UNICOM

Project acronym: UNICOM
Project full title: Up-scaling the global univocal identification of medicines
in the context of Digital Single Market strategy
Call identifier: H2020-SC1-DTH-2019

Deliverable 8.3:
Protocol and Statistical Analysis Plan for Proof of Concept Studies

Version:	1.0
Status:	Final
Dissemination Level[1]:	PU
Due date of deliverable:	30.11.2020
Actual submission date:	30.11.2020
Work Package:	WP8: Clinical care, patients, pharmacies and pharmacovigilance
Lead partner for this deliverable:	I~HD
Partner(s) contributing:	I~HD
Deliverable type[2]:	R

Main author(s):

Miriam Sturkenboom	i_HD
Carlos E. Durán	i~HD
Robert Vander Stichele	i~HD

---

[1] Dissemination level: PU: Public; CO: Confidential, only for members of the consortium (including the Commission Services); EU-RES: Classified Information: RESTREINT UE (Commission Decision 2005/444/EC); EU-CON: Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC); EU-SEC Classified Information: SECRET UE (Commission Decision 2005/444/EC)

[2] Type of the deliverable: R: Document, report; DEM: Demonstrator, pilot, prototype; DEC: Websites, patent fillings, videos, etc.; OTHER; ETHICS: Ethics requirement; ORDP: Open Research Data Pilot

Revision history

Version	Date	Changes made	Author(s)
0.1	26.10.2020	First draft	Miriam Sturkenboom, Robert Vander Stichele. Carlos E. Durán
0.2	29.10.2020	Second draft with diclofenac protocol	Miriam Sturkenboom, Robert Vander Stichele. Carlos E. Durán
0.3	02.11.2020	Third draft with addition of study abstracts	Miriam Sturkenboom, Robert Vander Stichele. Carlos E. Durán
0.4	30.11.2020	Fourth draft with corrections from internal revision	Miriam Sturkenboom, Robert Vander Stichele. Carlos E. Durán
1.0	30.11.2020	Last check and submission	EMP

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Deliverable abstract

This deliverable describes the approach plus a protocol of a potential proof of concept studies to demonstrate the value of IDMP in real world electronic health data.

IDMP will allow for better identification of the specificities of medicinal products, something that in current pharmacoepidemiological studies that are based on secondary use of electronic health care data is not feasible. Most analyses are on ATC level.

This protocol based on the EMA template for protocols aims to demonstrate the need to be able to distinguish between different salts of the same active compound. It is focusing on a multi-country study capitalizing on the IMI-funded ConcePTION project and common data model. In IMI-ConcePTION multiple data access providers to large electronic health care databases participate.

Implementation of such a protocol would yield information on
1) Therapeutic arsenal in different European countries
2) Ability and difficulties to identify specific products in health care databases
3) Added value to be able to compare different formulations of the same substance (therapeutic moiety).

Keywords: big data, common data model, pharmaceutical product, safety, pharmacovigilance

TABLE OF CONTENTS

LIST OF TABLES

List of abbreviations

Abbreviation	Complete form
ATC	Anatomic, Therapeutic and Chemical Classification of Drugs
CDM	Common Data Model
DPA	Database Performance Analyzer
ETL	extracted, transformed, and loaded
EMA	European Medicines Agency
IMI	Innovative Medicines Initiative
ISO	International Organization for Standardization
IDMP	the Identification of Medicinal Products
NSAID	Non-Steroidal Anti-inflammatory Drug
OMOP	Observational Medical Outcomes Partnership

1   Executive summary

In 2012, the International Organization for Standardization (ISO) published a series of standards and technical specifications called together the Identification of Medicinal Products (IDMP).
The European funded project Up-scaling the Global Univocal Identification of Medicines (UNICOM) aims to give an ultimate impulse to the implementation of ISO IDMP among drug databases of European countries and therefore allow a safe cross-national e-prescription and pharmacovigilance activities. As a first step, UNICOM's technical experts have selected a list of 35 medicinal products (pilot product list) in order to agree and define common identification concepts.

UNICOM's work package 8 (task 8.2: Application of IDMP in big data for science) aims to evaluate the added value of applying the ISO IDMP in biomedical research, mainly in the fields of pharmacoepidemiology, drug utilization research and pharmacovigilance. We provide an overview of the ongoing American and European initiatives for distributed analytics to re-use health data, and give their common data model structure and detail on the medicines tables.

Key questions to be answered for UNICOM in this proof of concept study are:

1) **What is the situation without IDMP: What is the current therapeutic arsenal/variety in Europe?**
   - For the UNICOM shortlisted products can we assess through article 57 database at EMA, national drug dictionaries, or the products table in the ConcePTION data sources how many different pharmaceutical alternatives are available across different countries?

2) **Would it be possible to map to 'IDMP' type level of detail in data sources using the ConcePTION Common Data Model (CDM) and can we demonstrate the added value of analyzing data at that level of detail?**
   - What are current possibilities in data sources to arrive at IDMP level detail. With existing data sources.
   - Would ICMP level of detail allow us to make more refined analysis and compare different medical products of the same active ingredient?

We have developed a full protocol based on re-use of big electronic health data (collected for health care administration purposes rather than dedicated research). We list data sources that have been mapped to the ConcePTION common date model, covering 11 countries and health data on 159 M subjects. These could be used for any of the proof of concept studies.

We have developed a full protocol according to the European Medicines Agency template for post authorization safety studies to look at question 2 (delivered as annex 1). We also developed outlines for other proof of concept studies related to amlodipine esters and effectiveness, COVID-19 vaccines and safety, and description of the national therapeutic arsenals. In the main text of the deliverable we describe the problem and current approaches to analysing a large data and the need for use of common data models as well as the type of data sources and organizations that have transformed their data in the ConcePTION CDM.

2 Need for collaboration & harmonization to evaluate medicines

The task 8.2: Application of IDMP in big data for science of UNICOM is charged with the creation of proof of concept studies that show the value of the IDMP.

Big data[3] is defined by the five V's: Veracity, Volume, Velocity, Value, Variety. Because of increasing computerization of health care, cheaper processing power and storage, big data analytics on routine health care data has become the promise to generate a true learning healthcare system. Large health care databases have been used to study the use and outcomes of therapeutics since the 1980s[4]. Their size allows the study of infrequent events, their representativeness of routine clinical care makes it possible to study real-world safety, effectiveness and utilization patterns, and their availability at relatively low cost without long delays makes them accessible[5] to many researchers[6].

A recent review of post-authorization safety studies registered in the European Postauthorization Study (EU PAS) register showed that between 30 and 50% of studies use these type of data[7], it is the assumption that this will increase according to a recent Head of Medicines Agency report. One of the key hurdles to more rapid implementation of real-world evidence into decision making is the limited penetration of interoperability standards and the difficulties encountered in heterogeneity of local data models and structures

2.1 Need to collaborate to evaluate use and effects of medicines

In many high-income countries regulatory agencies are realizing the benefits of accessing and use real world data. Below we describe examples of such systems

2.1.1 Europe

In Europe multinational collaboration started with the EU-ADR project (Exploring and Understanding Adverse Drug Reactions by Integrating Mining of Clinical Records and Biomedical Knowledge). This project was funded by the European Commission seventh framework program in 2008. This project built methods, tools and a simple common data model to implement distributed analysis of the use and safety of therapeutics in Europe. Based on the EU-ADR experience, the European Center for Disease Control & Prevention (ECDC) wished to bring the expertise to the vaccine area in the VAESCO project in 2009 to monitor pandemic vaccine safety (Vaccine Adverse Event Surveillance and Communication).

The VAESCO project demonstrated the ability of EU countries to use existing data and work in a distributed manner to assess vaccine safety issues. Because there was no subsequent funding the EU-ADR and VAESCO projects were both terminated in 2012. Since then several European Commission funded studies have utilized the approaches, while addressing real public health concerns such as safety of NSAIDS (SOS), arrhythmogenic characteristics of drugs (ARITMO), safety of diabetes drugs (SafeGUARD), across these projects and across infrastructural projects such as EMIF and ADVANCE, the tools and methodologies to deal with heterogeneous data in Europe has evolved[8], but none are yet implemented in a sustainable approach to support medicines monitoring.

---

[3] Refer to summary report of the HMA-EMA Joint Big Data Taskforce: Section 4, regulatory definition of 'Big Data' https://www.ema.europa.eu/en/documents/minutes/hma/ema-joint-task-force-big-data-summary-report_en.pdf

[4] Andrews, E.B., Margulis, A.V., Tennis, P. et al. Curr Epidemiol Rep (2014) 1: 194. https://doi.org/10.1007/s40471-014-0026-0

[5] Weng C, Appelbaum P, Hripcsak G, Kronish I, Busacca L, Davidson KW, et al. Using EHRs to integrate research with patient care: promises and challenges. J Am Med Inform Assoc JAMIA. 2012 Oct;19(5):684–7.

[6] J Clin Epidemiol. 2005 Apr;58(4):323-37. A review of uses of health care utilization databases for epidemiologic research on therapeutics. Schneeweiss S, Avorn J.

[7] Carroll R, Ramagopalan SV, Cid-Ruzafa J, et al. An analysis of characteristics of post-authorisation studies registered on the ENCePP EU PAS Register. F1000Res 2017;6:1447.

[8] Trifirò G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, Bonhoeffer J, Schuemie M, van der Lei J, Sturkenboom M. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? J Intern Med. 2014 Jun;275(6):551-61.

The EMA-HMA Big Data Taskforce published its report and recommendations on 20th January 2020 on, 'Evolving Data-Driven Regulation' and proposes ten recommendations, in particular a network proposition, DARWIN ('Data Analysis and Real World Interrogation Network')[9]

1. Deliver a sustainable platform to access and analyse healthcare data from across the EU (DARWIN)
2. Establish an EU framework for data quality and representativeness
3. Enable data discoverability
4. Develop EU network skills in Big Data
5. Strengthen EU network processes for Big Data submissions
6. Build EU network capability to analyse Big Data
7. Modernise the delivery of expert advice
8. Ensure data are managed and analysed within a secure and ethical governance framework
9. Collaborate with international initiatives on Big Data
10. Create an EU Big Data 'stakeholder implementation forum'

It shows that EMA is committed to implement a multi-country/database approach to inform regulatory decision making.

2.1.2   USA

The USA built the first distributed multisite system to monitor vaccine safety, the Vaccine Safety Datalink (VSD). The VSD is a collaboration between nine different health maintenance organization and the Center of Disease Controls Immunization Safety Office, in operation since 1990. The pioneering accomplishments of the VSD demonstrated the synergy of collaboration for methods and evidence generation, and the use of distributed data approaches for secure analysis of de-identified data.

Large scale re-use of health data to support medicines safety monitoring was further implemented in the USA through the FDA amendment act, following the reform in pharmacovigilance that was recommended by the US Institute of Medicine (IoM) during 2004-2008[10] after the rofecoxib safety concerns, which demonstrated that large scale active surveillance is needed). The IoM underlined the need for large scale analytics using existing big health care data and led to the building of US FDA Sentinel[11] and OMOP (Observational Medical Outcomes Project)[12].

In the fall of 2007, Congress passed the FDA Amendments Act (FDAAA), mandating the FDA to establish an Active Postmarket Risk Identification and Analysis (ARIA). FDAAA requires the FDA to develop, in collaboration with public, academic, and private entities, methods to obtain access to disparate data sources and validated methods for the establishment of a system to link and analyze safety data from multiple sources. In May 2008, the FDA launched the Sentinel Initiative to create a national electronic system, the Sentinel System, for medical product safety surveillance. Sentinel has the largest multisite distributed database in the world dedicated to medical product safety. It is constantly growing and improving to meet FDA's needs[13].

---

[9] https://www.ema.europa.eu/en/documents/other/hma-ema-joint-big-data-taskforce-phase-ii-report-evolving-data-driven-regulation_en.pdf

[10] Committee on the Assessment of the US Drug Safety System, Baciu A, Stratton K, Burke SP, eds. The future of drug safety: promoting and protecting the health of the public. Washington, DC: National Academies Press, 2006. (Accessed April 30, at http://www.iom.edu/CMS/3793/26341/37329.aspx.)

[11] Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new Sentinel Network--improving the evidence of medical-product safety. N Engl J Med. 2009 Aug 13;361(7):645-7.

[12] Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012 Dec 30;31(30):4401-15.

[13] As per Sentinel website: https://www.fda.gov/safety/fdas-sentinel-initiative/fdas-sentinel-initiative-background

---

2.1.3 Canada

The Canadian Institutes of Health Research have implemented the Drug Safety and Effectiveness Network (DSEN) a coordinated national network of over 200 researchers committed to the highest excellence in post-market drug safety and effectiveness research. examples of identification of medicinal products in multi-country studies. Before DSEN the majority of studies on drug safety and were conducted as individual initiatives on a single provincial health care database.

As these individual initiatives are often based on relatively small populations, they are limited in what they can achieve – especially when looking for rare serious adverse events, for the study of drugs used to treat infrequent diseases, or for the study of the effects of drugs in new users. DSEN decided to addressing these concerns by creating a pan-Canadian collaboration of researchers, the Canadian Network for Observational Drug Effect Studies (CNODES).The overarching aim of CNODES is to use collaborative, population-based approaches to provide rapid answers to questions about drug safety and effectiveness. The CNODES network includes the health and prescription records of over 40 million people to rapidly evaluate the risks and benefits of drugs on the health of Canadians[14].

2.2 Current medicinal product tables in Common Data Models

Since the format of data tables that are created for support of health care changes between organizations and across countries there is a need to harmonize both the structure(syntactic) as well as the meaning (syntactic harmonization) of the variables. This is also referred to as common data models. While there are countless study-specific common data models designed for one-time use, common data models designed for reuse within a network or community of researchers take on only a limited number of forms, each with one or two quintessential examples of the form in common usage. These are described below. We will focus on the way medicinal products are harmonized and can be identified.

2.2.1 Vaccine Safety Datalink

The common data model employed by the Vaccine Safety Datalink is an example of a syntactically (structurally) harmonized common data model with limited scope, in which only a limited set of variables relevant to vaccine safety are extracted, transformed, and loaded (ETL) to the CDM. The CDM comprises the following tables: Patient (Demographics and enrolment), Vaccination History (Vaccination dates, types, and manufacturers), Medical Visits (Healthcare encounters and diagnoses), Mortality (Death data), and Birth and Pregnancy (Pregnancy and birth data on mother and child). While high in derivation, it has proven utility to address vaccine safety concerns rapidly[15].

The vaccines table comprises the following information

Table 1: information in vaccines table of VSD CDM

Variable	explanation
StudyID:	study site
CDCsite:	HMO site
VACDATE:	date of vaccine administration
VAC:	Vaccine administered (numbered specific to number and type of antigens), no product or brand names
FACILITY:	facility in which vaccine was administered
SITE:	Body site of administration
MFR:	manufacturer
LOT:	Lot number
VACSOURC	location of administration

---

[14] https://www.ices.on.ca/Research/Research-programs/Chronic-Disease-and-Pharmacotherapy/CNODES

[15] https://www.cdc.gov/vaccinesafety/pdf/vsd-data.pdf

2.2.2   Sentinel

The Sentinel CDM is an example of a CDM which is high in reusability, low in derivation, and somewhat broad in scope. The Sentinel Common Data Model is a product of the United States Food and Drug Administration Sentinel Initiative (https://www.sentinelinitiative.org/) and comprises the following tables: Enrolment (periods of health plan enrolment), Demographic (demographic characteristics), Dispensing (outpatient pharmacy dispensing), Encounter (healthcare encounters), Diagnosis (in and outpatient diagnoses), Procedure (in and outpatient procedures), Death (Death records), Cause of Death (Causes of death related to a death record), Laboratory Result (Results of laboratory tests), Vital Signs (Results of measurements), Inpatient Pharmacy (Inpatient drug administrations), Inpatient Transfusion (Inpatient transfusion administration), and Mother-Infant Linkage (Linkage between mothers and live-born infants).

Data in the Sentinel CDM is developed for the United States and is primarily administrative and claims data from health insurers, collected for reimbursement purposes. Source data is harmonized to a common vocabulary for a subset of variables but for the most part the Sentinel CDM retains source data in its original format.

The Dispensing data in version 7.1.0 comprise the following information[16]

Table 2: Sentinel CDM dispensing table

Variable Name	Variable Type and Length (Bytes)	Values	Definition / Comments / Guideline	Example
PatID[3]	Char (Site specific length)	Unique member identifier	Arbitrary person-level identifier. Used to link across tables.	123456789012345
RxDate	Numeric (4)	SAS date	Dispensing date (as close as possible to date the person received the dispensing).	11/29/2005
NDC	Char (11)	National Drug Code	Please expunge any place holders (e.g., '-' or extra digit).	00006007431
RxSup[2]	Numeric (4)	Days supply	Number of days that the medication supports based on the number of doses as reported by the pharmacist. This amount is typically found on the dispensings record. It should not be necessary to calculate this variable for use in the SCDM. Positive integer values are expected.	30
RxAmt[2]	Numeric (4)	Amount dispensed	Number of units (pills, tablets, vials) dispensed. Net amount per NDC per dispensing. This amount is typically found on the dispensings record. It should not be necessary to calculate this variable for use in the SCDM. Positive values are expected.	60

---

[16]
    https://dev.sentinelsystem.org/projects/SCDM/repos/sentinel_common_data_model/browse/files/file0012_admin_dispensing.md?at=refs%2Fheads%2FSCDM7.1.0

2.2.3   OMOP

The OMOP CDM is a product of the Observational Medical Outcomes Partnership (OMOP), now OHDSI (Observational Health Data Sciences and Informatics). The semantic harmonization of this CDM to a common set of vocabularies, terminologies, and coding schemes allows for deployment of analysis scripts against data in the common data model with less extensive definition and construction of variables at the analysis stage. This means that the ETL of the source data to standardized concepts may not be conducted in a uniform and transparent manner across data sources. The OMOP CDM is extensive and includes the following tables among others: Person, Observation Period (time periods of observation), Specimen (Biological samples), Death (Causes of death), Visit Occurrence (Outpatient, inpatient, emergency, and long-term care visits), Visit detail (detailed data related to each visit occurrence), Procedure Occurrence (Procedures ordered or carried out), Drug exposure (drug utilization), Device exposure (device utilization), Condition Occurrence (Diagnoses), Measurement (Measurement results), Note (unstructured information), Observation (observations not recorded in other tables), Location (Physical location of care site), Care Site (Health care units), Provider (healthcare provider), and Drug Era (exposure periods).

The drug exposure table is comprising the following information

Table 3 Variables about medicines in OMOP drug exposure table[17]

CDM Field	User Guide
drug_exposure_id	The unique key given to records of drug dispensings or administrations for a person. Refer to the ETL for how duplicate drugs during the same visit were handled.
person_id	The PERSON_ID of the PERSON for whom the drug dispensing or administration is recorded. This may be a system generated code.
drug_concept_id	The DRUG_CONCEPT_ID field is recommended for primary use in analyses, and must be used for network studies. This is the standard concept mapped from the source concept id which represents a drug product or molecule otherwise introduced to the body. The drug concepts can have a varying degree of information about drug strength and dose. This information is relevant in the context of quantity and administration information in the subsequent fields plus strength information from the DRUG_STRENGTH table, provided as part of the standard vocabulary download.
drug_exposure_start_date	Use this date to determine the start date of the drug record.
drug_exposure_start_datetime	
drug_exposure_end_date	The DRUG_EXPOSURE_END_DATE denotes the day the drug exposure ended for the patient.
drug_exposure_end_datetime	
verbatim_end_date	This is the end date of the drug exposure as it appears in the source data, if it is given
drug_type_concept_id	You can use the TYPE_CONCEPT_ID to delineate between prescriptions written vs. prescriptions dispensed vs. medication history vs. patient-reported exposure, etc.
stop_reason	The reason a person stopped a medication as it is represented in the source. Reasons include regimen completed, changed, removed, etc. This field will be retired in v6.0.
refills	This is only filled in when the record is coming from a prescription written this field is meant to represent intended refills at time of the prescription.
quantity	
days_supply	

---

[17] https://ohdsi.github.io/CommonDataModel/cdm60.html#drug_exposure

sig	This is the verbatim instruction for the drug as written by the provider.
route_concept_id	
lot_number	
provider_id	The Provider associated with drug record, e.g. the provider who wrote the prescription or the provider who administered the drug.
visit_occurrence_id	The Visit during which the drug was prescribed, administered or dispensed.
visit_detail_id	The VISIT_DETAIL record during which the drug exposure occurred. For example, if the person was in the ICU at the time of the drug administration the VISIT_OCCURRENCE record would reflect the overall hospital stay and the VISIT_DETAIL record would reflect the ICU stay during the hospital visit.
drug_source_value	This field houses the verbatim value from the source data representing the drug exposure that occurred. For example, this could be an NDC or Gemscript code.
drug_source_concept_id	This is the concept representing the drug source value and may not necessarily be standard. This field is discouraged from use in analysis because it is not required to contain Standard Concepts that are used across the OHDSI community, and should only be used when Standard Concepts do not adequately represent the source detail for the Drug necessary for a given analytic use case. Consider using DRUG_CONCEPT_ID instead to enable standardized analytics that can be consistent across the network.
route_source_value	This field houses the verbatim value from the source data representing the drug route.
dose_unit_source_value	This field houses the verbatim value from the source data representing the dose unit of the drug given.

Data access providers are asked to map the drug concepts to RxNorm

2.2.4   ConcePTION

In 2019 the Innovative Medicines Initiative (IMI) funded the ConcePTION project which has a shared vision that there is a societal obligation to radically and rapidly reduce uncertainty about the safety of medication use in pregnancy and breastfeeding. It aims to build a sustainable ecosystem for generation around safety of medicines during pregnancy. Generic electronic health data across many countries is used to transform into evidence. A CDM that in its basics is quite similar to OMOP (in terms of tables) but it documents better the provenance of data and DAPs do only a structural harmonization and semantic harmonization (creation of study variables and harmonizing those) is done per study by the study team plus DAP, in a transparent manner[18].

Building on the 10 years of experience in working on common data models in Europe a generic, comprehensive common data model was created that is used in the IMI-ConcePTION project that is already is applied to all EMA tendered studies and is suitable for many other areas of pharmacoepidemiology. Currently 30 Data access providers across Europe are conducting an ETL to this CDM. It allows for a structured and re-usable approach to transforming health data from various sources into a common format and allowing for full transparency and flexibility in the creation of study variables as well as keeping focus on what is needed for the specific research questions. The fully CDM is publicly available(13).

---

[18] https://www.imi-conception.eu/wp-content/uploads/2020/10/ConcePTION-D7.5-Report-on-existing-common-data-models-and-proposals-for-ConcePTION.pdf

Table 4 The ConcePTION MEDICINES table collects data on drug prescriptions, dispensings or administrations occurred during routine healthcare for ConcePTION

Variable	Description	Format	Example1 (SNDS dispensing)
person_id	A foreign key to the person in "person" table who received the drug	Local code string	[PATIENT_ID] 35554
date_dispensing	Date when the drug that lead to the recording was dispensed or administrated to the patient	Character yyyymmdd	[DATE_MEDICAMENT_DELIV] 20130313
date_prescription	Date when the drug that lead to the recording was prescribed	Character yyyymmdd	[DATE_MEDICAMENT_PRESCRI] 20121204
disp_amount_drug	The quantity of drug dispensed or administrated as recorded in the original record.	Numeric	[QUANTITE_MED] 1
disp_amount_drug_unit	Unit caracterizing the quantity or drug dispensed or administrated	Character	package
presc_units_per_day	The posology for the drug as recorded in the original prescription	Character	
presc_duration	The duration of the treatment as defined in the original prescription in days	Numeric	
product_code	A foreign key to the drug details in "product_code" table	Character	[CIP13] 3400921620726
product_ATCcode	The code caracterizing the drug in the Anatomical Therapeutic Chemical classification	Character	[ATC] C09DA03
code_indication	Indication for which the drug was prescribed	Character	
code_indication_vocabulary	Vocabulary to which the 'code_indication' belongs to	Character	
meaning_of_drug_record	nature of the original record having originated the drug record	Character	dispensing_in_community_pharmacy
origin_of_drug_record	origin of the original record having originated the drug record	Character	community_pharmacy_reimbursement'
prescriber_type	Indicates the speciality of the physician who prescribed the drug	Character	[SPE_PRES] General practitioner
visit_occurrence_id	A foreign key linking this record to the "visit occurrence" table	Number	
product_lot_number	An identifier assigned to a particular quantity or lot of Drug product from the manufacturer.	Character	

ConcePTION also has a products table that can be linked and store more data

Table 5. ConcePTION CDM collects the information associated to each marketed product that may have been prescribed, dispensed or administered to a patient. It contains one row per product

Variable	Description	Format	Example1
product_code	Unique identifier of a drug package, from the national authorization code	Character	[CIP13] 3400921607079

full_product_name	The full label of the drug package defined by the product_code	Character	[PHA_PRD_LIB1] CLARITHROMYCINE PFIZER 250 MG 1 BOITE DE 10, COMPRIMES PELLICULES
box_size	Number of units per drug package, or total quantity in case of multi-dose single package	integer	[PHA_UNT_NBR_DSES] 10
box_size_unit	Unit of measure characterizing the box size (e.g. tablets or injections) or the total quantity (e.g. ml, g)	Character	derived from [PHA_UNT_NBR_DSES]
drug_form	Characterize the galenic form of the product unit	Character	[PHA_FRC_COD] 307 'COMPRIME PELLICULE'
route_of_administration	Caracterize the route of administration of the product unit	Character	
product_ATCcode	The quantity of the active ingredient in one single product unit, or the concentration for multi-dose single package	Character	[PHA_ATC_L07] J01FA09
ingredient1_ATCcode	The code characterizing the first active ingredient of the product in the Anatomical Therapeutic Chemical classification	Character	[PHA_ATC_L07] J01FA09
ingredient2_ATCcode	The code characterizing the second active ingredient of the product in the Anatomical Therapeutic Chemical classification	Character	
ingredient3_ATCcode	The code characterizing the third active ingredient of the product in the Anatomical Therapeutic Chemical classification	Character	
amount_ingredient1	The quantity of the first active ingredient in one single product unit, or the concentration for multi-dose single package	Character	[PHA_DOS_PRA_DSES] 250
amount_ingredient2	The quantity of the second active ingredient in one single product unit, or the concentration for multi-dose single package	Character	
amount_ingredient3	The quantity of the third active ingredient in one single product unit, or the concentration for multi-dose single package	Character	
amount_ingredient1_unit	Unit of measure characterizing the quantity of the first active ingredient in one single	Character	[PHA_UNT_PRA_DSES] mg

	product unit, or the concentration		
amount_ingredient2_unit	Unit of measure characterizing the quantity of the second active ingredient in one single product unit, or the concentration	Character	
amount_ingredient3_unit	Unit of measure characterizing the quantity of the third active ingredient in one single product unit, or the concentration	Character	
product_manufacturer	Manufacturer of the drug package defined by the product code	Character	

As you may note, difference with ConcePTION CDM and OMOP CDM for this table is the use of meaning_of_drug_record, origin_of_drug_record and the link to a products table that would have the details of the national products to be utilized.

3 How to demonstrate the value of IDMP without having it?

3.1 Questions to be addressed

In Europe most of the initial efforts in EU-ADR have focused on dealing with heterogeneity of event coding, since many different terminologies and languages are used throughout Europe[19]. Throughout subsequent EC funded projects a solution was found based on the use of the unified medical language system (UMLS) as a basis to map across terminologies with different levels of granularity. This has led to the development of a Codemapper tool[20].

Because of the availability of ATC codes in most data sources, less attention has been given to harmonization of medicines. Several ad hoc solutions have been implemented to overcome the heterogeneity, mostly by mapping drugs to ATC codes, but this means a tremendous loss of detail about the pharmaceutical product, such as route of administration, strength and excipients and difficulties in coding of combination products[21]. In the EC funded project on safety of NSAIDs, planned dose and duration analysis to compare gastro-intestinal and cardiovascular safety of NSAIDs could only be performed in few data sources because of lack of information, most analysis focused on active ingredient only[22], [23] [24]

The IMI-funded ADVANCE project that implemented a big data analytics system to monitor vaccine coverage, benefits and risks in Europe demonstrated that mapping of recorded vaccines to ATC codes was not even possible, since vaccines are represented very differently in different systems.

Implementation of the IDMP may greatly enhance and facilitate the abilities to conduct multisite studies on health care data across the EU and globally, its inclusion in an ontology will allow for increased opportunities for analysis and knowledge discovery e.g. aggregation and studying drugs/vaccines on different levels, e.g. based on common excipients, mechanisms of action, or unintended activities. Since the IDMP is not yet rolled out we can only demonstrate the potential value of the IDMP around several key questions

---

[19] Avillach P, Coloma PM, Gini R, Schuemie M, Mougin F, Dufour JC, Mazzaglia G, Giaquinto C, Fornari C, Herings R, Molokhia M, Pedersen L, Fourrier-Réglat A, Fieschi M, Sturkenboom M, van der Lei J, Pariente A, Trifirò G; EU-ADR consortium. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. J Am Med Inform Assoc. 2013 Jan 1;20(1):184-92. doi: 10.1136/amiajnl-2012-000933. Epub 2012 Sep 6. PMID: 22955495; PMCID: PMC3555316.

[20] Becker BFH, Avillach P, Romio S, van Mulligen EM, Weibel D, Sturkenboom MCJM, Kors JA; ADVANCE consortium. CodeMapper: semiautomatic coding of case definitions. A contribution from the ADVANCE project. Pharmacoepidemiol Drug Saf. 2017 Aug;26(8):998-1005. doi: 10.1002/pds.4245. Epub 2017 Jun 28. PMID: 28657162; PMCID: PMC5575526

[21] Trifirò G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, Bonhoeffer J, Schuemie M, van der Lei J, Sturkenboom M. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? J Intern Med. 2014 Jun;275(6):551-61.

[22] Arfè A, Scotti L, Varas-Lorenzo C, Nicotra F, Zambon A, Kollhorst B, Schink T, Garbe E, Herings R, Straatman H, Schade R, Villa M, Lucchi S, Valkhoff V, Romio S, Thiessard F, Schuemie M, Pariente A, Sturkenboom M, Corrao G; Safety of Non-steroidal Anti-inflammatory Drugs (SOS) Project Consortium. Non-steroidal anti-inflammatory drugs and risk of heart failure in four European countries: nested case-control study. BMJ. 2016 Sep 28;354:i4857. doi: 10.1136/bmj.i4857. PMID: 27682515.

[23] Masclee GMC, Straatman H, Arfè A, Castellsague J, Garbe E, Herings R, Kollhorst B, Lucchi S, Perez-Gutthann S, Romio S, Schade R, Schink T, Schuemie MJ, Scotti L, Varas-Lorenzo C, Valkhoff VE, Villa M, Sturkenboom MCJM. Risk of acute myocardial infarction during use of individual NSAIDs: A nested case-control study from the SOS project. PLoS One. 2018 Nov 1;13(11):e0204746. doi: 10.1371/journal.pone.0204746. PMID: 30383755; PMCID: PMC6211656.

[24] Schink T, Kollhorst B, Varas Lorenzo C, Arfè A, Herings R, Lucchi S, Romio S, Schade R, Schuemie MJ, Straatman H, Valkhoff V, Villa M, Sturkenboom M, Garbe E. Risk of ischemic stroke and the use of individual non-steroidal anti-inflammatory drugs: A multi-country European database study within the SOS Project. PLoS One. 2018 Sep 19;13(9):e0203362. doi: 10.1371/journal.pone.0203362. PMID: 30231067; PMCID: PMC6145581.

These questions are:

What is the situation without IDMP: What is the current therapeutic arsenal/variety in Europe?
- For the UNICOM shortlisted products can we assess through article 57 database at EMA or the products table in the ConcePTION data sources how many different pharmaceutical alternatives are available across different countries?

Would it be possible to map to 'IDMP' type level of detail in data sources using the ConcePTION CDM and can we demonstrate the added value of analyzing data at that level of detail?
- What are current possibilities in data sources to arrive at IDMP level detail. With existing data sources.
- Would ICMP level of detail allow us to make more refined analysis and compare different medical products of the same active ingredient?

To address these two key questions around the value of the IDMP we have developed:
- A full protocol based on the questions and outcomes of the Safety of NSAIDs study, but in the proposed study we would compare the cardiovascular outcomes of two different pharmaceutical alternatives of diclofenac (see Annex 1)
- Three outlines of studies one focusing on the therapeutic arsenal (question1) one on different pharmaceutical alternatives of amlodipine and effectiveness, and one on COVID-19 vaccines. (see Annex 2)

These studies could be implemented with data sources in Europe that are mapped to the ConcePTION CDM, or potentially also to the OMOP CDM, if source values have been retained.

3.2   Description of data sources mapping to ConcePTION CDM

Based on publicly available deliverables from IMI-ConcePTION and protocols for EMA tendered research in the EU PAS register we could retrieve information on the following data sources that are converting to ConcePTION CDM and using the model to address medicines utilizations and safety studies. The DAPs may be requested to participate in any of the proof of concept studies to address the UNICOM questions. In total data from 11 countries and a population size of 159 million citizens has been mapped to the ConcePTION CDM, by 24 different organizations

Table 6: Overview of European population based electronic health care data sources and DAPs that are already working with ConcePTION CDM in granted projects

Country	DAP Institute	Type of data source	Persons (*10^6)	Studies
Denmark	University Aarhus	Danish national registers & dispensing	5.8	ConcePTION, ACCESS
	University Copenhagen			Valproate
Germany	Leibniz Institute for Prevention Research and Epidemiology	Claims data health insurance & dispensing	20	ACCESS, ConcePTION
Finland	National Institute for Health and Welfare, Finland	Finnish registers & dispensing	5.5	ConcePTION

France	Centre Hospitalier Universitaire de Toulouse	French National Claims data (SNIIRAM) & dispensing	66	ConcePTION
	University of Bordeaux (population health)			Retinoids
	University of Bordeaux (BPE)			ConcePTION ACCESS
Italy	Università degli Studi di Ferrara – University of Ferrara	Record linkage of regional registers & claims data Emilia Romagna & dispensing	4.5	ConcePTION
	ARS Tuscany	Record linkage of regional registers & claims data Tuscany & dispensing	4	ConcePTION, ACCESS, Retinoids, Valproate,
	CNR Tuscany			ConcePTION
	Pedianet	Medical records primary care pediatrician & prescribing	0.2	ACCESS
	University Messina	Record linkage of regional registers & claims data province Caserta	1	RETINOIDS
Norway	University of Oslo	Record linkage of Norwegian registers	5.3	ConcePTION ACCESS
Netherlands	PHARMO Institute	PHARMO record linkage of Dutch registers & dispensings	3	ConcePTION, Retinoids, Valproate
	University Medical Center Utrecht			ACCESS
Sweden	Karolinska Institute	Swedish national registers & dispensings	9	ConcePTION, CONSIGN
Spain	Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana	Valencian GP and regional registers (dispensing & vaccines)	5	ConcePTION Retinoids ACCESS
	IDIAP-Jordi Gol	Catalunya GP & linked registers	5	ConcePTION ACCESS
	Spanish Medicines Agency	BIFAP Sample of GPs across the country	8	Retinoids, ACCESS Valproate
UK	University of Dundee/ISD	Scottish Registers	5.5	ConcePTION
	University of Swansea	SAIL data base GP & registers in Wales	5	ConcePTION
	GlaxoSmithKline	CPRD database of GPs & linked registers	7	ConcePTION
	London School of Hygiene & Tropical Medicine			Valproate
	University of Utrecht			ACCESS
11	**24 organizations**	**17 different data sources**	**159M**	**4 projects**

ConcePTION: IMI funded project https://www.imi-conception.eu D7.5
ACCESS: http://www.encepp.eu/encepp/viewResource.htm?id=37274
RETINOIDS: http://www.encepp.eu/encepp/viewResource.htm?id=31096
Valproate: http://www.encepp.eu/encepp/viewResource.htm?id=36586

Annex 1. Protocol & analysis plan

Version 0.4
Date: 30.11.2020

Research Protocol

Demonstrating the value of uniquely identify a medicinal product: the use and risk of diclofenac salts and the risk of cardiovascular events

UNICOM Project – WP 8
www.unicom-project.eu/

Title	Demonstrating the value of uniquely identify a medicinal product: the use and risk of diclofenac salts and the risk of cardiovascular events
Protocol version identifier	0.4
Date of last version of protocol	November 30, 2020
EU PAS register number	
Active substance	*Diclofenac*
Medicinal product	
Product reference	*NA*
Procedure number	*NA*
Marketing authorisation holder(s)	*NA*
Research question and objectives	Is there a difference in the incidence rate of acute myocardial infarction (AMI), heart failure (HF) and stroke after the oral intake of diclofenac sodium versus diclofenac potassium in adults ≥ 18 years old? The primary objective of this study is to compare the incidence rate of acute myocardial infarction, heart failure and stroke between adult users of oral diclofenac sodium and oral diclofenac potassium. The secondary study objective is to assess and describe the availability of different oral pharmaceutical alternatives of diclofenac in European countries where this research will be performed.
Country(-ies) of study	Electronic health care databases among ConcePTION partners.
Authors	Prof. dr. M.C.J.M. Sturkenboom, i~HD, University Utrecht Medical Center, Utrecht, The Netherlands. Dr. Carlos E. Durán, i~HD, Clinical Pharmacology Research Group. Ghent University, Belgium.

Table of Contents

Key words

IDMP Standards/research applications, Active substances, Salts and esters/active role, Diclofenac/sodium/potassium, cardiovascular risk/acute myocardial infarction/heart failure/stroke, electronic healthcare data sources/Europe.

Abbreviations

The following abbreviation will be used in this protocol:

- ATC – Anatomic, Therapeutic and Chemical Classification of Drugs
- COX2 – Cyclooxygenase 2
- DAP – Database access provider
- DDD – Defined Daily Doses
- EDQM - European Directorate for the Quality of Medicines and Healthcare
- IDMP – Identification of Medicinal Products
- ISO – International Organization for Standardization
- NSAID – Non-Steroidal Anti-inflammatory Drug
- PHPID – Pharmaceutical Product Identification
- PDD – Prescribed Daily Doses
- UMLS - Unified Medical Language System
- WHO – World Health Organization

1. Title

Demonstrating the value of uniquely identify a medicinal product: the use and risk of diclofenac salts and the risk of cardiovascular events

2. Marketing authorisation holder

Not applicable

3. Responsible parties

I~HD
Miriam Sturkenboom, PharmD, PhD, MSc, professor

Sponsor: European Commission, UNICOM project

Collaborating Institutions (by alphabetical order)	Study Sites	Key persons

4. Abstract

Title: Demonstrating the value of uniquely identify a medicinal product: the use and risk of diclofenac salts and the risk of cardiovascular events

Main authors:
Prof. dr. M.C.J.M. Sturkenboom, i~HD, University Utrecht Medical Center, Utrecht, The Netherlands.
Dr. Carlos E. Durán, i~HD, Clinical Pharmacology Research Group. Ghent University, Belgium.

Rationale and background: Many pharmacoepidemiological studies focus only on the active ingredient of a medicinal product and not the unique medicinal product, often because detail is not available nor easy to harmonize across countries. In this proof of concept study we will aim to assess whether we can identify different salts of diclofenac in electronic health care databases throughout Europe, and compare the risk of cardiovascular events. The main purpose is to show the value of the use of the IDMP.

Research question and objectives:
The primary objective of this study is to compare the incidence rate of acute myocardial infarction, heart failure and stroke between adult users of oral diclofenac sodium and oral diclofenac potassium.

The secondary study objective is to assess and describe the availability of different oral pharmaceutical alternatives of diclofenac in European countries where this research will be performed.

Study design: A retrospective dynamic cohort study will be conducted during the period 2010-2020. Simultaneously, a descriptive analysis of the availability of the pharmaceutical alternatives of oral diclofenac in several European countries during the year 2019 will be performed.

Population: The study population will include all individuals observed in one of the participating data sources for at least one day during the study period (01 January 2010 - last data availability) who have at least 1 year of data availability before cohort entry, except for individuals with data available since birth. Cohort entry occurs at first prescription of one of the study drugs.

Variables:
Variables of interest will be:
- Person-time: birth and death dates as well as periods of observation.
- Drug exposure: prescriptions/dispensings of diclofenac sodium and potassium.
- Events: dates of medical and/or procedure and/or prescription/dispensing codes to identify cardiovascular events: myocardial infraction, stroke and heart failure.

Data sources: The study will include data from data sources in European countries that will agree to participate and have converted their data in the ConcePTION CDM. Data sources may contain health insurance data, hospitalisation record linkage data or data from general practitioners.

Study size: The study population will comprise approximately all individuals complying with the inclusion criteria.

Data analysis: Incidence rates (and 95% CI) of cardiovascular events will be calculated by dividing the number of incident (new) cases by the total person-time.

Milestones:

Milestone	Planned date
Protocol submitted to UNICOM	2 November 2020
Submission to EC	30 November 2020
Identification of data sources	2021
Final report of study results	

5. Amendments and updates

Date	Amendment	Justification	Protocol Section

6. Deliverables and Milestones

Deliverable	Date
Protocol	30 November 2020
Results report	

7. Introduction and rationale

7.1 Introduction

Several attempts to develop a unique standard for identification of medicinal products have been carried out during the last decades. In 2012, the International Organization for Standardization (ISO) published a series of standards and technical specifications called the *Identification of Medicinal Products* (IDMP); it is built up from a set of 5 ISO standards and 4 technical specifications(1).

From 2012 on, several efforts have been putted in place to make possible the global implementation of the IDMP norm. Nowadays, two of the main world regulators (EMA and US FDA) have agreed a common implementation plan. For instance, the European Commission launched the Regulation No. 520/2012(2), making compulsory to European Medicines Agency, marketing authorization holders and member States, the implementation and use of ISO IDMP standards.

The European funded project *Up-scaling the Global Univocal Identification of Medicines* (UNICOM) aims to give an ultimate impulse to the implementation of ISO IDMP in European healthcare databases and therefore allow safety cross-national e-prescription practices as well as pharmacovigilance activities(3). As a first step, UNICOM's technical experts have selected a list of 35 medicinal products (pilot product list) in order to agree and define common identification concepts.

UNICOM's work package 8 (task 8.2) refers to the application of ISO IDMP in biomedical research, mainly in the fields of Pharmacoepidemiology, Drug Utilization Research and Pharmacovigilance. As a proof-of-concept, one active substance included in the pilot product list (diclofenac) has been selected to explore the usefulness of the unique identifier (IDMP) to solve a specific research question.

7.2 Rationale

Diclofenac

Diclofenac, a phenylacetic acid derivate, is a Nonsteroidal Anti-inflammatory Drug (NSAID). During decades, it has been widely used for the symptomatic treatment of chronic musculoskeletal pain and inflammatory conditions such as osteoarthritis, rheumatoid arthritis and ankylosing spondylitis; periarticular disorders, sprains and strains, and some acute painful conditions, mainly postoperative pain, gout, renal colic, migraine and dysmenorrhea(4).

Diclofenac has proven to be an effective drug to control of pain in chronic inflammatory conditions as well as for acute pain relief(5). However, during the last 15 years, diclofenac´s cardiovascular safety profile has raised lots of concerns. Several pharmacoepidemiologic studies have demonstrated the association between non-selective NSAIDs, such as

diclofenac, and cardiovascular adverse outcomes, i.e. acute myocardial infarction, heart failure and stroke(6–10). These adverse events also occurred after use of rofecoxib and etoricoxib, some COX-2 selective NSAIDs inhibitors. Until today, it has been proved that diclofenac intake increases the risk of acute myocardial infarction, heart failure and death in those who have previously suffered a major cardiovascular event(6,7,11,12), and do so by 50% in people without history of cardiovascular disease when compared with non-users, even after short periods of use(13).
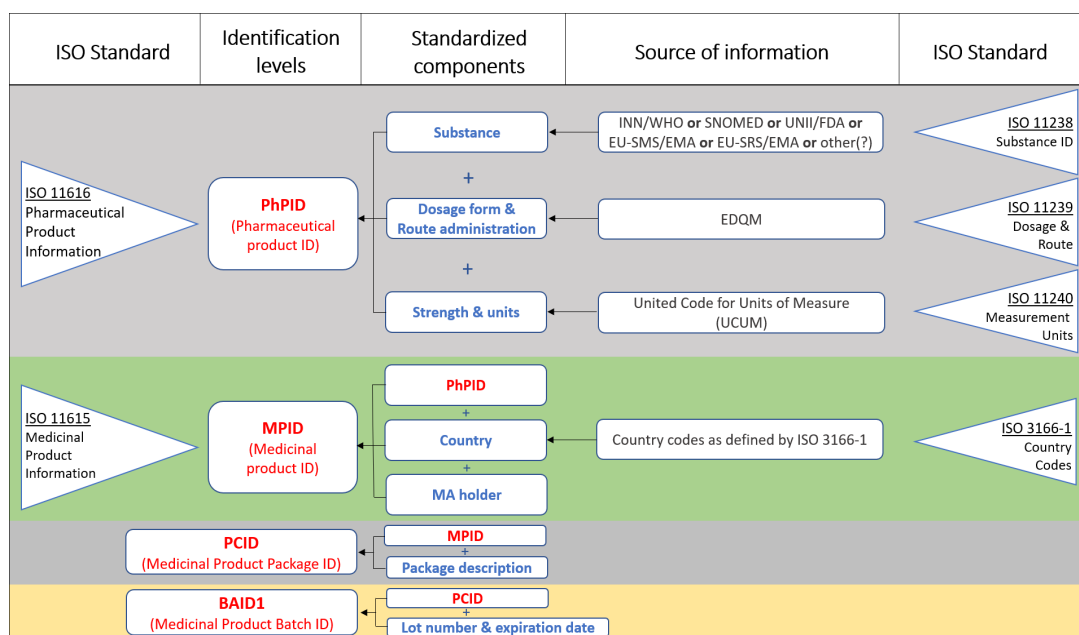

Figure 1. Main components of the Identification of Medicinal Products (IDMP) suite of standards.

Active substances, salts and esters

In general, drugs are weak organic acids or bases. Therefore, active molecular entities can be presented as different salts or esters. Every salt or ester drug is considered a different chemical entity with their own chemical and biological profile that could lead to differences in the clinical response and in the safety/therapeutic profile as well(14,15).

From the regulatory point of view, close to half of the active substances are marketed as salts(14). European Medicines Agency has defined them as *Pharmaceutical Alternatives,* being those "medicinal products with different salts, esters, ethers, isomers, mixtures of isomers, complexes or derivates of an active moiety, or which differ in dosage form or strength"(16). Pharmaceutical alternatives are considered to be equivalents if their bioavailability after administration of the same molar dose lie within the same predefined limits(16). In the case of diclofenac, it is mostly marketed in Europe as potassium and sodium salts for systemic use (oral and parenteral).

Although available diclofenac products in European countries are all bioequivalent products, there are no pharmacoepidemiologic studies that aimed to address specific questions

regarding the cardiovascular safety profile derived from the use of the available oral diclofenac alternatives. Basic research in animals have compared pharmacokinetics and bioavailability parameters between diclofenac sodium and potassium in normal and dehydrated rabbits, suggesting higher plasma concentrations for the potassium salt in both groups(17). A similar approach has been followed to compare normal *versus* diabetic rabbits, showing a decrease in plasma concentration of both, sodium and potassium diclofenac salts in diabetic rabbits, but also confirming a significant higher concentration of diclofenac potassium among them(18). In clinical settings, both pharmaceutical alternatives have been tested for the control of acute postoperative pain. Eighteen studies were evaluated in a Cochrane´s systematic review showing good rates of pain control after diclofenac potassium administration and limited efficacy of diclofenac sodium in this indication(5). Moreover, there were no differences on adverse events rates after a single dose administration of any of the alternatives.

Detailed identification of diclofenac oral salts in large observational databases will allow for a post-marketing study assessing the cardiovascular safety of specific pharmaceutical alternatives of oral diclofenac in European adults.

8. Research question and study objectives

8.1 Research question

Is there a difference in the incidence rate of acute myocardial infarction (AMI), heart failure (HF) and stroke after the oral intake of diclofenac sodium *versus* diclofenac potassium in adults ≥ 18 years old?

8.2 Study objectives

1. The primary objective of this study is to compare the incidence rate of acute myocardial infarction, heart failure and stroke between adult users of oral diclofenac sodium and oral diclofenac potassium.

2. The secondary study objective is to assess and describe the availability of different oral pharmaceutical alternatives of diclofenac in the European countries where this research will be performed.

9. Research methods

9.1 Study design

A retrospective dynamic cohort study will be conducted for the period 2010-2020.

Adults ≥ 18 years old will be included if they have been continuously registered in the database for a minimum of 12 months before the first oral diclofenac prescription during the study period (01 January 2010 - last data availability).

Start of follow up will be defined as the latest of the following dates:

- Start of study period
- Birth
- Registration in the database

Follow up will end at the earliest of:

- Death
- End of study period
- Moving out of the database
- End of registration
- Last data drawn down

The following exclusion criteria will be applied:

- History of myocardial infarction.
- History of heart failure, or any kind of coronary intervention.
- History of stroke.
- Chronic kidney disease, including hemodialysis and transplantation.
- Chronic liver disease.
- Cancer, past or present (excl. melanoma).

Comparator groups:

Two study cohorts of interest will be studied: initiators of oral diclofenac *sodium* and initiators of oral diclofenac *potassium*. Participants in the study cohorts will be matched by propensity scores measured prior one year of cohort entry.

Figure 2. Time-based cohort entry
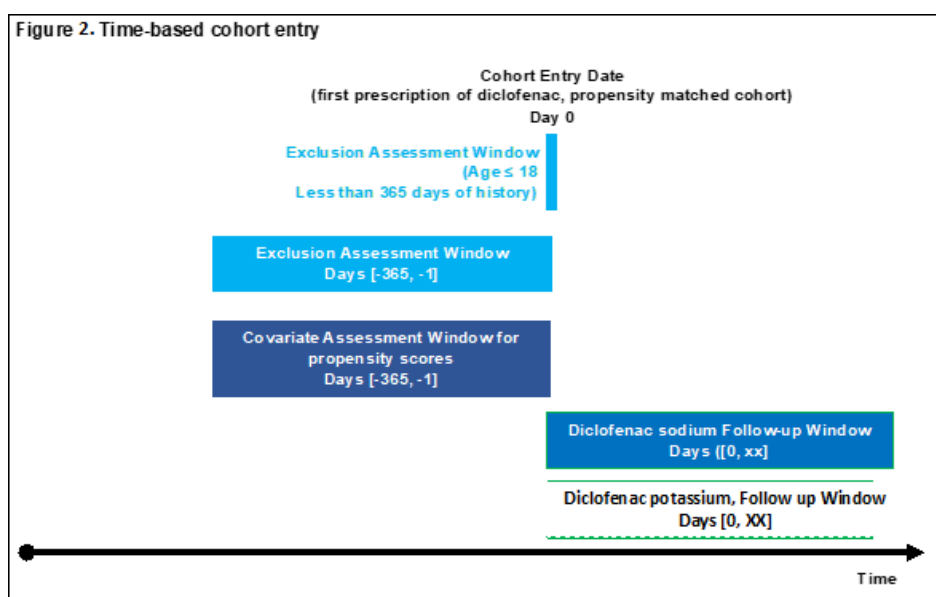
Cohort entry will start upon the first registered prescription of oral diclofenac, sodium or potassium. End of follow-up in a specific cohort will be the earliest of the following dates

- Occurrence of a non-fatal cardiovascular event, as defined in 9.4.2.
- Occurrence of a fatal cardiovascular event.
- Stopping of diclofenac: 30 days after the last diclofenac prescription.
- End of follow-up (see above).
- Occurrence of an exclusionary condition
- Start of another NSAIDs or switch to different diclofenac product.

9.2   Setting

The study will be conducted using population-based electronic health care data sources in Europe, that have converted their data in the ConcePTION common data model, and are willing to participate.

9.3   Variables

9.3.1   Exposure

From selected databases, we will retrieve the available prescription/dispensing information to construct PhPIDs of diclofenac sodium and potassium. Thereafter, the marketing authorization holder and country's name will be assigned to every PhPID, allowing to define MPIDs. In order to build up the PhPID and the MPID concepts, the following items will be retrieved from healthcare databases, if available:

- Modified substance name (moiety + salt) (ATC level V: M01AB05).
- Dosage form and route of administration.
- Strength and units of measure in milligrams.
- Name of the marketing authorization holder.
- Name of the country.

Figure 3 depicts the possible data collection scenarios to construct PHPIDs and MPIDs in European databases. In the particular case of diclofenac pharmaceutical alternatives, selected databases must allow the proper identification of sodium and potassium salts, for example, data sources with a moiety's identification number that allows the linkage to a drug dictionary containing detailed information.
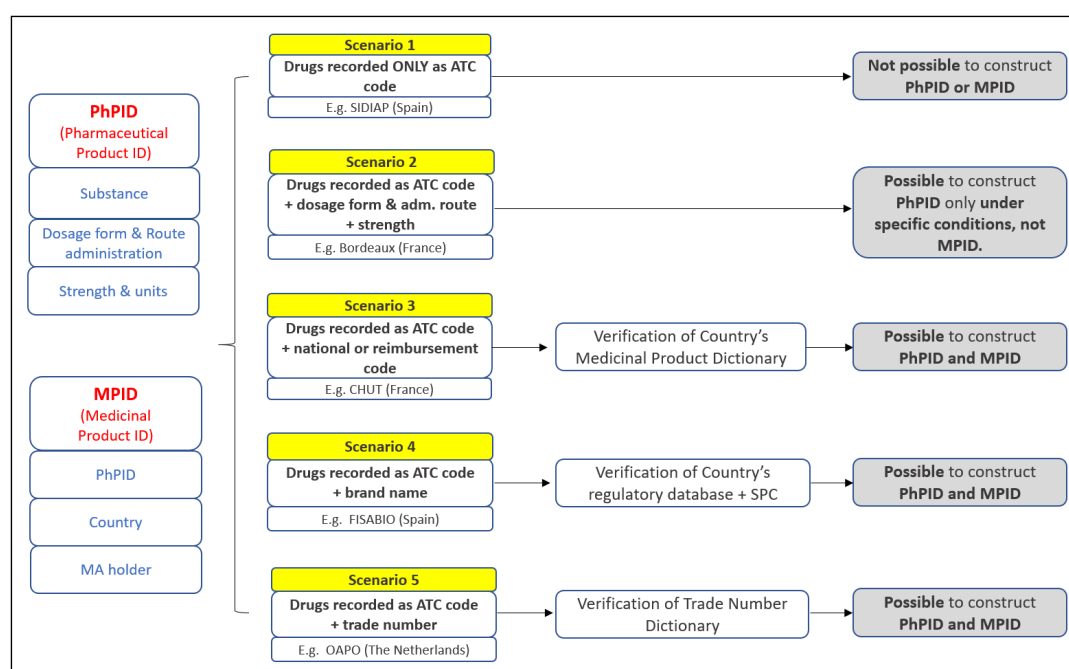
Figure 3. Possible scenarios to construct IDMP concepts according to the medicines´ registration items in European Healthcare databases.

Daily dose and duration

Diclofenac dose will be recoded according to the characteristics in the database. If detailed prescription information is recorded, the Prescribed Daily Dose (PDD) will be calculated. If not available, the Defined Daily Dose (DDD) (pack size) will be assigned.

Duration of diclofenac use will be retrieved as the prescription duration as registered in the database or by calculating the duration gaps of repeated prescription/redemptions.

9.3.2   Endpoints

Major cardiovascular events will be the primary endpoint. A major cardiovascular event will be a composite endpoint of non-fatal and fatal outcomes:

1. Non-fatal events will be defined as: in-patient diagnosis of acute myocardial infarction, heart failure and stroke.
2. Fatal events will be defined as any cardiovascular death.

The secondary endpoints will be the individual cardiovascular events.

9.3.2.1 *Acute myocardial infarction*

Myocardial infarction is defined as necrosis of the myocardium caused by an obstruction of the blood supply to the heart (coronary circulation) (National Library of Medicine, Medical Subject Headings). Blockage of a coronary artery deprives the heart muscle of blood and oxygen, causing injury to the heart muscle. Therefore, acute myocardial infarction is defined by the evidence of myocardial necrosis in a clinical setting consistent with myocardial ischemia, including ST elevation myocardial infarction and non-ST elevation myocardial infarction. Specific diagnosis and procedural codes in SNOMED, ICD-9, 10 and ICPC will be obtained using Codemapper according to the following Unified Medical Language System (UMLS) concepts for the outcomes of acute myocardial infarction, *see table 1*.

Table 1. UMLS concepts for acute myocardial infarction

Acute myocardial infarction
Acute transmural myocardial infarction of unspecified site
Acute transmural myocardial infarction of anterior wall
Acute transmural myocardial infarction of inferior wall
Acute transmural myocardial infarction of other sites
Acute subendocardial myocardial infarction
Acute myocardial infarction of anterolateral wall
Acute infarction of papillary muscle
Acute anteroseptal myocardial infarction
Acute myocardial infarction of inferior wall
Acute myocardial infarction of atrium
Acute subendocardial infarction
Other acute myocardial infarction NOS
Acute anteroapical infarction
Acute myocardial infarction of septum
Acute myocardial infarction of inferoposterior wall
Acute myocardial infarction of inferolateral wall
Acute Q wave myocardial infarction
Acute non-Q wave infarction
Acute myocardial infarction of anterior wall (disorder)
Acute myocardial infarction, of other anterior wall
Acute myocardial infarction, of other inferior wall
Acute myocardial infarction, of other lateral wall

Acute myocardial infarction, subendocardial infarction

Acute myocardial infarction, of other specified sites

True posterior myocardial infarction

Acute myocardial infarction, unspecified site, initial episode of care

Acute myocardial infarction, unspecified site, subsequent episode of care

ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction

Subsequent myocardial infarction of unspecified site

Subsequent myocardial infarction of other sites

Subsequent myocardial infarction of anterior wall

Subsequent myocardial infarction of inferior wall

9.3.2.2 *Heart failure*

Heart failure is a complex clinical syndrome in which structural or functional alterations of the heart lead to ventricular dysfunction resulting into the inability of pumping out sufficient blood to meet the metabolic need of the body. Heart failure can be caused by structural defects, functional abnormalities (ventricular dysfunction), or a sudden overload beyond its capacity. Chronic HF is more common than acute HF which results from sudden insult to cardiac function, such within the acute course of a coronary syndrome (National Library of Medicine, Medical Subject Headings). In this study, a HF event will be assessed based on HF being the main reason for hospitalization including congestive heart failure, left-sided heart failure and unspecified heart failure. Heart failure will be ascertained by specific inpatient diagnosis codes. HF incidental to a hospitalization for other cause will be excluded from the study case definition. Specific codes will the ascertained according to the following UMLS concepts for the outcomes of heart failure, *see table 2.*

Table 2. UMLS concepts for heart failure

Heart failure

Congestive heart failure

Left-Sided Heart Failure

Heart Failure, Right-Sided

Acute heart failure

Heart failure as a complication of care

Post cardiac operation functional disturbance

Compensated cardiac failure

Decompensated cardiac failure

Biventricular congestive heart failure

Heart Failure, Systolic

Heart Failure, Diastolic

Combined systolic and diastolic heart failure

benign hypertensive heart and renal disease with congestive heart failure

benign hypertensive heart and renal disease with congestive heart failure and renal failure

Hypertensive heart disease NOS with congestive cardiac failure

H/O: heart failure

> Hypertensive heart and renal disease with (congestive) heart failure
>
> Hypertensive heart and renal disease with both (congestive) heart failure and renal failure
>
> Malignant hypertensive heart disease with heart failure
>
> Benign hypertensive heart disease with heart failure
>
> Unspecified hypertensive heart disease with heart failure
>
> Hypertensive heart and kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
>
> Hypertensive heart and chronic kidney disease, malignant, with heart failure and chronic kidney failure stage V or end stage renal disease

9.3.2.3 *Stroke*

Stroke is defined as rapidly developed clinical signs of focal (or global) disturbance of cerebral function lasting more than 24 hours (unless interrupted by surgery or death), with no apparent cause other than a vascular origin: it includes patients presenting clinical signs and symptoms suggestive of subarachnoid haemorrhage, intracerebral haemorrhage or cerebral ischemic necrosis. It does not include transient cerebral ischemia or stroke events in cases of blood disease (e.g. leukemia, polycythaemia vera), brain tumour or brain metastases. Secondary stroke caused by trauma should also be excluded. Global applies to patients with subarachnoid haemorrhage or deep coma but excluding coma of systemic vascular origin such as shock, Stokes-Adams syndrome or hypertensive encephalopathy. In this study, a stroke event is defined as any form of stroke due to haemorrhage (subarachnoid, intracerebral) or infarction (i.e. ischemic) and stroke not specified as haemorrhage or infarction. Specific codes will the ascertained according to the following UMLS concepts for the outcomes of stroke, *see table 3.*

Table 3. UMLS concepts for stroke

> <u>Cerebrovascular accident</u>
>
>   Cerebral hemisphere hemorrhage
>
>   Cerebral Infarction
>
>   Cerebral Thrombosis
>
>   CVA - cerebrovascular accident due to cerebral artery occlusion
>
>   Right sided cerebral hemisphere cerebrovascular accident
>
>   Left sided cerebral hemisphere cerebrovascular accident
>
> <u>Perinatal Subarachnoid Hemorrhage</u>
>
> <u>Subarachnoid haemorrhage neonatal</u>
>
> <u>Subarachnoid Hemorrhage</u>
>
>   Subarachnoid hemorrhage from other intracranial arteries
>
>   Other subarachnoid hemorrhage
>
>   Subarachnoid hemorrhage from intracranial artery, unspecified
>
>   Subarachnoid hemorrhage from vertebral artery
>
>   Subarachnoid hemorrhage from middle cerebral artery aneurysm
>
>   Subarachnoid hemorrhage from anterior communicating artery aneurysm
>
>   Subarachnoid hemorrhage from posterior communicating artery aneurysm

Subarachnoid hemorrhage from basilar artery aneurysm
Subarachnoid hemorrhage from carotid siphon and bifurcation
Perinatal Subarachnoid Hemorrhage
Subarachnoid Hemorrhage, Traumatic
<u>Occlusion and stenosis of carotid artery</u>
Occlusion and stenosis of carotid artery, with cerebral infarction
<u>Persistent migraine aura with cerebral infarction</u>
Persistent migraine aura with cerebral infarction, with intractable migraine, so stated, without mention of status migrainosus
Persistent migraine aura with cerebral infarction, without mention of intractable migraine with status migrainosus
Persistent migraine aura with cerebral infarction, with intractable migraine, so stated, with status migrainosus
<u>Other nontraumatic intracranial hemorrhage</u>
Subdural hemorrhage (acute) (nontraumatic)
Intracranial hemorrhage (nontraumatic), unspecified
Nontraumatic extradural hemorrhage

9.3.3   Potential confounders

The following variables will be considered as potential confounders:

- Demographic characteristics: age and sex.
- Individual risk factors: obesity, alcohol abuse.
- Co morbidities: hypertension, diabetes, hyperlipidaemia, cardiac arrythmia and other conduction disorders, atrial fibrillation and flutter, transient stroke, cardiomyopathy, valvular disease and endocarditis, myocarditis and pericarditis, peripheral artery vascular disease, chronic ischemic heart disease, arterial embolism and thrombosis, blood coagulation disorders, renal failure, liver disease and chronic respiratory disorders.
- Drugs to treat the above indications as proxy for the co-morbid conditions.
- Number of different ATC codes in year prior as proxy for health care utilization.

UMLS concepts for individual risk factors and comorbidities are presented in *Annex 7.1.*

9.4   Data sources and data management

Several electronic healthcare databases are available for research purposes in Europe(19). Information for this study should be retrieved from those electronic databases allowing the proper characterization the PhPIDs and MDIPs of diclofenac pharmaceutical alternatives (sodium and potassium oral forms). In case, a given data source allows the construction of IDMP concepts by extracting the required information from other external sources, e.g. drug dictionaries (*see figure 3*), the lasts must also be available to the research team.

A description of some of the data sources that have been mapping to the ConcePTION CDM as part of the ACCESS Back ground rate study are described below (**EUPAS37273**). *Based on actual participation this section may need to be adapted, data sources have not yet confirmed their participation*

9.4.1. Germany: GePaRD

GePaRD is based on claims data from four statutory health insurance providers in Germany and currently includes information on approximately 25 million persons who have been insured with one of the participating providers since 2004 or later. Per data year, there is information on approximately 20% of the general population and all geographical regions of Germany are represented. In addition to demographic data, GePaRD contains information on dispensations of reimbursable prescription drugs as well as outpatient (i.e., from general practitioners and specialists) and inpatient services and diagnoses. GePaRD also contains information on influenza vaccinations and routine childhood immunizations and there is experience with studies on utilization and risk of vaccination and on background incidence of adverse events of vaccinations(20,21). GePaRD data have been used for vaccine safety studies. GePaRD is listed under the ENCePP resources database.

9.4.2 Netherlands: PHARMO Database Network

The PHARMO Database Network is a population-based network of electronic healthcare databases and combines anonymous data from different primary and secondary healthcare settings in the Netherlands. These different data sources, including data from general practices, in- and out-patient pharmacies, clinical laboratories, hospitals, the cancer registry, pathology registry and perinatal registry, are linked on a patient level through validated algorithms. To ensure the privacy of the data in the PHARMO Database Network, the collection, processing, linkage and anonymization of the data is performed by STIZON. STIZON is an independent, ISO/IEC 27001 certified foundation, which acts as a Trusted Third Party between the data sources and the PHARMO Institute. The longitudinal nature of the PHARMO Database Network system enables to follow-up more than 9 million persons of a well-defined population in the Netherlands for an average of twelve years. Currently, the PHARMO Database Network covers over 6 million active persons out of 17 million inhabitants of the Netherlands. Data collection period, catchment area and overlap between data sources differ. Therefore, the final cohort size for any study will depend on the data sources included. As data sources are linked on an annual basis, the average lag time of the data is one year. All electronic patient records in the PHARMO Database Network include information on age, sex, socioeconomic status and mortality. Other available information depends on the data source. A detailed description of the different data sources is given below. PHARMO is always seeking new opportunities to link with healthcare databases. Furthermore, it is possible to link additional data collections, such as data from chart reviews, patient-reported outcomes or data from general practice trials.

The General Practitioner database comprises data from electronic patient records registered by GPs. The records include information on diagnoses and symptoms, laboratory test results, referrals to specialists and healthcare product/drug prescriptions. The prescription records include information on type of product, prescription date, strength, dosage regimen, quantity and route of administration. Drug prescriptions are coded according to the WHO ATC Classification System [www.whocc.no]. Diagnoses and symptoms are coded according to the International Classification of Primary Care - ICPC [www.nhg.org], which can be mapped to the International Classification of Diseases - ICD codes, but can also be entered as free text. GP data cover a catchment area representing 3.2 million residents (~20% of the Dutch population).

The Out-patient Pharmacy Database comprises GP or specialist prescribed healthcare products dispensed by the out-patient pharmacy. The dispensing records include information on type of product, date, strength, dosage regimen, quantity, route of administration, prescriber specialty and costs. Drug dispensings are coded according to the WHO ATC Classification System. Out-patient pharmacy data cover a catchment area representing 4.2 million residents (~25% of the Dutch population). PHARMO is listed under the ENCePP resources database.

9.4.3 Denmark: Danish Registries

Denmark has a tax-funded health care system ensuring easy and equal access to health care for all its citizens, and with this system all contacts are recorded in administrative and medical registers(22). The records carry a unique personal identification number, called the CPR-number, assigned to every Danish citizen. Linkage between registers at an individual level is possible because this CPR-number is used in all Danish registers and assigned by the Danish Civil Registration System(23). All registers have a nationwide coverage and an almost 100% capture of contacts covering information on currently 5.8 million inhabitants plus historical information. For the purpose of the study we will obtain information from the following registries. The Danish National Prescription Registry (DNPR) includes data on all outpatient dispensing of medications and vaccines at Danish pharmacies from 1995 and onwards, including dispensing date, ATC code, product code and amount(24). The Danish National Health Service Register includes data on primary care services, including general practitioner contacts, examinations, procedures, and vaccinations; psychologist or psychiatrist and other primary care provider visits; etc. From the Danish Civil Registration System, data on demographics (sex, date of birth) and censoring (migration, vital status). The Danish National Patient Registry contains diagnoses and procedures from all hospitalizations since 1977 and contacts to hospital outpatient clinics since 1995(25). The Danish National Health Service Register contains information on referral for vaccine administration from GPs(26).

9.4.4 Spain: BIFAP

BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en Atencion Primaria), a computerized database of medical records of primary care (www.bifap.aemps.es) is a non-profit research project funded by the Spanish

Agency for Medicines and Medical Devices (AEMPS). The project started in 2001 and current version of the database with information until December 2019 includes clinical information of 6,419 GPs and 1,147 pediatricians. Ten participant autonomous regions send their data to BIFAP every year. BIFAP database currently includes anonymized clinical and prescription/dispensing data from around 14 million (8 active population) patients representing 85% of all patients of those regions participating in the database, and 25% of the Spanish population. Mean duration of follow-up in the database is 8.6 years. Diagnoses are classified according to the International Classification of Primary Care (ICPC)-2, SNOMED and ICD-9 code system. Information on hospital outpatient diagnosis is being progressively included.

9.4.5 Spain: SIDIAP

The Information System for Research in Primary Care (Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària' - SIDIAP; www.sidiap.org) was created in 2010 by the Catalan Health Institute (CHI) and the IDIAPJGol Institute. It includes information collected since 01 January 2006 during routine visits at 278 primary care centers pertaining to the CHI in Catalonia (North-East Spain) with 3,414 participating GPs. SIDIAP has pseudo-anonymized records for 5.7 million people (80% of the Catalan population) being highly representative of the Catalan population.

The SIDIAP data comprises the clinical and referral events registered by primary care health professionals (GPs, paediatricians and nurses) and administrative staff in electronic medical records, comprehensive demographic information, community pharmacy invoicing data, specialist referrals and primary care laboratory test results. It can also be linked to other data sources, such as the hospital discharge database, on a project by project basis. Health professionals gather this information using ICD-10 codes, ATC codes and structured forms designed for the collection of variables relevant for primary care clinical management, such as country of origin, sex, age, height, weight, body mass index, tobacco and alcohol use, blood pressure measurements, blood and urine test results. Encoding personal and clinic identifiers ensures the confidentiality of the information in the SIDIAP database. The SIDIAP database is updated annually at each start of the year. Nowadays, with the COVID-19 pandemic, there is the possibility to have shorter term updates in order to monitor the evolution of the pandemic. Recent reports have shown the SIDIAP data to be useful for epidemiological research. SIDIAP is listed under the ENCePP resources database www.encepp.eu/encepp/resourcesDatabase.jsp).

9.4.6 Spain: FISABIO

The region of Valencia, with 5 million inhabitants, is part of the Spanish National Health System, a universal public healthcare system. Information can be obtained from the population-based electronic information systems of the Valencia Health Agency (VHA) and the regional Government of Valencia: *i.* The Population Information System (SIP) provides an identification number for each person under Valencian Health Service (VHS) coverage, and registers some demographic

characteristics, and dates and causes of VHA discharge, including death. *ii.* The minimum basic dataset at hospital discharge is a synopsis of clinical and administrative information on all hospital discharges, including diagnoses and procedures (all electronic health systems in the VHS use the ICD- 9-CM). *iii.* The Emergency Department module (ED) including ED dates of visit and discharge and reason for discharge. *iv.* The electronic medical record (EMR) for ambulatory care, available in all primary healthcare centers and other ambulatory settings. It has all the information on patients regarding diagnoses, their personal and family medical history, laboratory results, lifestyle, etc. *v.* The pharmaceutical module (prescription information system), part of EMR, includes information about both physician prescriptions and dispensations from pharmacy claims. *vi.* The Corporate Resource Catalogue provides information about the geographical and functional organization of VHS, its health centers, health services provided and professionals in healthcare. Specific public health registries are available and linkable at an individual level (such as the perinatal registry and the congenital anomalies registry, from which pregnancy outcomes can be obtained). All the information in these systems can be linked at an individual level through the SIP number.

9.4.8 Italy: ARS database

The Italian National Healthcare System is organized at regional level: the national government sets standards of assistance and a tax-based funding for each region, and regional governments are responsible to provide to all their inhabitants. Tuscany is an Italian region, with around 3.6 million inhabitants. The Agenzia Regionale di Sanita' della Toscana (ARS) is a research institute of the Tuscany Region. The ARS database comprises all information that are collected by the Tuscany Region to account for the healthcare delivered to its inhabitants. Moreover, ARS collects data from regional initiatives. All the data in the ARS data source can be linked with each other at the individual level, through a pseudo-anonymous identifier. The ARS database routinely collects primary care and secondary care prescriptions of drugs for outpatient use, and is able to link them at the individual level with hospital admissions, admissions to emergency care, records of exemptions from co-payment, diagnostic tests and procedures, causes of death, mental health services registry, birth registry, spontaneous abortion registry, induced terminations registry. A pathology registry is available, mostly recorded in free text, but with morphology and topographic Snomed codes. Mother-child linkage is possible through the birth registry.

9.4.8 United Kingdom: CPRD & HES

The Clinical Practice Research Datalink (CPRD) from the UK collates the computerized medical records of general practitioners (GPs) in the UK who act as the gatekeepers of healthcare and maintain patients' life-long electronic health records. As such they are responsible for primary healthcare and specialist referrals, and they also store information stemming from specialist referrals, and hospitalizations. GPs act as the first point of contact for any non-emergency health-related issues, which may then be managed within primary care and/or referred to secondary care as necessary. Secondary care teams also feedback information to

GPs about their patients, including key diagnoses. The data recorded in the CPRD include demographic information, prescription details, clinical events, preventive care, specialist referrals, hospital admissions, and major outcomes, including death. The majority of the data are coded in Read Codes. Validation of data with original records (specialist letters) is also available.

The dataset is generalizable to the UK population based upon age, sex, socioeconomic class and national geographic coverage when GOLD & Aurum versions are used.

There are currently approximately 42 million patients (acceptable for research purposes) – of which 13 million are active (still alive and registered with the GP practice) – in approximately 1,700 practices (https://cprd.com/Data). Data include demographics, all GP/healthcare professional consultations (phone, letter, email, in surgery, at home), diagnoses and symptoms, laboratory test results, treatments, including all prescriptions, all data referrals to other care, hospital discharge summary (date and Read codes), hospital clinic summary, preventive treatment and immunizations, death (date and cause). For a proportion of the CPRD panel practices (>80%), the GPs have agreed to permit CPRD to link at patient level to the Hospital Episode Statistics (HES) data.

The HES database contains details of all admissions to National Health System (NHS) hospitals in England; approximately 60% of GP practices in the CPRD are linked to the HES database. Not all patients in the CPRD have linked data (e.g. if they live outside England or if their GP has not agreed that their data should be used in this way).

9.4.9 France: Système National des Données de Santé (SNDS)

The SNDS (Système National des Données de Santé) is the French nationwide healthcare database. It currently covers the overall French population (about 67 million persons) from birth (or immigration) to death (or emigration), even if a subject changes occupation or retires. Using a unique pseudonymized identifier, the SNDS merges all reimbursed outpatient claims from all French health care insurance schemes (SNIIRAM database), hospital-discharge summaries from French public and private hospitals (PMSI database), and the national death registry. SNDS data are available since 2006 and contains information on:

- General characteristics: gender, year of birth, area of residence, etc.
- Death: month, year and cause
- Long-Term Disease registration associated with an ICD-10 diagnostic codes
- Outpatient reimbursed healthcare expenditures with dates and codes (but not the medical indication nor result): visits, medical procedures, nursing acts, physiotherapy, lab tests, dispensed drugs and medical devices, etc. For each expenditure, associated costs, prescriber and caregiver information (specialty, private/public practice) and the corresponding dates are provided.
- Inpatients details: primary, related and associated ICD-10 diagnostic codes resulting from hospital discharge summaries with the date and duration of the hospital stay, the performed medical procedures, and the

related costs. Drugs included in the diagnosis related group cost are not captured. However, expansive drugs (i.e. the one charged in addition to the group cost) are.

Outpatient data (SNIIRAM) are uploaded to the SNDS throughout the year. It is admitted that a lag of around 6 months is required to catch 90% of the dispensings. Inpatient data (PMSI) are uploaded in one time, at the end of the following year. Hence, we consider that complete SNDS data of year Y are available in January of the year Y+2. SNDS access is regulated.

Each study and data extraction need approval from the CESREES (*Comité Ethique et Scientifique pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé*) in charge of assessing scientific quality of the project, and authorization from the CNIL (French data protection commission), and then contracts with the SNDS data holder (CNAM) for data extraction.

9.5 Study size

The study population will include all individuals registered with at least one year of data upon first prescription of diclofenac during the study period.

9.6 Data management

This study will be conducted in a distributed manner using a common protocol, common data model (CDM), and common analytics programs. This process was used successfully in several other European multi-database projects. The data pipeline has been further improved in the IMI-ConcePTION project (https://www.imi-conception.eu/) [25]This process maximizes the involvement of the data providers in the study by utilizing their knowledge on the characteristics and the process underlying the data collection which makes analysis more efficient.

1. First, to harmonize the structure of the data sets held by each partner, a shared syntactic foundation is utilized. In the ConcePTION common data model, data is represented in a common structure but the content of the data remain in their original format.

2. Second, to reconcile differences across terminologies a shared semantic foundation is built for the definition of events under study by collecting relevant concepts in a structured fashion using a standardized event definition template. The Codemapper tool will be used to create diagnosis code lists based upon the UMLS concepts(27). Based on the relevant diagnostic medical codes and keywords, as well as other relevant concepts (e.g. medications), one or more algorithms will be constructed (typically one sensitive, or broad, algorithm and one specific, or narrow algorithm) to operationalize the identification and measurement of each event. No validation will be done for this study, as there may be no resources for this within the budget of the UNICOM project. Scripts

---

for semantic harmonization will be developed in R, distributed to data access providers for local deployment, and shared on the catalogue. The impact of choices of different algorithms will be assessed quantitatively. This will result in a set of study variables which are both semantically and syntactically harmonized.

3. Third, following conversion to harmonized study variable sets, R programs for calculation of incidence and prevalence can be distributed to data access providers for local deployment. The aggregated results produced by these scripts can be uploaded to a Digital Research Environment (DRE) for pooled analysis and visualization (see f*igure 4*). As per (EUPAS37273) s*uch a DRE is available at UMC Utrecht, called anDREa, and will be used as example. If available other DRE can be used in the UNICOM project.*
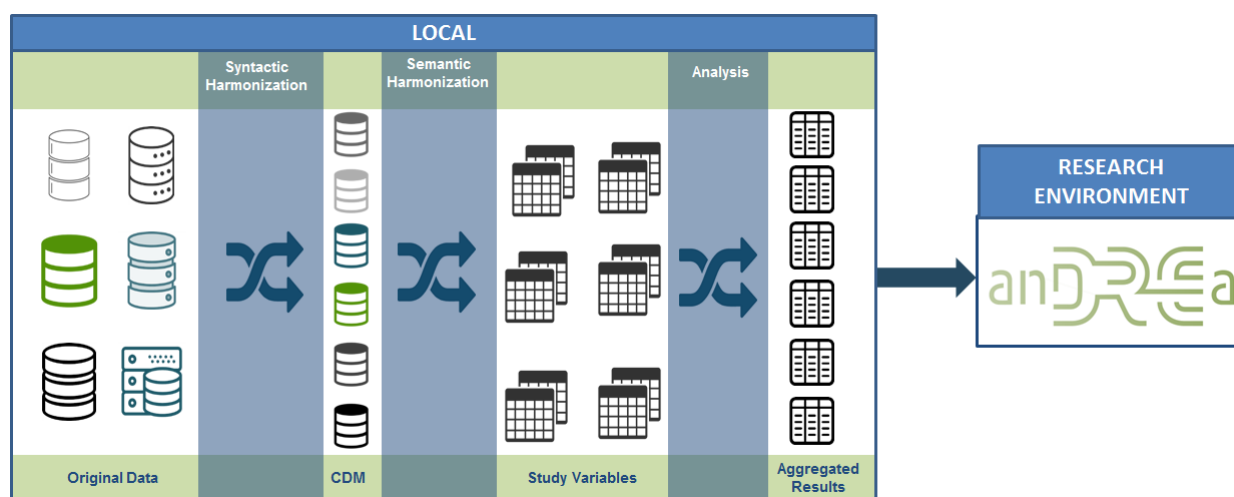

Figure 4. Data management plan.

9.6.1 Data extraction

Database access provider (DAP) listed above have already create ETL specifications using the standard ConcePTION ETL design template (accessible via this link: https://docs.google.com/document/d/1SWi31tnNJL7u5jJLbBHmoZa7AvfcVaqX7jiXgL9uAWg/edit.
ETL specifications may need to be adapted slightly based on the needs of the specific study and in case of updates of the tables of the original data sources. Upon completion of the template and review with study statisticians, each DAP can extract the relevant study data locally using their software (eg Stata, SAS, R, Oracle). This data can subsequently be loaded into the ConcepTION CDM structure in csv format. These data remain local, see *figure 4*.

9.6.2 Data Processing and transformation

Data processing and transformation should be conducted using R code against the syntactically harmonized CDM. The R scripts should first transform the data in the

syntactically harmonized CDM to semantically harmonized study variables (see *figure 4*). Following creation of study variables, the data should be characterized. This characterization will include calculation of code counts and incidence rates, as well as benchmarking within data source (over time), between data sources, and externally (against published estimates). ConcePTION makes code available on the GitHUB. Subsequently, R code to conduct analysis against semantically harmonized study variables should be distributed and run locally to produce aggregated results. The R scripts for these processing and analysis steps should be developed and tested centrally and sent to the DAPs. The R scripts will be structured in modular form in such a way that transparency is ensured. Functions to be used in the modules will be either standard R packages or packages designed, developed and tested on purpose for multi-database studies. As a result, scripts should be thoroughly documented and this will allow verification. The DAPs should run the R code locally and send aggregated analysis results to the digital research environment using a secure file transfer protocol. In the DRE, results will be further plotted, inspected (for quality assessment) and pooled (if needed) for final reporting.

9.6.3 Software and Hardware

All final statistical computations and pooling may be performed on the DRE using R and/or SAS. Data access providers should have access to the workspace for verification of the scripts.

9.6.4 Storage

Aggregated results, ETL specifications, and a repository of study scripts will be stored in the DRE.

9.6.5 Access

Within the DRE, each project-specific area consists of a separate, secure folder, called a 'workspace'. Each workspace is completely secure, so researchers are in full control of their data. Each workspace has its own list of users, which can be managed by its administrators.

The architecture of a DRE allows researchers to use a solution within the boundaries of data management rules and regulations. Although General Data Protection Regulation (GDPR) and Good (Clinical) Research Practice still rely on researchers, a DRE offers tools to more easily control and monitor which activities take place within projects.

All researchers who need access to DRE should be granted access to study-specific secure workspaces. Access to this workspace should only be possible with double authentication using an ID and password together with the user's mobile phone for authentication.

Upload of files should be possible for all researchers with access to the workspace within the DRE. Download of files is only possible after requesting and receiving permission from a workspace member with an 'owner' role.

9.6.6   Archiving and record retention

The final study aggregated results sets and statistical programs should be archived and stored on a central repository. The validation of the quality control (QC) of the statistical analysis should be documented. The final study protocol and possible amendments, the final statistical report, statistical programs and output files should be archived on a specific and secured drive centrally.

Documents that individually and collectively permit evaluation of the conduct of a study and the quality of the data produced should be retained for a period of 5 years in accordance with GPP guidelines. These documents could be retained for a longer period, however, if required by the applicable regulatory requirements or by an agreement between study partners. It is the responsibility of the principal investigator to inform the other investigators/institutions as to when these documents no longer need to be retained. Study records or documents may also include the analyses files, syntaxes (usually stored at the site of the database), ETL specifications, and output of data quality checks.

9.7 Data analysis

All analyses should be detailed in a Statistical Analysis Plan that should be developed ahead of data extraction.

9.7.1   Analysis of Demographics and Baseline Characteristics

Demographic characteristics and baseline characteristics such as co-morbidities and use of ATC codes in year prior will be summarized for each data source and for each cohort using descriptive statistics.

Frequency tables including numbers and percentages will be generated for categorical variables (age at cohort entry in categories, sex, and at-risk medical conditions).

Mean, standard error, median and range will be provided for continuous variables (e.g. age at cohort entry, duration of follow-up).

9.7.2   Hypotheses

The null hypothesis assumes that there is no difference in the risk of cardiovascular events between diclofenac potassium and diclofenac sodium.

9.7.3   Statistical Methods

In order to determine differences in the risk of major cardiovascular events associated with the use of two different pharmaceutical alternatives of oral diclofenac, two active comparator groups will be assessed: new users of diclofenac sodium *versus* new users of diclofenac potassium. Study subjects will be assigned to the cohorts by propensity scores matching. Propensity scores will be based con healthcare utilization and cardiovascular risk factors (see section on potential confounders). Propensity scores will be calculated using logistic

regression and the scores will be plotted to compare the overlap between the cohorts. Subjects in different cohorts will matched 1:1 on propensity score.

Incidence rates of acute myocardial infarction, heart failure and stroke will be compared between the study cohorts, adjusted for potential confounders if needed. A Cox proportional hazards model will be fitted to estimate the hazard ratios, their 95% confidence interval will be calculated as well. Primary and secondary endpoints will be estimated independently for each database to control for heterogeneity among databases.

The second study objective seeks to describe the availability of the pharmaceutical alternatives of oral diclofenac in the countries where this study will take place. The pathways to build up PhPIDs and MPIDs within the selected databases will be carefully described per country. The number of PhPIDs of diclofenac potassium and sodium will be aggregated per country and per marketing authorization holders (MPID).

9.7.4 Statistical Analysis

- Incidence rate (and 95% CI) of cardiovascular events will be calculated for each sub-cohort by data source and split by type of outcome: the numerator will be the number of incident cases after cohort entry calendar year in each data source and cohort. The denominator will be the total person-years at risk, i.e. from cohort entry to end of cohort time.
- Kaplan-Meier curves will be created to plot the hazard over time for the different sub-cohorts, followed by a proportional hazard model to estimate the hazard ratio.

9.7.5 Missing data

Since the underlying data represent attended medical care we generally assume that absence of information of clinical events means absence of that condition. No imputation will be done for missing data.

9.7.6 Sensitivity analysis

Sensitivity analysis will be conducted to investigate the impact of a different definition of the gap after the end of the last diclofenac prescription. The default is that with a gap of 30 days, the treatment episode ends, in a sensitivity analysis we will extend this to 90 days.

10. Protection of human subjects and data privacy

For this study, participants from various EU member states will process personal data from individuals which is collected in national/regional electronic health record databases. Due to the sensitive nature of this personal medical data it is important for to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy. According to the European Commission directive 95/46/EC, processing of personal data is legitimate for scientific purposes if adequate safeguards are provided and followed.

All member states have implemented this directive into their own national data protection legislation. All of the databases used in this study are currently already used for Pharmacoepidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to. To observe these regulations, rather than combining person level data and performing only a central analysis, local analyses will be run, which generate non-identifiable data with less information (e.g. no exact dates) that will be pooled across databases.

11. Limitations of the research methods

11.1 Limitations related to the data sources

This study will include different data sources which will be used to compute incidence rates of cardiovascular events. These data sources will be chosen based on availability, ability to run multisite studies and experience in using common data models.

Recorded disease diagnosis will be used as date to classify a case as incident, we will not be able to validate the events. No imputation will be performed for missing data, in this case assuming absence of the missing condition.

Prescribed regimens may not be available in the datasources, which would require to use a DDD based approach to estimate duration of use.

No imputation will be done for missing data, in this case assuming absence of the missing condition.

11.2 Limitations in the methodology

We use state-of-the-art methodology to create comparable cohorts of diclofenac sodium and potassium. However residual confounding that cannot be measured may remain.

12. Management and reporting of adverse events/adverse reactions

Aggregate analysis of database studies can identify an unexpected increase in risk associated with a particular exposure. Such studies may be reportable as study reports, but typically do not require reporting of individual cases. Moreover, access to automated databases does not confer a special obligation to assess and/or report any individual events contained in the databases. Formal studies conducted using these databases should adhere to these guidelines.

For non-interventional study designs that are based on secondary use of data, such as studies based on medical chart reviews or electronic health care records, systematic reviews or meta-analyses, reporting of adverse events/adverse drug reactions is not required. Reports of adverse events/adverse drug reactions should only be summarised in the study report, where applicable(28).

According to the EMA Guideline on Good Pharmacovigilance Practices (GVP), Module VI – Management and Reporting of Adverse Reactions to Medicinal Products(28). "All adverse events/reactions collected as part of [non-interventional post-authorisation studies with a design based on secondary use of data], the submission of suspected adverse reactions in the form of [individual case safety reports] is not required. All adverse events/reactions collected for the study should be recorded and summarised in the interim safety analysis and in the final study report."

Module VIII – Post-Authorisation Safety Studies, echoes this approach(29). The new legislation further states that for certain study designs such as retrospective cohort studies, particularly those involving electronic health care records, it may not be feasible to make a causality assessment at the individual case level.

13. Plans for disseminating and communicating study results

Once the data sources have been selected, the study protocol will be finalized and posted on the EU PAS register. Upon study completion and finalization of the study report, the results of this non-interventional study will be submitted for publication and posted in the EU PAS publicly accessible database of results. Publications will comply with the International Committee of Medical Journal Editors (ICMJE) guidelines.

14.   References

1.   Tranchard S. Revised IDMP standards to improve description of medicinal products worldwide. International Organization for Standardization. 2017. Available from: https://www.iso.org/news/ref2234.html

2.   The European Commission. Commission Implementing Regulation (EU) No 520/12 [Internet]. Official Journal of the European Union, 520/12. Available from: https://ec.europa.eu/health//sites/health/files/files/eudralex/vol-1/reg_2012_520/reg_2012_520_en.pdf

3.   Up-scaling the global univocal identification of medicines (UNICOM) [Internet]. Available from: https://unicom-project.eu/about/

4.   Diclofenac. Martindale: The Complete Drug Reference. Royal Pharmaceutical Society; 2020 [cited 2020 Oct 28]. Available from: https://www.medicinescomplete.com/#/content/martindale/13409-w?hspl=diclofenac

5.   Derry S, Wiffen PJ, Moore RA. Single dose oral diclofenac for acute postoperative pain in adults. Cochrane Database of Systematic Reviews 2015;2017..

6.   Masclee GMC, Straatman H, Arfè A, Castellsague J, Garbe E, Herings R, et al. Risk of acute myocardial infarction during use of individual NSAIDs: A nested case-control study from the SOS project. PLoS ONE. 2018 Nov 1;13(11).

7.   Arfè A, Scotti L, Varas-Lorenzo C, Nicotra F, Zambon A, Kollhorst B, et al. Non-steroidal anti-inflammatory drugs and risk of heart failure in four European countries: nested case-control study. BMJ. 2016 Sep 28;354:i4857.

8.   Schjerning Olsen AM, Fosbøl EL, Lindhardsen J, Andersson C, Folke F, Nielsen MB, et al. Cause-Specific Cardiovascular Risk Associated with Nonsteroidal Anti-Inflammatory Drugs among Myocardial Infarction Patients - A Nationwide Study. PLoS ONE. 2013;8(1).

9.   Bally M, Dendukuri N, Rich B, Nadeau L, Helin-Salmivaara A, Garbe E, et al. Risk of acute myocardial infarction with NSAIDs in real world use: Bayesian meta-analysis of individual patient data. BMJ. 2017;357.

10.   Baigent C, Bhala N, Emberson J, Merhi A, Abramson S, Arber N, et al. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: Meta-analyses of individual participant data from randomised trials. The Lancet. 2013;382(9894):769–79.

11.   Gislason GH, Jacobsen S, Rasmussen JN, Rasmussen S, Buch P, Friberg J, et al. Risk of death or reinfarction associated with the use of selective cyclooxygenase-2 inhibitors and nonselective nonsteroidal antiinflammatory drugs after acute myocardial infarction. Circulation. 2006;113(25):2906–13.

12.   Schjerning Olsen AM, Fosbøl EL, Lindhardsen J, Folke F, Charlot M, Selmer C, et al. Duration of treatment with nonsteroidal anti-inflammatory drugs and impact on risk of death and recurrent myocardial infarction in patients with prior myocardial infarction: A nationwide cohort study. Circulation. 2011 May 24;123(20):2226–35.

13.   Schmidt M, Sørensen HT, Pedersen L. Diclofenac use and cardiovascular risks: Series of nationwide cohort studies. BMJ. 2018;362.

14.   Verbeeck RK, Kanfer I, Walker RB. Generic substitution: The use of medicinal products containing different salts and implications for safety and efficacy. European Journal of Pharmaceutical Sciences. 2006;28:1–6.

15.   Bastin RJ, Bowker MJ, Slater BJ. Salt selection and optimisation procedures for pharmaceutical new chemical entities. Organic Process Research and Development. 2000;4(5):427–35.

16.   Committee for Medical Products for Human Use (CHMP). Guideline on the Investigation of Bioequivalence. London; 2010. Available from: http://www.ema.europa.eu

17.   Ahmad M, Iqbal M, Murtaza G. Comparison of bioavailability and pharmacokinetics of diclofenac sodium and diclofenac potassium in normal and dehydrated rabbits. Acta Pharmaceutica Sinica. 2009;44(1):80–4.

18.   Ahmad M IMMG. Comparison of bioavailability and pharmacokinetics of diclofenac sodium and diclofenac potassium in normal and alloxan-diabetic rabbits. Pak J Pharm Sci. 2012;25(2):301–6.

19.   Pacurariu A, Plueschke K, McGettigan P, Morales DR, Slattery J, Vogl D, et al. Electronic healthcare databases in Europe: Descriptive analysis of characteristics and potential for use in medicines regulation. BMJ Open. 2018;8(9).

20.   Hense S, Schink T, Kreisel SH, Marcelon L, Simondon F, Tahden M, et al. Estimation of Background Incidence Rates of Guillain-Barré Syndrome in Germany - A Retrospective Cohort Study with Electronic Healthcare Data. Neuroepidemiology. 2014;43(3–4):244-52.

21.   Schink T, Holstiege J, Kowalzik F, Zepp F, Garbe E. Risk of febrile convulsions after MMRV vaccination in comparison to MMR or MMR+V vaccination. Vaccine. 2014;32(6):645-650.

22.   Schmidt M, Schmidt SAJ, Adelborg K, Sundbøll J, Laugesen K, Ehrenstein V, et al. The Danish health care system and epidemiological research: from health care contacts to database records. Clinical Epidemiology. 2019;11:563-91.

23.   Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. European Journal of Epidemiology. 2014;29(8):541-9.

24.   Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. International Journal of Epidemiology. 2016;46(3):798–798f.

25.   Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clinical Epidemiology. 2015;7:449–90.

26.   Sahl Andersen J, de Fine Olivarius N, Krasnik A. The Danish National Health Service Register. Scandinavian Journal of Public Health. 2011;39(7_suppl):34-7.

27.   Becker BFH, Avillach P, Romio S, van Mulligen EM, Weibel D, Sturkenboom MCJM, et al. CodeMapper: semiautomatic coding of case definitions. A contribution from the ADVANCE project. Pharmacoepidemiology and Drug Safety. 2017;26(8):998-1005.

28.   European Medicines Agency. Guideline on good pharmacovigilance practices - Module VI – Collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2). 2017. Available from: www.ema.europa.eu

29.   European Medicines Agency. Guideline on good pharmacovigilance practices. Module-VIII-addendum-i-. 2020. Available from: www.ema.europa.eu

Annex - UMLS concepts for potential confounders (individual risk factors and comorbidities).

UMLS concepts for Obesity

Obesity Hypoventilation Syndrome

Other obesity

Simple obesity NOS

Obesity, Morbid

Central obesity

Generalized obesity

Body mass index 30+ - obesity

excessive weight (bmi>30)

Body Mass Index between 30-39, adult

Body Mass Index 40 and over, adult

Body Mass Index, pediatric, greater than or equal to 95th percentile for age

UMLS concepts for Alcohol Abuse

Alcoholic Intoxication, Chronic

    Chronic alcoholism in remission

    Episodic chronic alcoholism

    Continuous chronic alcoholism

    Other and unspecified alcohol dependence, unspecified drinking behavior

    Acute alcoholic intoxication

Alcohol abuse

    Alcoholic Intoxication, Chronic

    Nondependent alcohol abuse, unspecified

    Alcohol abuse, episodic drinking behavior

    Alcohol abuse, in remission

    Alcohol abuse, continuous drinking behavior

Liver Cirrhosis, Alcoholic

Hepatitis, Alcoholic

    Acute alcoholic liver disease

    Chronic Alcoholic Hepatitis

Alcoholic fibrosis and sclerosis of liver

Alcoholic hepatic failure

Alcoholic Liver Diseases

    Fatty Liver, Alcoholic

    Hepatitis, Alcoholic

    Liver Cirrhosis, Alcoholic

    Alcoholic fibrosis and sclerosis of liver

    Alcoholic hepatic failure

    Alcoholic liver damage

Fatty Liver, Alcoholic

Alcohol myopathy

Alcohol-induced chronic pancreatitis

Alcoholic Neuropathy

Cardiomyopathy, Alcoholic

Degeneration of nervous system due to alcohol

Alcoholic Liver Diseases

Alcohol units

    alcohol units/day

    alcohol units/week

UMLS concepts for Hypertension

Hypertensive disease

    Malignant Hypertension

    Essential Hypertension

    Secondary hypertension NOS

    Systolic hypertension

    Other specified hypertensive disease

    Hypertensive heart disease NOS

    Hypertensive heart and renal disease

hypertensive nephropathy

    hypertensive chronic kidney disease

    Hypertensive heart and chronic kidney disease

    hypertensive nephropathy

    Hypertensive renal disease with renal failure

    Hypertensive renal disease, unspecified, without mention of renal failure

    Malignant hypertensive renal disease

    Benign hypertensive renal disease

    Renal hypertension complicating pregnancy, childbirth and the puerperium unspecified

Uncomplicated hypertension

Hypertension complicated

Hypertensive heart disease NOS

    Hypertensive heart disease NOS without congestive cardiac failure

    Hypertensive heart disease NOS with congestive cardiac failure

    Malignant hypertensive heart disease NOS

    Benign hypertensive heart disease NOS

    Hypertensive heart and renal disease

    Pre-existing hypertensive heart disease complicating pregnancy, childbirth and the puerperium

    Unspecified hypertensive heart disease without heart failure

    Unspecified hypertensive heart disease with heart failure

Malignant Hypertension

Pre-existing essential hypertension complicating pregnancy, childbirth and the puerperium

Pre-existing hypertension complicating pregnancy, childbirth and puerperium

    Pre-existing hypertensive heart and renal disease complicating pregnancy, childbirth and the puerperium

    Pre-existing hypertensive heart disease complicating pregnancy, childbirth and the puerperium

    Pre-existing essential hypertension complicating pregnancy, childbirth and the puerperium

    Pre-existing hypertensive renal disease complicating pregnancy, childbirth and the puerperium

Pre-existing secondary hypertension complicating pregnancy, childbirth and puerperium

Unspecified pre-existing hypertension complicating pregnancy, childbirth and the puerperium

Pre-existing hypertensive disorder with superimposed proteinuria

UMLS concepts for Diabetes

Diabetes Mellitus

Diabetes Mellitus, Insulin-Dependent

Diabetes Mellitus, Non-Insulin-Dependent

Malnutrition related diabetes mellitus

Other specified diabetes mellitus

Diabetic peripheral angiopathy

Diabetic Neuropathies

Diabetes mellitus with no mention of complication

Secondary diabetes mellitus

Diabetes mellitus in mother complicating pregnancy, childbirth AND/OR puerperium

Diabetes mellitus NOS with ophthalmic manifestation

Diabetes mellitis with nephropathy NOS

Complications of Diabetes Mellitus

Diabetes mellitus with other specified manifestation

Diabetes mellitus, juvenile type, with no mention of complication

Diabetes mellitus, adult onset, with no mention of complication

Insulin-treated non-insulin-dependent diabetes mellitus

Brittle diabetes

Unspecified diabetes mellitus with renal complications

Other specified diabetes mellitus with other specified complications

Other specified diabetes mellitus with unspecified complications

Other specified diabetes mellitus with coma

Other specified diabetes mellitus with multiple complications

Unspecified diabetes mellitus with multiple complications

Pre-existing diabetes mellitus, unspecified

Diabetic gangrene

Unstable diabetes

Type I diabetes mellitus maturity onset

Unspecified diabetes mellitus with coma

Unspecified diabetes mellitus with ketoacidosis

Unspecified diabetes mellitus with other specified complications

Unspecified diabetes mellitus with ophthalmic complications

Unspecified diabetes mellitus with neurological complications

Unspecified diabetes mellitus with circulatory complication, unspecified

Unspecified diabetes mellitus with hyperosmolarity

Diabetes with other coma

Diabetic Ketoacidosis

Diabetic Nephropathy

Diabetes with other specified manifestations

Diabetes mellitus with hyperosmolarity

Diabetic arthropathy

Glomerular disorders in diabetes mellitus

Diabetic cataract

Diabetic Retinopathy

Nonproliferative diabetic retinopathy

Proliferative diabetic retinopathy

Moderate nonproliferative diabetic retinopathy

Severe nonproliferative diabetic retinopathy

Diabetic macular edema

mild nonproliferative diabetic retinopathy

UMLS concepts for Hyperlipidemia

Hyperlipidemia

hypercholesterolemia

Hyperlipidemia, Familial Combined

Hypertriglyceridemia

hypercholesterolemia

Other specified pure hypercholesterolemia

Pure hypercholesterolemia NOS

Disorders of lipoprotein metabolism and other lipidaemias

Hyperlipidemia

Hyperlipidemia, Familial Combined

Hyperlipoproteinemia Type IV

Hyperlipoproteinemia Type I

Other hyperlipidemia

Pure hypercholesterolemia NOS

UMLS concepts for Cardiac Arrhythmia

Cardiac arrhythmia

Sinus Arrhythmia

Tachycardia, Paroxysmal

Ectopic rhythm

Other specified cardiac arrhythmias

Other cardiac dysrhythmias

Cardiac Arrest

Paroxysmal supraventricular tachycardia

Paroxysmal ventricular tachycardia

Ventricular fibrillation and flutter NOS

Premature Cardiac Complex

Conduction disorder of the heart

Heart Block

Long QT Syndrome

Other specified conduction disorders

Accelerated atrioventricular conduction

Other conduction disorders NOS

Other and unspecified premature depolarization

Anomalous atrioventricular excitation NOS

Other heart block NOS

Right bundle branch block

Left bundle branch hemiblock NOS

Other left bundle branch block

Other and unspecified atrioventricular block

Bundle branch block, other and unspecified

Atrioventricular and left bundle-branch block

Left Bundle-Branch Block

First degree atrioventricular block

Second degree atrioventricular block

Left anterior fascicular block

Left posterior fascicular block

Other and unspecified atrioventricular block

Other and unspecified fascicular block

Cardiac Arrest

Cardiac arrest with successful resuscitation

Ventricular Fibrillation

Cardiac arrest as a complication of care

Electromechanical dissociation

Cardiopulmonary Arrest

UMLS concepts for Atrial Fibrillation and Flutter

Atrial Fibrillation

Paroxysmal atrial fibrillation

Non-rheumatic atrial fibrillation

Atrial Flutter

UMLS concepts for Transient Ischemic Attack

Transient Ischemic Attack

Subclavian Steal Syndrome

Middle Cerebral Artery Syndrome

Multiple and bilateral precerebral artery syndromes

Anterior Cerebral Artery Syndrome

Posterior Cerebral Artery Syndrome

Other transient cerebral ischemia

Other transient cerebral ischemic attacks and related syndromes

Carotid Circulation Transient Ischemic Attack

Transient Cerebral Ischemia

UMLS concepts for Cardiomyopathies

Cardiomyopathies

Cardiomyopathy, Dilated

Hypertrophic Cardiomyopathy

Primary eosinophilic endomyocardial restrictive cardiomyopathy

Chronic myocardial disorder due to chemical / external agent

Other hypertrophic cardiomyopathy

Other restrictive cardiomyopathy

Other cardiomyopathies

Restrictive cardiomyopathy

Other primary cardiomyopathy NOS

Ventricular hypertrophy

Endomyocardial Fibrosis

Obscure African cardiomyopathy

Other heart disease

Cardiomyopathies

UMLS concepts for Valvular Disorders and Endocarditis

Nonrheumatic heart valve disorder

Pulmonary valve disorder

Pulmonary Valve Insufficiency

Pulmonary Valve Stenosis

Other pulmonary valve disorders

Pulmonary valve stenosis with insufficiency

Rheumatic disease of pulmonary valve

Pulmonary valve anomaly, unspecified

Pulmonary valve disorders in diseases classified elsewhere

Non-rheumatic tricuspid valve disorder, unspecified

Other non-rheumatic tricuspid valve disorders

Tricuspid Valve Stenosis

Tricuspid valve regurgitation, non-rheumatic

Non-rheumatic tricuspid valve stenosis with insufficiency

Nonrheumatic aortic valve disorders
    Aortic Valve Insufficiency
    Aortic Valve Stenosis
    Other aortic valve disorders
    Aortic valve stenosis with insufficiency
    Aortic valve disorder
Valvular endocarditis
    Heart valve stenosis
    Chronic valvulitis
    Incompetence of unspecified heart valve
    Endocarditis, valve unspecified, OS
    Endocarditis in disease EC
    Endocarditis, valve unspecified, in diseases
classified elsewhere
    ENDOCARDITIS NEC in ICD9CM_2009
Nonrheumatic mitral valve disorder, unspecified
    Mitral Valve Insufficiency
    Mitral Valve Prolapse
    Other non-rheumatic mitral valve disorders
    Non-rheumatic mitral valve stenosis
    Rheumatic disease of heart valve
    Chronic rheumatic heart disease NOS
Rheumatic disease of mitral valve
Rheumatic aortic valve disease
    Rheumatic tricuspid valve disease
    Multiple valve disease
    diseases of other endocardial structures
    Mitral and aortic valve disease
    Aortic valve disorder
Rheumatic Heart Disease
    Acute rheumatic endocarditis
    Rheumatic endocarditis NOS
Acute and subacute endocarditis unspecified, NOS
    Acute endocarditis NOS
    Acute and subacute infective endocarditis
    Acute and subacute infective endocarditis
associated with another disorder
    Acute and subacute bacterial endocarditis NOS
    Subacute endocarditis
    Acute myoendocarditis NOS
    Subacute myoendocarditis NOS
    Acute periendocarditis NOS
    Subacute periendocarditis NOS

UMLS concepts for Myocarditis and Pericarditis

Myocarditis
    Acute myocarditis, unspecified

Toxic myocarditis

Myocarditis in other diseases classified elsewhere

Rheumatoid carditis

Rheumatoid myocarditis

Myocarditis due to infectious agent

Pericarditis

Acute pericarditis NOS

Chronic pericarditis

Infectious pericarditis

Pericarditis in bacterial diseases classified elsewhere

Pericarditis in other diseases classified elsewhere

UMLS concepts for Peripherial Arterial Diseases

Atherosclerosis

Generalized atherosclerosis

Atherosclerosis of aorta

Atherosclerosis of renal artery

Atherosclerosis of arteries of the extremities

Atherosclerosis of other arteries

Arteriosclerosis

Atherosclerotic Cardiovascular Disease

Other specified artery atheroma

Extremity artery atheroma NOS

Aortoiliac atherosclerosis

Monckeberg Medial Calcific Sclerosis

Atherosclerosis of other specified arteries

Of bypass graft of the extremities

Chronic total occlusion of artery of the extremities

UMLS concepts for Chronic Ischemic Heart Disease

Myocardial Ischemia

[Sign or Symptom] Angina Pectoris

Coronary Artery Vasospasm

Heart Aneurysm

Myocardial Infarction

Other forms of chronic ischemic heart disease

Microvascular Angina

Other acute and subacute ischemic heart disease NOS

Other specified ischemic heart disease

Coronary thrombosis not resulting in myocardial infarction

Other current complications following acute myocardial infarction

Single coronary vessel disease

Double coronary vessel disease

Other forms of acute ischemic heart disease

Coronary Arteriosclerosis

Procedures indicating carotid arterial disease

• carotid endarterectomy

• percutaneous coronary intervention (PCI) including percutaneous transluminal coronary

angioplasty (PTCA) and stent placement

• coronary artery bypass graft (CABG).

UMLS concepts for Arterial Embolism and Thrombosis

Arterial embolism

Fat embolism (disorder)

Arterial air embolus

Arterial thrombosis

Anterior spinal artery thrombosis

Thrombosis of renal artery

Superior mesenteric artery thrombosis

UMLS concepts for Blood Coagulation Disorders

Blood Coagulation Disorders

Blood Platelet Disorders

Hemophilia B

Disseminated Intravascular Coagulation

Hemophilia A

von Willebrand Disease

Purpura and other hemorrhagic conditions

Clotting or bleeding disorder NOS

Acquired coagulation disorder

Thrombophilia

Hemophilia carrier

Other specified coagulation defects

Hemophilia, NOS

Acquired coagulation factor deficiency NOS

Congenital deficiency of other clotting factor NOS

Factor XI Deficiency

Coagulation defects, other and unspecified

Hemorrhagic disorder due to intrinsic circulating anticoagulants

UMLS concepts for kidney failure

Kidney Failure

    Kidney Failure, Acute

    Kidney Failure, Chronic

    Renal failure as a complication of care

    Other acute renal failure

    Other chronic renal failure

    Postoperative renal failure

Renal Insufficiency

Procedures indicating kidney failure

    • Hemodialysis

    • Kidney transplantation

UMLS concepts for Liver Disease

Liver diseases

    Fatty Liver

    Hepatitis

    Hepatitis, Chronic

    Liver Cirrhosis

    Alcoholic Liver Diseases

    Other non-alcoholic chronic liver disease NOS

    Liver Failure

    Hepatic necrosis

    Liver abscess and chronic liver disease causing sequelae NOS

    Other sequelae of chronic liver disease

    Other liver disorders

    Subacute necrosis of liver NOS

    Chronic liver disease NOS

    Toxic liver disease with other disorders of liver

    Other and unspecified cirrhosis of liver

    Toxic liver disease

    Hepatic failure, not elsewhere classified

    Chronic hepatitis, not elsewhere classified

    Hepatic fibrosis and cirrhosis

UMLS concepts for Chronic Respiratory Disorders

Respiration Disorders

    Asthma

    Chronic Obstructive Airway Disease

    Other respiratory system diseases NOS

    Pulmonary Emphysema

    Complication of transplanted lung

Chronic lower respiratory diseases

Other disorders of lung

Respiratory Failure

CHRONIC OBSTRUCTIVE PULMONARY DISEASE AND ALLIED CONDITIONS

Annex 2: Additional study outlines

The potential of ISO/IDMP to allow a detailed description and comparison of therapeutic arsenals among European pharmaceutical markets.

Authors
Prof. dr. M.C.J.M. Sturkenboom, i~HD, University Utrecht Medical Center, Utrecht, The Netherlands.
Dr. Carlos E. Durán, i~HD, Clinical Pharmacology Research Group. Ghent University, Belgium.
Prof. dr. Robert Vander Stichele, i~HD, Clinical Pharmacology Research Group. Ghent University, Belgium.

Background
A detailed description of the availability of medicinal products in a given country, i.e. number of active substances within a therapeutic group, pharmaceutical forms, strengths, number and nature of marketing authorization (MA) holders, etc., is a fundamental step to allow a proper characterization of countries' pharmaceutical market and therefore, to correctly compare them. The implementation of ISO suite of standards for Identification of Medicinal Products (IDMP) could contribute to this goal.

Objective
To describe and compare the availability of medicinal products in [x] European countries through the implementation of ISO/IDMP concepts.

Methods
Information from European regulatory authorities and from the European Medicines Agency (Art. 57 database) will be collected to build up two fundamental IDMP concepts: Pharmaceutical Product Identification (PhPID) and Medicinal Product Identification (MPID). In every country's database, researchers will identify the active substance name, dosage form, route of administration, strength and units, in order to allow the subsequent description of PhPIDs. Simultaneously, the identification of the corresponding MA holders and country codes will be assigned to define MPIDs.

Results
The total number of PhPIDs and MPIDs (per country) will be presented. The number of MA holders per individual PhPIDs will be aggregated per country. Results will be organized according to the therapeutic level of the Anatomical, Therapeutic and Chemical (ATC) classification of drugs.

Discussion
This proof of concept study will allow a detailed characterization of the availability of medicinal products and thereby the granularity of countries´ therapeutic arsenal in Europe. Among others, it might have an impact on the quick identification of potential availability threats, e.g. (risk of) shortages of essential medicines. Furthermore, it may also allow a critical appraisal of the quality of the therapeutic arsenal in a given country.

Identification of pharmaceutical alternatives of amlodipine in European healthcare databases and incidence of cardiovascular events in hypertensive patients taking them.

Authors
Prof. dr. M.C.J.M. Sturkenboom, i~HD, University Utrecht Medical Center, Utrecht, The Netherlands.
Dr. Carlos E. Durán, i~HD, Clinical Pharmacology Research Group. Ghent University, Belgium.
Prof. dr. Robert Vander Stichele, i~HD, Clinical Pharmacology Research Group. Ghent University, Belgium.

Background
Amlodipine is a calcium-channel blocker indicated in the management of hypertension and angina pectoris. Pharmaceutical alternatives of amlodipine base have been marketed as besilate, maleate, benzoate and levamlodipine besilate, the isomer of amlodipine besilate. There is no published research exploring the occurrence of major cardiovascular events in users of different amlodipine pharmaceutical alternatives.

Objective
To assess if different pharmaceutical alternatives of amlodipine can be distinguished in European healthcare databases and to demonstrate the importance of that distinction by comparing the rate of cardiovascular events among hypertensive adult patients taking either one of the 4 amlodipine pharmaceutical alternatives.

Methods
Pharmaceutical Product Identification (PhPID) corresponding to amlodipine alternatives will be collected from selected European healthcare databases (active substance, route and form of administration, strength and units). Identified PhPIDs will be assigned according to the study cohorts. A retrospective dynamic cohort design (active comparator, new user) will be performed. European electronic health records and other applicable databases will be used. ≥ 18 years old hypertensive patients will be matched to one of the 4 study cohorts: amlodipine besilate initiators, amlodipine maleate initiators, amlodipine benzoate initiators and levamlodipine besilate initiators, and followed up until the occurrence of cardiovascular events, end of drug use or end of follow-up. Primary study outcome will be defined as a major cardiovascular event, fatal or not (heart failure, stroke and renal failure) occurred during the study period. Incidence rates of major and individual cardiovascular events will be calculated and compared among groups. A Cox proportional hazards model will be fitted to estimate the hazard ratio.

Results
PhPIDs of amlodipine pharmaceutical alternatives will be aggregated per country. Data collection pathways to get PhPIDs in every data source will be described. Rate of major and individual cardiovascular events will be reported for every amlodipine group [incidence rate, 95% confidence interval (CI)]. Outcome differences between study groups will be expressed as hazard ratios [95% CI].

Discussion
The ultimate goal of any antihypertensive treatment is to decrease the incidence of major cardiovascular complications. Although amlodipine has been evaluated in clinical trials and proven to be effective, different salts, esters or isomers (pharmaceutical alternatives) could influence its clinical performance. This study will bring evidence up regarding this aspect, while showing the strengths and limitations of amlodipine PhPID collection in European healthcare databases.

Identification of pharmaceutical alternatives of COVID-19 vaccines in European healthcare databases and the rate of severe COVID-19 between them

Authors
Prof. dr. M.C.J.M. Sturkenboom, i~HD, University Utrecht Medical Center, Utrecht, The Netherlands.
Dr. Carlos E. Durán, i~HD, Clinical Pharmacology Research Group. Ghent University, Belgium.
Prof. dr. Robert Vander Stichele, i~HD, Clinical Pharmacology Research Group. Ghent University, Belgium.

Background
COVID-19 vaccines have been developed in record time, and currently 10 products are in phase 3 testing. Severa products will be licensed. In order to speed up distribution, multi-dose vials will be used and only secondary packaging may contain the full identification of the product.

Objective
To assess if different pharmaceutical alternatives of COVID-19 vaccines can be distinguished in European healthcare databases/immunization registers and to demonstrate the importance of that distinction by comparing the rate of severe COVID-19 between them.

Methods
Pharmaceutical Product Identification (PhPID) corresponding to COVID-19 alternatives will be collected from selected European healthcare databases (brand, lot number). Identified PhPIDs will be assigned according to the study cohorts. A retrospective dynamic cohort design (active comparator, new user) will be performed according the ACCESS protocol to assess effectiveness in EHR data sources. European electronic health records and other applicable databases will be used. A Cox proportional hazards model will be fitted to estimate the hazard ratio.

Results
PhPIDs of COVID-19 vaccines will be aggregated per country. Data collection pathways to get PhPIDs in every data source will be described. Rate of sever COVID-19 will be reported for type of COVID-19 vaccine [incidence rate, 95% confidence interval (CI)], by time since vaccination. Outcome differences between study groups will be expressed as hazard ratios [95% CI].

Discussion
The ultimate goal of COVID-19 vaccines is to decrease the incidence of COVID-19 disease. Although COVID-19 vaccines are being evaluated in clinical trials and licensed when proven to be effective, different platforms, could influence its clinical performance. This study will generate evidence up regarding this aspect, while showing the strengths and limitations of COVID-19 PhPID collection in European healthcare databases.