Redefining comorbidity understanding in rheumatoid arthritis through novel approaches using real-world data

Diego Benavent[1]* , Chamaida Plasencia-Rodríguez[2]

[1]Rheumatology Department, Hospital Universitari de Bellvitge, 08907 Barcelona, Spain
[2]Rheumatology Department, Hospital Universitario La Paz, 28046 Madrid, Spain

*Correspondence: Diego Benavent, Rheumatology Department, Hospital Universitari de Bellvitge, Carrer de la Feixa Llarga, s/n, L'Hospitalet de Llobregat, 08907 Barcelona, Spain. d_benavent@hotmail.com

Abstract

Rheumatoid arthritis (RA) is a prevalent chronic disease that is associated with numerous comorbidities. Accurate assessment of these coexisting conditions, as reported by clinicians, is critical for an improved understanding of the impact of the disease and patient care. This perspective aims to evaluate the utility of real-world data (RWD) for enhancing the understanding of comorbidities in RA and to assess its potential in reshaping clinical management. RWD approaches, specifically the use of structured databases or data extracted from electronic health records, offer promising alternatives to overcome the limitations of traditional methodologies. Structured databases provide a systematic approach to data analysis, utilizing diagnosis codes to study large patient cohorts, revealing the prevalence of conditions, and demonstrating the potential for long-term disease trend analysis. Meanwhile, natural language processing (NLP) and artificial intelligence (AI) image analysis can bridge the gap between structured and unstructured data, by extracting meaningful information from unstructured fields such as free text or imaging. NLP has proven effective in the identification of RA patients and research outcomes, while AI image analysis has enabled the discovery of hidden findings in cardiovascular assessments, establishing a basis for the assessment of comorbidities in RA. However, while the benefits of using RWD are substantial, challenges remain. Ensuring comprehensive data capture, managing missing data, and improving data detection are key areas requiring attention. The involvement of clinicians and researchers in rheumatology is crucial in unlocking the potential of RWD studies, offering the promise of significant improvements in disease characterization and patient health outcomes.

Keywords

Rheumatoid arthritis, real-world evidence, structured databases, natural language processing

Introduction

The chronic nature of rheumatoid arthritis (RA) and its global prevalence underscores the importance of a comprehensive understanding that extends beyond the joints. RA, which affects hundreds of thousands of people worldwide, with prevalence estimates of approximately 0.5% [1], carries a heavy burden due to its association with an extensive array of comorbidities. This correlation is particularly relevant as the emergence of comorbidities leads to unfavorable health outcomes, including diminished functionality, worsening quality of life, and elevated rates of morbidity and mortality [2–4]. The chronic nature of RA generates a wealth of patient data over time, and leveraging this information can enhance the understanding of RA comorbidities. Nonetheless, clinical trials and traditional cohort studies present difficulties including broad samples of patients during long follow-up periods. Thus, the holistic approach that RA management requires may be facilitated by new research methodologies, including the use of big databases from clinical practice, that can help deepen our understanding of the disease and thereby tailor patient care. These approaches not only allow for a more nuanced understanding of disease progression and comorbidities but also enable the use of data from different sources for the formulation of predictive models for patient outcomes. While these advancements offer the potential to fundamentally reshape RA management landscape, they are not without challenges and limitations that warrant attention. This perspective article aims to analyze the existing evidence on these methodologies, specifically in relation to RA comorbidities, with the goal of providing a roadmap to navigate and address these challenges effectively.

Comorbidities in RA

The spectrum of comorbidities that accompany RA extends across multiple health domains, further complicating patient management. This was clearly evident in the international, population-based study COMOrbidities in RA (COMORA), which examined 3,920 patients across 17 countries, shedding light on the most observed comorbidities associated with RA [5]. Depression, reported in 15% of RA patients, was the most common, followed by asthma (7%), cardiovascular (CV) events (6%), and solid-organ malignancies (5%). It has consistently been shown that patients with RA face a heightened risk of CV disease compared to healthy controls [6]. Correspondingly, CV risk factors, such as hypertension, diabetes mellitus, and hyperlipidemia, exhibit a high prevalence among patients with RA [7, 8]. Susceptibility to infections is also notably higher in the RA population, with a reported two-fold rate of hospitalizations due to infections in patients with RA compared to age and sex-matched controls [9]. Similarly, the risk of lymphoma in RA patients is more than double that of the general population [10].

The presence of comorbidities significantly influences therapeutic decisions. For example, comorbidities impact initial treatment choices in patients with early RA [11]. Specifically, the presence of at least one comorbidity has been linked to the sole use of methotrexate (MTX) *versus* other treatment combinations, potentially affecting treatment efficacy and disease progression. A systematic review showed the impact of gastrointestinal and liver comorbidities on the choice of pain treatment in patients with RA, among other forms of inflammatory arthritis [12]. In fact, the link between an increased risk of infections in patients with RA comorbidities and treatment choices has been evaluated [13]. Accurate identification and management of these comorbidities are paramount not only for enhancing RA outcomes but also for optimizing disease control and prognosis. A study examining difficult-to-treat RA revealed that comorbidities, among other factors, can restrict treatment options and amplify disease burden [14]. Thus, investigating comorbidities in RA patients is vital for optimizing treatment decisions and enhancing patient care through tailored approaches.

Current approaches for the assessment of comorbidities

Building upon the numerous previously mentioned comorbidities linked to RA, it has become clear that an accurate and comprehensive evaluation of these associations is vital to understanding the global impact of the disease. Nonetheless, this endeavor is fraught with intricate challenges, reflected in the disparities of

reporting frequencies across studies [15]. Existing methodologies, such as randomized clinical trials (RCTs), post-commercialization surveillance, and clinical registries, each possess inherent advantages and disadvantages. For example, RCTs provide rigorous scientific data, but often fall short in representing the wider, real-world demographics of RA patients, limiting their external validity. Post-commercialization surveillance, which hinges upon spontaneous adverse event reports, is hampered by under-reporting biases, as it depends upon active participation from healthcare practitioners and patients. While traditional registry-based cohort studies somewhat address these limitations, their data collection typically involves a limited number of participants and variables over relatively brief study periods. This limits their ability to yield in-depth insights into the long-term and diverse nature of RA.

Novel research methodologies

In light of these challenges, there is rising interest in harnessing research that routinely collects data from a variety of sources, including patient experiences and electronic health records (EHRs), collectively referred to as real-world data (RWD) studies [16]. Among these, it is worth noting the value of structured databases, which include the use of diagnostic codes to encode information in a predefined format, as well as unstructured information, gathered as free text in EHRs or from patient images [17]. These innovative approaches aspire to provide a broader and more representative depiction of the comorbidities landscape in RA (Table 1). In fact, building on these innovative approaches, the European Medicines Agency has significantly expanded the utilization of RWD for regulatory decision-making and underscored the need for wider access to data sources [18]. A multi-faceted approach, utilizing both structured and unstructured data, would provide a more representative view of the diseases, thereby enhancing patient outcomes.

Table 1. Differences between structured databases and unstructured information in the research of RA

Characteristic	Structured databases	Unstructured information
Definition	Uses diagnostic codes and predefined formats	Found in free text or images
Data source	Claims; prescriptions and administrative databases	Clinical notes; imaging data
Data collection	International Classification of Diseases, 9th edition (ICD-9), ICD-10 codes	Natural language processing (NLP) for text; convolutional neural network (CNN) for imaging
Examples of RA research	Detailed study of comorbidities; treatment safety	Identification of RA patients; extraction of outcome measures
Limitations	Limited by predefined formats; requires systematic coding; possible missing variables and biases	Analytical challenges; require precision in data detection; design challenges in algorithms
Benefits	Systematic and standardized data; detection of long-term trends; prevalence in broad populations	Enhances collection of specific features; contributes to multimodal research

Structured databases

Analysis of structured databases offers a promising avenue for in-depth exploration of comorbidity-related events by examining large cohorts of patients over extended periods. This has already proven beneficial in the assessment of various rheumatic diseases, such as gout, lupus, and psoriatic arthritis [19–21]. Their strength lies in the systematic use of diagnostic codes, allowing researchers to tap into vast quantities of data in a consistent and standardized manner. Delving into specific examples within RA, Petri et al. [22] analyzed data from more than 60,000 patients with RA, encompassing more than 2,000 co-morbidities, and used a rank-order of relative risks to show the variety of conditions that are associated with RA. Ramos et al. [23] investigated the prevalence of comorbidities in a population-based cohort of 96,921 patients with RA, employing ICD-10 codes. Patients with RA were compared with 484,605 age- and sex-matched controls without RA, and 26 comorbidities were evaluated. The study revealed that all investigated comorbidities were notably more frequent in the RA cohort. In addition to CV risk factors, the study found osteoarthritis (44% *versus* 21%), depression (32% *versus* 20%), and osteoporosis (26% *versus* 9%) as the most prevalent comorbidities in patients with RA. Furthermore, increasing numbers of comorbidities correlated with worsening patient-reported outcomes in terms of tender and swollen joint counts, functional status, and overall well-being. Studies such as this underscore the potential of structured databases in providing comprehensive insights into the complex landscape of comorbidities in RA.

The in-depth and detailed data from structured databases can help obtain more accurate insights into comorbidities, such as greater detail in malignancies. An illustration of this potential is provided by a recent study exploring various cancer types in patients with RA [24]. It showed that the overall likelihood of a cancer diagnosis within a year was 2.57% for RA patients, compared to 2.12% for non-RA individuals [hazard ratio (HR) = 1.21, 95% confidence interval (CI): 1.14, 1.29]. Consistent with previous findings, this study confirmed the elevated risk of lymphoma, pulmonary, and skin cancer among RA patients [25]. Interestingly, the investigation explored the risks associated with particular cancers, by classifying ICD-9 and ICD-10 diagnostic codes into 17 categories of malignancies. Although the evolution of HR across different time horizons and cancer types yielded a downward trend as time elapsed, RA patients exhibited consistently higher risks for certain cancers, including lymphoid, hematopoietic tissue, and respiratory and intrathoracic organ cancers, even after extended periods. Focused analyses such as this meticulous cancer-specific approach underscore the substantial potential of structured databases to capture nuanced, long-term disease trends in RA patients.

The intersection of comorbidities and potential pharmacological side effects presents a complex challenge in the management of RA, where associated conditions may emerge or exacerbate following the administration of specific treatments. In this regard, monitoring drug safety is one of those areas where RWD may provide the most value. After the publication of the Oral Rheumatoid Arthritis Trial (ORAL)-surveillance trial, which assessed the safety of Janus kinase inhibitors (JAKis), real-life studies began to focus on the risk of associated conditions with the use of JAKis compared to tumor necrosis factor inhibitors (TNFi) [26]. The Safety of TofAcitinib in Routine care patients with RA (STAR-RA) trial analyzed two administrative databases including over 100,000 patients with RA, and found no significant differences in the risk of malignancy between patients using tofacitinib and TNFi [27]. As a limitation, follow-up periods were often less than a year, which may not rule out a time-dependent exposure effect. These data were confirmed in other data sets involving a great number of patients, including studies from Korea, Taiwan (China), and Sweden, and showed non-significant or slightly elevated risks of malignancies in patients treated with JAKis, particularly tofacitinib, compared to TNFi [28]. The CV safety profile of JAKis has also been assessed in RWD research. A recent study showed that tofacitinib carried a similar risk of CV events compared to TNFi in two cohorts numbering 102,263 RA patients (HR = 1.01; 95% CI: 0.83, 1.23) [27]. In the exploration of infection risks, a recent study conducted in Japan focused on the risk of hospitalized infections among different age groups receiving targeted therapy, including biological disease-modifying antirheumatic drugs (bDMARDs) and JAKis, *versus* MTX therapy [29]. It found that the incidence rate of infection-caused hospitalizations per 100 patient-years was 3.2, 5.0, and 10.1 in the young (aged 16–64), elderly (aged 65–74), and older elderly (aged ≥ 75) groups, respectively. Interestingly, the risk of hospitalized infection under targeted therapy was not elevated in elderly or older elderly patients compared to MTX, unlike in young patients; the odds ratio (OR) of targeted therapy *versus* MTX for hospitalized infections was 1.3 (1.0–1.7; P = 0.021), 0.79 (0.61–1.0; P = 0.084), and 0.73 (0.56–0.94; P = 0.015) for the young, elderly, and older elderly groups, respectively [29]. Another study utilizing Medicare data from 2006 to 2015, investigated the risk of serious infections linked to low-dose glucocorticoid use in RA treatment [30]. Among 163,603 treatment episodes involving 120,656 patients, those exposed to glucocorticoids ≤ 5 mg/day had an infection incidence of 11.7/100 person-years compared to 8.0/100 in unexposed patients. This revealed a link between low-dose glucocorticoids and a 26% increased risk of infection requiring hospitalization (HR = 1.26; 95% CI: 1.02, 1.56) [30]. Together, these data highlight the essential role of RWD in understanding and monitoring the safety profile of treatments for RA, ensuring more precise and patient-centered therapeutic decisions.

Structured databases have proven to be instrumental in organizing and analyzing information related to chronic diseases such as RA. However, identifying RA solely by using diagnosis billing codes can be difficult due to their limited accuracy, with a reported positive predictive value (PPV) of 22% when using only the detection of one ICD-9 code for RA [31]. In addition, most of the data in EHRs often exist in unstructured formats, which poses analytical challenges. Here, the role of NLP and artificial intelligence (AI)-driven image interpretation takes on added significance.

Unstructured information

NLP bridges linguistics and AI, focusing on the computational processing and analysis of human language to decipher, understand, and extract meaningful information from unstructured data sources such as EHRs [32]. Evidence from past research demonstrates the value of NLP in complementing structured coding systems in RA research [33]. Thus, incorporating unstructured narrative data from EHRs alongside codified data has resulted in a substantially improved PPV of 94%, as opposed to 88% using only codified data. In another seminal work by Maarseveen et al. [34], AI techniques were employed on format-free text entries in EHRs to accurately identify RA patients. By comparing various machine-learning methods and a naive word-matching algorithm, they were able to develop a highly efficient and precise classifier for identifying RA patients [34]. Humbert-Droz et al. [35] harnessed the power of NLP to extract mentions of RA outcome measures from free-text outpatient rheumatology notes within the Rheumatology Informatics System for Effectiveness (RISE) registry. Of note, 34 million notes from 854,628 patients from 158 practices were processed. Their NLP pipeline demonstrated a sensitivity, PPV, and F1 score of 95%, 87%, and 91%, respectively, suggesting its potential for enhancing research outcomes and clinical care. These methods hold promise not only for disease detection but also, given the demonstrated capability of NLP in detecting and better understanding not only conditions such as RA, but also comorbidities in the RA patient population.

The investigation of comorbidities in RA can also benefit from the utilization of image extraction techniques. Numerous studies have leveraged data derived from imaging, particularly using machine-learning methods for analysis [36]. Most of the algorithms have been developed for joint image extraction from X-rays, demonstrating the ability to identify RA patients with remarkable accuracy and precision [37]. Similarly, image-processing techniques have been developed for identifying areas of joint inflammation in patients with RA through thermal image analysis [38]. The investigation of comorbidities and associated conditions can benefit from machine-learning techniques, which have helped to better understand arterial tissues and atherosclerotic plaques in individuals with RA. By analyzing characteristics such as morphology and texture in ultrasonography images, changes associated with atherosclerosis in the arterial walls have been identified with an accuracy of 83% [37]. In regard to RA-associated conditions, some investigations are advancing the field by, for example, radiomic analysis, which has shown predictive value in the mortality of patients with RA-associated interstitial lung disease [39]. The integration of these technologies into clinical practice and research offers a promising avenue for the comprehensive and multi-modal investigation of disease complexities.

Clinical implications of novel research methodologies

The integration of RWD into routine practice may significantly enhance the clinical management of RA. Databases from real-world practice can help to monitor the effectiveness of drugs, thereby offering clinicians a dynamic overview beyond the data gleaned from randomized trials. Moreover, insights into healthcare resource utilization can also be extracted from these studies. By identifying inefficiencies in current treatment pathways or locating bottlenecks in healthcare delivery, resources for long-term disease management could potentially be optimized, leading to more cost-effective services. Additionally, the combination of structured databases and NLP-based data provides a large amount of data from different sources that allows the development of predictive models, helping to forecast disease courses and associated comorbidity risks. These models enable early intervention strategies, potentially altering disease progression and improving patient outcomes. Furthermore, these predictive models can be embedded into clinical decision support systems. This ensures that evidence-based, personalized recommendations are readily available to clinicians at the point of care. Such an approach not only augments the clinical decision-making process but also minimizes the risk of oversight, enhancing the overall quality of patient management in RA.

Limitations of RWD

Despite all the potential shown by RWD studies, the deployment of these new research methods involves intrinsic limitations. EHRs, which are devised with clinical or billing objectives in mind rather than meeting the precise research criteria, may lead to some uncertainty as to whether relevant covariates are being consistently recorded across time and patients. Potential selection bias may emerge from factors like missing clinical information or patient disenrollment during follow-up. Moreover, the absence of various data streams may not be randomly distributed; data often focuses on attributes that are directly relevant to management, a variable that can introduce statistical bias into the findings. On top of this, extracting data from unstructured sources requires high precision to achieve proper validity. One instructive instance in rheumatology can be gleaned from a study that scrutinized the efficacy of the Phenotype KnowledgeBase (PheKB) algorithm in RA patient identification [40]. Despite good specificity (95%), the algorithm's sensitivity was significantly lower (72%), which underlines the challenges facing algorithm design. Hence, addressing these challenges requires a rethinking of methodologies and a shift towards new frameworks that consider the complex nature of RWD. The analysis must decompose the data's origins into manageable components, reflecting the idiosyncrasies of the EHR environment. Strategies for optimizing variable detection and managing missing data are of the utmost importance to fully realize the potential of RWD for clinical applications.

Conclusions

The inherent chronicity of RA results in the accumulation of substantial volumes of patient data over time in EHRs. This wealth of information has the potential to unlock valuable insights into comorbidities that affect patients. Pooling this data from expanding databases significantly enhances our research capabilities, providing crucial information towards bettering understanding of disease characteristics. Tools such as NLP are pivotal to achieving this aim, facilitating automated access to a vast reservoir of information. Harnessing these technological innovations may catalyze significant improvements in the characterization and understanding of the disease, promising unprecedented improvements in our pursuit of better health outcomes.

Abbreviations

AI: artificial intelligence

CI: confidence interval

CV: cardiovascular

EHRs: electronic health records

HR: hazard ratio

ICD-9: International Classification of Diseases, 9th edition

JAKis: Janus kinase inhibitors

MTX: methotrexate

NLP: natural language processing

PPV: positive predictive value

RA: rheumatoid arthritis

RWD: real-world data

TNFi: tumor necrosis factor inhibitors

Declarations

References

1. Almutairi K, Nossent J, Preen D, Keen H, Inderjeeth C. The global prevalence of rheumatoid arthritis: a meta-analysis based on a systematic review. Rheumatol Int. 2021;41:863–77.
2. Norton S, Koduri G, Nikiphorou E, Dixey J, Williams P, Young A. A study of baseline prevalence and cumulative incidence of comorbidity and extra-articular manifestations in RA and their impact on outcome. Rheumatology (Oxford). 2013;52:99–110.
3. Radner H, Lesperance T, Accortt NA, Solomon DH. Incidence and prevalence of cardiovascular risk factors among patients with rheumatoid arthritis, psoriasis, or psoriatic arthritis. Arthritis Care Res (Hoboken). 2017;69:1510–8.
4. Taylor PC, Atzeni F, Balsa A, Gossec L, Müller-Ladner U, Pope J. The key comorbidities in patients with rheumatoid arthritis: a narrative review. J Clin Med. 2021;10:509.
5. Dougados M, Soubrier M, Antunez A, Balint P, Balsa A, Buch MH, et al. Prevalence of comorbidities in rheumatoid arthritis and evaluation of their monitoring: results of an international, cross-sectional study (COMORA). Ann Rheum Dis. 2014;73:62–8.
6. DeMizio DJ, Geraldino-Pardilla LB. Autoimmunity and inflammation link to cardiovascular disease risk in rheumatoid arthritis. Rheumatol Ther. 2020;7:19–33.

7. Crowson CS, Rollefstad S, Ikdahl E, Kitas GD, van Riel PLCM, Gabriel SE, et al.; A Trans-Atlantic Cardiovascular Consortium for Rheumatoid Arthritis (ATACC-RA). Impact of risk factors associated with cardiovascular outcomes in patients with rheumatoid arthritis. Ann Rheum Dis. 2018;77:48–54.

8. Nikiphorou E, de Lusignan S, Mallen CD, Khavandi K, Bedarida G, Buckley CD, et al. Cardiovascular risk factors and outcomes in early rheumatoid arthritis: a population-based study. Heart. 2020;106: 1566–72.

9. Doran MF, Crowson CS, Pond GR, O'Fallon WM, Gabriel SE. Frequency of infection in patients with rheumatoid arthritis compared with controls: a population-based study. Arthritis Rheum. 2002;46: 2287–93.

10. Klein A, Polliack A, Gafter-Gvili A. Rheumatoid arthritis and lymphoma: incidence, pathogenesis, biology, and outcome. Hematol Oncol. 2018;36:733–9.

11. Vega-Morales D, Treviño-Montes D, Wah-Suarez M. Age and comorbidities influence initial treatment of patients with early rheumatoid arthritis. J Rheumatol. 2016;43:827.

12. Radner H, Ramiro S, van der Heijde DM, Landewé R, Buchbinder R, Aletaha D. How do gastrointestinal or liver comorbidities influence the choice of pain treatment in inflammatory arthritis? A Cochrane systematic review. J Rheumatol Suppl. 2012;90:74–80.

13. Listing J, Gerhold K, Zink A. The risk of infections associated with rheumatoid arthritis, with its comorbidity and treatment. Rheumatology (Oxford). 2013;52:53–61.

14. Roodenrijs NMT, van der Goes MC, Welsing PMJ, Tekstra J, Lafeber FPJG, Jacobs JWG, et al. Difficult-to-treat rheumatoid arthritis: contributing factors and burden of disease. Rheumatology (Oxford). 2021; 60:3778–88.

15. Ranganath VK, Maranian P, Elashoff DA, Woodworth T, Khanna D, Hahn T, et al. Comorbidities are associated with poorer outcomes in community patients with rheumatoid arthritis. Rheumatology (Oxford). 2013;52:1809–17.

16. Concato J. FDA clinical investigator training course: real world evidence [Internet]. Maryland: U.S. Food and Drug Administration; 2022 [cited 2023 Aug 28]. Available from: https://www.fda.gov/media/167541/download

17. Knevel R, Liao KP. From real-world electronic health record data to real-world results using artificial intelligence. Ann Rheum Dis. 2023;82:306–11.

18. Use of real-world evidence in regulatory decision making–EMA publishes review of its studies [Internet]. The Netherlands: European Medicines Agency; ©1995–2023 [cited 2023 Aug 28]. Available from: https://www.ema.europa.eu/en/news/use-real-world-evidence-regulatory-decision-making-ema-publishes-review-its-studies

19. Shah K, Paris M, Mellars L, Changolkar A, Mease PJ. Real-world burden of comorbidities in US patients with psoriatic arthritis. RMD Open. 2017;3:e000588.

20. Benavent D, Peiteado D, Martinez-Huedo MÁ, Hernandez-Hurtado M, Balsa A, de Miguel E. Healthcare-related impact of gout in hospitalized patients in Spain. Sci Rep. 2021;11:13287.

21. Choi MY, Hahn J, Malspeis S, Stevens EF, Karlson EW, Sparks JA, et al. Association of a combination of healthy lifestyle behaviors with reduced risk of incident systemic lupus erythematosus. Arthritis Rheumatol. 2022;74:274–83.

22. Petri H, Maldonato D, Robinson NJ. Data-driven identification of co-morbidities associated with rheumatoid arthritis in a large US health plan claims database. BMC Musculoskelet Disord. 2010;11: 247.

23. Ramos AL, Redeker I, Hoffmann F, Callhoff J, Zink A, Albrecht K. Comorbidities in patients with rheumatoid arthritis and their association with patient-reported outcomes: results of claims data linked to questionnaire survey. J Rheumatol. 2019;46:564–71.

24. Wang F, Palmer N, Fox K, Liao KP, Yu KH, Kou SC. Large-scale real-world data analyses of cancer risks among patients with rheumatoid arthritis. Int J Cancer. 2023;153:1139–50.

25. Smitten AL, Simon TA, Hochberg MC, Suissa S. A meta-analysis of the incidence of malignancy in adult patients with rheumatoid arthritis. Arthritis Res Ther. 2008;10:R45.

26. Ytterberg SR, Bhatt DL, Mikuls TR, Koch GG, Fleischmann R, Rivas JL, et al.; ORAL Surveillance Investigators. Cardiovascular and cancer risk with tofacitinib in rheumatoid arthritis. N Engl J Med. 2022;386:316–26.

27. Khosrow-Khavar F, Desai RJ, Lee H, Lee SB, Kim SC. Tofacitinib and risk of malignancy: results from the safety of tofacitinib in routine care patients with rheumatoid arthritis (STAR-RA) study. Arthritis Rheumatol. 2022;74:1648–59.

28. Testa D, Calvacchi S, Petrelli F, Giannini D, Bilia S, Alunno A, et al. One year in review 2021: pathogenesis of rheumatoid arthritis. Clin Exp Rheumatol. 2021;39:445–52.

29. Sakai R, Tanaka E, Majima M, Harigai M. Unincreased risk of hospitalized infection under targeted therapies *versus* methotrexate in elderly patients with rheumatoid arthritis: a retrospective cohort study. Arthritis Res Ther. 2022;24:135.

30. George MD, Hsu JY, Hennessy S, Chen L, Xie F, Curtis JR, et al. Risk of serious infection with low-dose glucocorticoids in patients with rheumatoid arthritis: an instrumental variable analysis. Epidemiology. 2022;33:65–74.

31. Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, Cai T, Zink RM, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. J Am Med Inform Assoc. 2012;19:e162–9.

32. Masud JHB, Kuo CC, Yeh CY, Yang HC, Lin MC. Applying deep learning model to predict diagnosis code of medical records. Diagnostics (Basel). 2023;13:2297.

33. Liao KP, Cai T, Gainer V, Goryachev S, Zeng-treitler Q, Raychaudhuri S, et al. Electronic medical records for discovery research in rheumatoid arthritis. Arthritis Care Res (Hoboken). 2010;62:1120–7.

34. Maarseveen TD, Meinderink T, Reinders MJT, Knitza J, Huizinga TWJ, Kleyer A, et al. Machine learning electronic health record identification of patients with rheumatoid arthritis: algorithm pipeline development and validation study. JMIR Med Inform. 2020;8:e23930.

35. Humbert-Droz M, Izadi Z, Schmajuk G, Gianfrancesco M, Baker MC, Yazdany J, et al. Development of a natural language processing system for extracting rheumatoid arthritis outcomes from clinical notes using the National Rheumatology Informatics System for effectiveness registry. Arthritis Care Res (Hoboken). 2023;75:608–15.

36. Kim KJ, Tagkopoulos I. Application of machine learning in rheumatic disease research. Korean J Intern Med. 2019;34:708–22.

37. Kedra J, Gossec L. Big data and artificial intelligence: will they change our practice? Joint Bone Spine. 2020;87:107–9.

38. Wasilewska A, Pauk J, Ihnatouski M. Image processing techniques for ROI identification in rheumatoid arthritis patients from thermal images. Acta Mechanica et Automatica. 2018;12:49–53.

39. Venerito V, Manfredi A, Lopalco G, Lavista M, Cassone G, Scardapane A, et al. Radiomics to predict the mortality of patients with rheumatoid arthritis-associated interstitial lung disease: a proof-of-concept study. Front Med (Lausanne). 2023;9:1069486.

40. Zheng HW, Ranganath VK, Perry LC, Chetrit DA, Criner KM, Pham AQ, et al. Evaluation of an automated phenotyping algorithm for rheumatoid arthritis. J Biomed Inform. 2022;135:104214.