Predicting randomized clinical trial results with real-world evidence: A case study in the comparative safety of tofacitinib, adalimumab and etanercept in patients with rheumatoid arthritis

**Version:** 0.1

**Authors:**

Runsheng Wang 1, Patrick Ryan 1,2, Hamed Abedtash 3, Eldar Allakhverdiiev 4, Deepa Balraj 5, Juan Banda 5, Maura Beaton 1, Paul Biondich 6, Clair Blacketer 2, Richard Boyce 26, Alison Callahan 5, Ray Chen 1, Young-Geun Choi 7, Ivan John Clement 8, Michael Davies 8, Frank DeFalco 2, Sara Dempster 9, Stephen Deppen 10, 11, Jon Duke 12, Dmytro Dymshyts 4, Thomas Falconer 1, Kristin Feeney 13, Pavel Grafkin 4, Shaun Grannis 6, Jill Hardin 2, Ross Hayden 6, George Hripcsak 1, Tommy Huynh 14, Yeesuk Kim 15, 16, Christopher Knoll 2, Martin Lavallee 17, Evan Minty 18, Akihiko Nishimura 16, Paul Petraro 19, Melanie Philofsky 20, Aaron Potvien 12, Christian Reich 8, Jenna Reps 2, Peter Rijnbeek 21, Patrick Ryan 1,2, Paolas Saroufim 22, Lisa Schilling 20, Trey Schneider 12, Martijn Schuemie 2, Anthony Sena 2, Nigam Shah 5, Andrey Soares 20, David Sontag 23, Marc Suchard 16, Joel Swerdel 2, Devin Tian 22, Mui Van Zandt 8, Rohit Vashisht 5, Runsheng (Bridget) Wang 1, James Weaver 2, Chunhua Weng 1, Andrew Williams 24, Ross Williams 21, Jin Zhou 25, George Hripcsak 1

1 Columbia University, 2 Janssen Research and Development, 3 Eli Lilly, 4 Odysseus Data Services, 5 Stanford University, 6 Regenstrief Institute, 7 Fred Huchinson Cancer Research Center, 8 IQVIA, 9 sdempsterconsulting, 10 Vanderbilt University Medical Center, 11 US Veterans Affairs, 12 Georgia Tech Research Institute, 13 Deloitte, 14 Blue Cross Blue Shield of South Carolina, 15 Hanyang University, 16 UCLA, 17 Bayer, 18 University of Calgary, 19 Novo Nordisk Inc, 20 University of Colorado, Denver Anschutz Medical Campus, 21 Erasmus MC, 22 Case Western Reserve University, 23 MIT, 24 Tufts Medical Center, 25 University of Arizona, 26 University of Pittsburgh

**Date:** 2 May 2018

**Acknowledgment:** The analysis is based in part on work from the Observational Health Sciences and Informatics collaborative. OHDSI (<http://ohdsi.org>) is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics.

The authors declare the following disclosures: PBR are employees of Janssen Research and Development.

# Table of contents

[1 Table of contents 3](#_Toc512777561)

[2 List of abbreviations 4](#_Toc512777562)

[3 Abstract 4](#_Toc512777563)

[4 Amendments and Updates 4](#_Toc512777564)

[5 Milestones 5](#_Toc512777565)

[6 Rationale and Background 5](#_Toc512777566)

[7 Study Objectives 5](#_Toc512777567)

[7.1 Primary Hypotheses 5](#_Toc512777568)

[7.2 Secondary Hypotheses 5](#_Toc512777569)

[7.3 Primary Objectives 6](#_Toc512777570)

[7.4 Secondary Objectives 7](#_Toc512777571)

[8 Research methods 7](#_Toc512777572)

[8.1 Study Design 7](#_Toc512777573)

[8.2 Data Source(s) 8](#_Toc512777574)

[8.3 Study population 9](#_Toc512777575)

[8.3.1 Criteria common to both target and comparator cohorts 9](#_Toc512777576)

[8.4 Exposures 10](#_Toc512777577)

[8.4.1 Target: New users of tofacitinib with rheumatoid arthritis satisfying RCT criteria 10](#_Toc512777578)

[8.4.2 Target: New users of adalimumab with rheumatoid arthritis satisfying RCT criteria 10](#_Toc512777579)

[8.4.3 Target: New users of etanercept with rheumatoid arthritis satisfying RCT criteria 10](#_Toc512777580)

[8.4.4 Target: New users of tofacitinib with rheumatoid arthritis 10](#_Toc512777581)

[8.4.5 Target: New users of adalimumab with rheumatoid arthritis 10](#_Toc512777582)

[8.4.6 Target: New users of etanercept with rheumatoid arthritis 10](#_Toc512777583)

[8.5 Outcomes 10](#_Toc512777584)

[8.5.1 Malignancies excluded non-melanoma skin cancer (NMSC) 10](#_Toc512777585)

[8.5.2 Major adverse cardiovascular events (MACE) 10](#_Toc512777586)

[8.5.3 Myocardial infarction 10](#_Toc512777587)

[8.5.4 Ischemic stroke 10](#_Toc512777588)

[8.5.5 Cardiovascular death 10](#_Toc512777589)

[8.5.6 Opportunistic infections 10](#_Toc512777590)

[8.5.7 Hepatic events 11](#_Toc512777591)

[8.5.8 Cardiovascular events other than MACE 11](#_Toc512777592)

[8.5.9 All-cause mortality 11](#_Toc512777593)

[8.5.10 Negative control outcomes 11](#_Toc512777594)

[8.5.11 Positive control outcomes 11](#_Toc512777595)

[8.6 Covariates 12](#_Toc512777596)

[8.6.1 Propensity score covariates 12](#_Toc512777597)

[9 Data Analysis Plan 12](#_Toc512777598)

[9.1 Calculation of time-at risk 12](#_Toc512777599)

[9.2 Model Specification 13](#_Toc512777600)

[9.2.1 Pooling effect estimates across databases 13](#_Toc512777601)

[9.3 Analyses to perform 13](#_Toc512777602)

[9.4 Output 14](#_Toc512777603)

[9.5 Evidence Evaluation 14](#_Toc512777604)

[10 Study Diagnostics 15](#_Toc512777605)

[10.1 Sample Size and Study Power 15](#_Toc512777606)

[10.2 Cohort Comparability 15](#_Toc512777607)

[10.3 Systematic Error Assessment 15](#_Toc512777608)

[11 Strengths and Limitations of the Research Methods 15](#_Toc512777609)

[12 Protection of Human Subjects 16](#_Toc512777610)

[13 Management and Reporting of Adverse Events and Adverse Reactions 16](#_Toc512777611)

[14 Plans for Disseminating and Communicating Study Results 16](#_Toc512777612)

[15 Appendix 1 16](#_Toc512777613)

[15.1 Concepts excluded from covariates 16](#_Toc512777614)

[15.2 Negative control outcomes 16](#_Toc512777615)

[16 References 16](#_Toc512777616)

# List of abbreviations

ATC Anatomic Therapeutic Chemical

CYCLOPS Cyclic coordinate descent for logistic, Poisson and survival analysis

SNOMED Systematized Nomenclature of Medicine

OHDSI Observational Health Data Sciences and Informatics

OMOP Observational Medical Outcomes Partnership

T Target cohort

C Comparator cohort

O Outcome cohort

PS Propensity Scores

LASSO Least absolute shrinkage and selection operator

MDRR Maximum detectable relative risk

CI Confidence Interval

DOD Date of Death

MACE Major Adverse Cardiovascular Event

NMSC Non-melanoma skin cancer

ADA Adalimumab

ETN Etanercept

TOF Tofacitinib

RA Rheumatoid arthritis

# Abstract

This study aims to compare the safety of tofacitinib with adalimumab and etanercept in patients with rheumatoid arthritis (RA). We will replicate the design and population inclusion criteria of an ongoing phase 3b/4 randomized clinical trial (NCT02092467), with the aim of predicting the RCT results using real-world evidence. In this study, we will analyze data from observational databases across the OHDSI network using the OHDSI CohortMethod package framework to perform this comparative study.

# Amendments and Updates

|  |  |  |  |
| --- | --- | --- | --- |
| 0.1 | 2 May 2018 | P. Ryan | First draft |
| 1.1 | 15 March 2019 | R. Wang | Second draft |
|  |  |  |  |
|  |  |  |  |

# Milestones

|  |  |
| --- | --- |
| Milestone | Planned / Estimated Date |
| Start of analysis | 2 May 2018 |
| End of initial analysis | 3 May 2018 |
| End of final analysis | 30 May 2019 |
| Submission of manuscript | 31 July 2019 |

# Rationale and Background

Rheumatoid arthritis (RA) is a chronic inflammatory condition that primarily affecting peripheral joints, causing joint damage and loss of function. Treatment options for patients with RA have significantly expanded in the past two decades, including conventional synthetic disease-modifying antirheumatic drugs (csDMARDs), biologic DMARDs (bDMARDs) and targeted synthetic DMARDs. When patients have inadequate response to csDMARDs, a second agent, either bDMARDs or tsDMARDs, is added, based on the expected efficacy and safety profile of the drug.

Tofacitinib, an oral Janus kinase inhibitor, was approved for use in patients with moderate-to-severe rheumatoid arthritis by the FDA in 2012. As part of the approval, FDA required a postmarketing study to examine the long-term effects of tofacitinib on cardiovascular events, malignancies, and serious infections; the Phase 3b/4 randomized clinical trial uses adalimumab and etanercept as active comparators and is scheduled to complete in 2019.

Regulatory approvals with post-marketing commitments for obtaining additional evidence are commonplace, as it is understood that the complete safety and efficacy product of a medicine cannot be fully understood during a product’s initial development. Additional prospective studies, such as randomized clinical trials and clinical registries, are time-consuming and resource-intensive, and may still not provide sufficient evidence to resolve all uncertainties about the effects of a medical product in clinical practice. Real-world evidence, including population-level effect estimation from observational databases about the safety and comparative effectiveness of medical products, provides a complementary perspective toward – and potentially offers the possibility to supplant in select circumstances – randomized clinical trials. It would be the shared interest of regulators, product manufacturers, health systems, clinicians and patients if the desired evidence to understand a product profile and its appropriate use in medical practice could be obtained more quickly and with less resources, so long as the evidence was of sufficient quality to make appropriate decisions.

Regulatory authorities seek the highest quality available evidence to inform their regulatory decision-making, including the initial approval of products, ongoing pharmacovigilance, and extension to supplemental indications. While observational evidence has been used in post-marketing safety evaluations, there is not currently consensus across the research community as to whether the reliability of real-world evidence has been sufficiently demonstrated to constitute “substantial evidence” when evaluating the effectiveness of new drugs or determining the need for further prospective studies. Retrospective analyses have shown mixed results about the ability of observational studies to replicate the findings from clinical trials. A systematic review comparing results from observational studies with randomized trials found the expected difference in effect estimates to be small. Some discordant findings, such as the association between hormone therapies and cardiovascular risk have been explained away through re-analysis, while other replication failures have been left unresolved. To our knowledge, there has been little work to prospectively evaluate the performance of retrospective observational database analyses in their ability to predict randomized trial results.

In this study, using the tofacitinib Phase 4 clinical trial as a case study, we will conduct a retrospective observational database analysis to replicate the clinical trial design and study population. We will report the results in anticipation of the trial completion, such that an objective assessment of observational study-RCT concordance can be made. We will provide an a priori framework for the study-RCT comparison, to avoid the risk of posthoc rationalization clouding the interpretation of the assessment. Further, we will examine the impact of trial’s inclusion criteria on the generalizability of the results across the real-world population currently using these treatments in practice. In this regard, we aim to not only provide specifically actionable information to clinicians about the comparative safety of tofacitinib, adalimumab and etanercept prior to the RCT read-out, but also provide a demonstration project to more generally support further policy about when retrospective observational databases can appropriately be used to inform regulatory decision-making.

# Study Objectives

## Primary Hypotheses

This study’s primary hypotheses are:

* There is no difference in the incidence of malignancies excluding non-melanoma skin cancer between tofacitinib and adalimumab in patients with rheumatoid arthritis
* There is no difference in the incidence of malignancies excluding non-melanoma skin cancer between tofacitinib and etanercept in patients with rheumatoid arthritis
* There is no difference in the incidence of major adverse cardiovascular events (MACE) between tofacitinib and adalimumab in patients with rheumatoid arthritis
* There is no difference in the incidence of major adverse cardiovascular events (MACE) between tofacitinib and etanercept in patients with rheumatoid arthritis

## Secondary Hypotheses

* There is no difference in the incidence of opportunistic infections between tofacitinib and adalimumab in patients with rheumatoid arthritis
* There is no difference in the incidence of opportunistic infections between tofacitinib and etanercept in patients with rheumatoid arthritis
* There is no difference in the incidence of hepatic events between tofacitinib and adalimumab in patients with rheumatoid arthritis
* There is no difference in the incidence of hepatic events between tofacitinib and etanercept in patients with rheumatoid arthritis
* There is no difference in the incidence of cardiovascular events other than MACE between tofacitinib and adalimumab in patients with rheumatoid arthritis
* There is no difference in the incidence of cardiovascular events other than MACE between tofacitinib and etanercept in patients with rheumatoid arthritis
* There is no difference in the incidence of all-cause mortality between tofacitinib and adalimumab in patients with rheumatoid arthritis
* There is no difference in the incidence of all-cause mortality between tofacitinib and etanercept in patients with rheumatoid arthritis

## Primary Objectives

* To compare the risk of **O: malignancies excluding non-melanoma skin cancer** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of adalimumab with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to cohort end date (‘on treatment’ time-at-risk).
* To compare the risk of **O: malignancies excluding non-melanoma skin cancer** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of adalimumab with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to observation period end date (‘intent-to-treat’ time-at-risk).
* To compare the risk of **O: malignancies excluding non-melanoma skin cancer** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of etanercept with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to cohort end date (‘on treatment’ time-at-risk).
* To compare the risk of **O: malignancies excluding non-melanoma skin cancer** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of etanercept with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to observation period end date (‘intent-to-treat’ time-at-risk).
* To compare the risk of **O: major adverse cardiovascular events (MACE)** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of adalimumab with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to cohort end date (‘on treatment’ time-at-risk).
* To compare the risk of **O: major adverse cardiovascular events (MACE)** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of adalimumab with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to observation period end date (‘intent-to-treat’ time-at-risk).
* To compare the risk of **O: major adverse cardiovascular events (MACE)** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of etanercept with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to cohort end date (‘on treatment’ time-at-risk).
* To compare the risk of **O: major adverse cardiovascular events (MACE)** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of etanercept with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to observation period end date (‘intent-to-treat’ time-at-risk).

## Secondary Objectives

* To compare the risk of **secondary outcomes: O3: opportunistic infections, O4: hepatic events, O5: cardiovascular events other than MACE, O6: all-cause mortality** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of adalimumab with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to cohort end date (‘on treatment’ time-at-risk).
* To compare the risk of **secondary outcomes: O3: opportunistic infections, O4: hepatic events, O5: cardiovascular events other than MACE, O6: all-cause mortality** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of adalimumab with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to observation period end date (‘intent-to-treat’ time-at-risk).
* To compare the risk of **secondary outcomes: O3: opportunistic infections, O4: hepatic events, O5: cardiovascular events other than MACE, O6: all-cause mortality** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of etanercept with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to cohort end date (‘on treatment’ time-at-risk).
* To compare the risk of **secondary outcomes: O3: opportunistic infections, O4: hepatic events, O5: cardiovascular events other than MACE, O6: all-cause mortality** between **T: new users of tofacitinib with rheumatoid arthritis** and **C: new users of etanercept with rheumatoid arthritis**, we will estimate the population-level effect of exposure on the time to event during the period from cohort start date to observation period end date (‘intent-to-treat’ time-at-risk).
* To compare the relative risks mentioned in the primary objectives in the subgroups of subjects meeting all criteria in the clinical trials to the risk in the total sets of subjects having the indication for which the drug was approved based on the trials.

# Research methods

## Study Design

This study will follow a retrospective, observational, comparative cohort design. We define ‘retrospective’ to mean the study will be conducted using data already collected prior to the start of the study. We define ‘observational’ to mean there is no intervention or treatment assignment imposed by the study. We define 'cohort' to mean a set of patients satisfying one or more inclusion criteria for a duration of time. We define ‘comparative cohort design’ to mean the formal comparison between two cohorts, a target cohort and comparator cohort, for the risk of an outcome during a defined time period after cohort entry [[1](#_ENREF_1)]. The design will be conducted in one administrative claims database in the US, as described in section 8.2. The specific exposure cohorts are described in section 8.3 and 8.4. The time-at-risk definitions are described in section 9.1. The statistical analysis plan for population-level effect estimation is described in section 9.2.

## Data Source(s)

The analyses will be performed across one observational database. This database has been transformed into the OMOP Common Data Model, version 5.0 or higher. The complete specification for OMOP Common Data Model is available at: <https://github.com/OHDSI/CommonDataModel>.

Data sources expected to participate to include:

Optum ClinFormatics Extended Datamart

Optum Pan-therapeutic HER

Truven MarketScan CCAE

Truven MarketScan MDCD

Truven MarketScan MDCR

Iqvia PhamMetrics

Iqvia Ambulatory EHR

Iqvia LRx/Dx

Columbia / New York Presbyterian

Stanford STRIDE

Regenstrief

US Veterans Affairs VINCI

Each database is described below:

* Optum’s Clinformatics® Extended Data Mart – Date of Death (DOD)

Optum Clinformatics® Extended DataMart is an adjudicated US administrative health claims database for members of private health insurance, who are fully insured in commercial plans or in administrative services only (ASOs), Legacy Medicare Choice Lives (prior to January 2006), and Medicare Advantage (Medicare Advantage Prescription Drug coverage starting January 2006). The population is primarily representative of commercial claims patients (0-65 years old) with some Medicare (65+ years old) however ages are capped at 90 years. It includes data captured from administrative claims processed from inpatient and outpatient medical services and prescriptions as dispensed, as well as results for outpatient lab tests processed by large national lab vendors who participate in data exchange with Optum. This dataset also provides date of death (month and year only) for members with both medical and pharmacy coverage from the Social Security Death Master File (however after 2011 reporting frequency changed due to changes in reporting requirements) and location information for patients is at the US state level. Family identifiers are provided and utilized to infer mother-child linkages.

The major data elements contained within this database are outpatient pharmacy dispensing claims (coded with National Drug Codes (NDC), inpatient and outpatient medical claims which provide procedure codes (coded in CPT-4, HCPCs, ICD-9-CM or ICD-10-PCS) and diagnosis codes (coded in ICD-9-CM or ICD-10-CM). The data also contain selected laboratory test results (those sent to a contracted thirds-party laboratory service provider) for a non-random sample of the population (coded with LOINC codes).

For this study we will use the version of this database at Janssen Research and Development, which spans June 2000 up to and including September 2017.

## Study population

The overall study population consists of patients who enter either the target cohort or comparator cohort. Patients who qualify for both the target cohort and comparator cohort are only considered for whichever cohort occurs first. Both cohorts were further restricted to those subjects whose index date was in the period when both T and C were observed. In this case, this mean that subjects in the comparator cohort (adalimumab and etanercept) prior to November 2012 were removed, since no subjects were observed in the target cohort (tofacitinib) prior to that date.

### Criteria common to both target and comparator cohorts

For the T and C cohorts pertaining, we impose the following requirements in our observational study, compared to those in the clinical trial:

|  |  |
| --- | --- |
| CLINICAL Trial (NCT02092467) | Observational study |
| Moderate to severe rheumatoid arthritis | At least one diagnosis of rheumatoid arthritis one or prior to index exposure (moderate-to-severe inferred by initiation of biologic therapy)  At least one diagnostic code for rheumatoid arthritis (code set 1), 6 months run-in period before the index exposure (ADA, ETN, or TOFA). |
| Adults >/= 50 years | Adults >/= 50 years \* |
| Taking methotrexate without adequate control of symptoms | At least one record of prior exposure to methotrexate \* |
| Have at least one cardiovascular risk factor (eg, current smoker, high blood pressure, high cholesterol levels, diabetes mellitus, history of heart attack, family history of coronary heart disease, extra-articular RA disease) | At least one condition record for hypertension, hyperlipidemia, diabetes mellitus, or myocardial infarction prior to index exposure or at least one record suggesting current smoker status \* |
| No Current or recent infection | At least one condition record of infection on or in prior 60 days from index exposure \* |
| No Clinically significant laboratory abnormalities | Exactly zero measurement records with numeric value outside normal range on or in 60 days prior to index exposure \* |
| No Pregnancy | Exactly zero condition, procedure, measurement or observation records indicating pregnancy in 270 days before to 90 days after index exposure \* |
|  | Add an exclusion code set (code set 2): at least one code for any of the following conditions during the 6-month run in period: psoriasis, psoriatic arthritis, ankylosing spondylitis, crohn’s disease, ulcerative colitis (any approved indications for TNFi and TOFA) |

Index date is the first date that either ADA, ETN, or TOFA was prescribed. Patient needs to be continuously enrolled for at least 6 months prior to the index date. Patients are followed until they 1) meet the primary endpoint by having MACE; 2) exit the data source (censored); 3)discontinuation of the drug (defined by either a different biologic was prescribed), or 60-90 days after the last dispense of study drugs (censored).

For the secondary analyses, the criteria marked with a star (\*) will be removed one at a time, starting with the last one, to explore the influence of these restrictions on the estimated effect.

Both cohorts were further restricted to those subjects whose index date was in the time period when both T and C were observed.

## Exposures

### Target: New users of tofacitinib with rheumatoid arthritis satisfying RCT criteria

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Target: New users of adalimumab with rheumatoid arthritis satisfying RCT criteria

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Target: New users of etanercept with rheumatoid arthritis satisfying RCT criteria

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Target: New users of tofacitinib with rheumatoid arthritis

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Target: New users of adalimumab with rheumatoid arthritis

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Target: New users of etanercept with rheumatoid arthritis

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

## Outcomes

### Malignancies excluded non-melanoma skin cancer (NMSC)

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Major adverse cardiovascular events (MACE)

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Myocardial infarction

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Ischemic stroke

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Cardiovascular death

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Opportunistic infections

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Hepatic events

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Cardiovascular events other than MACE

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### All-cause mortality

URL: <https://ohdsi.org/web/atlas/#/cohortdefinition/XXX>

### Negative control outcomes

Negative controls are concepts known to not be associated with the target or comparator cohorts, such that we can assume the true relative risk between the two cohorts is 1. Negative controls are selected using a similar process to that outlined by Voss et al. [[5](#_ENREF_5)]. Person counts of all potential drug-condition pairs are reviewed in observational data; this person count data helps determine which pairs are even probable for use in calibration. Given the list of potential drug-condition pairs, the concepts in the pairs must meet the following requirements to be considered as negative controls: (1) that there is no Medline abstract where the MeSH terms suggest an association between the drug and the condition [[6](#_ENREF_6)], (2) that there is no mention of the drug-condition pair on a US Product Label in the “Adverse Drug Reactions” or “Postmarketing” section [[7](#_ENREF_7)], (3) there are no US spontaneous reports suggesting that the pair is in an adverse event relationship [[8](#_ENREF_8), [9](#_ENREF_9)], (4) that the OMOP Vocabulary does not suggest that the drug is indicated for the condition, (5) that the concepts are usable (i.e. not too broad, not suggestive of an adverse event relationship, not pregnancy related), and (6) the exact concept itself is utilized in patient level data (i.e. concepts that are not usually used within the data are usually indicative a broad concept that has a child that is more specific). The remaining concepts are “optimized”, meaning parent concepts remove children as defined by the OMOP Vocabulary (e.g. if both “Non-Hodgkin’s Lymphoma” and “B-Cell Lymphoma” we selected, child concept “B-Cell Lymphoma would be removed for its parent “Non-Hodgkin’s Lymphoma”). Once potential negative control candidates were selected, manual clinical review to exclude any pairs that may still be in a causal relationship or similar to the study outcome will be performed to select the top concepts by patient exposure. The final list can be found in appendix 15.2.

For each negative control outcome, a patient enters the negative control outcome cohort at the occurrence of a diagnose code identified by the concepts listed above, or any one of its descendant codes.

### Positive control outcomes

In addition to negative control outcomes, we will also include synthetic positive control outcomes. These are outcomes based on the real negative controls, but where the true effect size is artificially increased to a desired effect size by injection of additional, simulated outcomes [[10](#_ENREF_10)]. To preserve confounding, these additional outcomes are sampled from predicted probabilities generated using a fitted predictive model. For each negative control outcome, three positive control outcomes will be generated with true relative risk is 1.5, 2, and 4. Using both negative and positive controls, we will fit a systematic error model and perform confidence interval calibration [[10](#_ENREF_10)].

## Covariates

### Propensity score covariates

Propensity scores (PS) will be used as an analytic strategy to reduce potential confounding due to imbalance between the target and comparator cohorts in baseline covariates. The propensity score is the probability of a patient being classified in the target cohort vs. the comparator cohort, given a set of observed covariates.

The types of baseline covariates used to fit the propensity score model will be:

* Demographics
  + Gender
  + Age group (5-year bands)
  + Index year
  + Index month
* Conditions
  + In prior 30d
  + In prior 365d
* Condition aggregation
  + SNOMED
* Drugs
  + In prior 30d
  + In prior 365d
  + Overlapping index date
* Drug aggregation
  + Ingredient
  + ATC Class
* Risk scores
  + Charlson comorbidity index

Specific covariates to be excluded from the propensity score model are labelled **concepts to exclude** as detailed in Appendix 15.1.

All covariates that occur in fewer than 0.1% of the persons between the target and comparator cohorts combined will be excluded prior to model fitting for computational efficiency.

# Data Analysis Plan

## Calculation of time-at risk

The time-at-risk will be defined in two ways: 1) as the ‘intent to treat’ period, defined as the time from the cohort start date to observation period end date, and 2) as a ‘on treatment’ period, defined as the time from cohort start date to cohort end date (representing the period of persistent exposure without switching or augmentation of therapy).

As previously described, the “on-treatment” period would be: 1) meet the primary endpoint by having MACE; 2) exit the data source (censored); 3)discontinuation of the drug (defined by a different biologic (code set 4) was prescribed), or 60-90 days after the last dispense of study drugs (censored).

## Model Specification

In this study, we compare the target cohort with the comparator cohort for the hazards of outcome during the time-at-risk by applying a Cox proportional hazards model.

The time-to-event of outcome among patients in the target and comparator cohorts is determined by calculating the number of days from the start of the time-at-risk window (the cohort start date), until the earliest event among 1) the first occurrence of the outcome, 2) the end of the time-at-risk window, and 3) the end of the observation period that spans the time-at-risk start.

Patients with the outcome, observed prior to target or comparator cohort entry are not excluded from consideration.

Propensity scores will be used as an analytic strategy to reduce potential confounding due to imbalance between the target and comparator cohorts in baseline covariates. The propensity score is the probability of a patient being classified in the target cohort vs. the comparator cohort, given a set of observed covariates. In this study, the propensity score is estimated for each patient, using the predicted probability from a regularized logistic regression model, fit with a Laplace prior (LASSO) and the regularization hyperparameter selected by optimizing the likelihood in a 10-fold cross validation using 10 replications per fold, a starting variance of 0.01 and a tolerance of 2e-7. Covariates to be used in the propensity score model are listed in section 8.5.

The target cohort and comparator cohorts will be stratified into five quantiles of the propensity score distribution. The final outcome model will apply a conditional Cox proportional hazard model, conditions on the propensity score strata.

### Pooling effect estimates across databases

Random-effects meta-analytic estimates were generated using the Hartung-Knapp-Sidik-Jonkman and Dersimonian Laird methods to pool evidence across the databases for all comparison-outcome-analyses where there was sufficient homogeneity (I2 < 40%).

## Analyses to perform

The following analyses will be performed:

* A = 4 comparisons: 2 main comparisons (tofacitinib vs. adalimumab; tofacitinib vs. etanercept) using all inclusion criteria, and removing RCT criteria for each main comparison.
* B = 9 outcomes: 1) malignancies excluding non-melanoma skin cancer, 2) major adverse cardiovascular events (MACE), 3) myocardial infarction, 4) ischemic stroke, 5) cardiovascular death, 6) opportunistic infections, 7) hepatic events, 8) cardiovascular events over than MACE, 9) all-cause mortality
* C = 2 time-at-risk definitions: 1) intent-to-treat, 2) on treatment
* D = 1 model: Cox regression using propensity score stratification
* E = 12 databases: Optum ClinFormatics Extended Datamart, Optum Pan-therapeutic HER, Truven MarketScan CCAE, Truven MarketScan MDCD, Truven MarketScan MDCR, Iqvia PhamMetrics, Iqvia Ambulatory HER, Iqvia LRx/Dx, Columbia / New York Presbyterian, Stanford STRIDE, Regenstrief, US Veterans Affairs VINCI

The total number of analyses is therefore A/1 x B/3 x C/2 x D/1 x E/? = ? analyses.

## Output

Covariate balance will be summarized in tabular form by showing the mean value for all baseline covariates in the target and comparator cohort, with the associated standardized mean difference computed for each covariate.

Once the propensity score model is fit, we will plot the propensity score distribution of the target and comparator cohorts to evaluate the comparability of the two cohorts. The plot will be scaled to the preference score, normalizing for any imbalance in cohort size. The covariates selected within the propensity score model, with associated coefficients will also be reported.

A plot showing the propensity score distributions for both cohorts after stratification will be provided, with each quantile cut point shown as a vertical line. Covariate balance will be evaluated by plotting the standardized mean difference of each covariate before propensity score stratification against the standardized mean difference for each covariate after propensity score stratification.

An attrition diagram will be provided to detail the loss of patients from the original target cohort and comparator cohort to the subpopulations that remain after all design considerations have been applied.

The final outcome model, a conditional Cox proportional hazards model, will be summarized by providing the hazards ratio and associated 95% confidence interval. The number of persons, amount of time-at-risk, and number of outcomes in each cohort will also be reported.

## Evidence Evaluation

We have executed diagnostics to determine if the analysis can be appropriately conducted. The diagnostics include:

* Propensity score distribution
* Covariate balance before and after propensity score matching
* Estimation for negative and positive controls, to assess residual error
* Negative and positive control exposures and outcomes will be used to evaluate the potential impact of residual systematic error in the study design, and to facilitate empirical calibration of the p-value and confidence interval for the exposures and outcome of interest.

Negative control outcomes in the context of this study are outcomes that are not believed to be caused by neither tofacitinib nor adalimumab or etanercept, and where therefore the true hazard ratio is equal to 1. We will execute the same analysis used for the primary hypothesis to produce hazard ratio estimates for the negative controls. The distribution of effect estimates across all negative controls will be used to fit an empirical null distribution which models the observed residual systematic error. The empirical null distribution will then be applied to the target exposures and outcome of interest to calibrate the p-value [[11](#_ENREF_11)].

Positive control exposures and outcomes are pairs of exposures and outcomes where the hazard ratio is known to be of some magnitude greater than 1. We will synthesize positive controls by starting with the negative controls defined earlier, and adding additional, simulated outcomes during the time-at-risk until the desired true hazard ratio is achieved. The target hazard ratios are 1.5, 2 and 4. The negative and positive controls together will be used to estimate an empirical systematic error model, which will inform whether systematic error changes as a function of true effect size. The empirical systematic error model will then be applied to the target the target exposures and outcome of interest to calibrate the confidence interval [[10](#_ENREF_10)].

Empirical calibration serves as an important diagnostic tool to evaluate if the residual systematic error is sufficient to cast doubt on the accuracy of the unknown effect estimate. The calibration effect plot and calibration probability plots will be generated for review. We will report the traditional and empirically calibrated p-value and confidence interval for each negative control, as well as the hypothesis of interest.

# Study Diagnostics

## Sample Size and Study Power

## Cohort Comparability

## Systematic Error Assessment

# Strengths and Limitations of the Research Methods

Strength

* Cohort studies allow direct estimation of incidence rates following exposure of interest, and the new-user design can capture early events following treatment exposures while avoiding confounding from previous treatment effects. New use allows for a clear exposure index date.
* PS matching allow balancing on a large number of baseline potential confounders.
* Use of negative and positive control outcomes allow for evaluating the study design as a whole in terms of residual bias.

Limitations

* Even though many potential confounders will be included in this study, there may be residual bias due to unmeasured or misspecified confounders.

# Protection of Human Subjects

The use of the Optum Extended DoD database were reviewed by the New England Institutional Review Board (IRB) and was determined to be exempt from broad IRB approval, as this research project did not involve human subjects research.

# Management and Reporting of Adverse Events and Adverse Reactions

This study uses coded data that already exist in an electronic database. In this type of database, it is not possible to link (i.e., identify a potential causal association between) a particular product and medical event for any individual. Thus, the minimum criteria for reporting an adverse event (i.e., identifiable patient, identifiable reporter, a suspect product, and event) are not available and adverse events are not reportable as individual adverse events reports. The study results will be assessed for medically important results.

# Plans for Disseminating and Communicating Study Results

The study results will be posted on the OHDSI website after completion of the study. At least one paper describing the study and its results will be written and submitted for publication to a peer-reviewed scientific journal. Currently, the goal is to publish in the New England Journal of Medicine.

# Appendix 1

## Concepts excluded from covariates

These concepts and their descendants are excluded when creating covariates:

|  |  |  |
| --- | --- | --- |
| Concept ID | Concept Name |  |

## Negative control outcomes

|  |  |
| --- | --- |
| Concept ID | outcomeName |

# References