# OUR RESEARCH

EHDEN Zenodo

tinyurl.com/t688a8tu

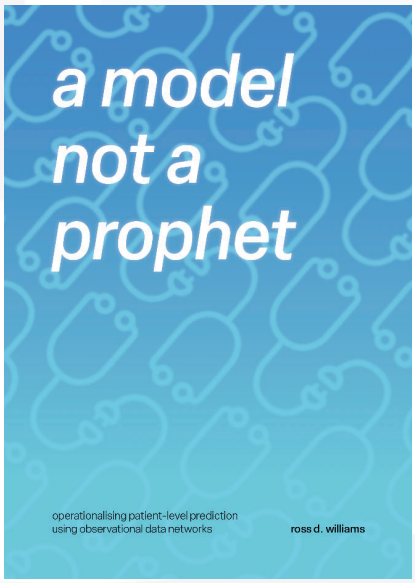
From Day 1, EHDEN has been increasingly focused on delivering results against the second “E” in EHDEN: Evidence. Thanks to the creation of its federated network 187 Data Partners from 29 countries across Europe, the transformational potential of real-world data harmonised to the OMOP Common Data Model is being realised. This network is not only generating real-world evidence that is improving patient treatment and outcomes, and is doing this in days, weeks and months, instead of years. To date, more than seventy publications and many other presentations in scientific conferences covering a broad range of methodological and clinical research questions have been disseminated in many of the world’s leading scientific journals and congresses.

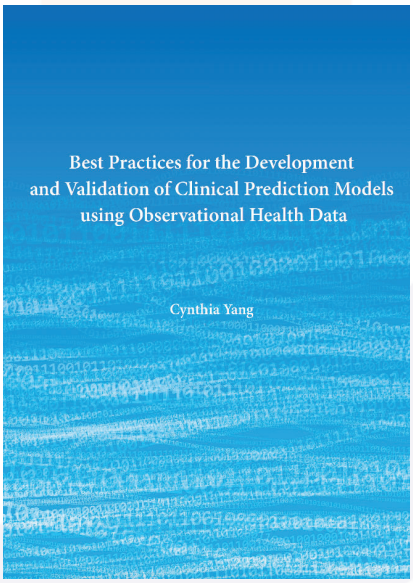
Some examples of our methodological and clinical research are included here. The up-to-date listing can be found on the EHDEN Zenodo page.

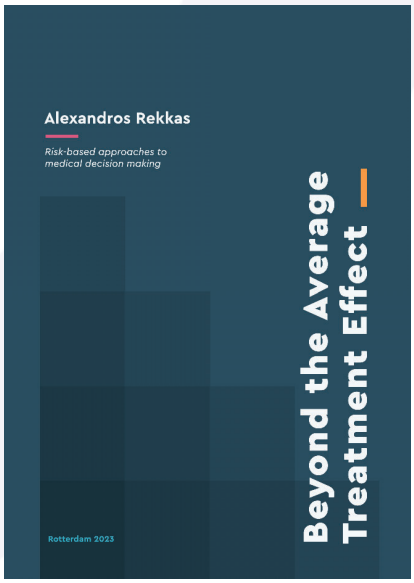
## Our Methodological Research

To perform methodological research at scale we can maximally benefit from the standardisation of the data via the OMOP CDM and the use of common analytics that can be parameterised.

This work in EHDEN had a significant impact on the careers of young scientists, three of whom finished their EHDEN-funded PhD trajectories, and four are close to finalising their thesis.



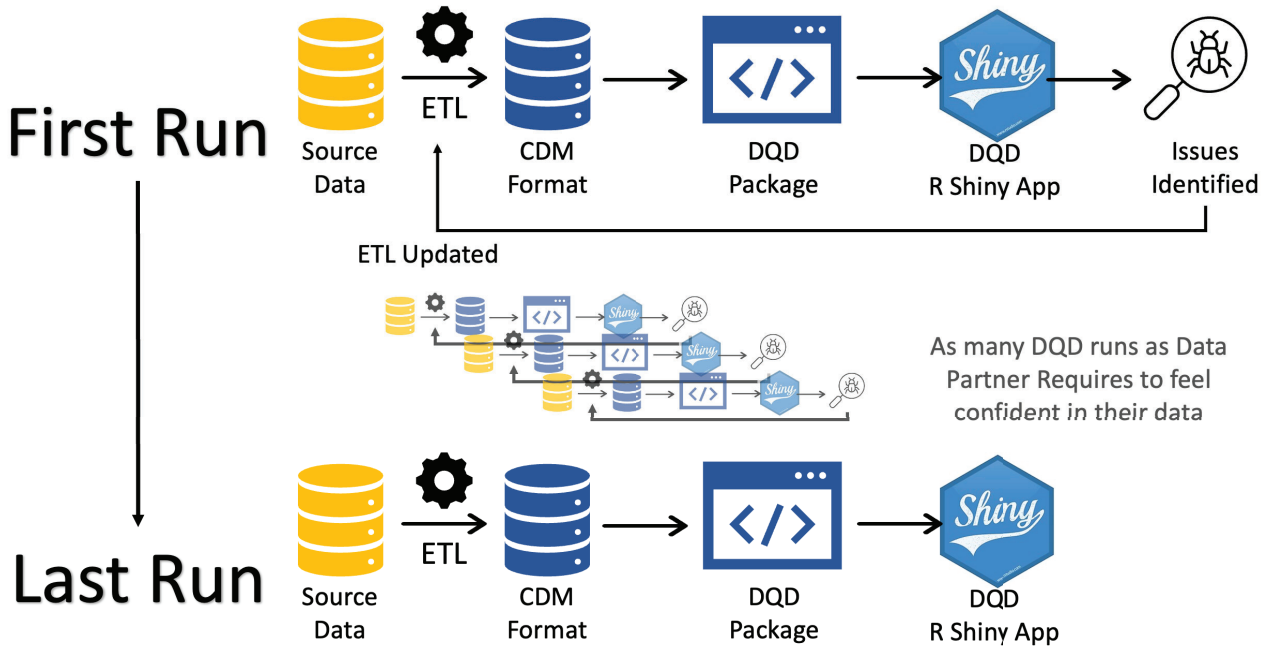

Examples of EHDEN-funded PhD trajectories at Erasmus MC, Rotterdam

## Quality Control Mechanisms

Clair Blacketer, Erasmus MC and Johnson & Johnson

Blacketer C, et al. **Using the Data Quality Dashboard to Improve the EHDEN Network.** Applied Sciences. 2021; 11(24):11920. https://doi.org/10.3390/app112411920

Federated networks of observational health databases have the potential to be a rich resource to inform clinical practice and regulatory decision making. However, the lack of standard data quality processes makes it difficult to know if these data are research ready. The EHDEN COVID-19 Rapid Collaboration Call presented the opportunity to assess how the newly developed open-source tool Data Quality Dashboard (DQD) informs the quality of data in a federated network. Fifteen Data Partners (DPs) from 10 different countries worked with the EHDEN taskforce to map their data to the OMOP CDM. Throughout the process at least two DQD results were collected and compared for each DP. All DPs showed an improvement in their data quality between the first and last run of the DQD. The DQD excelled at helping DPs identify and fix conformance issues but showed less of an impact on completeness and plausibility checks. This is the first study to apply the DQD on multiple, disparate databases across a network. While study-specific checks should still be run, we recommend that all data holders converting their data to the OMOP CDM use the DQD as it ensures conformance to the model specifications and that a database meets a baseline level of completeness and plausibility for use in research.




The Extract, Transform, & Load (ETL) to Common Data Model (CDM) to Data Quality Dashboard (DQD) Feedback Loop used in the European Health Data & Evidence Network (EHDEN) COVID-19 Rapid Collaboration Call.

## ETL Best Practices

Erica Voss, Erasmus MC and Johnson & Johnson

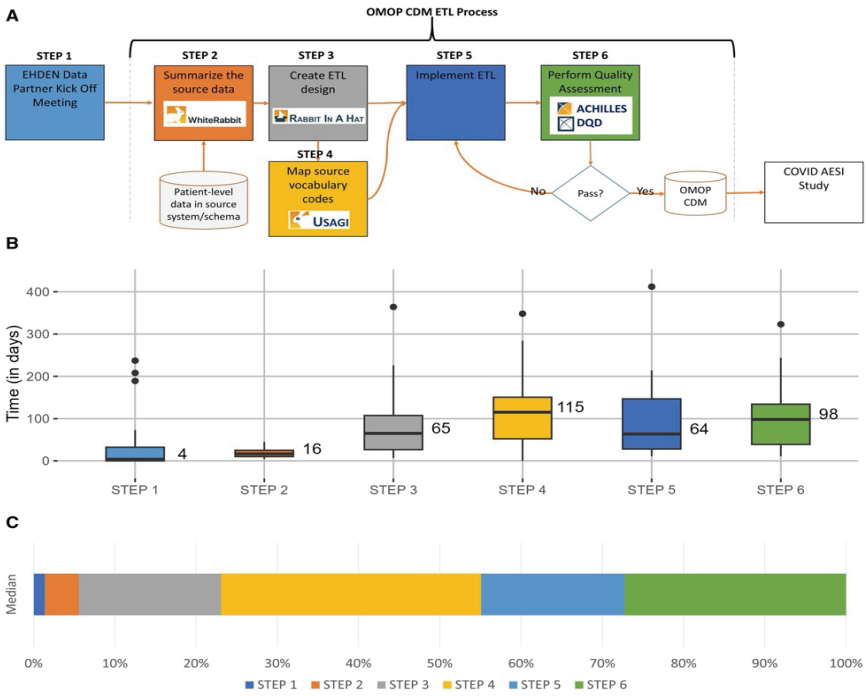
Erica A Voss, et al. **European Health Data & Evidence Network—learnings from building out a standardized international health data network**, Journal of the American Medical Informatics Association, Volume 31, Issue 1, January 2024, Pages 209–219

Health data standardised to a common data model (CDM) simplifies and facilitates research. This study examines the factors that make standardising observational health data to the Observational Medical Outcomes Partnership (OMOP) CDM successful.

Twenty-five data partners (DPs) from 11 countries received funding from the European Health Data Evidence Network (EHDEN) to standardise their data. Three surveys, DataQualityDashboard results, and statistics from the conversion process were analysed qualitatively and quantitatively. Our measures of success were the total number of days to transform source data into the OMOP CDM and participation in network research.

This study shows that the consistent workflow used by EHDEN proves appropriate to support the successful standardisation of observational data across Europe. Over the 25 successful transformations, we confirmed that getting the right people for the ETL is critical and vocabulary mapping requires specific expertise and support of tools. Additionally, we learned that teams that proactively prepared for data governance issues were able to avoid considerable delays improving their ability to finish on time.

This study provides guidance for future DPs to standardise to the OMOP CDM and participate in distributed networks. We demonstrated that the Observational Health Data Sciences and Informatics community must continue to evaluate and provide guidance and support for what ultimately develops the backbone of how community members generate evidence.



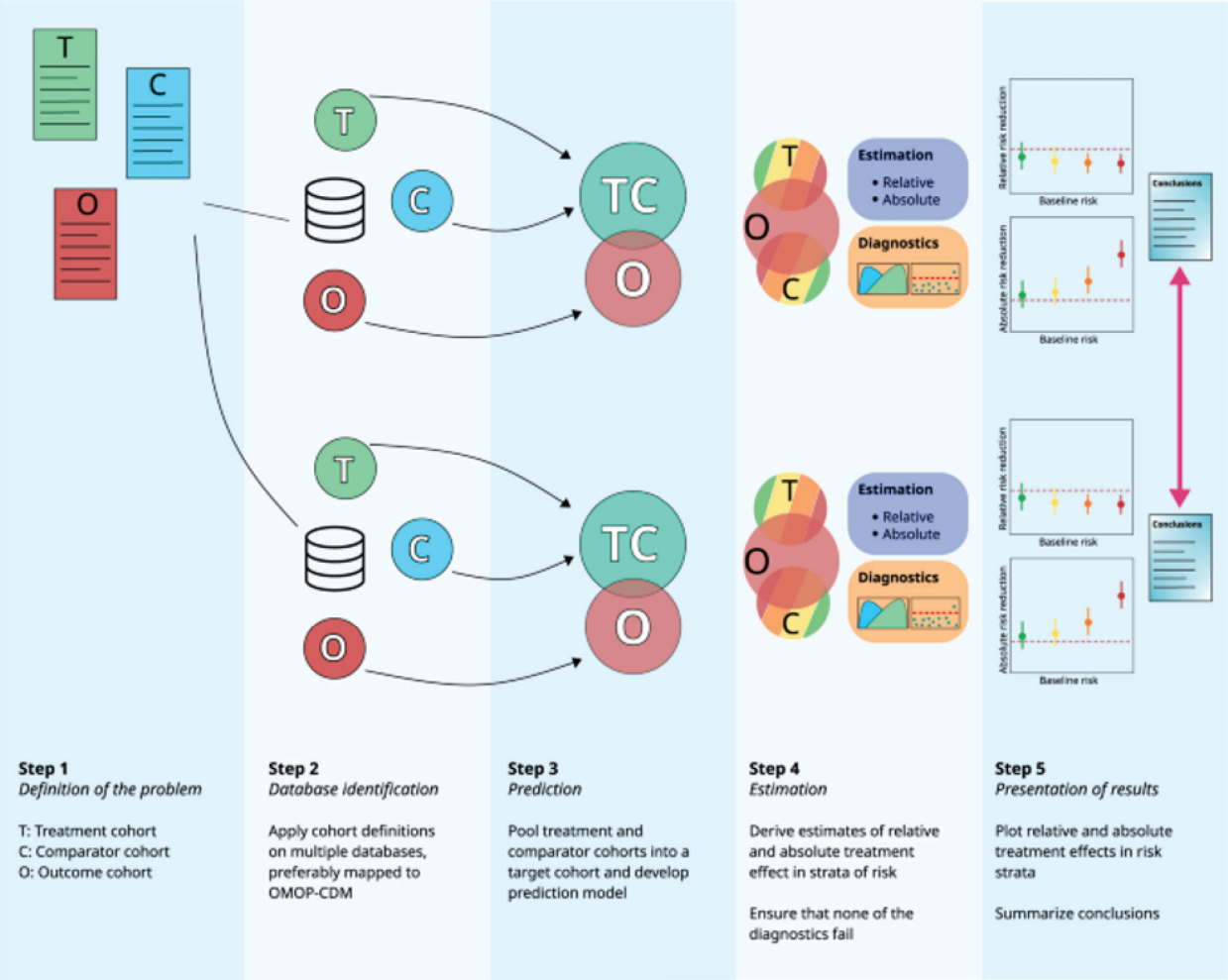

OMOP CDM ETL development process: (A) represents the ETL process map, (B) is a box plot of median length in days for each step across all Data Partners, and (C) is a stacked bar chart showing the percentage of median time each step took. CDM, common data model; COVID AESI Study, "Adverse Events of Special Interest within COVID-19 Subjects" study; DQD, DataQualityDashboard; EHDEN, European Health Data & Evidence Network; ETL, extract, transform, and load; OMOP, outcomes partnership common data model.

# Risk Stratified Effect Estimaton

Alexandros Rekkas, Erasmus MC

Rekkas, A., et al. **A standardized framework for risk-based assessment of treatment effect heterogeneity in observational healthcare databases.** npj Digit. Med. 6, 58 (2023). https://doi.org/10.1038/s41746-023-00794-y

Treatment effects are often anticipated to vary across groups of patients with different baseline risk. The Predictive Approaches to Treatment Effect Heterogeneity (PATH) statement focused on baseline risk as a robust predictor of treatment effect and provided guidance on risk-based assessment of treatment effect heterogeneity in a randomised controlled trial. The aim of this study is to extend this approach to the observational setting using a standardised scalable framework. The proposed framework consists of five steps: (1) definition of the research aim, i.e., the population, the treatment, the comparator and the outcome(s) of interest; (2) identification of relevant databases; (3) development of a prediction model for the outcome(s) of interest; (4) estimation of relative and absolute treatment effect within strata of predicted risk, after adjusting for observed confounding; (5) presentation of the results. We demonstrated our framework by evaluating heterogeneity of the effect of thiazide or thiazide-like diuretics versus angiotensin-converting enzyme inhibitors on three efficacy and nine safety outcomes across three observational databases. We provided a publicly available R software package for applying this framework to any database mapped to the OMOP CDM. In our demonstration, patients at low risk of acute myocardial infarction receive negligible absolute benefits for all three efficacy outcomes, though they are more pronounced in the highest risk group, especially for acute myocardial infarction. Our framework allows for the evaluation of differential treatment effects across risk strata, which offers the opportunity to consider the benefit-harm trade-off between alternative treatments.




# Disease trajectories

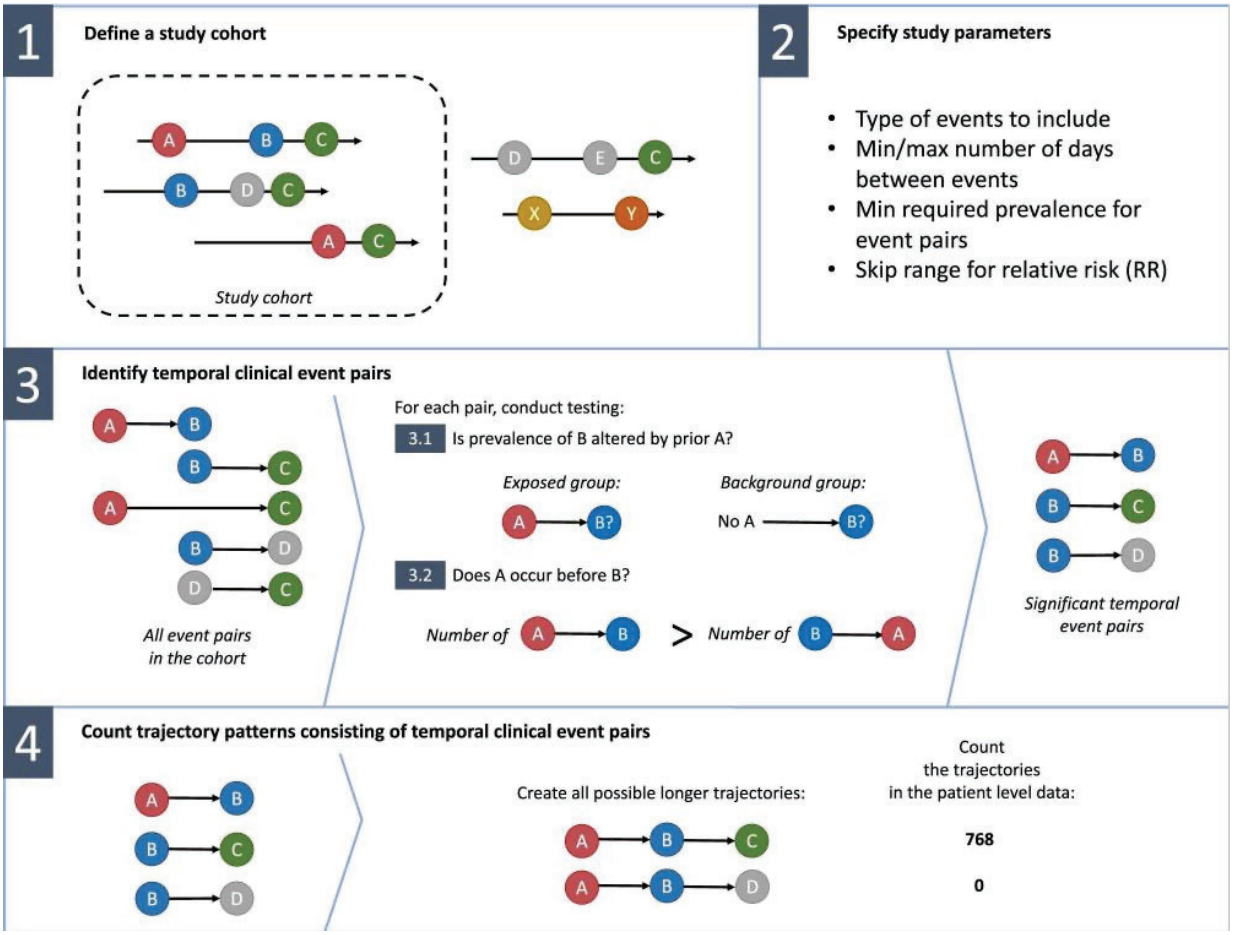
Sulev Reisberg, OHDSI Estonia Lead, Tartu University

Künnapuu K, Ioannou S, Ligi K, Kolde R, Laur S, Vilo J, Rijnbeek PR, Reisberg S. **Trajectories: a framework for detecting temporal clinical event sequences from health data standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model.** JAMIA Open. 2022 Mar 16;5(1):ooac021. doi: 10.1093/jamiaopen/ooac021. PMID: 35571357; PMCID: PMC9097714.

Patient journeys/trajectories – sequences of health events, including diseases, procedures, and visits, which patients follow – have gained more and more attention during recent years. Detecting the sequences of health events may help us to better understand disease aetiology (what happened before the disease) and predict events for the future (what happened after). It can also reveal and enable how to better analyse different treatment options that are used in practice for the same diseases.

Observational health data is a great source for analysing such trajectories. Visits, diagnoses, lab analyses, drugs/prescriptions, etc. are all common elements in most of these datasets. They are also the key elements of the OMOP CDM. This makes it extremely useful to run health event trajectory analyses also on OMOP CDM data, as the same analysis could be easily run on various databases.

In the EHDEN project, we have performed methods research on disease trajectories and a four-step framework based on significant temporal event pair detection is described and implemented as an open-source R package. It is used on a population-based Estonian dataset to first replicate a large Danish population-based study and second, to conduct a disease trajectory detection study for type 2 diabetes patients in the Estonian and Dutch databases as an example.



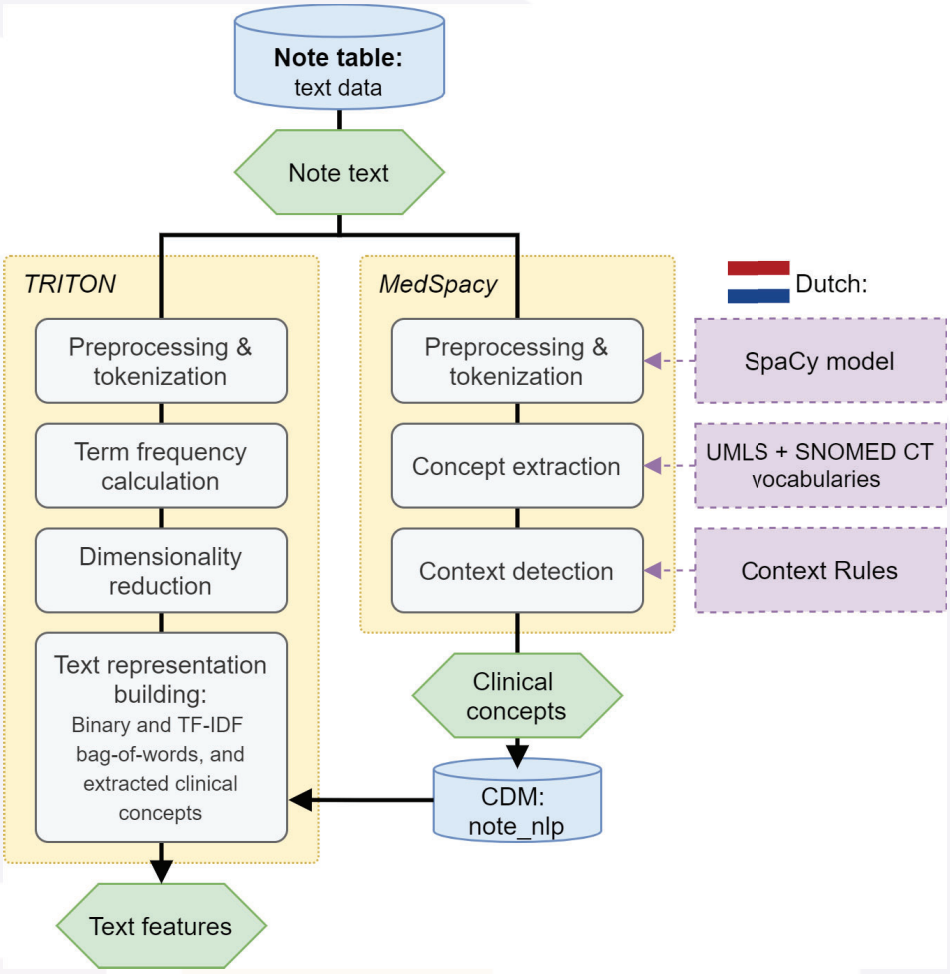

# Natural Language Processing

Tom Seinen, Erasmus MC

Seinen TM, Kors JA, van Mulligen EM, Fridgeirsson E, Rijnbeek PR, 2023. **The added value of text from Dutch general practitioner notes in predictive modeling.** Journal of the American Medical Informatics Association. https://doi.org/10.1093/jamia/ocad160

Electronic health record (EHR) databases are a rich source of data for building patient level prediction models. Currently, most prediction models use only the structured data in the EHR, such as coded conditions, measurements, vital signs, and drug prescriptions, as features. However, EHRs commonly also store vast amounts of unstructured textual data (e.g., physicians' and nurses' notes and discharge letters).

Using natural language processing (NLP) methods, the information hidden in the unstructured clinical text can be extracted and incorporated in PLP models. We developed a standardised NLP pipeline tool, within the OHDSI framework, for extracting textual features in a data-driven and language-independent manner. This tool extends the FeatureExtraction framework in the form of a custom covariate builder and constructs a set of text-based covariates. The tool contains a modular NLP pipeline for the pre-processing, tokenisation and vectorisation of text documents, that can be fully customised to specific needs. The pipeline settings and customisations are saved with the result for sharing and reproducibility. The tool is called Text Represented In Terms Of Numeric-features (TRITON) and is now publicly available on GitHub at github.com/mi-erasmusmc/Triton.




Visualisation of the TRITON and MedSpacy pipelines to process raw text data from the note table in the OMOP CDM and generate text-based features for prediction models.

We used this package to explore the value of Dutch unstructured data, in combination with structured data, for the development of prognostic prediction models in a general practitioner (GP) setting. We trained and validated prediction models for four common clinical prediction problems using various sparse text representations, common prediction algorithms, and observational GP electronic health record (EHR) data. We trained and validated 84 models internally and externally on data from different EHR systems. Our study highlights the potential benefits of incorporating unstructured data in clinical prediction models in a GP setting. Although the added value of unstructured data may vary depending on the specific prediction task, our findings suggest that it has the potential to enhance patient care.

## Learning Curves

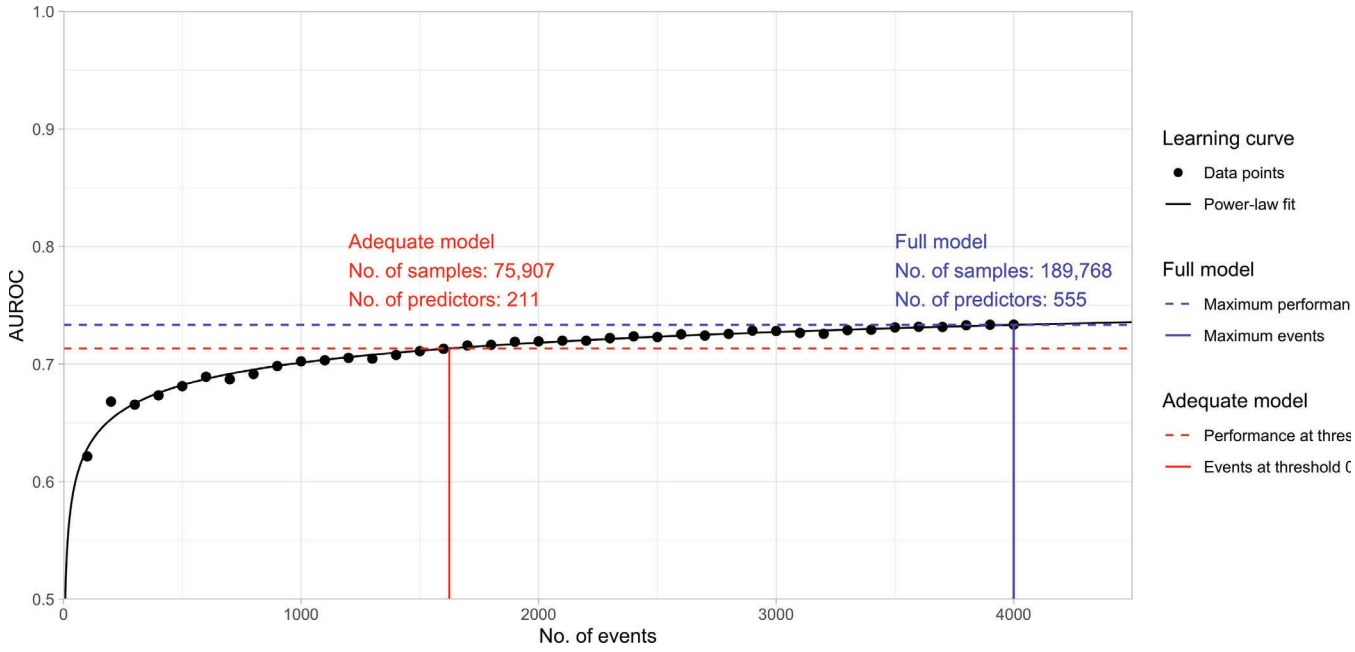
Henrik John, Erasmus MC

John LH, Kors JA, Reps JM, Ryan PB, Rijnbeek PR. **Logistic regression models for patient-level prediction based on massive observational data: Do we need all data?** Int J Med Inform. 2022 Jul;163:104762. doi: 10.1016/j.ijmedinf.2022.104762. Epub 2022 Apr 12. PMID: 35429722.

EHDEN's federated data network opens up possibilities to develop clinical prediction models on massive amounts of patient data which can serve large patient populations in a timely manner. In practice, this could manifest in the development of several hundreds or even thousands of prediction models for the various target-outcome pairs and for the many different databases that are currently being mapped to the OMOP CDM.

However, models developed on these large amounts of observational health data run the risk of being more complex than needed. These models can include many more features without achieving substantially better discrimination than smaller models. As a result, these models may become harder to interpret, more difficult to implement in clinical practice, and more susceptible to overfitting. In addition, developing prediction models on such large data sources can put strong demands on computing resources and may require computation times that can become prohibitive. Reducing the sample size of a large and unwieldy dataset to an "adequate" sample size that is still sufficient to achieve nearly the same performance as the full dataset, may facilitate the development of less complex clinical prediction models with less computing resources.

The objective of this study was to provide guidance on sample size considerations for developing predictive models by empirically establishing the adequate sample size, which balances the competing objectives of improving model performance and reducing model complexity as well as computational requirements. Our results suggest that in most cases only a fraction of the available data was sufficient to produce a model close to the performance of one developed on the full data set, but with a substantially reduced model complexity.




Learning curve for the prediction of venous thromboembolic events in patients with hypertension using data from MDCR. The horizontal lines indicate the maximum performance of the fitted curve (blue) and the performance at a threshold of 0.02 (red). The vertical lines denote the maximum number of events (blue) and the adequate number of events (red). Number of samples and predictors shown for the adequate model pertain to the model at the closest data point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

# Class Imbalance

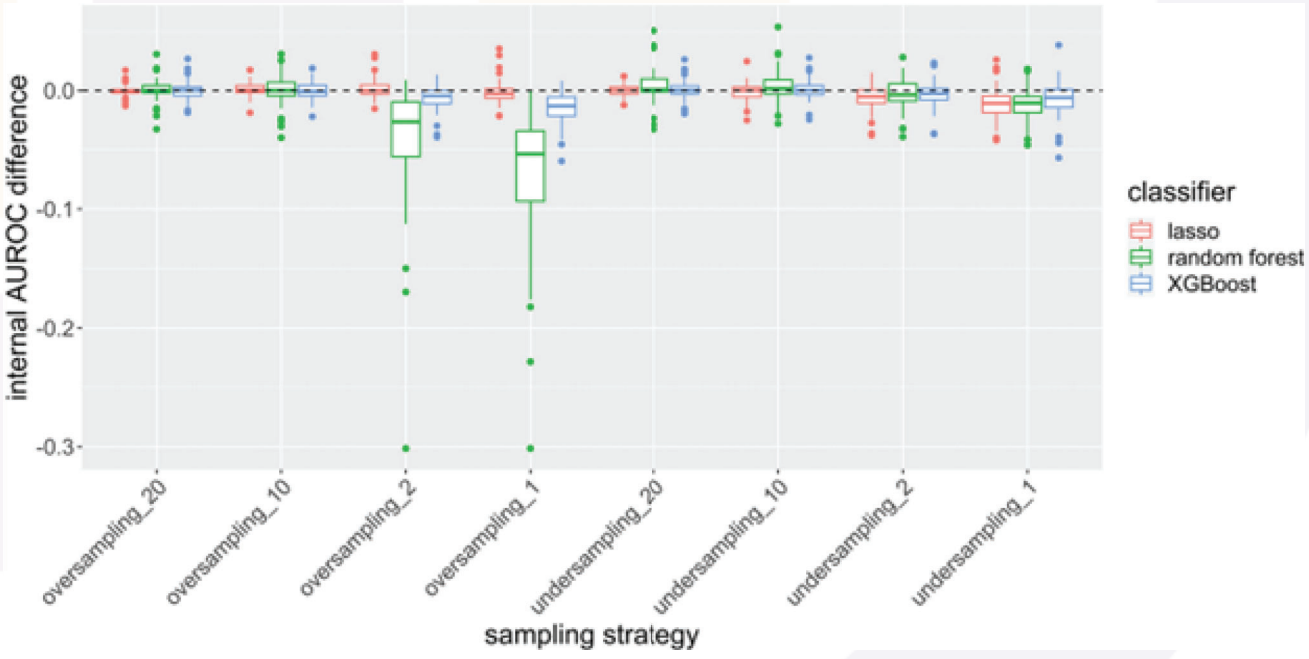
Cynthia Yang, Erasmus MC

Yang, C., Fridgeirsson, E.A., Kors, J.A. et al. **Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data.** J Big Data 11, 7 (2024). https://doi.org/10.1186/s40537-023-00857-7

There is currently no consensus on the impact of class imbalance methods on the performance of clinical prediction models. We aimed to empirically investigate the impact of random oversampling and random undersampling, two commonly used class imbalance methods, on the internal and external validation performance of prediction models developed using observational health data.

We developed and externally validated a total of 1,566 prediction models. On internal and external validation, random oversampling and random undersampling generally did not result in higher AUROCs. Moreover, we found overestimated risks, although this miscalibration could largely be corrected by recalibrating the models towards the imbalance ratios in the original dataset.

Overall, we found that random oversampling or random undersampling generally does not improve the internal and external validation performance of prediction models developed in large observational health databases. Based on our findings, we do not recommend applying random oversampling or random undersampling when developing prediction models in large observational health databases




Internal AUROC differences across all prediction problems and databases for each sampling strategy and classifier. A positive difference means original data model had a lower AUROC, and a negative difference means original data model had a higher AUROC.

# Treatment Patterns

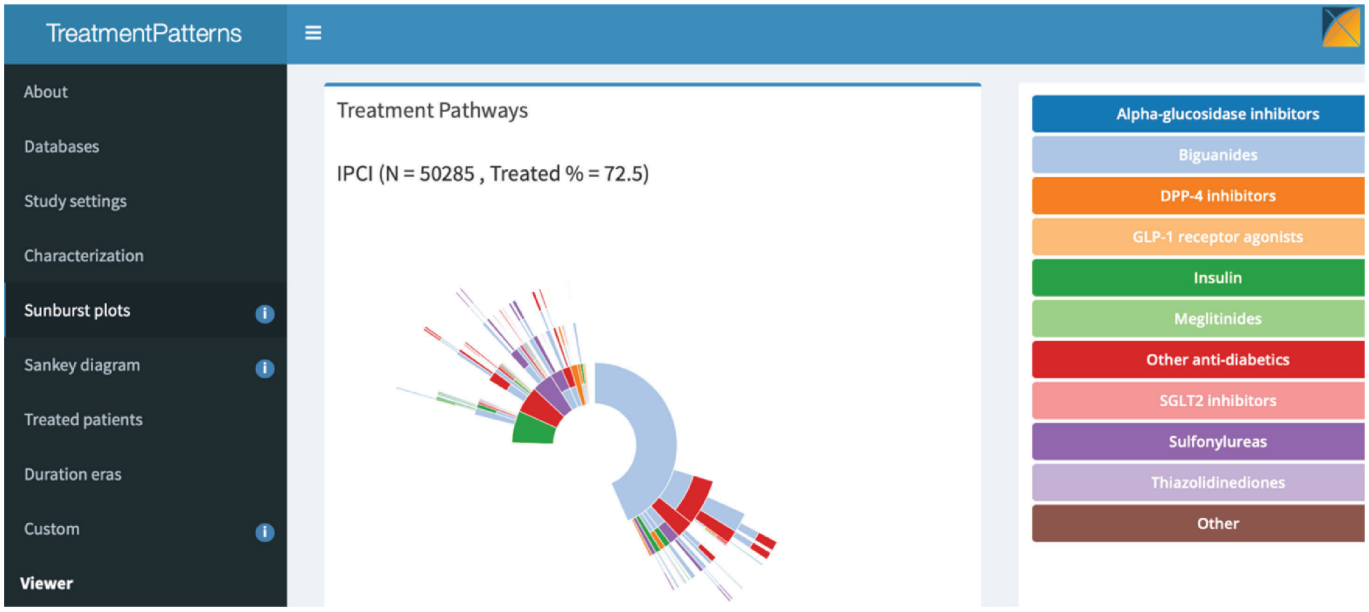
Aniek Markus, Erasmus MC

Markus, A. F., Verhamme, K. M., Kors, J. A., & Rijnbeek, P. R. **TreatmentPatterns: An R package to facilitate the standardized development and analysis of treatment patterns across disease domains.** Computer Methods and Programs in Biomedicine (2022). https://doi.org/10.1016/j.cmpb.2022.107081

There is an increasing interest to use real-world data to illustrate how patients with specific medical conditions are treated in real life. Insight into the current treatment practices helps to improve and tailor patient care, but is often held back by a lack of data interoperability and a high-level of required resources. We aimed to provide an easy tool that overcomes these barriers to support the standardised development and analysis of treatment patterns for a wide variety of medical conditions.

We formally defined the process of constructing treatment pathways and implemented this in an open-source R package TreatmentPatterns (github.com/mi-erasmusmc/TreatmentPatterns) to enable a reproducible and timely analysis of treatment patterns.

TreatmentPatterns is a tool to make the analysis of treatment patterns more accessible, more standardised, and more interpretation-friendly. We hope it thereby contributes to the accumulation of knowledge on real-world treatment patterns across disease domains. We encourage researchers to further adjust and add custom analysis to the R package based on their research needs.




Sunburst plot visualising the treatment pathways of patients with type II diabetes mellitus in the Integrated Primary Care Information (IPCI) database.

# Iterative Pairwise External Validation

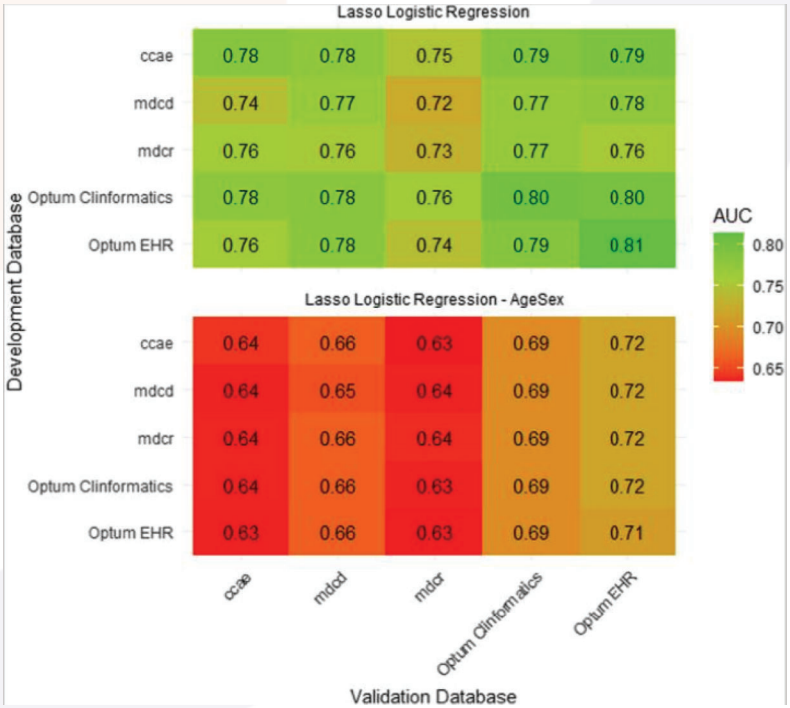
Ross Williams, Erasmus MC

Williams, R.D., Reps, J.M., Kors, J.A. et al. **Using Iterative Pairwise External Validation to Contextualize Prediction Model Performance: A Use Case Predicting 1-Year Heart Failure Risk in Patients with Diabetes Across Five Data Sources.** Drug Saf 45, 563–570 (2022).

External validation of prediction models is increasingly being seen as a minimum requirement for acceptance in clinical practice. However, the lack of interoperability of healthcare databases has been the biggest barrier to this occurring on a large scale. Recent improvements in database interoperability enable a standardised analytical framework for model development and external validation. External validation of a model in a new database lacks context, whereby the external validation can be compared with a benchmark in this database. Iterative pairwise external validation (IPEV) is a framework that uses a rotating model development and validation approach to contextualise the assessment of performance across a network of databases. As a use case, we predicted 1-year risk of heart failure in patients with type 2 diabetes mellitus.

The method follows a two-step process involving (1) development of baseline and data-driven models in each database according to best practices and (2) validation of these models across the remaining databases. We introduced a heatmap visualisation that supports the assessment of the internal and external model performance in all available databases. As a use case, we developed and validated models to predict 1-year risk of heart failure in patients initialising a second pharmacological intervention for type 2 diabetes mellitus. We leveraged the power of the OMOP CDM to create an open-source software package to increase the consistency, speed, and transparency of this process.

Using IPEV lends weight to the model development process. The rotation of development through multiple databases provides context to model assessment, leading to improved understanding of transportability and generalisability. The inclusion of a baseline model in all modelling steps provides further context to the performance gains of increasing model complexity. The CCAE model was identified as a candidate for clinical use. The use case demonstrates that IPEV provides a huge opportunity in a new era of standardised data and analytics to improve insight into and trust in prediction models at an unprecedented scale.




A heatmap of the area under the concentration–time curve values across internal validation (values on the lead diagonal) and external validations of the developed prediction models. The colour scale runs from red (low discriminative ability) to green (high discriminative ability). The upper section details the performances for the data-driven model. The lower half details the same but then for the age and sex model. AUCarea under the concentration–time curve, ccae Commercial Claims and Encounters, EHR electronic health records, mdcd Medicaid, mdcrMedicare

# Calculating Daily Dose

Theresa Burkard, Oxford University

Burkard T, López-Güell K, Gorbachev A, et al. **Calculating daily dose in the Observational Medical Outcomes Partnership Common Data Model.** Pharmacoepidemiol Drug Saf. 2024 Jun;33(6):e5809. doi: 10.1002/pds.5809. PMID: 38773798.

In this methodological work, we aimed to develop a standardised method to calculate daily dose (i.e., the amount of drug a patient was exposed to per day) of any drug on a global scale using only drug information of typical observational data in the OMOP CDM and a single reference table from Observational Health Data Sciences And Informatics (OHDSI).

The OMOP DRUG_STRENGTH reference table contains information on the strength or concentration of drugs, whereas the OMOP DRUG_EXPOSURE table contains information on patients' drug prescriptions or dispensations/claims. Based on DRUG_EXPOSURE data from the primary care databases Clinical Practice Research Datalink GOLD (United Kingdom) and Integrated Primary Care Information (IPCI, The Netherlands) and healthcare claims from PharMetrics® Plus for Academics (USA), we developed four formulas to calculate daily dose given different DRUG_STRENGTH reference table information.

We tested the dose formulas by comparing the calculated median daily dose to the World Health Organization (WHO) Defined Daily Dose (DDD) for six different ingredients in those three databases and an additional four international databases representing a variety of healthcare settings: MAITT (Estonia, healthcare claims and discharge summaries), IQVIA Disease Analyzer Germany (outpatient data), IQVIA Longitudinal Patient Database Belgium (outpatient data), and IMASIS Parc Salut (Spain, hospital data). Finally, in each database, we assessed the proportion of drug records for which daily dose calculations were possible using the suggested formulas.

Applying the dose formulas, we obtained median daily doses that generally matched the WHO DDD definitions. Our dose formulas were applicable to >85% of drug records in all but one of the assessed databases.

We have established and implemented a standardised daily dose calculation in OMOP CDM providing reliable and reproducible results.

# Our Clinical Research

Find below some examples of published clinical studies from EHDEN researchers in collaboration with our Data Partners and external partners from the global OHDSI community.

The up-to-date listing can be found on the EHDEN Zenodo page.

Gauffin O, et al. **Supporting Pharmacovigilance Signal Validation and Prioritization with Analyses of Routinely Collected Health Data: Lessons Learned from an EHDEN Network Study.** Drug Saf. 2023 Dec;46(12):1335-1352. doi: 10.1007/s40264-023-01353-w. Epub 2023 Oct 7.

**Introduction**
Individual case reports are the main asset in pharmacovigilance signal management. Signal validation is the first stage after signal detection and aims to determine if there is sufficient evidence to justify further assessment. Throughout signal management, a prioritisation of signals is continually made. Routinely collected health data can provide relevant contextual information but are primarily used at a later stage in pharmacoepidemiological studies to assess communicated signals.

**Objective**
The aim of this study was to examine the feasibility and utility of analysing routine health data from a multinational distributed network to support signal validation and prioritisation and to reflect on key user requirements for these analyses to become an integral part of this process.

**Methods**
Statistical signal detection was performed in Vigi Base,the WHO global database of individual case safety reports, targeting generic manufacturer drugs and 16 prespecified adverse events. During a five-day study-a-thon, signal validation and prioritisation were performed using information from VigiBase, regulatory documents and the scientific literature alongside descriptive analyses of routine health data from ten EHDEN Data Partners . Databases included in the study were from the UK, Spain, Norway, the Netherlands and Serbia, capturing records from primary care and/or hospitals.

**Results**
Ninety-five statistical signals were subjected to signal validation, of which eight were considered for descriptive analyses in the routine health data. Design, execution and interpretation of results from these analyses took up to a few hours for each signal (of which 15–60 minutes were for execution) and informed decisions for five out of eight signals. The impact of insights from the routine health data varied and included possible alternative explanations, potential public health and clinical impact and feasibility of follow-up pharmacoepidemiological studies. Three signals were selected for signal assessment; two of these decisions were supported by insights from the routine health data. Standardisation of analytical code, availability of adverse event phenotypes including bridges between different source vocabularies, and governance around the access and use of routine health data were identified as important aspects for future development.

**Conclusions**
Analyses of routine health data from a distributed network to support signal validation and prioritisation are feasible in the given time limits and can inform decision-making. The cost–benefit of integrating these analyses at this stage of signal management requires further research.

Voss EA, et al. **Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study.** EClinicalMedicine. 2023 Apr;58:101932. doi: 10.1016/j.eclinm.2023.101932. Epub 2023 Apr 4.


**Background**
Adverse events of special interest (AESIs) were pre-specified to be monitored for the COVID-19 vaccines. Some AESIs are not only associated with the vaccines, but with COVID-19. Our aim was to characterise the incidence rates of AESIs following SARS-CoV-2 infection in patients and compare these to historical rates in the general population.

**Methods**
A multi-national cohort study with data from primary care, electronic health records, and insurance claims mapped to a common data model. This study's evidence was collected between Jan 1, 2017 and the conclusion of each database (which ranged from July 2020 to May 2022). The 16 prespecified prevalent AESIs were: acute myocardial infarction, anaphylaxis, appendicitis, Bell's palsy, deep vein thrombosis, disseminated intravascular coagulation, encephalomyelitis, Guillain- Barré syndrome, haemorrhagic stroke, non-haemorrhagic stroke, immune thrombocytopenia, myocarditis/pericarditis, narcolepsy, pulmonary embolism, transverse myelitis, and thrombosis with thrombocytopenia. Age-sex standardised incidence rate ratios (SIR) were estimated to compare post-COVID-19 to pre-pandemic rates in each of the databases.

**Findings**
Substantial heterogeneity by age was seen for AESI rates, with some clearly increasing with age but others following the opposite trend. Similarly, differences were also observed across databases for same health outcome and age-sex strata. All studied AESIs appeared consistently more common in the post-COVID-19 compared to the historical cohorts, with related meta-analytic SIRs ranging from 1.32 (1.05 to 1.66) for narcolepsy to 11.70 (10.10 to 13.70) for pulmonary embolism.

**Interpretation**
Our findings suggest all AESIs are more common after COVID-19 than in the general population. Thromboembolic events were particularly common, and over 10-fold more so. More research is needed to contextualise post-COVID-19 complications in the longer term.

Prats-Uribe A, et al. **Use of repurposed and adjuvant drugs in hospital patients with COVID-19: multinational network cohort study.** BMJ. 2021 May 11;373:n1038. doi: 10.1136/bmj.n1038. Erratum in: BMJ. 2021 May 21;373:n1277. doi: 10.1136/bmj.n1277.


**Objective**
To investigate the use of repurposed and adjuvant drugs in patients admitted to hospital with COVID-19 across three continents.

**Design**
Multinational network cohort study.

**Setting**
Hospital electronic health records from the United States, Spain, and China, and nationwide claims data from South Korea.

**Participants**
303,264 patients admitted to hospital with COVID-19 from January 2020 to December 2020.

**Main Outcome Measures**
Prescriptions or dispensations of any drug on or 30 days after the date of hospital admission for covid-19.

**Results**
Of the 303,264 patients included, 290,131 were from the US, 7,599 from South Korea, 5,230 from Spain, and 304 from China. 3,455 drugs were identified. Common repurposed drugs were hydroxychloroquine (used in from <5 (<2%) patients in China to 2,165 (85.1%) in Spain), azithromycin (from 15 (4.9%) in China to 1,473 (57.9%) in Spain), combined lopinavir and ritonavir (from 156 (<2%) in the VA-OMOP US to 2,652 (34.9%) in South Korea and 1,285 (50.5%) in Spain), and umifenovir (0% in the US, South Korea, and Spain and 238 (78.3%) in China). Use of adjunctive drugs varied greatly, with the five most used treatments being enoxaparin, fluoroquinolones, ceftriaxone, vitamin D, and corticosteroids. Hydroxychloroquine use increased rapidly from March to April 2020 but declined steeply in May to June and remained low for the rest of the year. The use of dexamethasone and corticosteroids increased steadily during 2020.

**Conclusions**
Multiple drugs were used in the first few months of the COVID-19 pandemic, with substantial geographical and temporal variation. Hydroxychloroquine, azithromycin, lopinavir-ritonavir, and umifenovir (in China only) were the most prescribed repurposed drugs. Antithrombotics, antibiotics, H2 receptor antagonists, and corticosteroids were often used as adjunctive treatments. Research is needed on the comparative risk and benefit of these treatments in the management of COVID-19.

Burn E, et al. **Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study.** Nat Commun. 2020 Oct 6;11(1):5009. doi: 10.1038/s41467-020-18849-z.

Comorbidity conditions appear to be common among individuals hospitalised with coronavirus disease 2019 (COVID-19) but estimates of prevalence vary and little is known about the prior medication use of patients. Here, we describe the characteristics of adults hospitalised with COVID-19 and compare them with influenza patients. We include 34,128 (US: 8362, South Korea: 7,341, Spain: 18,425) COVID-19 patients, summarising between 4,811 and 11,643 unique aggregate characteristics. COVID-19 patients have been majority male in the US and Spain, but predominantly female in South Korea. Age profiles vary across data sources. Compared to 84,585 individuals hospitalised with influenza in 2014-19, COVID-19 patients have more typically been male, younger, and with fewer comorbidities and lower medication use. While protecting groups vulnerable to influenza is likely a useful starting point in the response to COVID-19, strategies will likely need to be broadened to reflect the particular characteristics of individuals being hospitalised with COVID-19.

Burn E, et al. **Opioid use, postoperative complications, and implant survival after unicompartmental versus total knee replacement: a population-based network study.** Lancet Rheumatol. 2019 Dec;1(4):e229-e236. doi: 10.1016/S2665-9913(19)30075-X. Epub 2019 Nov 7.


**Background**
There is uncertainty around whether to use unicompartmental knee replacement (UKR) or total knee replacement (TKR) for individuals with osteoarthritis confined to a single compartment of the knee. We aimed to emulate the design of the Total or Partial Knee Arthroplasty Trial (TOPKAT) using routinely collected data to assess whether the efficacy results reported in the trial translate into effectiveness in routine practice, and to assess comparative safety.

**Methods**
We did a population-based network study using data from four US and one UK health-care database, part of the Observational Health Data Sciences and Informatics network. The inclusion criteria were the same as those for TOPKAT; briefly, we identified patients aged at least 40 years with osteoarthritis who had undergone UKR or TKR and who had available data for at least one year prior to surgery. Patients were excluded if they had evidence of previous knee arthroplasty, knee fracture, knee surgery (except diagnostic), rheumatoid arthritis, inflammatory arthropathies, or septic arthritis. Opioid use from 91-365 days after surgery, as a proxy for persistent pain, was assessed for all participants in all databases. Postoperative complications (i.e., venous thromboembolism, infection, readmission, and mortality) were assessed over the 60 days after surgery and implant survival (as measured by revision procedures) was assessed over the five years after surgery.

Outcomes were assessed in all databases, except for readmission, which was assessed in three of the databases, and mortality, which was assessed in two of the databases. Propensity score matched Cox proportional hazards models were fitted for each outcome. Calibrated hazard ratios (cHRs) were generated for each database to account for observed differences in control outcomes, and cHRs were then combined using meta-analysis.

**Findings**
33,867 individuals who received UKR and 557,831 individuals who received TKR between Jan 1, 2005, and April 30, 2018, were eligible for matching. 32,379 with UKR and 250,377 with TKR were propensity score matched and informed the analyses. UKR was associated with a reduced risk of postoperative opioid use (cHR from meta-analysis 0·81, 95% CI 0·73-0·90) and a reduced risk of venous thromboembolism (0·62, 0·36-0·95), whereas no difference was seen for infection (0·85, 0·51-1·37) and readmission (0·79, 0·47-1·25). Evidence was insufficient to conclude whether there was a reduction in risk of mortality. UKR was also associated with an increased risk of revision (1·64, 1·40-1·94).

**Interpretation**
UKR was associated with a reduced risk of postoperative opioid use compared with TKR, which might indicate a reduced risk of persistent pain after surgery. UKR was associated with a lower risk of venous thromboembolism but an increased risk of revision compared with TKR. These findings can help to inform shared decision-making for individuals eligible for knee replacement surgery.

Li X, et al. **Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study.** BMJ. 2021 Jun 14;373:n1435. doi: 10.1136/bmj.n1435.


**Objective**
To quantify the background incidence rates of 15 prespecified adverse events of special interest (AESIs) associated with COVID-19 vaccines.

**Design**
Multinational network cohort study.

**Setting**
Electronic health records and health claims data from eight countries: Australia, France, Germany, Japan, the Netherlands, Spain, the United Kingdom, and the United States, mapped to a common data model.

**Participants**
126,661,070 people observed for at least 365 days before 1 January 2017, 2018, or 2019 from 13 databases.

**Main outcome measures**
Events of interests were 15 prespecified AESIs (non-haemorrhagic and haemorrhagic stroke, acute myocardial infarction, deep vein thrombosis, pulmonary embolism, anaphylaxis, Bell's palsy, myocarditis or pericarditis, narcolepsy, appendicitis, immune thrombocytopenia, disseminated intravascular coagulation, encephalomyelitis (including acute disseminated encephalomyelitis), Guillain-Barré syndrome, and transverse myelitis). Incidence rates of AESIs were stratified by age, sex, and database. Rates were pooled across databases using random effects meta-analyses and classified according to the frequency categories of the Council for International Organizations of Medical Sciences.

**Results**
Background rates varied greatly between databases. Deep vein thrombosis ranged from 387 (95% confidence interval 370 to 404) per 100,000 person years in UK CPRD GOLD data to 1,443 (1,416 to 1,470) per 100,000 person years in US IBM MarketScan Multi-State Medicaid data among women aged 65 to 74 years. Some AESIs increased with age. For example, myocardial infarction rates in men increased from 28 (27 to 29) per 100,000 person years among those aged 18-34 years to 1,400 (1,374 to 1,427) per 100 000 person years in those older than 85 years in US Optum electronic health record data. Other AESIs were more common in young people. For example, rates of anaphylaxis among boys and men were 78 (75 to 80) per 100,000 person years in those aged 6-17 years and 8 (6 to 10) per 100,000 person years in those older than 85 years in Optum electronic health record data. Meta-analytic estimates of AESI rates were classified according to age and sex.

**Conclusion**
This study found large variations in the observed rates of AESIs by age group and sex, showing the need for stratification or standardisation before using background rates for safety surveillance. Considerable population level heterogeneity in AESI rates was found between databases.

Li X, et al. **Association between Covid-19 vaccination, SARS-CoV-2 infection, and risk of immune mediated neurological events: population based cohort and self-controlled case series analysis.** BMJ. 2022 Mar 16;376:e068373. doi: 10.1136/bmj-2021-068373.


**Objective**
To study the association between COVID-19 vaccines, SARS-CoV-2 infection, and risk of immune mediated neurological events.

**Design**
Population-based historical rate comparison study and self-controlled case series analysis.

**Setting**
Primary care records from the United Kingdomand  Spain linked to hospital data.

**Participants**
8,330,497 people who received at least one dose of COVID-19 vaccines ChAdOx1 nCoV-19, BNT162b2, mRNA-1273, or Ad.26.COV2.S between the rollout of the vaccination campaigns and end of data availability (UK: 9 May 2021; Spain: 30 June 2021). The study sample also comprised a cohort of 735,870 unvaccinated individuals with a first positive reverse transcription polymerase chain reaction test result for SARS-CoV-2 from 1 September 2020, and 14,330,080 participants from the general population.

**Main outcome measures**
Outcomes were incidence of Bell's palsy, encephalomyelitis, Guillain-Barré syndrome, and transverse myelitis. Incidence rates were estimated in the 21 days after the first vaccine dose, 90 days after a positive test result for SARS-CoV-2, and between 2017 and 2019 for background rates in the general population cohort. Indirectly standardised incidence ratios were estimated. Adjusted incidence rate ratios were estimated from the self-controlled case series.

**Results**
The study included 4,376,535 people who received ChAdOx1 nCoV-19, 3,588,318 who received BNT162b2, 244 913 who received mRNA-1273, and 120,731 who received Ad26.CoV.2; 735,870 people with SARS-CoV-2 infection; and 14,330,080 people from the general population. Overall, post-vaccine rates were consistent with expected (background) rates for Bell's palsy, encephalomyelitis, and Guillain-Barré syndrome. Self-controlled case series was conducted only for Bell's palsy, given limited statistical power, but with no safety signal seen for those vaccinated. Rates were, however, higher than expected after SARS-CoV-2 infection. For example, in the data from the UK, the standardised incidence ratio for Bell's palsy was 1.33 (1.02 to 1.74), for encephalomyelitis was 6.89 (3.82 to 12.44), and for Guillain-Barré syndrome was 3.53 (1.83 to 6.77). Transverse myelitis was rare (<5 events in all vaccinated cohorts) and could not be analysed.

**Conclusions**
No safety signal was observed between COVID-19 vaccines and the immune mediated neurological events of Bell's palsy, encephalomyelitis, Guillain-Barré syndrome, and transverse myelitis. An increased risk of Bell's palsy, encephalomyelitis, and Guillain-Barré syndrome was, however, observed for people with SARS-CoV-2 infection.

Català M, et al. **The effectiveness of COVID-19 vaccines to prevent long COVID symptoms: staggered cohort study of data from the UK, Spain, and Estonia.** Lancet Respir Med. 2024 Mar;12(3):225-236. doi: 10.1016/S2213-2600(23)00414-9. Epub 2024 Jan 11.


**Background**
Although vaccines have proved effective to prevent severe COVID-19, their effect on preventing long-term symptoms is not yet fully understood. We aimed to evaluate the overall effect of vaccination to prevent long COVID symptoms and assess comparative effectiveness of the most used vaccines (ChAdOx1 and BNT162b2).

**Methods**
We conducted a staggered cohort study using primary care records from the UK (Clinical Practice Research Datalink [CPRD] GOLD and AURUM), Catalonia, Spain (Information System for Research in Primary Care [SIDIAP]), and national health insurance claims from Estonia (CORIVA database). All adults who were registered for at least 180 days as of Jan 4, 2021 (the UK), Feb 20, 2021 (Spain), and Jan 28, 2021 (Estonia) comprised the source population. Vaccination status was used as a time-varying exposure, staggered by vaccine rollout period. Vaccinated people were further classified by vaccine brand according to their first dose received. The primary outcome definition of long COVID was defined as having at least one of 25 WHO-listed symptoms between 90 and 365 days after the date of a PCR-positive test or clinical diagnosis of COVID-19, with no history of that symptom 180 days before SARS-Cov-2 infection. Propensity score overlap weighting was applied separately for each cohort to minimise confounding. Sub-distribution hazard ratios (sHRs) were calculated to estimate vaccine effectiveness against long COVID, and empirically calibrated using negative control outcomes. Random effects meta-analyses across staggered cohorts were conducted to pool overall effect estimates.

**Findings**
A total of 1,618,395 (CPRD GOLD), 5,729 800 (CPRD AURUM), 2,744,821 (SIDIAP), and 77,603 (CORIVA) vaccinated people and 1,640 371 (CPRD GOLD), 5,860,564 (CPRD AURUM), 2,588,518 (SIDIAP), and 302,267 (CORIVA) unvaccinated people were included. Compared with unvaccinated people, overall HRs for long COVID symptoms in people vaccinated with a first dose of any COVID-19 vaccine were 0·54 (95% CI 0·44-0·67) in CPRD GOLD, 0·48 (0·34-0·68) in CPRD AURUM, 0·71 (0·55-0·91) in SIDIAP, and 0·59 (0·40-0·87) in CORIVA. A slightly stronger preventative effect was seen for the first dose of BNT162b2 than for ChAdOx1 (sHR 0·85 [0·60-1·20] in CPRD GOLD and 0·84 [0·74-0·94] in CPRD AURUM).

**Interpretation**
Vaccination against COVID-19 consistently reduced the risk of long COVID symptoms, which highlights the importance of vaccination to prevent persistent COVID-19 symptoms, particularly in adults.

Yang C, et al. **Development and external validation of prediction models for adverse health outcomes in rheumatoid arthritis: A multinational real-world cohort analysis.** Semin Arthritis Rheum. 2022 Oct;56:152050. doi: 10.1016/j.semarthrit.2022.152050. Epub 2022 Jun 15.


**Background**
Identification of rheumatoid arthritis (RA) patients at high risk of adverse health outcomes remains a major challenge. We aimed to develop and validate prediction models for a variety of adverse health outcomes in RA patients initiating first-line methotrexate (MTX) monotherapy.

**Methods**
Data from 15 claims and electronic health record databases across nine countries were used. Models were developed and internally validated on Optum® De-identified Clinformatics® Data Mart Database using L1-regularised logistic regression to estimate the risk of adverse health outcomes within three months (leukopenia, pancytopenia, infection), two years (myocardial infarction (MI) and stroke), and five years (cancers [colorectal, breast, uterine] after treatment initiation. Candidate predictors included demographic variables and past medical history. Models were externally validated on all other databases. Performance was assessed using the area under the receiver operator characteristic curve (AUC) and calibration plots.

**Findings**
Models were developed and internally validated on 21,547 RA patients and externally validated on 131,928 RA patients. Models for serious infection (AUC: internal 0.74, external ranging from 0.62 to 0.83), MI (AUC: internal 0.76, external ranging from 0.56 to 0.82), and stroke (AUC: internal 0.77, external ranging from 0.63 to 0.95), showed good discrimination and adequate calibration. Models for the other outcomes showed modest internal discrimination (AUC < 0.65) and were not externally validated.

**Interpretation**
We developed and validated prediction models for a variety of adverse health outcomes in RA patients initiating first-line MTX monotherapy. Final models for serious infection, MI, and stroke demonstrated good performance across multiple databases and can be studied for clinical use.

# Study-a-thons

EHDEN and OHDSI have been working together on finding innovative ways to carry out collaborative research. One of those was the use of what we called a study-a-thon.

The premise of the study-a-thon is simple: bring together a diverse group of researchers aligned on a common question and focus together on collaboratively designing research protocols, executing analyses across databases, and interpreting results over an intense, but fun-filled few days. This approach significantly speeds up research time from what historically took years - to now just months, weeks and days.

Oxford study-a-thon Long COVID/PASC April 2023

EHDEN and OHDSI collaborators have held multiple study-a-thons on a wide array of topics, including rheumatoid arthritis, cancer treatments, cardiovascular prediction, Alopecia Areata, methods for pharmacovigilance, and COVID-19. Each event has demonstrated our collective ability to accomplish in a short time what may be unimaginable alone, and it has provided further reinforcement of the power of community and the value of multi-disciplinary collaboration.

Our most memorable study-a-thon happened at the beginning of the COVID pandemic in which the OHDSI community worked closely together with EHDEN and many other collaborators from across the world.

## 88 Hours

When the COVID-19 pandemic began to close down much of the world and the planned OHDSI Europe Symposium was cancelled, our global collaborators came together for four days of rigorous work about a disease with limited available data. These 88 hours set the foundation for years of COVID research, as well as ongoing research around vaccine surveillance. Read the feature on this memorable community event by visiting ohdsi.org/88-hours.

### What You Should Know About the COVID Study-A-Thon

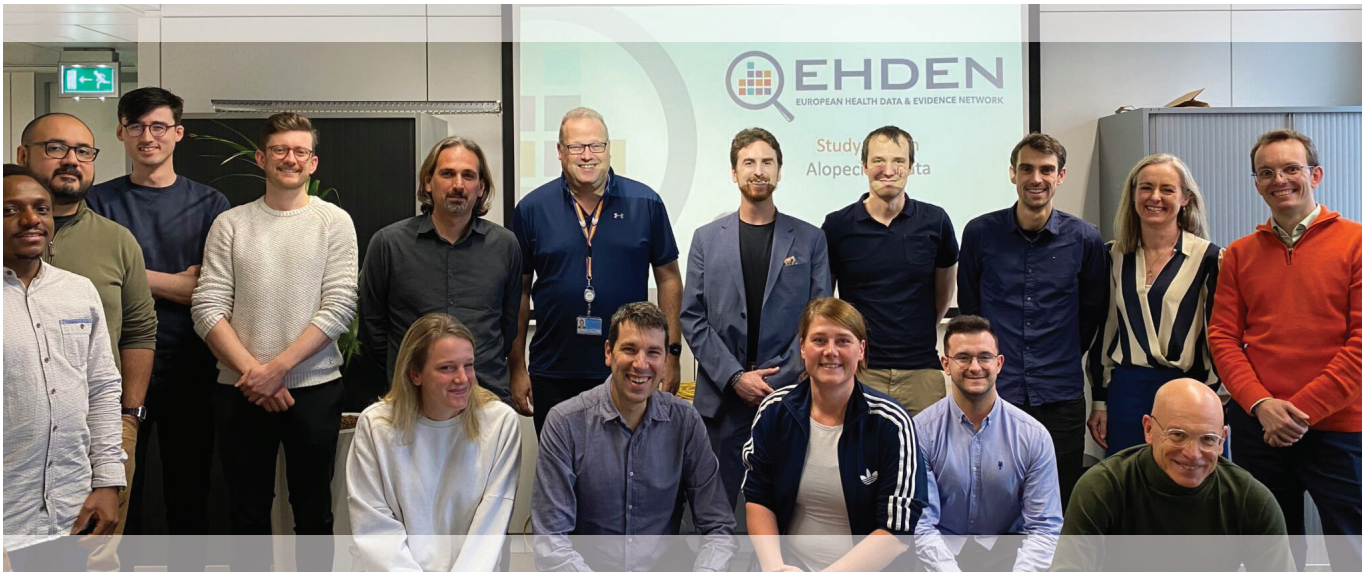
- More than 330 people from across 30 countries (six continents) registered for the event.
- The event took place over 88 hours between March 26-29, 2020 and it was coordinated by the Erasmus University Medical Center.
- There were 17 concurrent channels on the overall Teams platform, and those channels hosted more than 100 collaborator calls.
- There were 12 global huddles, spaced out so collaborators from around the world would have a daily opportunity to hear about community progress.
- More than 10,000 publications were reviewed both prior and during the event.
- There were 13,000+ chat messages that helped design both 355 cohort definitions and nine protocols, as well as the release of 13 study packages.
- The closing call has been viewed almost 1,800 times since it was posted to YouTube.
- The OHDSI community has published multiple COVID-19 studies (including in Lancet Rheumatology, Nature Communications, Lancet Digital Health, and The BMJ), and it continues to do research in many areas of COVID-19 and vaccine surveillance.

Study-a-thon on rheumatoid arthritis in Barcelona

Study-a- thon on pharmacovigilance in Uppsala

Study-a-thon on cancer survival in Barcelona

Study-a-thon on osteoarthritis and knee replacement in Oxford




Study-a-thon on Alopecia Areata in Rotterdam

# Safeguarding against critical medicines' shortage

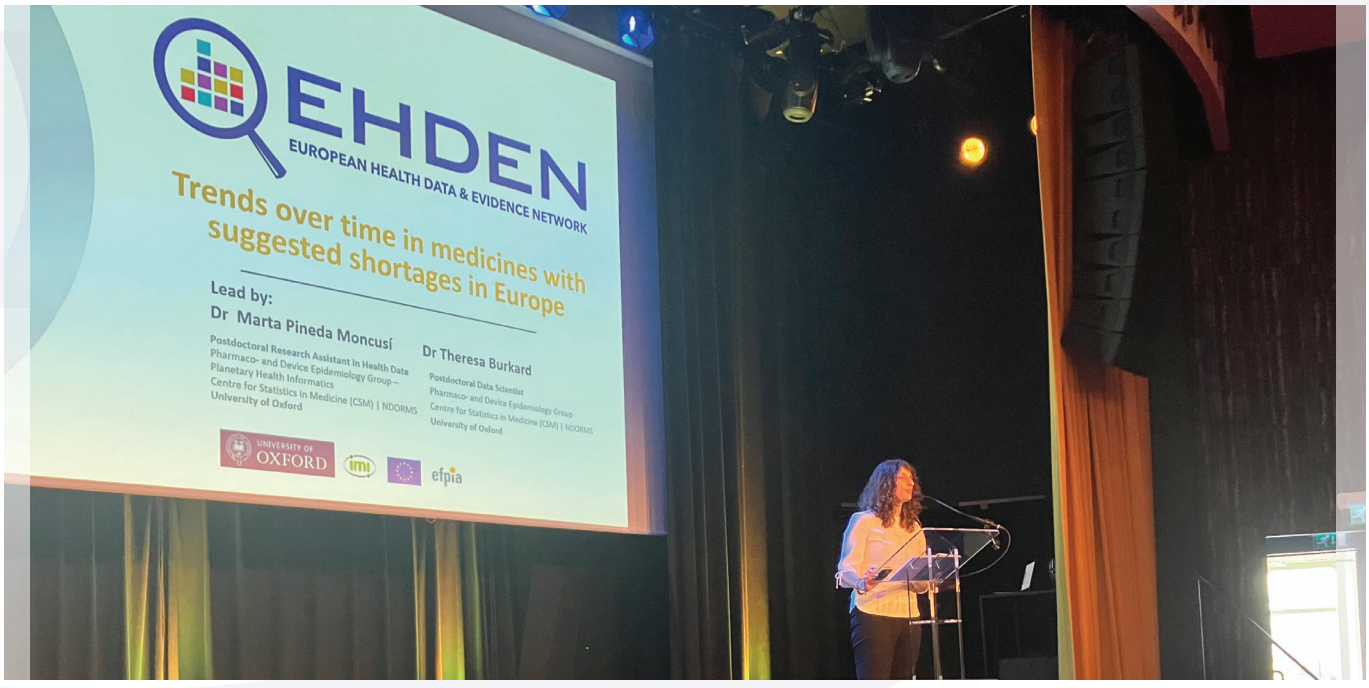
Access to essential medicines is on the decline in Europe and this is having a harmful impact on patient health by increasing mortality, adverse events, and therapy errors. Since 2016, the European Medicines Agency (EMA) has published a public catalogue of medicines under surveillance due to shortage in more than one European country. The list includes information such as the reason for the shortage, and recommendations for patients, healthcare professionals and other stakeholders.

One of the most worrying and recurring reasons for their shortage is an increased demand for the product in Europe, for example, antibiotics or insulin. Thus, the EMA shortages catalogue and further development of a centralised electronic platform for pharmaceutical companies and countries to report on shortages, the European Shortages Monitoring Platform (ESMP), will, by 2025, enable the monitoring of consumption, defining targets and the effects of stewardship policies that aim to optimise the use of medicines in shortage.

In response to this important issue, EHDEN, in collaboration with multiple partners, launched a federated network study in November 2023 led by Oxford University entitled: ***Incidence, prevalence and characterisation of medicines with suggested drug shortages in Europe***. This study will improve our understanding of the use of medicines with reported shortages in routine healthcare delivery depicting trends in use, including indication, treatment duration, and doses. A federated data network such as EHDEN has the advantage of providing a more generalisable and complete picture of drug shortages in Europe. The results will contribute to the European efforts to monitor the use of critical medicines as part of the global fight against medicine shortage.

# Unprecedented study scope

While this study is still ongoing, more than fifty Data Partners collaborated to send results on the use of medicines with suggested shortages. Of these, the IRB approval/data-sharing waiver was received for almost all Data Partners. For the characterisation of the drug users, we received data from forty Data Partners (all for whom we can share data) **– making this the largest observational database study conducted across Europe in terms of the number of databases and geographic distribution.**




Marta Pineda Moncusi at OHDSI Europe Symposium, June 2024

# Key learnings to date

Preprint

tinyurl.com/2u5ym7f7

Lessons learned from coordinating this large study:

- We have produced the most comprehensive information on drug usage among drugs with reported shortage and their alternatives in European data.

- European databases show that the impact of drug shortages can be different within and between countries, and across different types of health care providers such as general practices and hospitals.

- A central contact point for communication and organisation of the steering committee was important. Generally, a positive was that we had a task force, so the study leads had a group of experienced and engaged people to share ideas/check back on process.

- Obtaining ethical approvals from Data Partners was a tedious process that needs to be initiated early on to obtain the green light for results-sharing in time. Receiving results but no ethical approvals in time is a waste of resources.

- Webinars for joint interpretation of study results to handle the large number of participating Data Partners was a good solution and was well received.

- Having a study co-lead was important to share workload and be more creative and solution oriented by having two pairs of eyes and ears.

- Sending out the code to a small group of Data Partners for trial rounds assured a smooth code-sharing experience and we did not have to re-run code.

Thus far, preliminary findings were presented at the OHDSI Europe Symposium in June 2024 and the International Conference on Pharmacoepidemiology & Therapeutic Risk Management (IPCE) in August 2024. A preprint with incidence/prevalence results were made publicly available in August 2024 (doi.org/10.1101/2024.08.28.24312695), and the addition of drug utilisation results to the preprint by October 2024. The publication of results will be shared with the public in a high impact, peer-reviewed journal.

**"This study demonstrates how impactful the EHDEN project truly is. It would not have been possible to run a study at this scale without the efforts of many in the EHDEN consortium that played a significant role in the standardisation of all these data sources to the OMOP CDM via our open calls, those that worked on methods and tools, and our fantastic project management team."**

*Peter Rijnbeek, Erasmus MC*

***"We have accumulated more results than we can ever easily process. Thus, the collection of data in aggregated form available to the public through open science data-sharing will serve future researchers as a go-to point on medicine use with reported shortages, thereby helping them to source study questions (and descriptive answers)."***

*Theresa Burkard, study co-lead*

# OUR PUBLICATIONS

EHDEN Zenodo

tinyurl.com/t688a8tu

Find below an overview of our current list of peer-reviewed publications. For an up-to-date overview and our preprints see our Zenodo Page.


1. Yang C, Fridgeirsson, E. A., Kors, J. A., Reps, J. M., Rijnbeek, P. R. Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data. Journal of Big Data. 2024;11(1).

2. Sibert NT, Soff J, La Ferla S, Quaranta M, Kremer A, Kowalski C. Transforming a Large-Scale Prostate Cancer Outcomes Dataset to the OMOP Common Data Model-Experiences from a Scientific Data Holder's Perspective. Cancers (Basel). 2024;16(11).

3. Seinen TM, Kors JA, van Mulligen EM, Rijnbeek PR. Annotation-preserving machine translation of English corpora to validate Dutch clinical concept extraction tools. J Am Med Inform Assoc. 2024;31(8):1725-34.

4. Seinen TM, Kors JA, van Mulligen EM, Fridgeirsson EA, Verhamme KM, Rijnbeek PR. Using clinical text to refine unspecific condition codes in Dutch general practitioner EHR data. Int J Med Inform. 2024;189:105506.

5. Schuemie M, Reps J, Black A, Defalco F, Evans L, Fridgeirsson E, et al. Health-Analytics Data to Evidence Suite (HADES): Open-Source Software for Observational Research. Stud Health Technol Inform. 2024;310:966-70.

6. Markus AF, Rijnbeek PR, Kors JA, Burn E, Duarte-Salles T, Haug M, et al. Real-world treatment trajectories of adults with newly diagnosed asthma or COPD. BMJ Open Respir Res. 2024;11(1).

7. Lo Re Iii V, Cocoros NM, Hubbard RA, Dutcher SK, Newcomb CW, Connolly JG, et al. Risk of Arterial and Venous Thrombotic Events Among Patients with COVID-19: A Multi-National Collaboration of Regulatory Agencies from Canada, Europe, and United States. Clin Epidemiol. 2024;16:71-89.

8. Liu S, Golozar A, Buesgens N, McLeggon JA, Black A, Nagy P. A framework for understanding an open scientific community using automated harvesting of public artifacts. JAMIA Open. 2024;7(1):ooae017.

9. Krastev E, Markov E, Abanos S, Krasteva R, Tcharaktchiev D. Mapping the Bulgarian Diabetes Register to OMOP CDM: Application Results. Stud Health Technol Inform. 2024;313:28-33.

10. John LH, Fridgeirsson EA, Kors JA, Reps JM, Williams RD, Ryan PB, et al. Development and validation of a patient-level model to predict dementia across a network of observational databases. BMC Med. 2024;22(1):308.

11. Haug M, Oja M, Pajusalu M, Mooses K, Reisberg S, Vilo J, et al. Markov modeling for cost-effectiveness using federated health data network. J Am Med Inform Assoc. 2024;31(5):1093-101.

12. Gomes M, Turner, A.J., Sammon, C., Dawoud, D., Ramagopalan, S., Simpson, A. and Siebert, U. Acceptability of using Real-World Data to Estimate Relative Treatment Effects in Health Technology Assessments: Barriers and Future Steps. . Value in Health. 2024;27(5):623-32.

13. Fridgeirsson EA, Williams R, Rijnbeek P, Suchard MA, Reps JM. Comparing penalization methods for linear models on large observational health data. J Am Med Inform Assoc. 2024;31(7):1514-21.

14. Burkard T, Lopez-Guell K, Gorbachev A, Bellas L, Jodicke AM, Burn E, et al. Calculating daily dose in the Observational Medical Outcomes Partnership Common Data Model. Pharmacoepidemiol Drug Saf. 2024;33(6):e5809.

15. Brauner KB, Tsouchnika A, Mashkoor M, Williams R, Rosen AW, Hartwig MFS, et al. Prediction of 30-day, 90-day, and 1-year mortality after colorectal cancer surgery using a data-driven approach. Int J Colorectal Dis. 2024;39(1):31.

16. Barclay NL, Burkard T, Burn E, Delmestri A, Miquel Dominguez A, Golozar A, et al. The Impact of the COVID-19 Pandemic on Incidence and Short-Term Survival for Common Solid Tumours in the United Kingdom: A Cohort Analysis. Clin Epidemiol. 2024;16:417-29.

17. Almeida JR, Zuquete A, Pazos A, Oliveira JL. A federated authentication schema among multiple identity providers. Heliyon. 2024;10(7):e28560.

18. Pera V, van Vaerenbergh F, Kors JA, van Mulligen EM, Parry R, de Wilde M, et al. Descriptive analysis on disproportionate medication errors and associated patient characteristics in the Food and Drug Administration's Adverse Event Reporting System. Pharmacoepidemiol Drug Saf. 2024;33(1):e5743.

19. Yang C, Fridgeirsson EA, Kors JA, Reps JM, Rijnbeek PR, Wong J, et al. Does Using a Stacking Ensemble Method to Combine Multiple Base Learners Within a Database Improve Model Transportability? Stud Health Technol Inform. 2023;302:129-30.

20. Williams RD, den Otter S, Reps JM, Rijnbeek PR. The DELPHI Library: Improving Model Validation, Transparency and Dissemination Through a Centralised Library of Prediction Models. Stud Health Technol Inform. 2023;302:139-40.

21. Voss EA, Shoaibi A, Yin Hui Lai L, Blacketer C, Alshammari T, Makadia R, et al. Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study. EClinicalMedicine. 2023;58:101932.

22. Voss EA, Blacketer C, van Sandijk S, Moinat M, Kallfelz M, van Speybroeck M, et al. European Health Data & Evidence Network-learnings from building out a standardized international health data network. J Am Med Inform Assoc. 2023;31(1):209-19.

23. Seinen TM, Kors JA, van Mulligen EM, Fridgeirsson E, Rijnbeek PR. The added value of text from Dutch general practitioner notes in predictive modeling. J Am Med Inform Assoc. 2023;30(12):1973-84.

24. Sarri G, Liu W, Zabotka L, Freitag A, Claire R, Wangge G, et al. Prognostic Factors of COVID-19:

An Umbrella Review Endorsed by the International Society for Pharmacoepidemiology. Clin Pharmacol Ther. 2023;114(3):604-13.

25. Rekkas A, van Klaveren D, Ryan PB, Steyerberg EW, Kent DM, Rijnbeek PR. A standardized framework for risk-based assessment of treatment effect heterogeneity in observational healthcare databases. NPJ Digit Med. 2023;6(1):58.

26. Rekkas A, Rijnbeek PR, Kent DM, Steyerberg EW, van Klaveren D. Estimating individualized treatment effects from randomized controlled trials: a simulation study to compare risk-based approaches. BMC Med Res Methodol. 2023;23(1):74.

27. Raventos B, Fernandez-Bertolin S, Aragon M, Voss EA, Blacketer C, Mendez-Boo L, et al. Transforming the Information System for Research in Primary Care (SIDIAP) in Catalonia to the OMOP Common Data Model and Its Use for COVID-19 Research. Clin Epidemiol. 2023;15:969-86.

28. Puttmann D, de Groot R, de Keizer N, Cornet R, Elbers PWG, Dongelmans D, et al. Assessing the FAIRness of databases on the EHDEN portal: A case study on two Dutch ICU databases. Int J Med Inform. 2023;176:105104.

29. Oja M, Tamm S, Mooses K, Pajusalu M, Talvik HA, Ott A, et al. Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned. JAMIA Open. 2023;6(4):ooad100.

30. Meszaros A, Kovacs S, Heja T, Bagyura Z, Zemplenyi A. Mapping Hungarian procedure codes to SNOMED CT. BMC Med Res Methodol. 2023;23(1):240.

31. Markus AF, Strauss VY, Burn E, Li X, Delmestri A, Reich C, et al. Characterising the treatment of thromboembolic events after COVID-19 vaccination in 4 European countries and the US: An international network cohort study. Front Pharmacol. 2023;14:1118203.

32. Markus AF, Fridgeirsson EA, Kors JA, Verhamme KMC, Rijnbeek PR. Challenges of Estimating Global Feature Importance in Real-World Health Care Data. Stud Health Technol Inform. 2023;302:1057-61.

33. Makadia R, Shoaibi A, Rao GA, Ostropolets A, Rijnbeek PR, Voss EA, et al. Evaluating the impact of alternative phenotype definitions on incidence rates across a global data network. JAMIA Open. 2023;6(4):ooad096.

34. Höllig J, Markus, A.F., de Slegte, J., Bagave, P. . Semantic Meaningfulness: Evaluating Counterfactual Approaches for Real-World Plausibility and Feasibility. . Longo, L (eds) Explainable Artificial Intelligence xAI 2023 Communications in Computer and Information Science,. 2023;1902.

35. Gauffin O, Brand JS, Vidlin SH, Sartori D, Asikainen S, Catala M, et al. Supporting Pharmacovigilance Signal Validation and Prioritization with Analyses of Routinely Collected Health Data: Lessons Learned from an EHDEN Network Study. Drug Saf. 2023;46(12):1335-52.

36. Fridgeirsson EA, Sontag D, Rijnbeek P. Attention-based neural networks for clinical prediction modelling on electronic health records. BMC Med Res Methodol. 2023;23(1):285.

37. Claire R, Elvidge J, Hanif S, Goovaerts H, Rijnbeek PR, Jonsson P, et al. Advancing the use of real world evidence in health technology assessment: insights from a multi-stakeholder workshop. Front Pharmacol. 2023;14:1289365.

38. Pera V, Brusselle GG, Riemann S, Kors JA, Van Mulligen EM, Parry R, et al. Parasitic infections related to anti-type 2 immunity monoclonal antibodies: a disproportionality analysis in the food and drug administration's adverse event reporting system (FAERS). Front Pharmacol. 2023;14:1276340.

39. Yang C, Williams RD, Swerdel JN, Almeida JR, Brouwer ES, Burn E, et al. Development and external validation of prediction models for adverse health outcomes in rheumatoid arthritis: A multinational real-world cohort analysis. Semin Arthritis Rheum. 2022;56:152050.

40. Yang C, Kors JA, Ioannou S, John LH, Markus AF, Rekkas A, et al. Trends in the conduct and reporting of clinical prediction model development and validation: a systematic review. J Am Med Inform Assoc. 2022;29(5):983-9.

41. Wong J, Prieto-Alhambra D, Rijnbeek PR, Desai RJ, Reps JM, Toh S. Applying Machine Learning in Distributed Data Networks for Pharmacoepidemiologic and Pharmacovigilance Studies: Opportunities, Challenges, and Considerations. Drug Saf. 2022;45(5):493-510.

42. Williams RD, Reps JM, Kors JA, Ryan PB, Steyerberg E, Verhamme KM, et al. Using Iterative Pairwise External Validation to Contextualize Prediction Model Performance: A Use Case Predicting 1-Year Heart Failure Risk in Patients with Diabetes Across Five Data Sources. Drug Saf. 2022;45(5):563-70.

43. Williams RD, Reps JM, Group OEKA, Rijnbeek PR, Ryan PB, Prieto-Alhambra D. 90-Day all-cause mortality can be predicted following a total knee replacement: an international, network study to develop and validate a prediction model. Knee Surg Sports Traumatol Arthrosc. 2022;30(9):3068-75.

44. Williams RD, Markus AF, Yang C, Duarte-Salles T, DuVall SL, Falconer T, et al. Seek COVER: using a disease proxy to rapidly develop and validate a personalized risk calculator for COVID-19 outcomes in an international network. BMC Med Res Methodol. 2022;22(1):35.

45. Seinen TM, Fridgeirsson EA, Ioannou S, Jeannetot D, John LH, Kors JA, et al. Use of unstructured text in prognostic clinical prediction models: a systematic review. J Am Med Inform Assoc. 2022;29(7):1292-302.

46. Reps JM, Williams RD, Schuemie MJ, Ryan PB, Rijnbeek PR. Learning patient-level prediction models across multiple healthcare databases: evaluation of ensembles for increasing model transportability. BMC Med Inform Decis Mak. 2022;22(1):142.

47. Papez V, Moinat M, Voss EA, Bazakou S, Van Winzum A, Peviani A, et al. Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond. J Am Med Inform Assoc. 2022;30(1):103-11.

48. Moreno-Martos D, Verhamme K, Ostropolets A, Kostka K, Duarte-Sales T, Prieto-Alhambra D, et al. Characteristics and outcomes of COVID-19 patients with COPD from the United States, South Korea, and Europe. Wellcome Open Res. 2022;7:22.

49. Markus AF, Verhamme KMC, Kors JA, Rijnbeek PR. TreatmentPatterns: An R package to facilitate the standardized development and analysis of treatment patterns across disease domains. Comput Methods Programs Biomed. 2022;225:107081.

50. Markus AF, Rijnbeek, P. R., & Reps. J. M. Why predicting risk can't identify 'risk factors': empirical assessment of model stability in machine learning across observational health databases. Machine Learning for Healthcare 2022. 2022:1-25.

51. Kadri Künnapuu SI, Kadri Ligi, Raivo Kolde, Sven Laur, Jaak Vilo, Peter R Rijnbeek, Sulev Reisberg. Trajectories: a framework for detecting temporal clinical event sequences from health data standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model. JAMIA open. 2022;5(1).

52. John LH, Kors JA, Reps JM, Ryan PB, Rijnbeek PR. Logistic regression models for patient-level prediction based on massive observational data: Do we need all data? Int J Med Inform. 2022;163:104762.

53. John LH, Kors JA, Fridgeirsson EA, Reps JM, Rijnbeek PR. External validation of existing dementia prediction models on observational health data. BMC Med Res Methodol. 2022;22(1):311.
54. Hughes N, Rijnbeek PR, van Bochove K, Duarte-Salles T, Steinbeisser C, Vizcaya D, et al. Evaluating a novel approach to stimulate open science collaborations: a case series of "study-a-thon" events within the OHDSI and European IMI communities. JAMIA Open. 2022;5(4):ooac100.

55. Gomes M, Latimer, N., Soares, M., Dias, S., Baio, G., Freemantle, N., Dawoud, D., Wailoo, A. and Grieve, R. Target trial emulation for transparent and robust estimation of treatment effects for health technology assessment using real-world data: opportunities and challenges. . Pharmacoeconomics. 2022;40(6):577-86.

56. Delanerolle G, Williams R, Stipancic A, Byford R, Forbes A, Tsang RSM, et al. Methodological Issues in Using a Common Data Model of COVID-19 Vaccine Uptake and Important Adverse Events of Interest: Feasibility Study of Data and Connectivity COVID-19 Vaccines Pharmacovigilance in the United Kingdom. JMIR Form Res. 2022;6(8):e37821.

57. Burn E, Li X, Kostka K, Stewart HM, Reich C, Seager S, et al. Background rates of five thrombosis with thrombocytopenia syndromes of special interest for COVID-19 vaccine safety surveillance: Incidence between 2017 and 2019 and patient profiles from 38.6 million people in six European countries. Pharmacoepidemiol Drug Saf. 2022;31(5):495-510.

58. Reps JM, Rijnbeek P, Cuthbert A, Ryan PB, Pratt N, Schuemie M. An empirical analysis of dealing with patients who are lost to follow-up when developing prognostic models using a cohort design. BMC Med Inform Decis Mak. 2021;21(1):43.

59. Prieto-Alhambra D, Kostka K, Duarte-Salles T, Prats-Uribe A, Sena A, Pistillo A, et al. Unraveling COVID-19: a large-scale characterization of 4.5 million COVID-19 cases using CHARYBDIS. Res Sq. 2021.

60. Prats-Uribe A, Sena AG, Lai LYH, Ahmed WU, Alghoul H, Alser O, et al. Use of repurposed and adjuvant drugs in hospital patients with covid-19: multinational network cohort study. BMJ. 2021;373:n1038.

61. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021;113:103655.

62. Khalid S, Yang C, Blacketer C, Duarte-Salles T, Fernandez-Bertolin S, Kim C, et al. A standardized analytics pipeline for reliable and rapid development and validation of prediction models using observational health data. Comput Methods Programs Biomed. 2021;211:106394.

63. Kent S, Burn E, Dawoud D, Jonsson P, Ostby JT, Hughes N, et al. Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment. Pharmacoeconomics. 2021;39(3):275-85.

64. Burn E, Tebe C, Fernandez-Bertolin S, Aragon M, Recalde M, Roel E, et al. The natural history of symptomatic COVID-19 during the first wave in Catalonia. Nat Commun. 2021;12(1):777.

65. Blacketer C VE, DeFalco F, Hughes N, Schuemie MJ, Moinat M, Rijnbeek PR. Using the Data Quality Dashboard to Improve the EHDEN Network. Applied Sciences 2021;11(24).

66. Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR. Increasing trust in real-world evidence through evaluation of observational data quality. J Am Med Inform Assoc. 2021;28(10):2251-7.

67. Wang Q, Reps JM, Kostka KF, Ryan PB, Zou Y, Voss EA, et al. Development and validation of a prognostic model predicting symptomatic hemorrhagic transformation in acute ischemic stroke at scale in the OHDSI network. PLoS One. 2020;15(1):e0226718.

68. Reps JM, Williams RD, You SC, Falconer T, Minty E, Callahan A, et al. Feasibility and evaluation of a large-scale external validation approach for patient-level prediction in an international data network: validation of models predicting stroke in female patients newly diagnosed with atrial fibrillation. BMC Med Res Methodol. 2020;20(1):102.

69. Rekkas A, Paulus JK, Raman G, Wong JB, Steyerberg EW, Rijnbeek PR, et al. Predictive approaches to heterogeneous treatment effects: a scoping review. BMC Med Res Methodol. 2020;20(1):264.

70. Duarte-Salles T, Vizcaya D, Pistillo A, Casajust P, Sena AG, Lai LYH, et al. Baseline characteristics, management, and outcomes of 55,270 children and adolescents diagnosed with COVID-19 and 1,952,693 with influenza in France, Germany, Spain, South Korea and the United States: an international network cohort study. medRxiv. 2020.

71. Burn E, You SC, Sena AG, Kostka K, Abedtash H, Abrahao MTF, et al. Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. Nat Commun. 2020;11(1):5009.

72. Burn E, Weaver J, Morales D, Prats-Uribe A, Delmestri A, Strauss VY, et al. Opioid use, postoperative complications, and implant survival after unicompartmental versus total knee replacement: a population-based network study. Lancet Rheumatol. 2019;1(4):e229-e36.