



# Probabilistic Linkage of Randomized Controlled Trial Data to Administrative Claims: A Case Study of Patients from Baricitinib Clinical Trials

Catherine B. McGuiness · Natalie N. Boytsov · Xiang Zhang ·  
Xin Wang · Carol L. Kannowski · Rolin L. Wade

Received: November 30, 2020 / Accepted: March 18, 2021 / Published online: April 2, 2021  
© The Author(s) 2021

## ABSTRACT

**Introduction:** The aim of this work is to assess the feasibility of probabilistically linking randomized controlled trial (RCT) data to claims data in a real-world setting to inform future rheumatoid arthritis (RA) research.

**Methods:** This retrospective cohort study utilized IQVIA's Patient Centric Medical Claims (Dx) Database, IQVIA's Longitudinal Prescription Claims (LRx) Database, and Lilly's baricitinib RCT data from a sample of patients that consented to the linkage of their de-identified insurance claims to their de-identified RCT data. Patients were initially matched on age, gender, and three-digit ZIP code of the provider and further matched according to a point scoring system using additional clinical variables.

**Results:** A total of 245 patients from 49 US clinical trial sites were eligible for the study and 78 (31.8%) of these patients consented to participate. Of the 78 consented patients, 69 (88%) were successfully matched on age, gender, and three-digit ZIP code of the provider. Of the 69 patients successfully matched on age, gender, and three-digit ZIP code of the provider, 44

(63.8%) had at least one sufficient match using the point scoring system. Of these 44, 23 (52.3%) patients matched at a ratio of one RCT patient to one Dx/LRx patient, 11 (25.0%) at a ratio of 1:2, 7 (15.9%) at a ratio of 1:3 and three (6.8%) at a ratio of 1:4 or greater. To further improve match ratios, a variable hierarchy was applied to the 18 RCT patients with 2–3 matches. Overall, 38 of the 78 (48.7%) consented RCT patients were successfully matched 1:1 to claims database patients.

**Conclusions:** This probabilistic linkage methodology demonstrates the feasibility, at a moderate linkage rate, of linking patients from RCTs to real-world data, which can provide a means to assess additional information not usually collected within or following a clinical trial.

**Keywords:** Rheumatoid arthritis; Clinical trial data; Administrative claims data; Probabilistic linkage

---

C. B. McGuiness (✉) · X. Wang · R. L. Wade  
IQVIA, Plymouth Meeting, PA, USA  
e-mail: katy.mcguiness@iqvia.com

N. N. Boytsov · X. Zhang · C. L. Kannowski  
Eli Lilly and Company, Indianapolis, IN, USA

## Key Summary Points

### Why carry out this study?

There is a large demand for early real-world evidence (RWE), however real-world data (RWD) for newly launched drugs takes time to accumulate and depends on a number of factors, including drug uptake in the market.

Linkage of RCT data to administrative claims can provide an opportunity to study the clinical trial population in greater detail, which can also serve as an early and timely assessment of real-world performance of newly launched molecules.

### What was learned from the study?

The probabilistic linkage methodology presented in this study demonstrates the feasibility of linking patients from clinical trials to real-world data, which can provide a means to assess additional information not usually captured/collected within a clinical trial.

DMARDs (bDMARDs), which include tumor necrosis factor-alpha (TNF) inhibitors and non-TNF bDMARDs, and Janus kinase inhibitors (JAK)—the newest class of RA therapy [2, 5–7]. Baricitinib is an oral selective inhibitor of JAK1 and JAK2 and has been shown to improve RA signs and symptoms in phase II/III randomized controlled trials (RCTs) in adult patients with moderately to severely active RA despite treatment with TNF inhibitors (RA-BEACON trial), csDMARDs (RA-BUILD trial) and methotrexate (MTX; RA-BEAM trial), and in csDMARD-naive patients (RA-BEGIN trial) [8]. Patients who completed these trials were eligible to participate in a long-term extension study (i.e., the RA-BEYOND trial) [9]. Patients enrolled in the RA-BEYOND study were consented to have their RCT data linked to administrative claims data.

There is a large demand for early real-world evidence (RWE), however real-world data (RWD) for newly launched drugs takes time to accumulate and depends on a number of factors, including drug uptake in the market. This delay limits the timely assessment of real-world outcomes. To assess the feasibility of mitigating this limitation, we conducted an experiment where US patients enrolled in the RA-BEYOND long-term extension study were consented to have their RCT data probabilistically linked to health plan claims. Although Protected Health Information (PHI) provides an option for direct linkage to RWD sources, patient privacy concerns may contribute to low patient-participation rates. In addition, there are restrictions placed on the re-identification of patients in many RWD data sources, precluding the use of direct identifiers. As such, a probabilistic linkage method was selected to match RCT patients in the respective datasets. Probabilistic linkage includes methods to incorporate indirect identifiers, assigning a higher priority to variables upon which linking is less likely.

Administrative data from insurance claims databases can be an important resource for clinical studies, particularly when linked to clinically rich and detailed RCT data. To date, several studies have linked RCT data to administrative claims data with linkage rates ranging from 55.6 to 90.8% [10–12]. Strom et al. [11] assessed the accuracy of identifying

## DIGITAL FEATURES

This article is published with digital features, including a summary slide, to facilitate understanding of the article. To view digital features for this article go to <https://doi.org/10.6084/m9.figshare.14230274>.

## INTRODUCTION

Rheumatoid arthritis (RA) is a chronic, systemic inflammatory disease characterized by progressive joint damage that can lead to loss of physical function and disability [1–5]. Multiple treatment options are available for RA, including conventional synthetic disease-modifying antirheumatic drugs (csDMARDs), biologic

cardiovascular events through billing claims in place of clinical event adjudication in structural heart disease trials by linking clinical trial data to inpatient Medicare claims. Their results indicated that claims could be used to trigger evaluation of neurological events, potentially improving the efficiency of the evaluation of techniques and devices designed to reduce such events. Brennan et al. [12] assessed the accuracy and completeness of claims- versus site-based follow-up with clinical event committee adjudication of cardiovascular outcomes and found similar randomized treatment effects using the two methods, also suggesting that claims data can be used to support clinical research by leveraging routinely collected data.

Linkage of RCT data to administrative claims can provide an opportunity to study the clinical trial population in greater detail, which can also serve as an early and timely assessment of real-world performance of newly launched molecules. This study aimed to assess the feasibility of probabilistic linkage of RCT data to US claims data using a sample of patients from baricitinib clinical trials.

## METHODS

### Data sources

This retrospective cohort study utilized IQVIA's Patient Centric Medical Claims (Dx) Database, IQVIA's Longitudinal Prescription Claims (LRx) Database and Lilly's baricitinib RCT data from the sample of patients that consented to the linkage of their de-identified RCT data to de-identified insurance claims.

The Dx database is derived from professional fee claims using the CMS-1500 billing form. It provides patient-level diagnoses and procedures for visits to US office-based physicians, ambulatory, and general health care sites. It includes claims paid by commercial insurance, Medicare, Medicaid and those paid in cash. The Dx data provides patient level information including the following key analysis variables: age, gender, three-digit ZIP code, geography, international classification of diseases (ICD) diagnostic codes, current procedural terminology (CPT)

procedure codes, healthcare common procedure coding system (HCPCS) products, date of service, location of care, reported cost of service, payer type (e.g., third party, Medicare), and rendering practitioner. The Dx database is composed of approximately 1 billion outpatient medical claims per year submitted by over 860,000 practitioners in the US. All data are rendered non-identified to protect patient privacy.

The LRx database consists of pharmacy claims for dispensed prescriptions collected from pharmacies. The database covers approximately 90% of all dispensed prescriptions from US retail pharmacies and over 1.4 billion prescriptions per year, representing claims from all payer types, including self-pay/cash, Medicare, Medicaid, and other third-party payers. The LRx database captures information starting in 2001 on adjudicated dispensed prescriptions sourced from retail, mail, long-term care, and specialty pharmacies. The database provides the following patient-level information: prescriber, payer type, products, age, gender, three-digit ZIP code, prescription fill date, quantity dispensed, days of supply, number of refills, and cost information. All data are rendered non-identified to protect patient privacy.

RA-BEYOND is a clinical trial that was designed to investigate the long-term safety and any side effects of baricitinib in participants who completed a previous baricitinib rheumatoid arthritis study, providing additional treatment with baricitinib for 7 years [9].

The study protocol and patient consent form were reviewed and approved by an Institutional Review Board (IRB). Investigators participating in RA-BEYOND were informed of the purpose of this study and trained before they approached patients for informed consent. The study protocol was approved by Quorum Review Institutional Review Board (file number 28020). The study was conducted in agreement with requirements of registered clinical trials (ClinicalTrials.gov identifier NCT01885078) and with the Declaration of Helsinki.

Patients in the RA-BEYOND trial were invited to participate in the current study during an RA-BEYOND trial visit. All participants in the current study consented to have their de-identified

RCT data linked to de-identified insurance claims. De-identified patient information of the consented patient cohort was transferred from Lilly to IQVIA (who has access to de-identified insurance claims data) via secure transmission using a secure file transfer protocol (SFTP) method. Data were stored on secure servers and only necessary members of the study team had access to the data.

## Study Design

Patients who completed the RA-BEACON, RA-BUILD, RA-BEAM, or RA-BEGIN trial were eligible to participate in the RA-BEYOND trial. All patients who participated in the RA-BEYOND trial and consented to have their de-identified RCT data linked to de-identified insurance claims were included in the analysis. Exclusion criteria were not applied a priori. Patients were initially matched on age, gender, and three-digit ZIP code of the RCT-participating provider. To further improve matching, a point scoring system was implemented where  $\geq 3$  points constituted a sufficient match (Fig. 1). The scoring system described below was developed based on the available RCT data:

- Investigator/Provider match using National Provider Identifier (NPI) (1 point)
- Prior DMARD exposure. DMARDs of interest included auranofin, azathioprine, cyclophosphamide, cyclosporine, gold sodium thiomalate, hydroxychloroquine,

leflunomide, methotrexate, minocycline, mycophenolate, penicillamine, sulfasalazine, abatacept, adalimumab, certolizumab pegol, etanercept, golimumab, infliximab, rituximab, sarilumab, tocilizumab, and tofacitinib.

- No prior DMARDs used (1 point)
- Number of prior DMARDs used (1 point for one drug matched, 2 points for two drugs matched, etc.)
- Prior non-DMARD exposure. Non-DMARDs were defined as any drug that is not a DMARD of interest
  - No prior non-DMARDs used (1 point)
  - Number of prior non-DMARDs used (1 point for one drug matched, 2 points for two drugs matched, etc.)
- RCT informed consent date matched ( $\pm 30$  days) to a Dx or LRx claim with the same date of service (1 point)
- Clinical trial code (i.e., ICD-9-CM code V70.7; ICD-10-CM code Z00.6) observed on a Dx claim at any time from the informed consent date to last RCT visit date (1 point)
- Visit dates
  - At least three visit dates matched (1 point)
  - At least 75% visit dates matched (1 additional point)

To identify a 1:1 match ratio among RCT patients with a “one to many” match ratio, a variable hierarchy (i.e., NPI > DMARD

• Investigator/provider match using NPI	1 point
• Prior DMARD exposure	
○ No prior DMARDs used	1 point
○ Number of prior DMARDs matched	1 point for each matched drug
• Prior non-DMARD exposure	
○ No prior non-DMARDs used	1 point
○ Number of prior non-DMARDs used	1 point for each matched drug
• RCT informed consent date (+/- 30 days) matched to a Dx or LRx claim with the same date of service	1 point
• Clinical trial code (i.e., ICD-9-CM code V70.7; ICD-10-CM code Z00.6) observed on a Dx claim at any time from the informed consent date to last RCT visit date	1 point
• Visit dates (all Dx visit dates with an RA diagnosis on the claim including the informed consent date were checked)	
○ At least 3 visit dates matched	1 point
○ At least 75% visit dates matched	1 additional point

NPI: National Provider Identifier; DMARD: Disease-modifying antirheumatic drug; Dx: IQVIA's Patient Centric Medical Claims Database; LRx: IQVIA's

Longitudinal Prescription Claims Database; RCT: randomized clinical trial; ICD: International Classification of Diseases; RA: rheumatoid arthritis

**Fig. 1** Matching point system

use > visits > non-DMARD use > RCT code) was applied to the baricitinib patients with 2–3 matches. The hierarchy order was based on the sensitivity and specificity of each variable. For example, NPI was ranked as the most important variable in the hierarchy because NPI is the most specific variable followed by DMARD use, visits, and non-DMARD use. The RCT code variable was ranked as the least important variable in the hierarchy because it is the least specific (all RCTs are included under the same ICD codes) and not particularly sensitive due to under reporting of RCT participation in claims data.

### Statistical Analyses

Patient demographic and clinical characteristics were reported as descriptive statistics. Categorical variables were reported as percentages. Continuous variables were reported as means with standard deviations. Age distributions were compared via *t* test, gender via Chi-square, with an alpha level of 0.05 to assess statistical significance. All statistical analyses were performed using SAS<sup>®</sup> version 9.4 (SAS Institute Inc., Cary, NC, USA).

## RESULTS

A total of 245 patients from 49 RA-BEYOND participating US sites were eligible for this study. A total of 78 (31.8%) of the eligible patients consented to participate in the study. Mean (SD) age of the total study sample ( $N = 78$ ) was 54.8 (10.9) years and 69.2% were female. Of these 78 consented patients, 69 (88%) were successfully matched on age, gender, and three-digit ZIP code of the provider (Table 1). Among the 268 sufficiently matched pairs (i.e., total matching points  $\geq 3$ ; Table 2), 212 pairs remained after applying RCT DMARD inclusion/exclusion criteria. Using the point scoring system to further improve matching ratios, 44 consented RCT patients had at least one sufficient match (Table 3) as follows: 23 of 44 (52.3%) patients matched at a ratio of one RCT patient to one Dx/LRx patient (no duplication), 11 (25.0%) at a ratio of 1:2, seven

(15.9%) at a ratio of 1:3 and three (6.8%) at a ratio of 1:4 or greater. Two RCT patients matched to the same Dx/LRx patient and both pairs were subsequently excluded. The variable hierarchy (i.e., NPI > DMARD use > visits > non-DMARD use > RCT code) was applied to the 18 baricitinib patients with 2–3 Dx/LRx matches resulting in 38 of 78 (48.7%) baricitinib patients' successful 1:1 match to patients from the claims database. The mean (SD) age of the 38 successfully matched patients was 57.5 (9.8) years and 81.6% were female. The matching ratio of three RCT patients with 2–3 Dx/LRx matches could not be reduced to 1:1.

The mean (SD) age of the 38 successfully matched patients was 57.5 (9.8) years and 81.6% were female. Age and gender comparisons revealed significant differences between the final matched sample and the initial sample of consented patients with the final matched sample ( $n = 38$ ) being significantly older and with more females compared to the initial sample of consented patients ( $n = 78$ ) [57.5 vs. 54.8,  $p = 0.0318$  and 81.6 vs. 69.2%,  $p = 0.0213$ , respectively].

## DISCUSSION

Long-term follow-up is often an essential component of clinical trial research in chronic diseases including RA as both the benefits and the adverse effects of treatment may take years to emerge in the real world. The linkage of clinical trial data to administrative claims data allows for leveraging of existing data to create a unique linked claims-clinical database that contains rich clinical information paired with long-term medical and prescription data to increase the depth and breadth of the research capabilities (including assessment of early real-world evidence of treatment outcomes).

This study explored a method of probabilistic linking of RA RCT patients to claims data in the real world. A total of 44 of the 78 (56.4%) consented patients had at least one sufficient match using the point scoring system; 23 of 78 (29.5%) patients matched at a ratio of one RCT patient to one Dx/LRx patient, 11 of 78 (14.1%) at a ratio of 1:2, seven of 78 (9.0%) at a ratio of

**Table 1** Probabilistic linkage details

Step	Criteria	RCT patient count	%	Dx/LRx patient count
1	Overall RCT subjects	78	100.0	–
2	Patients from step 1 who matched on year of birth and sex	78	100.0	1,706,722
3	3-digit ZIP code (provider)	69	88.5	935 patients (944 total pairs)
<b>The step 3 sample was assessed on the following matching variables*</b>				
	Match on NPI (1 point)	27	39.1	72
	Prior DMARD use			
	No prior DMARDs used (1 point)	13	18.8	46
	Number of prior DMARDs used (1 point for 1 drug matched, 2 points for 2 drugs matched, etc.)			
	1 drug matched	45	65.2	341
	2 drugs matched	10	14.5	45
	3 drugs matched	1	1.4	1
	Prior non-DMARDs use			
	No prior non-DMARDs used (1 point)	20	29.0	78
	Number of prior non-DMARDs used (1 point for 1 drug matched, 2 points for 2 drugs matched, etc.)			
	1 drug matched	35	50.7	196
	2 drugs matched	15	21.7	53
	3 drugs matched	4	5.8	12
	4 drugs matched	2	2.9	2
	≥ 1 visits with an RA diagnosis ± 30 days from the informed consent date (1 point)	69	100.0	934
	Clinical trial code observed in claims any time from the informed consent date to the last RCT visit date (1 point)	5	7.2	5
	RA visit dates (all visit dates including the informed consent date were checked)			
	At least three visit dates matched (1 point)	24	34.8	45
	At least 75% visit dates matched (1 additional point)	0	0.0	–

*RCT* randomized clinical trial, *NPI* National Provider Identifier, *RA* rheumatoid arthritis, *DMARDs* disease-modifying antirheumatic drugs, *Dx* IQVIA's Patient Centric Medical Claims Database, *LRx* IQVIA's Longitudinal Prescription Claims Database

\*Proportions within this section are based on the count of patients remaining after step 3 (i.e.,  $n = 69$ ) as the denominator

**Table 2** Distribution of total points among RCT-Dx/LRx patient pairs

Count of points	Count of RCT-Dx/LRx patient pairs	% of RCT-Dx/LRx patient pairs matched on age, sex, and provider three-digit ZIP code (N = 944)
0	1	0.1
1	295	31.3
2	380	40.3
3	186	19.7
4	57	6.0
5	20	2.1
6	5	0.5

1:3 and three of 78 (3.8%) at a ratio of 1:4 or greater. To further improve matching, a variable hierarchy was applied to the 18 RCT patients with three or more matches. Overall, 38/78 (48.7%) baricitinib patients from the RCT data were successfully matched 1:1 to patients from claims data demonstrating the feasibility of linkage, to a moderate degree, of RCT data to claims data. Probabilistic data linkage has also been demonstrated to inform research in other disease areas (e.g., cancer, chronic obstructive pulmonary disease, stroke, pediatric cardiovascular medicine) [10–16]. Hay et al. successfully demonstrated the feasibility of linking RCT data to administrative data in Canada with a linkage rate of 90.8% [10]. Another study by Strom et al. achieved a match rate of 79.8% by linking clinical trial data to Medicare inpatient discharge data [11]. Brennan et al. linked Medicare claims data to the clinical data from seven randomized cardiovascular trials and achieved a linkage rate of 55.6% [12].

Compared to previous studies using similar probabilistic linking approaches without relying on personal identifiers, our study achieved a slightly lower linkage rate (48.7%). One of the potential reasons for the observed difference could be related to the number of matching variables that were utilized. For example, in the study by Hay et al., probabilistic data linkage was based on birth date, patient initials, gender,

**Table 3** RCT-Dx/LRx matching ratios

Dx/LRx patient count	RCT patient count	% of unique pairs who met RCT DMARD inclusion/exclusion criteria* and had ≥ 3 matching points (N = 44)
1	23	52.3
2	11	25.0
3	7	15.9
4	1	2.3
5	1	2.3
9	1	2.3

\*Dx/LRx patients were excluded if they matched to trial patients meeting any of the following DMARD history exclusion criteria per trial protocols (the applicable measurement window included the consent date and all history before the consent date): With a cDMARD within the 4 weeks before the consent date (RA-BEAM trial patients); Ever received a biologic DMARD (RA-BEAM trial patients); Ever received a cDMARD other than MTX before the consent date (RA-BEGIN trial patients); Ever received a bDMARD before the consent date (RA-BEGIN trial patients); With any three cDMARDs within the 3 m before the consent date. Drug overlap was not assessed, only occurrence during the measurement window; With a bDMARD before the consent date (RA-BUILD trial patients)

With any three cDMARDs within 30 days before the consent date (RA-BEACON trial patients). Drug overlap was not examined, only occurrence during the measurement window; With any prior bDMARD use (JADA trial patients); With any prior oral JAKi use

diagnosis site, diagnosis date, and histology. The only variable the current study had in common with the Hay study was gender, which was utilized in the matching process. In addition to the three required variables used in the current study (age, gender, and three-digit ZIP code of the RCT provider), another six variables of interest were also selected for further matching according to a point scoring system in order to increase the matching accuracy though this may have resulted in a relatively lower linkage rate. The difference in the source of the claims data may have been another driving

factor in the observed differences in linkage rates. The administrative data used in most of the literature included inpatient claims, however, in the current study, inpatient data variables were not available potentially yielding a lower linkage rate as hospitalizations and their associated dates may be more unique/specific compared to outpatient visits and prescriptions.

Results of the current study provide a framework for future research to obtain RWE through linkage of RCT patients' non-identified information to real-world data (e.g., non-identified insurance claims databases). This framework may provide an opportunity to assess additional information from treated patients not collected within a clinical trial such as healthcare resource utilization and associated costs. These types of additional information allow for early RWE to inform future studies on the treatment's associated outcomes (e.g., comparative cost-effectiveness, healthcare cost-offsets).

This study had several limitations and challenges. Obtaining informed consent from patients after the trial initiation limited the number of participants in this study and may have led to a sample that was not representative of the original RCT population. Differences in mean age and gender distribution were observed between the final matched sample and the initial sample of consented patients though these differences may be explained by the small sample sizes. Limited familiarity of clinical investigators with RWE studies contributed to challenges in enrolling greater numbers of patients for this particular study. This challenge highlights the importance of having a robust plan for patient consent/enrollment at the start of a clinical trial in case this type of subsequent research is of interest. The results of this study are limited further due to the inability to validate the matches. Validation was not feasible as PHI was not collected in this study (to preserve patient privacy). Unlike deterministic linkage methods, mismatches are possible with probabilistic linkage methods. For example, two different patients could be linked (i.e., a false-positive link) resulting in incorrectly reported outcomes, or two records from the same patient may not be linked (i.e., a false-negative link),

resulting in potential missing information. Minimizing mismatches depends on the quality and unique identification of the information used in the linkage process. Lastly, unlike claims databases that are "closed" (i.e., with the exception of claims paid for in cash, all of the claims are captured for each patient from one data collection source), "open-source" claims databases (i.e., databases with multiple data collection sources), such as those used in the current study, often do not contain information related to patient enrollment into pharmacy and medical benefits limiting the ability to ensure complete capture of a patient's claim history.

This study yielded a few recommendations for future research. Expanding the originator trial consent to include the possibility of this type of subsequent research during the trial design is recommended to accelerate the generation of RWE of newly launched molecules. Additionally, if linkage studies were incorporated into the original clinical study protocol, inclusion of comparative patient groups (i.e., data flags to indicate placebo or active treatment) could provide an opportunity to investigate the real-world impact of the treatment. The development of algorithms to validate the probabilistic linkage is needed. In addition, utilizing administrative claims data with patient enrollment information would help to minimize the limitation of missing patients claim history.

## CONCLUSIONS

The probabilistic linkage methodology presented in this study demonstrates the feasibility, at a moderate rate, of linking patients from clinical trials to real-world data, which can provide a means to assess additional information not usually captured/collected within a clinical trial (e.g., healthcare resource utilization and associated costs).



## ACKNOWLEDGEMENTS

We thank Yi-Chien Lee, Associate Data Science & Advanced Analytics director at IQVIA for data analysis and statistical support, and Scott Beatlie, Sr. Research Advisor in Statistics at Eli Lilly and Company for his assistance with the clinical trial patient sample data extraction and transfer.

**Funding.** This study was funded by Eli Lilly and Company (including funding of the journal's Rapid Service Fee) who had a role in the study design, interpretation of the study results, decision to submit the results for publication, and manuscript preparation.

**Authorship.** All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

**Disclosures.** Catherine B. McGuiness, Xin Wang, and Rolin L. Wade are employees of IQVIA. IQVIA was paid by Lilly to conduct the current study. Natalie N. Boytsov and Xiang Zhang were employees and stockholders of Eli Lilly and Company at the time the study was conducted. Carol L. Kannowski is an employee and stockholder of Eli Lilly and Company.

**Compliance with Ethics Guidelines.** The study protocol was approved by Quorum Review Institutional Review Board (file number 28020). Patients were required to provide written informed consent. The study was conducted in agreement with requirements of registered clinical trials (ClinicalTrials.gov identifier NCT01885078) and with the Declaration of Helsinki.

**Data Availability.** Lilly provides access to all individual participant data collected during the trial, after anonymization, with the exception of pharmacokinetic or genetic data. Data are available to request 6 months after the indication studied has been approved in the US and EU and after primary publication

acceptance, whichever is later. No expiration date of data requests is currently set once data are made available. Access is provided after a proposal has been approved by an independent review committee identified for this purpose and after receipt of a signed data sharing agreement. Data and documents, including the study protocol, statistical analysis plan, clinical study report, blank or annotated case report forms, will be provided in a secure data sharing environment. For details on submitting a request, see the instructions provided at [www.vivli.org](http://www.vivli.org). The claims data that were linked to Lilly's clinical trial data are not publicly available per contracted agreements.

**Open Access.** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

1. Biggioggero M, Crotti C, Becciolini A, Favalli EG. Tocilizumab in the treatment of rheumatoid arthritis: an evidence-based review and patient selection. *Drug Des Dev Ther.* 2019;13:57.
2. Aletaha D, Smolen JS. Diagnosis and management of rheumatoid arthritis: a review. *JAMA.* 2018;320(13):1360–72.

3. Lin Y-J, Anzaghe M, Schülke S. Update on the pathomechanism, diagnosis, and treatment options for rheumatoid arthritis. *Cells*. 2020;9(4):880.
4. Matcham F, Scott IC, Rayner L, et al. The impact of rheumatoid arthritis on quality-of-life assessed using the SF-36: a systematic review and meta-analysis. *Semin Arthritis Rheum*. 2014;44:123–30.
5. Singh JA, Saag KG, Bridges SL Jr, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol*. 2016;68(1):1–26.
6. Abbasi M, Mousavi MJ, Jamalzahi S, et al. Strategies toward rheumatoid arthritis therapy; the old and the new. *J Cell Physiol*. 2019;234(7):10018–31.
7. Kumar P, Banik S. Pharmacotherapy options in rheumatoid arthritis. *Clin Med Insights Arthritis Musculoskelet Disord*. 2013;6:CMAMD-S5558.
8. Fleischmann R, Alam J, Arora V, et al. Safety and efficacy of baricitinib in elderly patients with rheumatoid arthritis. *RMD Open*. 2017;3(2):e000546.
9. National Institutes of Health USNLoM. An Extension Study in Participants With Moderate to Severe Rheumatoid Arthritis (RA-BEYOND). 2020. <https://clinicaltrials.gov/ct2/show/NCT01885078?term=RA-BEYOND&draw=2&rank=1>. Accessed 1 April 2020.
10. Hay AE, Pater JL, Corn E, et al. Pilot study of the ability to probabilistically link clinical trial patients to administrative data and determine long-term outcomes. *Clin Trials*. 2019;16(1):14–7.
11. Strom JB, Zhao Y, Faridi KF, et al. Comparison of clinical trials and administrative claims to identify stroke among patients undergoing aortic valve replacement: findings from the EXTEND study. *Circ Cardiovasc Interv*. 2019;12(11):e008231.
12. Brennan JM, Wruck L, Pencina MJ, et al. Claims-based cardiovascular outcome identification for clinical research: Results from 7 large randomized cardiovascular clinical trials. *Am Heart J*. 2019;218:110–22.
13. Blanchette CM, DeKoven M, De AP, Roberts M. Probabilistic data linkage: a case study of comparative effectiveness in COPD. *Drugs Context*. 2013. <https://doi.org/10.7573/dic.212258>.
14. Ido MS, Bayakly R, Frankel M, Lyn R, Okosun IS. Peer reviewed: administrative data linkage to evaluate a quality improvement program in acute stroke care, Georgia, 2006–2009. *Prev Chronic Dis*. 2015. <https://doi.org/10.5888/pcd12.140238>.
15. Kim TJ, Lee JS, Kim J-W, et al. Building linked big data for stroke in Korea: linkage of stroke registry and national health insurance claims data. *J Korean Med Sci*. 2018. <https://doi.org/10.3346/jkms.2018.33.e343>.
16. Pasquali SK, Li JS, Jacobs ML, Shah SS, Jacobs JP. Opportunities and challenges in linking information across databases in pediatric cardiovascular medicine. *Prog Pediatr Cardiol*. 2012;33(1):21–4.