COVID EHDEN Use Case Protocol

Github Repository: https://github.com/ohdsi-studies/EhdenCovidUseCase

List of Abbreviations

CDM	Common Data Model
COVID-19	Coronavirus disease 2019
EHDEN	European Health Data and Evidence Network
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership

Responsible Parties

Responsible party/person	Institution/Affiliation
Ravinder Claire	National Institute for Health and Care Excellence, Manchester, M1 4BT United Kingdom
Dalia Dawoud	National Institute for Health and Care Excellence, Manchester, M1 4BT United Kingdom
Thomas Debray	Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Heidelberglaan 100, Utrecht, 3515GB, Netherlands
Jamie Elvidge	National Institute for Health and Care Excellence, Manchester, M1 4BT United Kingdom
Christina Read	Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Heidelberglaan 100, Utrecht, 3515GB, Netherlands

Rationale and Background

The use of a common data model (CDM) can improve observational healthcare data research by allowing standardised analyses to be conducted across databases. The European Health Data and Evidence Network (EHDEN) provides federated access to databases across Europe mapped to the Observational Medical Outcomes Partnership (OMOP) CDM. It has been used for a number of COVID-19 studies, including Project SCYLLA (SARS-Cov-2 Large-scale Longitudinal Analyses on the comparative safety and effectiveness of treatments under evaluation for COVID-19 across an international observational data network) (1).

This study builds upon the results of SCYLLA to use real-world observational data to compare intervention effects for Aspirin vs Heparin and Baricitinib vs Tocilizumab. Each of these comparisons are recommended treatments for COVID-19 patients at similar time periods in the treatment pathway (2). The study will combine observational results with the results of randomised studies to evaluate treatment effects.

Research Questions and Objective

The aim of this study is to estimate treatment effects for COVID-19 treatments using data from the EHDEN network and to combine these observational results with data from randomised studies.

Primary objectives:

1. To assess comparative effectiveness and safety among Tocilizumab and Baricitinib in hospitalised patients
2. To assess comparative effectiveness and safety among Aspirin and Heparin in hospitalised patients

Subgroup Analyses:

3. Effectiveness and safety in ICU patients
4. Effectiveness and safety in patients receiving corticosteroids
5. Effectiveness and safety in patients receiving oxygen

Research Methods

Study Design

This study will be a multinational, multi-database network comparative cohort study. It will use new user comparative study cohorts to compare outcomes between cohort groups treated with Tocilumizab versus Baricitinib and Aspirin versus Heparin and combine these results with published randomised control trials.

Setting

Participants from organisations in the EHDEN network will be included. Electronic health records (EHR) and administrative claims from primary care and secondary care will be utilised. Data will be linked to additional data sources including biobanks (e.g. UK Biobank), laboratory test data (e.g. Catalan central registry of COVID-19 PCR tests), national audits (e.g. UK National Audit of Intensive Care Units), and other relevant data sources where relevant and possible. The study will be conducted using data from multiple real world data (RWD) sources previously mapped to the OMOP CDM in collaboration with the OHDSI and European Health Data and Evidence Network (EHDEN) initiatives.

Study Period

The study period, when index events and outcomes of interest can be observed, will start from 01/01/2020 and end at the latest available date for all data sources. This study period is based on the SCYLLA study.

Study Population (Inclusion/Exclusion Criteria)

Participants will be identified using pre-specified concept sets implemented in the SCYLLA study. These concept sets were reviewed by a core team of clinicians, epidemiologists, vocabulary experts, and health data scientists with extensive expertise in the use of the OMOP CDM and the OHDSI tools.

New Drug User Study Cohorts:

Cohort criteria is taken from the SCYLLA study.

1) To assess comparative effectiveness and safety among treatments administered during hospitalisation:

   Study cohorts will be defined, to include subjects for all cohorts that, at time of initiation of an index treatment:

   - Are aged 18 or over at cohort entry
   - Have at least 365 days of continuous observation time prior to cohort entry
   - Have 0 prior exposures to index treatment in the 365 days prior to index
   - Have at least 1 COVID-19 diagnosis or positive test results in the 30 days prior to or on index
   - Are hospitalised on index, defined by an inpatient visit with an admission date in the 30 days prior to or on index and no corresponding discharge date prior to or on index

2) For subgroup analysis of ICU patients

   Study cohorts will be defined, to include subjects for all cohorts that, at time of initiation of an index treatment meet all above criteria and:

   - Have at least 1 intensive services (mechanical ventilation OR tracheostomy OR ECMO) in the 30 days prior to or on index

We will not attempt to construct an 'untreated' cohort, because of concerns of intractable confounding by indication and risk of immortal time bias that can be induced by looking across a hospital admission to classify exposure.

Follow Up

Index date is defined by the first prescription/dispensation of a treatment.

Two periods of follow-up will be considered for all outcomes:

- In a fixed 30-day time-at-risk analysis, the analysis follow-up starts 1 day after therapy initiation and continues up until the first of: 30 days after therapy initiation, death, or end of observation period.
- In an on-treatment analysis, the analysis follow-up starts 1 day after therapy initiation and continues until the first of: discontinuation of treatment, death, or end of observation period.

Variables

Exposures

Each treatment will be defined based on standard concepts in the OMOP Standardised Vocabularies. The analysis package will provide the final cohort definitions, conceptsets and associated source codes.

Exposure assessment:

Exposure to a treatment will commence on the date of the first qualifying record, subject to satisfying all inclusion criteria. Each drug exposure record has a start date

and inferred end date, which is either explicitly entered or derived from other available information, such as days supply or refills. Exposure will be inferred to continue until treatment discontinuation, defined as the end date following the last of one or more records for the treatment that occur within a persistent window. A persistence window of 7 days between drug utilisation records for each study drug will be allowed considered as continuous exposure. We will not perform analyses of dose.

Outcomes

30-day all-cause mortality will be assessed as a primary outcome and length of stay will be assessed as a secondary outcome. Safety outcomes will include bleeding, respiratory tract infection, venous thromboembolic (pulmonary embolism and deep vein thrombosis) events, and total cardiovascular disease events. The analysis package will provide the final cohort definitions, concept sets and associated source codes. Safety outcomes were selected to align with RCTs.

Negative Controls

We will use a sample of negative controls as a study diagnostic to evaluate and calibrate for residual systematic error in each analysis strategy applied to each database. Negative controls are exposure-outcome pairs for which there is no expected causal relationship, such that unbiased analyses can be expected to generate effect estimates consistent with relative risk = 1. Candidate negative controls will be identified empirically through characterization of data partners, and evaluated for lack of causal relationship through clinical review.

Covariates

Baseline covariates will be defined by observations prior to the index date. Specific pre-index characteristics to use for confounding adjustment include:

- Age = year (cohort start date) – year of birth
- Biologic sex
- Race
- Ethnicity
- Index year
- Index month
- Condition groups (SNOMED + descendants), >=1 occurrence during the 30 days prior to 1 day prior to index
- Condition groups (SNOMED + descendants), >=1 occurrence during the 365 days prior to 1 day prior to index
- Drug era groups (ATC/RxNorm + descendants), >=1 day during the interval of 30 day prior to 1 day prior to index which overlaps with at least 1 drug era
- Drug era groups (ATC/RxNorm + descendants), >=1 day during the interval of safety outcomes365 day prior to 1 day prior to index which overlaps with at least 1 drug era
- Procedures, >=1 occurrence during the 30 days prior to 1 day prior to index

- Procedures, >=1 occurrence during the 365 days prior to 1 day prior to index
- Measurements (values and range groups) during the 30 days prior to 1 day prior to index
- Measurements (values and range groups) during the 365 days prior to 1 day prior to index
- Observations during the 30 days prior to 1 day prior to index
- Observations during the 365 days prior to 1 day prior to index
- Devices during the 30 days prior to 1 day prior to index
- Devices during the 365 days prior to 1 day prior to index

Data Sources

This study will be conducted using routinely collected data from different data sources that participate in the OHDSI or EHDEN initiatives. These databases will provide representative clinical information as collected in actual routine practice conditions in different European, North America, and Asia-Pacific healthcare settings, including electronic health record and administrative claims data from primary and secondary care. All interested researchers with access to data sources with qualifying patient-level data converted to OMOP CDM v5.3 or above will be encouraged to participate and selection will be made for only data partners with records involving relevant treatment comparisons.

Study Size

Since this study will be undertaken using routinely collected data, all patients meeting the eligibility criteria above will be included. No a priori sample size calculation was performed; instead, a minimum detectable rate ratio (MDRR) will be estimated for each outcome analysis in each of the available databases. All exposure cohorts must have greater than 5 patients for outcome analyses to be performed. Analyses will not produce effect estimates if there are no outcomes in either arm, otherwise will produce estimates even if unpowered.

Data Management

All data extraction and curation will be conducted using the ATLASTM tool, an open access software generated by the OHDSI community, as well as the OHDSI Methods Library, a set of R packages developed and maintained by the OHDSI community. (https://ohdsi.github.io/MethodsLibrary)

The process will follow the steps described here:

1. Identification of the study populations
2. Identification of the comparator and treatment cohorts
3. Identification of the different outcome cohorts
4. Review of cohort diagnostics including age and sex-specific incidence rates for face validity

The different study cohorts will be identified using cohort definitions for the SCYLLA study.

Cohort definitions will be exported from ATLAS, packaged in study R packages relying on the OHDSI Methods Library, and shared with each of the data partners for a consistent extraction and curation of the population, exposures and outcomes of interest.

Data Analysis

Comparative Cohort Analysis

This study will follow the design established by the SCYLLA study and will be conducted after appropriate diagnostics to rule out power and/or confounding issues.

For each research question, we will make a pairwise comparison of each treatment. We will estimate and compare incidence and time-to-event analyses for each outcome during the time-at-risk windows when possible.

Only the first exposure per person will be included. We will require 365 days as the minimum continuous observation time prior to index date for a person to be included in the cohort. If persons qualify for both the Tocilumizab and Baricitinib or both the Aspirin and Heparin cohorts, they will be only included in the first cohort they qualify for, and their time-at-risk be censored when the new time-at-risk starts to prevent overlap. We will restrict the analysis to the period when both exposures are observed. We remove persons that have the outcome prior to the risk window start.

Two alternative outcome models will be used for each outcome-'time at risk' combination:

1) Logistic regression to estimate the incidence odds ratio based on the proportion of persons observed to experience the outcome during the time-at-risk
2) Cox proportional hazards model to estimate the hazard ratio based on the time-to-event, in those databases where outcome dates are accurately captured.

For each outcome model, patients will be excluded from the cohorts if they have previously experienced the outcome in 30 days prior to index.

The primary confounding adjustment strategy that will be employed is propensity score adjustment through large-scale modelling, using demographics and all condition and drug groupings in the 30 days and 365 days pre-index intervals as baseline covariates. A similar approach has been recently used to test the comparative effectiveness and safety of antihypertensives. The propensity score will not be trimmed.

Propensity scores will be used for exposure 1:1 matching (using a caliper = 0.2 * standardised logit) with unconditional outcome modelling to estimate the ATE.

Study diagnostics will be applied blinded to outcome results to all analysis variants to evaluate confidence in the results. These diagnostics will include:

- Empirical equipoise: % of persons in target and comparator with preference score (scaled propensity score) between 0.25 and 0.75: passed diagnostic if equipoise >= 50%
- Covariate balance: Standardized mean difference (SMD) for all pre-index covariates after adjustment; passed diagnostic if SMD for all pre-index covariates <= 0.1
- Negative control calibration: % of negative controls with statistically significant estimates ($p<0.05$); passed diagnostics if <=5%
- Empirical null distribution: the mean and standard deviation for the error distribution, as estimated from the negative control sample; pass diagnostic if abs(mean) < 0.1

Additional diagnostics may be considered.

If diagnostics do not pass or there is otherwise concern about residual systematic error, then post-hoc sensitivity analyses that employ alternative confounding adjustment strategies may be considered. All results will be publicly disclosed, including estimates where study diagnostics failed, but analyses with failed diagnostics will be noted as such.

All comparative cohort analyses we will rely on the CohortMethod package (https://ohdsi.github.io/CohortMethod/).

All analysis code will be completed and version controlled at https://github.com/ohdsi-studies prior to unblinding estimation results. All study diagnostics will be available for exploration at https://data.ohdsi.org/.

Meta-analysis of non-randomised evidence

All the proposed analyses will be conducted for each database separately, with estimates combined in random effects meta-analysis with Sidik-Jonkman Hartung-Knapp adjustment and restricted maximum likelihood estimation.

Network Meta-analysis of the randomised evidence

Treatment effect estimates from different randomised control trials will be combined in random effects meta-analysis with Sidik-Jonkman Hartung-Knapp adjustment and restricted maximum likelihood estimation.

Network Meta-analysis of the randomised and non-randomized evidence

Treatment effect estimates from randomised control trials and analyses from each database will be combined in random effects meta-analysis with Sidik-Jonkman Hartung-Knapp adjustment and restricted maximum likelihood estimation.

Limitations of the Research Methods

Selection bias

Selection bias might arise as the consequence of including subjects with a specific period of prior observation time available in the data. Attrition tables will be provided to report on the impact of such exclusion criteria.

Information bias

Information bias may occur due to the incorrect identification of exposure, outcomes or covariates. With regards to exposure, misclassification may occur due to the patient not fulfilling the prescription (primary non-adherence) or in relation with non-compliance. Hence an overestimate of utilization of the study drugs can happen, expectedly leading to non-differential misclassification. Experimental treatments used during randomized trials may not be captured in source data, which would result in exposure misclassification. In addition, lack or incomplete recording of safety events may lead to misclassification of the proposed safety endpoints.

Confounding

Confounding may occur if there are differences in (observed or unmeasured) baseline characteristics between the comparator and target cohorts which are also associated with outcome. Analysis strategies outlined in 9.7 offer approaches to identify and adjust for confounding, but may be insufficient if there is model misspecification or insufficient sample and/or data to precisely ascertain baseline covariates or to ascertain some factors at all, which can result in residual confounding and some remaining confounding bias. There may be unmeasured confounding associated with COVID severity.

Protocol for Addressing Study Failures

Because of the nature of this study, there is likely to be some proportion of sites with zero cases or failed analyses. We therefore here pre-specify our approach to address study failures, while still conforming to the LEGEND principles

1) On discovery that more than two sites fail to recover cases despite knowledge that such cases exist in the databases, the sites will investigate the source of the case loss and report back centrally for the phenotyping group to reassess the phenotype definition and adjust the common central definition, asking all sites to rerun the study on the new phenotypes. Site-specific phenotype definitions will not be permitted for this estimation study.
2) If a site fails to produce estimates due to low sample size and cases are accumulating at the site, then a reanalysis will be permitted with a 50% increase in the number of cases.
3) If a particular analytic method fails to produce estimates for the majority of sites, that method will be pulled from the analysis
4) If almost all analytic methods fail to produce estimates for almost all hypotheses at almost all sites, then the analytic methods will be redesigned centrally for adjustment to the study protocol and a restart of the study

Protection of Human Subjects

For this study, participants from numerous healthcare databases will be studied. The use of the OMOP common data model and OHDSI tools will enable the federated analysis of these different databases without changing access rights to patient-level data. Each data partner is required to provide a statement about IRB approval or exemption to participate.

Management and Reporting of Adverse Events/Adverse Reactions

According to the new guidelines for good pharmacovigilance practice (EMA/873138/2011) there is no requirement for expedited reporting of adverse drug reactions from studies with secondary use of data (such as electronic health care databases). All the identified adverse events/reactions will be summarised in the resulting manuscript/s and/or interactive web-based report of all conducted analyses

Plans For Disseminating And Communicating Study Results

Dissemination activities will be of a scientific nature (publish articles, etc). Our aim is for these studies to be made available in order to support treatment decisions in the global COVID-19 pandemic.

Updates to Protocol

5/8/2022: Covariates list modified to include all default OHDSI covariates.

References

(1)
https://www.ohdsi.org/wp-content/uploads/2020/05/OHDSI-SCYLLA-PLE-COVID-effectiveness-protocol-v1.0-final-with-appendix.pdf
(2) NICE. COVID-19 rapid guideline: critical care in adults. NICE guideline [NG159]. 2020.