



EHDEN

EUROPEAN HEALTH DATA & EVIDENCE NETWORK

806968 – EHDEN

European Health Data & Evidence Network

WP1 – Evidence Workflow Development

D1.5 Results of "validation" Use Case 1 study

Lead contributor	Katia Verkamme (1 – EMC) Alan Lamb (NICE)
Lead contributor email	k.verhamme@erasmusmc.nl ; Alan.lamb@nice.org.uk
Other contributors	Aniek Markus (1-EMC) Daniel Prieto-Alhambra (3-UOXF) Dalia Dawoud (NICE)
Due date	30/11/2021
Delivery date	14/01/2022
Deliverable type	R
Dissemination level	PU
DoA - Version	V2
Date	15/10/2021



	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development	Version: v3.0 – Final	
	Author(s): Katia Verhamme; Alan Lamb	Security: PU	2/14



TABLE OF CONTENTS


Table of contents	2
Document History	3
Definitions	4
Publishable Summary.....	5
1. Introduction	6
2. Methods	7
2.1 Study period	7
2.2 Study population	7
2.3 Exposure cohort.....	7
3. Results	9
3.1 total number of patients in the source population	9
3.2 total number of patients in the study population.....	10
3.3 total number of patients in the exposure cohorts	10
4. Discussion	12
5. Conclusion	14
References.....	14

	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development		Version: v3.0 – Final
	Author(s): Katia Verhamme; Alan Lamb	Security: PU	3/14



DOCUMENT HISTORY

Version	Date	Description
V1.0	30 Nov 2021	First Draft for formal review
V2.0	6 Jan 2022	Second draft for consortium review
V3.0	14 Jan 2022	Final version for submission

 EHDEN <small>EUROPEAN HEALTH DATA & EVIDENCE NETWORK</small>	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development	Version: v3.0 – Final	
	Author(s): Katia Verhamme; Alan Lamb	Security: PU	4/14




DEFINITIONS

Participants of the EHDEN Consortium are referred to herein according to the following codes:

EMC	Erasmus Universitair Medisch Centrum Rotterdam- The Netherlands (Project Coordinator)
Synapse	Synapse Research Management Partners S.L. - Spain
UOXF	The Chancellor, Masters and Scholars of the University of Oxford - United Kingdom
UTARTU	Tartu Ulikool - Estonia
UAVR	Universidade de Aveiro – Portugal
The Hyve	The Hyve BV – the Netherlands
Odysseus	Odysseus Data Services SRO – Czech Republic
EPF	European Patients’ Forum (EPF) - Belgium
NICE	National Institute for Health and Care Excellence – United Kingdom
UMC	Stiftelsen WHO Collaborating Centre for International Drug Monitoring - Sweden
ICHOM	International Consortium for Health Outcomes measurement LTD - United Kingdom
Janssen	Janssen Pharmaceutica NV - Belgium (Project Lead)
Pfizer	Pfizer Limited – United Kingdom
Abbvie	AbbVie Inc - United States
IRIS	Institut De Recherches Internationales Servier – France
SARD	Sanofi Aventis Recherche & Developpement – France
Bayer	Bayer Aktiengesellschaft – Germany
Lilly	Eli Lilly and Company Limited – United Kingdom
AZ	AstraZeneca AB – Sweden
Novartis	Novartis Pharma AG – Switzerland
UCB	UCB Biopharma SPRL – Belgium
Celgene	Celgene Management SARL – Switzerland
BI	Boehringer Ingelheim – Germany


Grant agreement	The agreement signed between the beneficiaries and the IMI JU for the undertaking of the EHDEN project (806968).
Project	The sum of all activities carried out in the framework of the Grant Agreement.
Consortium	The EHDEN Consortium, comprising the above-mentioned legal entities.
Consortium agreement	Agreement concluded amongst EHDEN participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties’ obligations to the Community and/or to one another arising from the Grant Agreement.

	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development		Version: v3.0 – Final
	Author(s): Katia Verhamme; Alan Lamb		Security: PU 5/14



PUBLISHABLE SUMMARY

Healthcare data, generated through the delivery of normal clinical care is increasingly being used to investigate the use of drugs in real life. As part of previous research with EHDEN, we had investigated treatment patterns of patients with asthma, COPD or asthma-COPD overlap. For this research, data from the IPCI database, a Dutch primary care database, was used. IPCI data was mapped to the OMOP CDM which allows the generation of one software script to explore treatment patterns which can then be run on other data sources. As a follow-up to our research on the treatment patterns of asthma & COPD, we wanted to explore whether the study population, as selected using IPCI data mapped to the OMOP CDM was similar to IPCI source data (thus not mapped to the OMOP CDM). To do so, we defined criteria for our study period, study population and exposure cohorts and extracted the data (IPCI source & IPCI CDM). Some differences in number of patients were identified but these could be attributed to the fact that IPCI CDM applies quality standards which result in some patients being excluded. For example, diagnoses with missing dates and multiple prescriptions of the same drug on the same day are disregarded. The remaining differences between IPCI source and IPCI CDM were related to discrepancies in mapping drugs from ATC code to RxNorm, which is challenging for inhalation therapy.

	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development	Version: v3.0 – Final	
	Author(s): Katia Verhamme; Alan Lamb	Security: PU	6/14



1. INTRODUCTION

Healthcare data, generated through the delivery of normal clinical care is increasingly being proposed as a source of evidence to support not only drug development and regulatory decision-making but also to understand the physiology and pathogenesis of diseases.


Use of multiple electronic health care databases is important not only to increase sample size but also to investigate country specific differences, differences by type of databases (e.g., primary vs. secondary care) or to replicate findings. One of the challenges, however, are the differences between the databases regarding the underlying structures and semantic mapping. A common data model (CDM) could help harmonise healthcare data across multiple data sets and provide a mechanism to allow the conduct of multi-database, international studies. ⁽¹⁾

The European Health Data and Evidence Network (EHDEN) project (<https://www.ehden.eu/>) is an international project supported by the Innovative Medicines Initiative (IMI) that standardises health care data to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM). In addition, EHDEN develops and implements tools to facilitate research on large electronic health care databases, and offers free, on-demand courses via the EHDEN Academy that address the tools, skills and method training for researchers working with real world observational data in generating RWE.

In addition to testing existing methodologies, one of the objectives of the EHDEN project is to develop new methodologies and analytical tools to conduct (pharmaco) epidemiological research using electronic health care databases mapped to the OMOP CDM. To investigate the validity and functionality of this approach, we conducted a drug-utilisation study using electronic health records (EHR) data. As proof of concept, we performed drug utilisation studies on drug use in respiratory patients with asthma and chronic obstructive pulmonary disease (COPD) with a focus on treatment patterns.

For this research, source data mapped to the OMOP CDM were used. To investigate whether mapping of data to the OMOP CDM would affect study results, we conducted a validation study using data from the IPCI (Integrated Primary Care Information) database <https://www.ipci.nl/>.

IPCI (Integrated Primary care Information database) data is collected from EHR records of patients registered with their General Practitioner (GP) throughout the Netherlands. ⁽²⁾ The selection of 391 GPs is representative of the entire country. The database contains records from 2.5 million patients out of a Dutch population of 17M (14.7%) starting in 1996. The median follow-up is 2.2 years. The observation period for a patient is determined by the date of registration at the GP and the date of leave/death. The observation period start date is refined by many quality indicators, e.g. exclusion of peaks of conditions when registering at the GP. All data before the observation period is kept as historical data. Drugs are captured as prescription records

	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development	Version: v3.0 – Final	
	Author(s): Katia Verhamme; Alan Lamb	Security: PU	7/14



with product, quantity, dosing directions, strength and indication. Drugs not prescribed in the GP setting might be underreported. Indications are available as diagnoses by the GPs and, indirectly, from secondary care providers but the latter might not be complete. Approval needs to be obtained for each study from the Governance Board. ⁽²⁾

For the research as presented in this deliverable, we explored whether there were differences in the number of patients included in the respective cohorts (exposure cohort, disease cohort) when using data mapped to the OMOP CDM vs. source data.

2. METHODS

To investigate differences in the cohort sizes (exposure cohorts and disease cohorts) data were extracted from the IPCI source cohort as well as from IPCI mapped to the OMOP CDM. For this study, the IPCI data with last data extraction up to 1/1/2020 was used (both for source and mapped data).

2.1 Study period

The study period was from 1/1/2010 until 1/1/2020.

2.2 Study population

The study population consisted of adult patients (≥ 18 years during the study period) with active follow-up time during the study period and diagnosed with asthma (either prevalent or incident). For the IPCI source, presence of asthma was based on an ICPC (International Classification of Primary Care) disease code of "R96" and for the IPCI data mapped to the OMOP CDM, adult patients diagnosed with asthma OHDSI concept ID '317009' OR concept ID = '4191479'.


2.3 Exposure cohort

Exposure to respiratory drugs was identified via an automated Anatomical Therapeutic Chemical (ATC) search for IPCI source data and via exposure concept IDs for the data mapped to the OMOP CDM.⁽³⁾ Exposure to respiratory drugs meant that the patient was prescribed any of the respiratory drugs of interest during the study period. Exposure was assessed in two ways, namely by the number of patients receiving respiratory drugs and by a count of the number of prescriptions.

The description of the ATC codes and respective concept IDs of interest are described in Table 1.

Table 1: Identification of respiratory drugs via ATC code or concept ID

Drug classes	ATC codes	Concept ID
SABA (short acting B2 agonists)	R03AC02, R03AC03, R03AC04	<p>Concepts with following ingredients:</p> <p>1154343, 1236744, 19053979, 19063387, 19068969, 1181809, 1125449, 35198052, 40798689, 19035396, 19028950, 1138050</p> <p>And dose forms:</p> <p>Inhalant, Inhalation Powder, Inhalation Solution, Inhalation Spray, Inhalation Suspension, Metered Dose Inhaler, Dry Powder Inhaler, Gas for Inhalation</p>
ICS (inhaled corticosteroids)	R03BA	<p>Concepts with following ingredients:</p> <p>1115572, 939259, 1196514, 1149380, 905233, 902938, 920458, 903963</p> <p>And dose forms:</p> <p>Inhalant, Inhalation Powder, Inhalation Solution, Inhalation Spray, Inhalation Suspension, Metered Dose Inhaler, Dry Powder Inhaler, Gas for Inhalation</p>
Systemic glucocorticoids	H02AB	<p>Concepts with following ingredients:</p> <p>920458, 1518254, 19055344, 1506270, 19027186, 1550557, 1551099, 903963, 975125, 1507705, 19011127, 977421, 19086888, 19050907, 19009116, 19061907, 19055156</p> <p>And dose forms:</p> <p>Injectable Solution, Injectable Suspension, Auto-Injector, Intravenous Solution, Prefilled Syringe, Injection, Intramuscular Solution, Buccal Tablet, Delayed Release Oral Tablet, Disintegrating Oral Tablet, Chewable Tablet, Delayed Release Oral Capsule, Extended Release Oral Tablet, Extended</p>

	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development	Version: v3.0 – Final	
	Author(s): Katia Verhamme; Alan Lamb	Security: PU	9/14




		Release Suspension, Oral Capsule, Oral Gel, Oral Granules, Oral Lozenge, Oral Ointment, Oral Paste, Oral Powder, Oral Solution, Oral Suspension, Effervescent Oral Tablet, Extended Release Oral Capsule, Powder for Oral Suspension, Sublingual Tablet, Tablet for Oral Suspension, Oral Tablet, Pack
LABA (long acting B2 agonists)	R03AC12, R03AC13, R03AC18, R03AC19	<p>Concepts with following ingredients:</p> <p>19043191, 1137529, 1196677, 19097824, 40240664, 45775116, 43532539</p> <p>And dose forms:</p> <p>Inhalant, Inhalation Powder, Inhalation Solution, Inhalation Spray, Inhalation Suspension, Metered Dose Inhaler, Dry Powder Inhaler, Gas for Inhalation</p>
LAMA (long-acting muscarinic antagonists)	R03BB04, R03BB05, R03BB06, R03BB07, R03BB08)	<p>Concepts with following ingredients:</p> <p>1106776, 42873639, 45775571, 44785907</p> <p>And dose forms:</p> <p>Inhalant, Inhalation Powder, Inhalation Solution, Inhalation Spray, Inhalation Suspension, Metered Dose Inhaler, Dry Powder Inhaler, Gas for Inhalation</p>

3. RESULTS

3.1 total number of patients in the source population

The total number of individuals with active follow-up during the study period was identical between the IPCI source vs. the IPCI data mapped to the OMOP CDM, namely 2,369,989 individuals.

	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development	Version: v3.0 – Final	
	Author(s): Katia Verhamme; Alan Lamb	Security: PU	10/14



3.2 total number of patients in the study population

Next, the number of individuals with a medical diagnosis of asthma was investigated. As we only selected individuals with a disease code of asthma (ICPC code in source vs. OHDSI concept ID in IPCI mapped to OMOP CDM) it is likely that we underestimated the true population of patients with asthma. Initially, there were some (small) discrepancies between the IPCI source and IPCI data mapped to the OMOP CDM. These differences could be explained by the following:

- IPCI source data also has patients with a diagnosis code of asthma, but where the date of diagnosis is empty. Events with missing diagnosis dates were excluded in the IPCI data mapped to the OMOP CDM.
- When data were extracted via the OMOP CDM, disease codes after the end of follow-up (IPCI end date in IPCI source data) were still selected whereas this is not the case for the IPCI source data.


When excluding patients without a date of asthma diagnosis and excluding patients where asthma was diagnosed after the end of follow-up, the number of adult patients with asthma and with active follow-up during the study period was 248,776 (which was similar for the IPCI source data compared to the IPCI OMOP CDM).

3.3 total number of patients in the exposure cohorts

Next, the number of patients treated with any of the respiratory drugs of interest (during the study period and since date of asthma diagnosis) was investigated. Some considerable differences between the number of patients receiving prescriptions for respiratory drugs as well as differences in the total number of prescriptions were observed. This is documented in Table 2:

Table 2: Number of patients treated with respiratory drugs and number of prescriptions (source IPCI data vs. data mapped to the OMOP CDM)

	IPCI-Source	IPCI-CDM	Comments
Drug classes	# prescriptions (# patients)	# prescriptions (# patients)	Problem mapping RxNorm vs ATC?
SABA	661,641 (116,381)	369,706 (111,556)	Source much more prescriptions
ICS	415,151 (67,659)	229,412 (70,878)	Source much higher prescriptions, but CDM more people
Systemic glucocorticoids	248,820 (51,841)	154,456 (62,789)	Source much higher prescriptions, but CDM more people
LABA	99,536 (12,274)	50,648 (14,603)	Source much higher prescriptions, but CDM more people

	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development		Version: v3.0 – Final
	Author(s): Katia Verhamme; Alan Lamb		Security: PU 11/14



	IPCI-Source	IPCI-CDM	Comments
Drug classes	# prescriptions (# patients)	# prescriptions (# patients)	Problem mapping RxNorm vs ATC?
LAMA	168,849 (16,105)	48,342 (16,173)	Source much higher prescriptions


This difference could be explained because in IPCI, there were patients with more than one prescription of the same drug on the same day which was not the case for data mapped to the OMOP CDM. In addition, for the IPCI source data, respiratory drugs prescribed during the study period but prior to the date of asthma diagnosis were also considered. To compare IPCI source vs. IPCI CDM, data from IPCI source were again extracted but now only selecting drugs prescribed during the study period and after the date of asthma diagnosis and disregarding subsequent prescriptions of the same drug on the same day. Table 3 provides the number of patients and number of prescriptions following this new extraction.

Table 3: Number of patients treated with respiratory drugs and number of prescriptions (source IPCI data vs. data mapped to the OMOP CDM) – only considering drugs during study period and following asthma diagnosis

	Source	CDM	Comments
Drug classes	# prescriptions (# patients)	# prescriptions (# patients)	
SABA (= ATC R03AC02, R03AC03, R03AC04)	645,289 (116,381)	369,706 (111,556)	Source much higher prescriptions
ICS (= ATC R03BA)	412,575 (65,089)	229,412 (70,878)	Source much higher prescriptions, but CDM more people
Systemic glucocorticoids (= ATC H02AB)	227,676 (51,708)	154,456 (62,789)	Source much higher prescriptions, but CDM more people
LABA (= ATC R03AC12, R03AC13, R03AC18, R03AC19)	99,238 (12,274)	50,648 (14,603)	Source much higher prescriptions, but CDM more people
LAMA (= ATC R03BB04, R03BB05, R03BB06, R03BB07, R03BB08)	163,500 (16,105)	48,342 (16,173)	Source much higher prescriptions

Following this new extraction, the total number of prescriptions was still much lower for IPCI mapped to OMOP CDM vs. IPCI source (table 3)

To further explore the differences between IPCI source and IPCI CDM, a sample of some individual patients was compared. In IPCI, no treatment periods were created based on subsequent (overlapping) prescriptions, whereas ATLAS (a web-based OHDSI tool that automatically generates exposure eras)

	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development	Version: v3.0 – Final	
	Author(s): Katia Verhamme; Alan Lamb	Security: PU	12/14



(<https://www.ohdsi.org/software-tools/atlas/>) combines prescriptions into treatment periods explaining why the number of prescriptions (per patient) was lower for IPCI CDM compared to IPCI source. Other discrepancies could be explained by differences in mapping to RxNorm for ICS and systemic steroids. For these drug ingredients, it is indeed not easy to differentiate whether drugs are administered via inhalation or systemically.

Of the individual patients that were finally being checked, patients indeed belonged to the same drug class and, after applying to the Source data the same rule of combining prescriptions into exposure eras, the number of prescriptions were similar between IPCI source and IPCI CDM.

4. DISCUSSION

This validation study was conducted to investigate important differences if data were extracted using the source data compared to data mapped to the OMOP CDM.

No differences in source population could be observed, however, some differences between the disease cohorts (number of patients with asthma and with active follow-up during the study period) and the exposure cohorts were observed.


These differences could be attributed to the following:

- Quality control in IPCI mapped to the OMOP CDM where disease codes with empty date of diagnosis were discarded.
- Diseases occurring after end of follow-up which were discarded in IPCI source but not in the IPCI CDM.

These types of differences can easily be resolved by first of all improving the quality of source data (i.e., not allowing events with empty dates) and improving quality of data extraction by not considering events/drugs occurring after the study period.

More important differences were observed for the exposure cohorts which could be explained by the following:


- Difficulties in mapping respiratory drugs to RxNorm especially for ICS and systemic steroids as way of administration are not always documented in the data source.
- If individual prescriptions overlap, ATLAS generates treatment episodes which was not the case for IPCI source.

	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development	Version: v3.0 – Final	
	Author(s): Katia Verhamme; Alan Lamb	Security: PU	13/14



ATLAS already considers overlap in the creation of treatment episodes which is fine, and this would be the next step a researcher would do when individual prescriptions are extracted. Flexibility in the definition of overlap would be recommended where the researcher, using ATLAS, has the possibility to define the size of gaps and size of overlap when creating drug eras (treatment episodes). As already described by other research groups, mapping of respiratory drugs to RxNorm proves to be challenging especially if information on strength and way of administration is missing.⁽⁴⁾ Within OHDSI, continuous efforts are ongoing to optimise mapping of drugs to RxNorm (and check for consistency with ATC codes).

In this exercise, we did not investigate whether the differences in exposure cohorts would affect the interpretation of treatment patterns as the related “treatment pattern software package” only ran on the IPCI CDM data and not on the source data. This is an interesting question for further research.

	D1.5- Validation Study Results of "validation" Use Case 1 study		
	WP1 -Evidence Workflow Development	Version: v3.0 – Final	
	Author(s): Katia Verhamme; Alan Lamb	Security: PU	14/14



5. CONCLUSION

Some differences between data extracted via source data or via data mapped to the OMOP CDM were observed. Upon further investigation these differences could be explained by data quality measures where events with missing data and multiple prescriptions of the same drug on the same day were disregarded in IPCI CDM. The remaining differences between IPCI source and IPCI CMD were related to difficulties in mapping to RxNorm especially for ICS and systemic steroids if the route of administration is unclear. As part of subsequent research, in joint collaboration with OHDSI coding teams, we are working together to optimize mapping from ingredient level (and/or ATC code) to RxNorm.

REFERENCES

1. Gini R, Schuemie M, Brown J, Ryan P, Vacchi E, Coppola M, Cazzola W, Coloma P, Berni R, Diallo G, Oliveira JL, Avillach P, Trifiro G, Rijnbeek P, Bellentani M, van Der Lei J, Klazinga N, Sturkenboom M. Data Extraction and Management in Networks of Observational Health Care Databases for Scientific Research: A Comparison of EU-ADR, OMOP, Mini-Sentinel and MATRICE Strategies. EGEMS (Wash DC). 2016;4(1):1189.
2. Vlug AE, van der Lei J, Mosseveld BM, van Wijk MA, van der Linden PD, Sturkenboom MC, van Bemmelen JH. Postmarketing surveillance based on electronic patient records: the IPCI project. Methods of information in medicine. 1999;38(4-5):339-44.
3. Lagerlund O, Strese S, Fladvad M, Lindquist M. WHODrug: A Global, Validated and Updated Dictionary for Medicinal Information. Ther Innov Regul Sci. 2020;54(5):1116-22.
4. Freimuth RR, Wix K, Zhu Q, Siska M, Chute CG. Evaluation of RxNorm for Medication Clinical Decision Support. AMIA Annu Symp Proc. 2014;2014:554-63.