Concordance between primary and secondary electronic healthcare databases: A multi-database self-controlled case series study

First published: 18/10/2022

Last updated: 23/04/2024

Study Planned

Administrative details

EU PAS number

EUPAS49386

Study ID

49387

DARWIN EU® study

No

Study countries

- ☐ Netherlands
- ☐ United Kingdom

Study description

There is often mismatch between the recording of diagnoses in primary and secondary electronic healthcare data. Differences may exist in the recorded date of the event or whether it is recorded at all. For example, around two-fifths of all recorded stroke events are in both UK primary and secondary healthcare databases (within 120 days of each other) and around half of these had same-day recordings. The lack of concordance between different electronic health care records, which capture the same population, could lead to outcome misclassification and therefore bias, depending on which data domain is correct and then used in the epidemiologic study. Here we will describe the concordance between primary and secondary electronic healthcare data in the United Kingdom and the Netherlands in the occurrence of major bleeding. Agreement between the data settings, time gap between recordings and occurrence of recordings after recorded death date will be assessed. We will also compare the outcomes identified from different healthcare settings when applied to a self-controlled case series (SCCS) study. This will assess the association of major bleeding and use of direct oral anticoagulants or vitamin K antagonists for atrial fibrillation patients. The incidence rate of the outcome in exposed versus non-exposed time (incidence rate ratio) will be assessed, comparing outcomes derived from the different data domains. The aims of this study are to better inform pharmacoepidemiologic decision making.

Study status

Planned

Research institutions and networks

Institutions

Division of Pharmacoepidemiology & Clinical Pharmacology (PECP), Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University

Netherlands

First published: 01/03/2010

Last updated: 23/05/2024

Institution | Educational Institution | ENCePP partner

Contact details

Study institution contact

Klungel Olaf o.h.klungel@uu.nl

Study contact

o.h.klungel@uu.nl

Primary lead investigator

Hunt Nicholas

Primary lead investigator

Study timelines

Date when funding contract was signed

Planned: 07/10/2022

Study start date
Planned: 01/12/2022

Date of final study report
Planned: 31/08/2023

Sources of funding

- Other

More details on funding

none

Study protocol

databaseconcordance_SCCS_protocol_v2.0.pdf (217.85 KB)

Regulatory

Was the study required by a regulatory body?

No

Is the study required by a Risk Management Plan (RMP)?

Not applicable

Methodological aspects

Study type

Study type list

Study type:

Non-interventional study

Scope of the study:

Assessment of risk minimisation measure implementation or effectiveness

Main study objective:

Objective 1: Describe the concordance between primary and secondary care data in both the United Kingdom and the Netherlands, Objective 2: Compare the incidence of outcomes identified from primary and/or secondary care data in a self-controlled case series study (SCCS) design

Study Design

Non-interventional study design

Other

Non-interventional study design, other

Self-controlled case series

Study drug and medical condition

Study drug International non-proprietary name (INN) or common name

ACENOCOUMAROL

APIXABAN

DABIGATRAN

EDOXABAN

PHENINDIONE

PHENPROCOUMON

RIVAROXABAN

WARFARIN

Medical condition to be studied

Haemorrhage

Population studied

Age groups

- Adults (18 to < 46 years)
- Adults (46 to < 65 years)
- Adults (65 to < 75 years)
- Adults (75 to < 85 years)
- Adults (85 years and over)

Estimated number of subjects

54000

Study design details

Outcomes

1) Percentage overlap of bleeding events occurring in the primary and secondary healthcare data domains, 2) incidence rates of major bleeding using primary and/or secondary care data and 3) Incidence rate ratios of major bleeding in the exposed time (first 30 days or including the remaining length of prescription) versus unexposed (baseline) time comparing primary and/or secondary care data.

Data analysis plan

The baseline characteristics will be stratified by treatment group (DOAC or VKA) and by data source (CPRD Aurum or PHARMO). The baseline period is defined as the unexposed reference period 30 days prior to use of a one of the exposures and unexposed time begins 30 days after the last calculated exposure. Means, standard deviations (SD) and (percentage) totals will be calculated. Median follow-up will be calculated per treatment group in each data source. Incidence rates (IRs) for events occurring within exposed and unexposed intervals will be calculated, along with incidence rate ratios (IRRs) comparing these two periods. The IRR and corresponding 95% confidence interval (CI) will be calculated using conditional Poisson regression. Time-varying confounders which are associated with the exposure and the outcome, such as age, will be accounted for in the adjusted model. The analysis will be stratified by sex (effect modifier).

Data management

ENCePP Seal

The use of the ENCePP Seal has been discontinued since February 2025. The ENCePP Seal fields are retained in the display mode for transparency but are no longer maintained.

Signed checklist for study protocols

Appendix 3 ENCePPChecklistforStudyProtocols_databaseconcordance.pdf (177.56 KB)

Data sources

Data source(s)

Clinical Practice Research Datalink

PHARMO Data Network

Data sources (types)

Drug dispensing/prescription data

Electronic healthcare records (EHR)

Other

Data sources (types), other

Routine secondary care electronic patient registry

Use of a Common Data Model (CDM)

CDM mapping

No

Data quality specifications

Check conformance

Unknown

Check completeness

Unknown

Check stability

Unknown

Check logical consistency

Unknown

Data characterisation

Data characterisation conducted

No