21 June 2023

Protocol

Title: Association between use of direct oral anticoagulants (DOACs) and increased risk of interstitial lung disease

Administrative details of the data analysis	
Substance(s)	Edoxaban, Apixaban, Rivaroxaban, Dabigatran
Condition/ADR(s)	Interstitial lung disease
Short title of topic	ILD and DOACs
Study team	Valentijn de Jong, María Clara Restrepo-Méndez and Daniel Morales

1. List of abbreviations

MAH	*Marketing Authorisation Holder*
EMA	*European Medicines Agency*
PRAC	*Pharmacovigilance Risk Assessment Committee*

2. Milestones

Milestone	**Planned date**
Draft analysis plan prepared	24/03/2023
Revised analysis plan following review	21/06/2023
Final analysis report	TBD

3. Rationale and background

Interstitial lung disease (ILD) describes a heterogenous group of respiratory disorders affecting the interstitium of the lungs. [1,2] ILD may occur when an injury to the lungs triggers an abnormal healing response. The repair process is disrupted, and the tissue around the alveoli becomes scarred and thickened. Prolonged ILD may result in pulmonary fibrosis, but this is not always the case. [1,2]

Cases of interstitial lung disease (ILD) have been recently described, especially in Asian subjects, in people using direct oral anticoagulants (DOACs) [3-5]. In addition, a recently published observational study suggests an increased risk of ILD in patients treated with factor Xa inhibitors (edoxaban, apixaban, rivaroxaban) as compared to warfarin users among subjects with non-valvular atrial fibrillation in Taiwan [6, 7], which triggered a new potential safety signal. A thorough review of this issue within the ongoing and upcoming PSUSA procedures for edoxaban, apixaban and rivaroxaban is expected. In this context, a real-world data (RWD) analysis has been requested to better inform the PRAC on the relationship between increased risk of ILD and use of DOACs in the European population.

4. Research question and objectives

Is the use of factor Xa (FXa) inhibitors (edoxaban, apixaban and rivaroxaban) or direct thrombin inhibitor (dabigatran) associated with an increased risk of interstitial lung disease (ILD) when compared with patients treated with vitamin K antagonists (VKA), among patients with Atrial Fibrillation (AF), Deep Vein Thrombosis (DVT) and Pulmonary Embolism (PE)?

5. Research methods

5.1. Study design

Propensity score matched comparative cohort studies with two-stage meta-analysis.

5.2. Data sources

The following databases will be used: IQVIA™ Medical Research Data (IMRD) UK, IQVIA™ Disease Analyser Germany, and THIN® Spain and Italy. These databases have been mapped to the OMOP (Observational Medical Outcomes Partnership) CDM (Common Data Model) version 5.3. Brief descriptions of these databases are provided in **Annex 1**.

5.3. Setting and study population

The included study population will consist of adult new users of the anticoagulants of interest (see section on Exposures), as recorded for patients visiting general practices (UK, Germany, Spain, and Italy). Patients will require to have been observed at least once at minimum one year (365 days) prior to entering the cohort. In the UK database, this will be defined as a minimum of 365 days between registration with the practice and the index date. For the other databases, this will be defined as a record of health care system use recorded in the databases at least 365 days prior to the index date.

The population will be stratified by indication for anticoagulation, namely for the management of atrial fibrillation (AF) or for the management of venous thromboembolism (VTE, namely deep vein thrombosis or pulmonary embolism). Indications will be identified by the presence of diagnosis codes within the individual's electronic health record in the year up to the index date. Patients initiating anticoagulation with no diagnostic code for AF or VTE will not be included in the analysis. For patients with multiple indications, the most recent indication (before index date) will be used. Patients with a history of cardiac mechanical valve or mitral stenosis prior to cohort entry will be excluded, as they have an indication for warfarin, not DOACs.

Individuals with a diagnosis of ILD prior to cohort entry (anytime in the patient's prior history) will be excluded from the cohort.

5.4. Study period

The study observation period to identify new users will be from 1st of January 2010, to 31st of December 2019 (i.e., the period before COVID-19 cases emerged).

Cohort entry will be the date of anticoagulant initiation (**index date**). New anticoagulant use will be defined as no prescription for any anticoagulant of interest in the 365 days prior to cohort entry.

Cohort exit will be the earliest of ILD, death (for databases where death is captured), end of individual observability (deregistration or end of practices contributing data where these data are captured), treatment switching (for the on-treatment analysis) or end of the study period (31st December 2019). For the analysis of any DOAC vs VKA, switching to a different DOAC will not be considered a switch of exposure.

5.5. Variables

Outcome:

The outcome of interest will consist of an incident diagnosis of ILD. ILD consists of a heterogenous group of conditions. The ILD phenotype is therefore a composite of different codes representing ILD in general, or as specific conditions that are classed as an ILD. Exclusion of patients with pre-existing ILD will be evaluated using a composite phenotype consisting of any ILD related code. For the outcome in the analysis, a narrower definition of ILD will be used, namely pulmonary fibrosis, pneumonitis, ILD specific codes, and drug induced ILD (please see Appendix for exemplar codes lists). As codes for drug induced ILD are infrequently used, inclusion of this as a specific subgroup may provide limited information. We followed the "narrow" scope list of terms (i.e., very specific preferred terms) documented in the Introductory Guide for Standardised MedDRA Queries (SMQs) Version 25.0 [8]. We mapped these SMQ codes to SNOMED codes and used their OMOP concept IDs, and subsequently removed any codes that indicate another well-defined cause of ILD. Detailed list of codes included are shown in **Annex 2**.

Exposure:

Exposures of interest will consist of VKAs (comparator, consisting of either warfarin, acenocoumarol, phenindione, or phenprocoumon), and DOACs (target, consisting of apixaban, dabigatran, edoxaban and rivaroxaban). Exposures will be identified through prescriptions recorded in the electronic health record. Target exposures will be evaluated by class (any DOAC (direct oral anticoagulants)) and as individual drug substances.

For the on-treatment analysis, patients will be considered continuously exposed from cohort entry (index date) up to 30 days after the date of the last anticoagulant prescription in the patients record. In order to account for different treatment episodes over time, gaps of >90 days between anticoagulant prescriptions will be considered as different treatment episodes and continuous exposure will be censored 30 days after the last prescription before the gap. Detailed list of codes and terms included to identify exposure to DOACs are shown in **Annex 2**.

Other variables:

Baseline risk factors for ILD will be measured including: age, sex, history of pneumonia, chronic obstructive pulmonary disease, chronic kidney disease, hepatitis C, history of tuberculosis sarcoidosis, history of autoimmune diseases, history of connective tissue disorders, history of sarcoidosis, history of cancer, exposure to medications (nitrofurantoin, sulfonamides, amiodarone, methotrexate, cyclophosphamide, dronedarone, SSRIs/SNRI antidepressants , anticonvulsants, antiplatelets, statins, angiotensin converting enzyme inhibitors (ACEI), angiotensin-II receptor blockers (ARB), thiazide diuretics and systemic fluoroquinolone antibiotics [Detailed list is shown in Annex 2]), smoking status. [9, 10] Smoking status may be missing. We will include current smoking status as a binary variable. An individual will be classed as a current smoker (1) or non-current smoker (0, i.e., other smoking status or the absence of a smoking status) based on the most recent smoking status prior to their index date. Risk factors that will not be reliably identifiable include environment and occupational exposures, family history, and biologics, and will not be considered in the analysis. Information on prior radiotherapy and chemotherapy may not be available for some patients but will be included in the analysis when available.

Additionally, the incidence of ILD has changed substantially over time as well as the prescription patterns of the exposures. We will therefore also account for calendar year in the propensity score analysis.

5.6. Statistical analysis

5.6.1. Brief summary of the analysis method (for publication)

For the primary analysis, propensity scores will be used to match new users of the target exposure (DOAC) to new users of the comparator exposure (VKAs), to control for confounding. For the secondary analysis, propensity scores will be used to match new users of the individual target exposures to the comparator exposure (VKAs). For all analyses, cause-specific Cox proportional hazards models will be used to estimate the relative hazard ratio of ILD after exposure. Finally, meta-analysis methods will be used to synthesise results across databases.

5.6.2. Main statistical methods

Propensity score matching will be performed using a 1:1 matching of warfarin users to DOAC users using a caliper width of 0.2 of the standard deviation of the logit of the propensity score. [11] We will include the baseline risk factors and calendar year as predictor variables in the propensity score. Matching will occur by indication (AF and VTE). Absolute standardised mean differences (SMD) for each variable will be calculated before and after propensity score matching, with SMD of less than 0·1 representing successful matching. We will visually check the distributions of the propensity scores across treatment arms, within strata (indication, database). If the sample size appears insufficient after matching, we will use Inverse Probability Treatment Weighting (IPTW) instead of matching.

In the propensity-matched cohort, the incidence of each study outcome will be estimated between target (DOAC) and comparator (warfarin) groups. Cox models will be used to calculate hazard ratios for the instantaneous risk of the ILD outcome. Where death is available in the data source, competing

mortality risk will be accounted for by estimating cause-specific hazard functions. We will apply two analytical approaches: an **on-treatment analysis** (OT) and an **intention to treat analysis** (ITT as per the study by Chan et al.).

Summary of key analysis factors:

- 4 Countries
- 4 DOACs + 1x all DOACS = 5 (warfarin is comparator)
- 2 indications

This brings the number of analyses to 4 * 5 * 2 = 40 (this number will vary depending on the analytical approach and principal findings).

5.6.3. **Meta-analysis**

Hazard ratios will be reported for each database, exposure and indication separately. Log hazard ratios for will also be meta-analysed across databases using a random-effects model using REML (Restricted Maximum Likelihood). 95% Confidence intervals will be provided using the HKSJ method. [12, 13] 95% approximate prediction intervals will be provided. [14] I^2 will be estimated to describe the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error.

5.6.4. **Missing data**

The analysis will be conducted on the assumption that an individual has a condition or prescription exposure if it is recorded in their electronic health record. No multiple imputation methods will be performed. Therefore, patients with missing records for an indication will be excluded from the analyses.

5.6.5. **Sample size**

The sample size will be driven by the availability of individuals with exposures and outcomes within each database and no a priori sample size requirement will be stipulated.

5.6.6. Sensitivity analysis

If we observe imbalances in propensity score distributions among the cohorts after matching that includes calendar year in the model, we will examine whether undertaking calendar time stratified propensity score matching results in improvement for example by undertaking the propensity score matching within calendar year(s) (such as 2010 to 2014 and 2015 to 2019).

5.7. Quality control

The study will be conducted according to the ENCePP code of conduct (European Medicines Agency 2018).

Standard operating procedures or internal process guidance will be adhered to for the conduct of the study. These procedures include rules for secure and confidential data storage, quality-control procedures for all aspects of the study from protocol development to the reporting of the results.

All documents will undergo at least one round a review by an experienced reviewer, while the results from the statistical analysis will be either reviewed or checked via double coding.

The quality control of the data is the responsibility of the data holder.

5.8. Limitations of the research methods

Limitations of this study include the potential for missing data if exposure, outcomes, or covariates are not recorded in an individual's electronic health record. Only primary care databases will be used with no linkage to hospital diagnoses that could lead to under ascertainment of outcomes.

Known risk factors for ILD are used as covariates to estimate the propensity score. However, some risk factors will not be possible to include (e.g., environmental exposure) and unknown risk factors may remain. The analysis will control for baseline observed confounders. Complex time-varying analysis of confounders (e.g., drug exposure) will not be performed. Thus, residual confounding remains possible.

Diagnostic coding for ILD is not known to have been validated in the primary care databases available. Although confirmation of the diagnoses requires specialist input, the nature of the diagnosis means that its recording in primary care records could be reasonably accurate. However, this assumption should be treated cautiously. We will assess narrow and broad definitions in the analyses to account for uncertainties around potential misclassification of ILD.

For privacy of the patients, the exact dates of birth and death are not recorded in the used databases. Instead, year of birth is available, and for patients whose death is registered, year of death is available. This increases the estimates for exposure time.

6. Protection of human subjects

Patient confidentiality will be protected according to the EU General Data Protection Regulation (GDPR) on the protection of individuals.

7. Management and reporting of adverse events/adverse reactions

Pursuant to the requirements for reporting of adverse events for secondary data (GVP module VI, VI.C.1.2.1.2), adverse event reporting will not be conducted as part of this study given the study objectives will be met through the use of secondary data.

8. Plans for disseminating and communicating study results

The analysis plan and study results will be published in EUPAS registries upon completion.

9. References

[1] Wijsenbeek M, Suzuki A, Maher TM. Interstitial lung diseases. Lancet. 2022 Sep 3;400(10354):769-786. doi: 10.1016/S0140-6736(22)01052-2.

[2] Kaul B, Cottin V, Collard HR and Valenzuela C (2021) Variability in Global Prevalence of Interstitial Lung Disease. Front. Med. 8:751181.

[3] Tomari S, Homma K, Noguchi T, Aiba T, Matsuki T, Suzuki R, Koga M, Takigami M, Tagawa H, Hashimoto T, Toyoda K. Development of Interstitial Lung Disease after Initiation of Apixaban Anticoagulation Therapy. J Stroke Cerebrovasc Dis. 2016 Jul;25(7):1767-1769. doi: 10.1016/j.jstrokecerebrovasdis.2016.03.036.

[4] Yanagihara T, Yamamoto N, Kotetsu Y, Hamada N, Harada E, Suzuki K, Ijichi K, Oda Y, Nakanishi Y. Interstitial pneumonia caused by dabigatran. Respir Med Case Rep. 2017 Oct 31;23:10-12. doi: 10.1016/j.rmcr.2017.10.009.

[5] Raschi E, Fusaroli M, Diemberger I, Poluzzi E. Direct Oral Anticoagulants and Interstitial Lung Disease: Emerging Clues from Pharmacovigilance. Drug Saf. 2020 Nov;43(11):1191-1194. doi: 10.1007/s40264-020-00990-9.

[6] Chan YH, Chao TF, Chen SW, Lee HF, Chen WM, Li PR, Yeh YH, Kuo CT, See LC, Lip GYH. Development of Interstitial Lung Disease Among Patients With Atrial Fibrillation Receiving Oral Anticoagulants in Taiwan. JAMA Netw Open. 2022 Nov 1;5(11):e2243307. doi: 10.1001/jamanetworkopen.2022.43307.

[7] Raschi E. Interstitial Lung Disease With Non–Vitamin K Oral Anticoagulants—A Clinical Concern? JAMA Netw Open. 2022;5(11):e2243316. doi:10.1001/jamanetworkopen.2022.43316

[8] MedDRA®. Introductory guide version 25.0. March, 2022.

[9] Spagnolo P, Bonniaud P, Rossi G, Sverzellati N, Cottin V. Drug-induced interstitial lung disease. European Respiratory Journal [Internet]. 2022 Oct 1 [cited 2023 Mar 16];60(4). Available from: https://erj.ersjournals.com/content/60/4/2102776

[10] Choi WI, Dauti S, Kim HJ, Park SH, Park JS, Lee CW. Risk factors for interstitial lung disease: a 9-year Nationwide population-based study. BMC Pulm Med [Internet]. 2018 Jun 4 [cited 2023 Mar 16];18:96. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5987651/

[11] Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharmaceutical statistics. 2011;10(2):150–61.

[12] Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. Statistics in Medicine. 2002 21(21):3153–9. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.1262

[13] Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Statistics in Medicine. 2003;22(17):2693–710. Available from: https://doi.org/10.1002/sim.1482

[14] Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2009;172(1):137–59. Available from: https://https://doi.org/10.1111/j.1467-985X.2008.00552.x

Annexes

Annex 1 - Information on Databases and Healthcare systems included

IQVIA™ Medical Research Data (IMRD) UK

IQVIA™ Medical Research Data (IMRD) UK is a primary care database from the UK. GPs play a gatekeeper role in the healthcare system in the UK, as they are responsible for delivering primary health care and specialist referrals. Over 98% of the UK-resident population is registered with a GP, so that GP patient records are broadly representative of the UK population in general. Patients are affiliated to a practice, which centralizes the medical information from GPs, specialist referrals, hospitalizations, and tests.

IQVIA™ Disease Analyzer Germany

IQVIA™ Disease Analyzer Germany collects computerised information from specialised and general primary care practices throughout Germany since 1992. Around 3% of general practitioners (GP) practices are included, which covers all patients consulting a practice. Data from IQVIA™ Disease Analyzer Germany have been shown to be reasonably representative of German healthcare statistics for demographics and certain diseases and is considered one of the largest national medical databases worldwide. IQVIA™ Disease Analyzer Germany includes more than 2,500 practices and 3,100 physicians (13 speciality groups) representing over 15,000,000 patients. This database used to be named IMS® Disease Analyzer Germany and some use of this terminology may persist.

The quality of IQVIA™ Disease Analyzer data is ensured by a series of continuous QA controls and data refinement. These include checking incoming data for criteria such as completeness and correctness (e.g., linkage between diagnoses and prescriptions) and standardizing certain data values such as laboratory test results to enable reliable analysis.

The Health Improvement Network (THIN®) Italy

In THIN® Italy data collection started in 2000 and this database is currently able to provide clinical monitoring data of anonymised patients managed by 500 GPs in primary care (including patients' history). The data source of THIN® Italy is electronic health care records. The entire database reaches 900,000 patients (active and non-active), from which 500,000 are currently actively followed. In order to be representative at national and macroregional level, physicians have been recruited in accordance with their universe distribution in terms of geography, age and gender.

THIN® is an unobtrusive European medical data collection scheme that collects anonymized patient data from the Electronic Health Records of GPs and specialists, including information on patient's diagnoses, test results and medication. The databases follow a very strict anonymization process. In all countries patients are informed about the collection and anonymization of the data and they are able to opt out, in which case no data are subsequently transmitted to the THIN® database.

The Health Improvement Network (THIN®) Spain

THIN® Spain is mainly a primary care healthcare database, including practitioners (GP), specialists and pediatricians & nurses. It contains data from approximately 2,000 GPs and 2,400 specialists (cardiology, pulmonology, urology, etc.). THIN® Spain also includes partial activities related to the

hospital. THIN® Spain is globally representative of the whole national demographics and prevalence on the main chronic health pathologies. THIN® Spain includes 3,000,000 individuals out of the overall population. Among these, 1,050,000 are active in the previous year and 1,800,000 are active from 2014. Number of deceased patients globally varies between 8 and 9 thousand individuals per year, and number of new-borns ranges between 10 and 12 thousand individuals. New patients are automatically included into the database, and deceased patients identified in a specific field.

THIN® is an unobtrusive European medical data collection scheme that collects anonymized patient data from the Electronic Health Records of GPs and specialists, including information on patient's diagnoses, test results and medication. The databases follow a very strict anonymization process. In all countries patients are informed about the collection and anonymization of the data and are able to opt out, in which case no data are subsequently transmitted to the THIN® database.

The study protocol for this study will be submitted to and approved by Hospital X ethics committee, who reviewed the data collection, protection, and anonymization processes. In addition to this study, THIN® Spain had previously received protocol approval in the scope of other observational study by two ethics committees (Hospital Ramón Cajal, Madrid, and Hospital Clinic, Barcelona).

Annex 2 – Code lists

- List of codes and terms that will be used for narrow and broad ILD definition:

ILD_broad and narrow list_21.06.20

- List of codes and terms that will be used to identify DOACs and Vitamin K Antagonist:

DOACs - RxNorm and RxNorm Extensi

VitKAntagonists_21.06.2021.xlsx

- List of codes and terms that will be used to identify AF, DVT and PE patients:

Indications - 2023 08 03.xlsx

- List of codes and terms that will be used to identify exposure to co-medications:

DOAC-ILD-Covariate-Drugs_21.06.2023.

- List of codes and terms that will be used to identify comorbidities and smoking:

DOAC-ILD-Covariate-Codes 2023 08 03.