, we estimated four key measures of diagnostic accuracy (SE, SP, PPV, and NPV) while according to a systematic review of validation studies to identify rheumatic diseases published in 2013, all of those four key measures were estimated only in 4 of 23 studies [26]. In this systematic review, the values of those measures varied according to different sampling population sources, sources of data for case definition, and reference standard definitions [26].

Of 26 claims-based definitions estimated in the current validation study, we found that the definition of [condition code of RA in 1 or 2 monthly claims] and [any DMARDs in 1 or 2 monthly claims] during the study period of 1 year (Definitions 3–6) had a high PPV (> 85%) but a low SE (< 60%). These definitions

may be useful when patients without RA should be excluded as much as possible as in the study where the effectiveness or safety of a drug and other intervention is evaluated by comparing the incidence of an outcome between the exposed and unexposed patients (or between those who had Drug A and Drug B). However, for the study to estimate the prevalence of RA, Definition 21 (or 22) is recommended as it has a relatively high sensitivity (around 73%) and a high PPV (around 80%). We also recommend conducting the additional quantitative bias analysis to estimate the adjusted prevalence because bias due to the difference of the population in the "hospital-based" validation study and the population in the future study can be mitigated to some extent.

In the primary analysis, we excluded 39 patients with RA receiving the care for RA in a different hospital. In the Japanese health care system, the precise definition of the population covered by one hospital is difficult [16]. As in the definition of "secondary bases" in a case-control study [27], the study population in the current validation study would be defined as "all people who would have the care for RA and observed for 365 days or longer in the study hospital if they had RA". We excluded 6,712 patients whose observation period was less than 365 days as well as 39 patients who had RA care in a different hospital as they were not thought to be included in the study population. Indeed, 39 patients who had RA care in a different hospital differed considerably from the 73 patients with RA who received care in the study hospital. For example, in 73 patients with RA treated in the study hospital, 63 (86.3%) had a condition code of RA in a claim and 43 (58.9%) patients had a DMARD, while in 39 patients receiving RA care in a different hospital, only 17 (43.6%) had a condition code of RA and only 4 (10.2%) had a DMARD. When those 39 patients with RA are included, the sensitivity and PPV are also considerably different from those in the primary analysis (Table S2 in Additional File 1). We believe that excluding those 39 patients was a proper strategy to evaluate the valid claims-based definitions.

Figure 2 shows the results of another sensitivity analysis assuming some of patients with RA did not meet the definition of "possible cases of RA". However, it is unlikely that many patients with RA receiving care for RA in the study population did not meet any component of "possible cases of RA" (Table 1), including RA-related phrases in free text of the medical chart. The electronic medical chart was used in all of the 64 hospitals in the Tokushukai Medical Group and the search for RA-related phrases was almost perfect (except for the phrases handwritten in the reference letter in PDF format).

A strength of the current study was that we used data from 64 hospitals located in various areas in Japan and thus the results are likely more representative of the general population compared to studies covering only one or a few hospitals [6–8]. As shown in Table 3, the age—sex distribution was similar between small non-DPC hospitals and DPC hospitals located in East and West Japan. The boundary between primary care and second/tertiary care is indistinct in the Japanese health care system [16] and both small and large hospitals maintain large outpatient departments and provide outpatient care to nearby residents. For example, 86% out of a total of 1,590,669 patients had outpatient care only during 1 year of the study period. The similarity of the age—sex distribution between small and large hospitals in East and West Japan reflects the fact that outpatient care in Japanese hospitals is usually open to all nearby residents although this does not necessarily mean that the population in the current validation

study is representative of the people in the nation. Another strength of the current study was that two rheumatologists were able to review the original electronic medical charts in the various hospitals through the network.

There were, however, several limitations in the current study. As record-linkage access in the study using the administrative database is strictly prohibited, the chart review could not be done for random samples directly selected from the administrative database. Thus, the population in the validation study is likely to differ from those in the future study where the validated definition is used. However, the adjusted prevalence estimated by the method of the quantitative bias analysis will reduce, to some extent, bias due to the difference of the population in the validation study compared to the future study, provided that the sensitivity and specificity are the same between the two studies. Another limitation was that as many as 39 patients receiving RA care in a different hospital were excluded in the main analysis leading to reduced statistical power. Lastly, X-ray radiographs were not available through the network when two rheumatologists reviewed the medical charts in the current study, which could reduce the accuracy of the judgement on whether the patient had RA.

Conclusion

We conducted a validation study of the claims-based definition of RA in about 1.6 million patients who had inpatient or outpatient care in 64 hospitals located in various areas of Japan and found a suitable claims-based definition which may be used to estimate the prevalence of RA in future studies. We recommend that (in the future study to estimate the prevalence of RA) the best claims-based definition found in the current validation study (Definition 21 or 22) is used. We also recommend estimating the adjusted prevalence using the quantitative bias analysis method because the population in the “hospital-based” validation study is different from that in future studies.

Abbreviations

ACR/EULAR:American College of Rheumatology Board of Directors and the European League Against Rheumatism

anti-CCP: anti-cyclic citrullinated peptide.

DMARD:disease-modifying antirheumatic drug

DPC: Diagnosis Procedure Combination/Per-Diem Payment System

ICD-10: International Classification of Diseases, 10th revision

NDB Japan:the National Database of Health Insurance Claims and Specific Health Checkups of Japan

PMR: polymyalgia rheumatica

RA: Rheumatoid arthritis

TIS: Tokushukai Information System

FN: false negative

FP: false positive

TN: true negative

TP: true positive

SE: sensitivity

SP: specificity

PPV: positive predictive value

NPV: negative predictive value

Declarations

Ethics approval

The study was approved by the Tokushukai Group Ethics Committee where obtaining the consent from study subjects was waived for the current study but the Committee indicated that the conduct of the study be announced through the internet

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

Not applicable

Funding

This work was supported by a grant from AbbVie GK, which has not been involved in any processes of the study including the protocol development, data analysis, interpretation, and manuscript submission.

Authors’ contribution

KK: Study design, data analysis, manuscript preparation. MY and HK: Evaluation of survey form, Review of medical charts, Interpreting the data. ST: Overseeing the entire study, YF, and HN: Data acquisition through the network. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Ms. Ayako Yano of EP-CRUSU Co., Ltd for helping the current study as a secretariat of the study office.

Authors’ information

NPO Drug Safety Research Unit Japan, Chiyoda-ku, Tokyo, Japan

Kiyoshi Kubota

Department of Rheumatology, Shonan Kamakura General Hospital, Kamakura, Kanagawa, Japan

Masaki Yoshizawa

Tokushukai General Incorporated Association Tokyo Headquarters, Chiyoda-ku, Tokyo, Japan

Satoru Takahashi

Tokushukai Information System, Inc., Osaka-shi, Osaka, Japan

Yoshiaki Fujimura

Tokushukai General Incorporated Association Osaka Headquarters, Osaka-shi, Osaka, Japan

Hiroko Nomura

Rheumatology Center, Chiba-Nishi General Hospital, Matsudo, Chiba, Japan

Hitoshi Kohsaka

References

1. Kao KT, Islam N, Fox DA, Amed S. Incidence Trends of Diabetic Ketoacidosis in Children and Adolescents with Type 1 Diabetes in British Columbia, Canada. J Pediatr 2020; 221:165-173.e2. doi: 10.1016/j.jpeds.2020.02.069.

2. Amankwah N, Marrie RA, Bancej C, Garner R, Manuel DG, Wall R, et al. Multiple sclerosis in Canada 2011 to 2031: results of a microsimulation modelling study of epidemiological and economic impacts. Health Promot Chronic Dis Prev Can 2017; 37(2):37-48. doi: 10.24095/hpcdp.37.2.02.
3. Kubota K, Kamijima Y, Sato T, Ooba N, Koide D, Iizuka H, et al. Epidemiology of psoriasis and palmoplantar pustulosis: a nationwide study using the Japanese national claims database. BMJ Open 2015;5(1):e006450. doi: 10.1136/bmjopen-2014-006450.
4. Widdifield J, Paterson JM, Bernatsky S, Tu K, Tomlinson G, Kuriya B, et al. The epidemiology of rheumatoid arthritis in Ontario, Canada. Arthritis Rheumatol 2014;66(4):786-93. doi: 10.1002/art.38306.
5. Shichikawa K, Inoue K, Hirota S, Maeda A, Ota H, Kimura M, et al. Changes in the incidence and prevalence of rheumatoid arthritis in Kamitonda, Wakayama, Japan, 1965-1996. Ann Rheum Dis 1999;58(12):751-6. doi: 10.1136/ard.58.12.751.
6. Kojima M, Nakayama T, Tsutani K, Igarashi A, Kojima T, Suzuki S, et al. Epidemiological characteristics of rheumatoid arthritis in Japan: Prevalence estimates using a nationwide population-based questionnaire survey. Mod Rheumatol 2020;30(6):941-947. doi: 10.1080/14397595.2019.1682776.
7. Yamanaka H, Sugiyama N, Inoue E, Taniguchi A, Momohara S. Estimates of the prevalence of and current treatment practices for rheumatoid arthritis in Japan using reimbursement data from health insurance societies and the IORRA cohort (I) . Mod Rheumatol 2014;24(1):33-40. doi: 10.3109/14397595.2013.854059.
8. Nakajima A, Sakai R, Inoue E, Harigai M. Prevalence of patients with rheumatoid arthritis and age-stratified trends in clinical characteristics and treatment, based on the National Database of Health Insurance Claims and Specific Health Checkups of Japan. Int J Rheum Dis. 2020; Oct 5. doi: 10.1111/1756-185X.13974.
9. Sato I, Yagata H, Ohashi Y. The accuracy of Japanese claims data in identifying breast cancer cases. Biol Pharm Bull. 2015;38(1):53-7. doi: 10.1248/bpb.b14-00543.
10. Ando T, Ooba N, Mochizuki M, Koide D, Kimura K, Lee SL, et al. Positive predictive value of ICD-10 codes for acute myocardial infarction in Japan: a validation study at a single center. BMC Health Serv Res 2018;18(1):895. doi: 10.1186/s12913-018-3727-0.
11. Yamana H, Moriwaki M, Horiguchi H, Kodan M, Fushimi K, Yasunaga H. Validity of diagnoses, procedures, and laboratory data in Japanese administrative data. J Epidemiol 2017;27(10):476-482. doi: 10.1016/j.je.2016.09.009.
12. Palmsten K, Huybrechts KF, Kowal MK, Mogun H, Hernandez-Diaz S. Validity of maternal and infant outcomes within nationwide Medicaid data. Pharmacoepidemiol Drug Saf 2014; 23(6):646-655. doi: 10.1002/pds.3627.
13. Cea Soriano L, Soriano-Gabarró M, García Rodríguez LA. Validity and completeness of colorectal cancer diagnoses in a primary care database in the United Kingdom. Pharmacoepidemiol Drug Saf. 2016;25(4):385-91. doi: 10.1002/pds.3877.

14. Ingeman A, Andersen G, Hundborg HH, Johnsen SP. Medical complications in patients with stroke: data validity in a stroke registry and a hospital discharge registry. Clin Epidemiol. 2010. PMID: 20865097.
15. Widdifield J, Bombardier C, Bernatsky S, Paterson JM, Green D, Young J, et al. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. BMC Musculoskelet Disord 2014;15:216. doi: 10.1186/1471-2474-15-216.
16. Kato D, Ryu H, Matsumoto T, Abe K, Kaneko M, Ko M, et al. Building primary care in Japan: Literature review. J Gen Fam Med 2019; 20(5): 170-179. doi: 10.1002/jgf2.252.
17. Amed S, Vanderloo SE, Metzger D, Collet JP, Reimer K, McCrea P, et al. Validation of diabetes case definitions using administrative claims data. Diabet Med. 2011 Apr;28(4): 424-7. doi: 10.1111/j.1464-5491.2011.03238.x.
18. Krysko KM, Ivers NM, Young J, O'Connor P, Tu K. Identifying individuals with multiple sclerosis in an electronic medical record. Mult Scler 2015; 21(2): 217-24. doi: 10.1177/1352458514538334.
19. The Tokushukai Group Ethics Committee. Tokushukai Group Institutional Review Board. http://www.mirai-iryo.com/rinri/rinri.html (in Japanese). Accessed Dec 14, 2020.
20. Tokushukai Medical Group. On “the validation study of claims-based definition for identifying rheumatoid arthritis”. https://www.tokushukai.or.jp/research/database/ (in Japanese). Accessed Dec 14, 2020.
21. Matsuda S, Fujimori K. The claim database in Japan. Asian Pac J Dis Manag 2012;6:55–9.
22. Cutrona SL, Toh S, Iyer A, Foy S, Cavagnaro E, Forrow S, et al. Design for validation of acute myocardial infarction cases in Mini-Sentinel. Pharmacoepidemiol Drug Saf 2012; 21 Suppl 1(0 1): 274-81. doi: 10.1002/pds.2314.
23. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum. 2010; 62(9): 2569-81. doi: 10.1002/art.27584.
24. Lash TL, Fox MP, Fink AK. Applying Quantitative Bias Analysis to Epidemiological Data. New York: Springer, 2009
25. Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL (eds). Modern Epidemiology. 3rd edn. Philadelphia, PA: Lippincott Williams and Wilkins; 2008.
26. Widdifield J, Labrecque J, Lix L, Paterson JM, Bernatsky S, Tu K, et al. Systematic review and critical appraisal of validation studies to identify rheumatic diseases in health administrative databases. Arthritis Care Res (Hoboken) 2013;65(9):1490-503. doi: 10.1002/acr.21993.
27. Rothman KJ, Greenland S, Lash TL. Case-control studies. In: Rothman KJ, Greenland S, Lash TL (eds). Modern Epidemiology. 3rd edn. Philadelphia, PA: Lippincott Williams and Wilkins; 2008.

Tables

Due to technical limitations, tables are only available as a download in the Supplemental Files section.

Figures

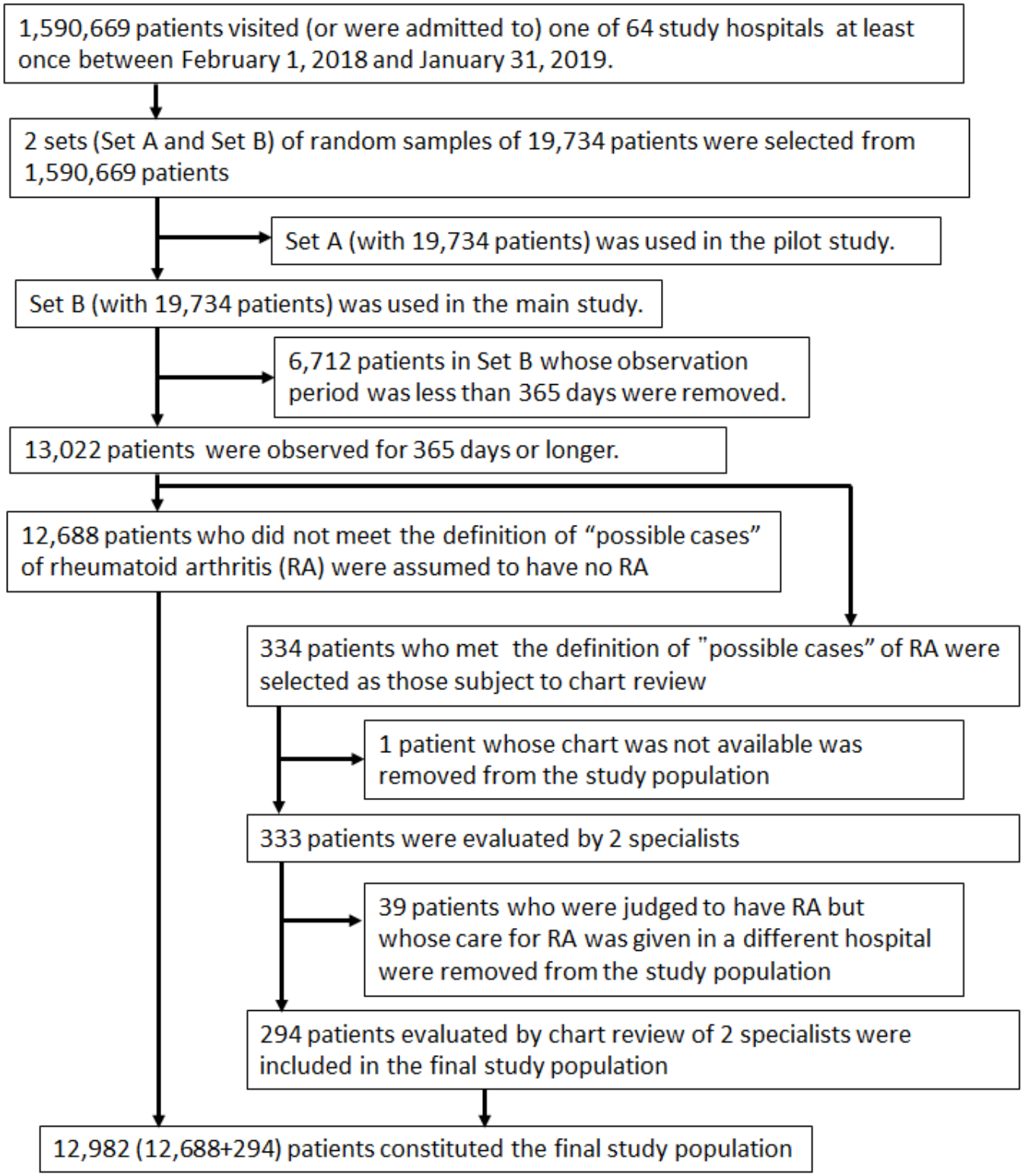

Figure 1

Selection of the study population and "possible cases" of RA

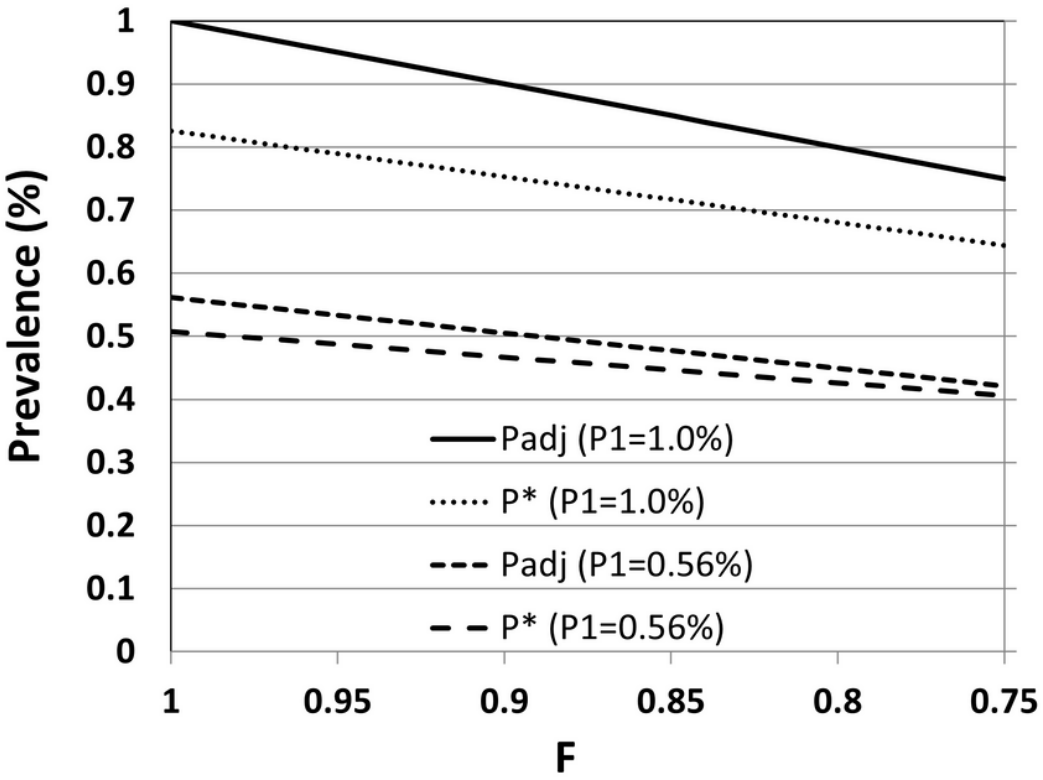

Figure 2

Sensitivity analysis to estimate the effect of the misclassification of patients with RA who did not meet the definition of “possible cases of RA” as those having no RA on the prevalence of definition-positive patients and the adjusted prevalence F: the proportion of patients with RA included in “possible cases of RA”; P1: true prevalence of patients with RA in the future study to estimate the prevalence of RA; P*: the prevalence of the definition-positive patients with RA; Padj: the adjusted prevalence of the patients with RA using the method of the quantitative bias analysis; RA: rheumatoid arthritis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BMCAdditionalFile1.docx
- Tables.docx