LLM Innovations Throughout OHDSI

OHDSI Community Call
Aug. 26, 2025 • 11 am ET

Upcoming Community Calls

Date	Topic
Sept. 2	Standardized Vocabulary Summer Refresh Update
Sept. 9	Global Symposium Preview
Sept. 16	OHDSI/OMOP Research Spotlight
Sept. 23	Educating on OHDSI: Lessons Learned
Sept. 30	OHDSI 2025 Poster Preview Mad Minutes / Symposium Logistics
Oct. 7	No Call – OHDSI Symposium
Oct. 14	Welcome to OHDSI
Oct. 21	Meet the Titans

Three Stages of The Journey

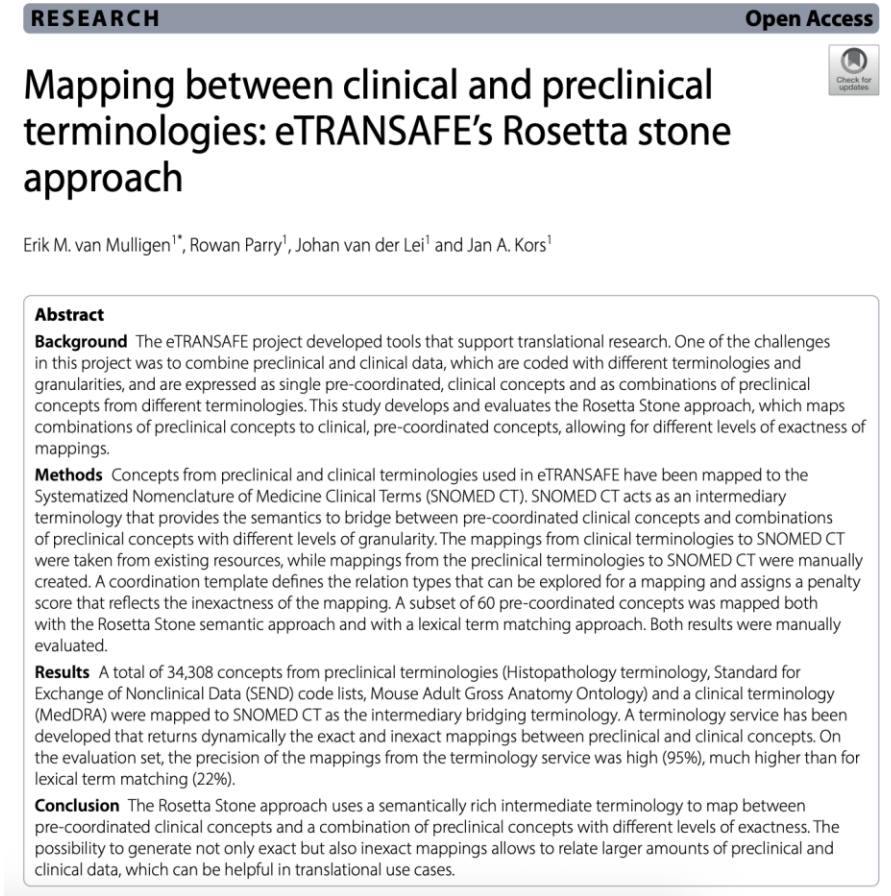
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Erik M. van Mulligen, Rowan Parry, Johan van der Lei, and Jan A. Kors** on the publication of **Mapping between clinical and preclinical terminologies: eTRANSAFE's Rosetta stone approach** in the *Journal of Biomedical Semantics.*



van Mulligen *et al. Journal of Biomedical Semantics* (2025) 16:15
https://doi.org/10.1186/s13326-025-00337-2

Journal of Biomedical Semantics

RESEARCH Open Access

Mapping between clinical and preclinical terminologies: eTRANSAFE's Rosetta stone approach

Erik M. van Mulligen[1*], Rowan Parry[1], Johan van der Lei[1] and Jan A. Kors[1]

Abstract

Background The eTRANSAFE project developed tools that support translational research. One of the challenges in this project was to combine preclinical and clinical data, which are coded with different terminologies and granularities, and are expressed as single pre-coordinated, clinical concepts and as combinations of preclinical concepts from different terminologies. This study develops and evaluates the Rosetta Stone approach, which maps combinations of preclinical concepts to clinical, pre-coordinated concepts, allowing for different levels of exactness of mappings.

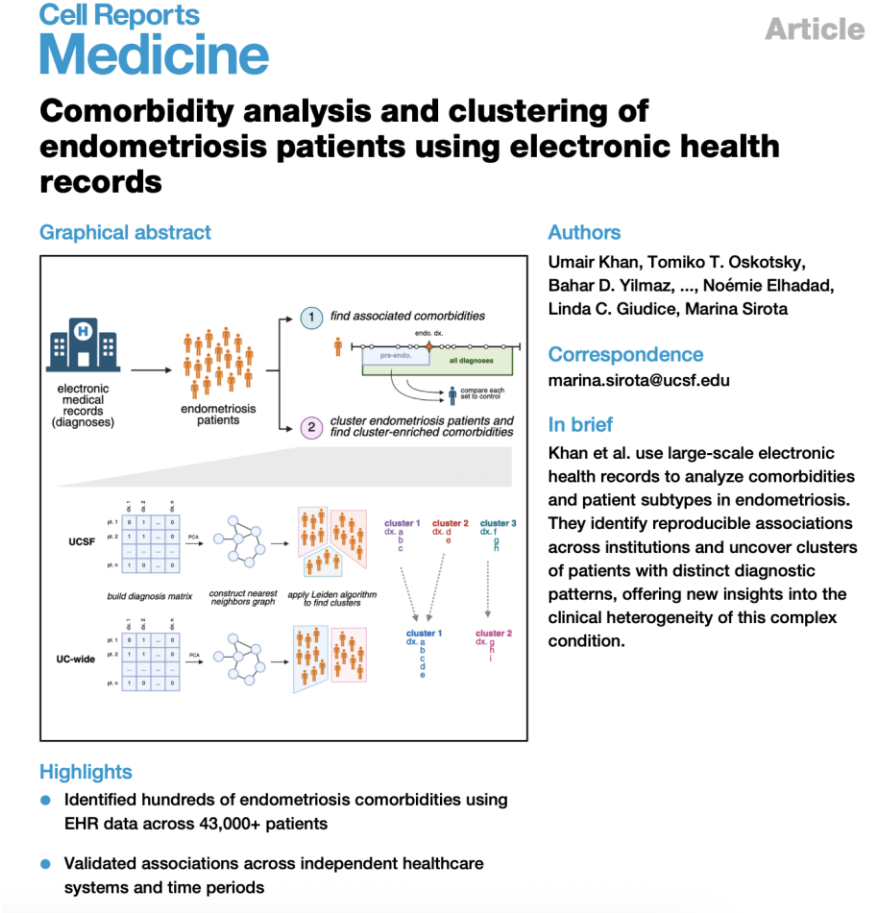
Methods Concepts from preclinical and clinical terminologies used in eTRANSAFE have been mapped to the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT). SNOMED CT acts as an intermediary terminology that provides the semantics to bridge between pre-coordinated clinical concepts and combinations of preclinical concepts with different levels of granularity. The mappings from clinical terminologies to SNOMED CT were taken from existing resources, while mappings from the preclinical terminologies to SNOMED CT were manually created. A coordination template defines the relation types that can be explored for a mapping and assigns a penalty score that reflects the inexactness of the mapping. A subset of 60 pre-coordinated concepts was mapped both with the Rosetta Stone semantic approach and with a lexical term matching approach. Both results were manually evaluated.

Results A total of 34,308 concepts from preclinical terminologies (Histopathology terminology, Standard for Exchange of Nonclinical Data (SEND) code lists, Mouse Adult Gross Anatomy Ontology) and a clinical terminology (MedDRA) were mapped to SNOMED CT as the intermediary bridging terminology. A terminology service has been developed that returns dynamically the exact and inexact mappings between preclinical and clinical concepts. On the evaluation set, the precision of the mappings from the terminology service was high (95%), much higher than for lexical term matching (22%).

Conclusion The Rosetta Stone approach uses a semantically rich intermediate terminology to map between pre-coordinated clinical concepts and a combination of preclinical concepts with different levels of exactness. The possibility to generate not only exact but also inexact mappings allows to relate larger amounts of preclinical and clinical data, which can be helpful in translational use cases.

OHDSI Shoutouts!

Congratulations to the team of **Umair Khan, Tomiko T Oskotsky, Bahar D Yilmaz, Jacquelyn Roger, Ketrin Gjoni, Juan C Irwin, Jessica Opoku-Anane, Noémie Elhadad, Linda C Giudice, and Marina Sirota** on the publication of **Comorbidity analysis and clustering of endometriosis patients using electronic health records** in the *Cell Reports Medicine.*



Cell Reports Medicine

Article

Comorbidity analysis and clustering of endometriosis patients using electronic health records

Graphical abstract

Authors

Umair Khan, Tomiko T. Oskotsky, Bahar D. Yilmaz, ..., Noémie Elhadad, Linda C. Giudice, Marina Sirota

Correspondence

marina.sirota@ucsf.edu

In brief

Khan et al. use large-scale electronic health records to analyze comorbidities and patient subtypes in endometriosis. They identify reproducible associations across institutions and uncover clusters of patients with distinct diagnostic patterns, offering new insights into the clinical heterogeneity of this complex condition.

Highlights

- Identified hundreds of endometriosis comorbidities using EHR data across 43,000+ patients
- Validated associations across independent healthcare systems and time periods

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	ATLAS/WebAPI
Wednesday	10 am	Surgery and Perioperative Medicine
Wednesday	10 am	Women of OHDSI
Thursday	12 pm	Latin America
Thursday	9:30 am	Network Data Quality
Friday	10 am	GIS – Geographic Information System
Friday	11:30 am	Steering
Monday	9 am	Vaccine Vocabulary
Monday	10 am	Africa Chapter
Tuesday	9:30 am	CDM Survey WG

OHDSI 2025 Agenda Posted

Agenda • Tuesday, Oct. 7

Time (ET)	Session/Topic
7:00 am - 8:00 am	Lite Breakfast and Registration, Exhibits
8:00 am - 12:00 pm	Introductory Tutorial: An Introduction to the Journey from Data to Evidence Using OHDSI Vocabulathon 2025
12:00 pm - 1:00 pm	Buffet Lunch for Tutorial Registrants, Exhibits
1:00 pm - 5:00 pm	Advanced Tutorial: Developing and Evaluating Your Extract, Transform, Load (ETL) Process to the OMOP Common Data Model Advanced Tutorial: Using the OHDSI Standardized Vocabularies for Research Advanced Tutorial: Clinical Characterization Applications to Generate Reliable Real-World Evidence Advanced Tutorial: Population-Level Effect Estimation Applications to Generate Reliable Real-World Evidence Advanced Tutorial: Patient-Level Prediction Applications to Generate Reliable Real-World Evidence
5:00 pm - 6:00 pm	Collaborator Showcase Poster Placement
6:00 pm - 8:00 pm	Networking Reception; Collaborator Showcase Preview; Pre-Registration

Agenda • Wednesday, Oct. 8

Time (ET)	Topic
7:00 am - 8:00 am	Lite Breakfast and Registration, Exhibits
7:15 am - 7:45 am	Newcomer Orientation
8:00 am - 9:00 am	State of the Community: Welcome to OHDSI
9:00 am - 9:30 am	Group Networking Activity
9:30 am - 10:15 am	Collaborator Showcase Poster/Software Demo Session #1
10:15 am - 12:00 pm	Plenary
12:00 pm - 1:00 pm	Buffet Lunch, Exhibits
1:00 pm - 2:00 pm	Presentation
2:00 pm - 2:45 pm	Collaborator Showcase Lightning Talk Session #1
2:45 pm - 3:30 pm	Collaborator Showcase Poster/Software Demo Session #2
3:30 pm - 4:15 pm	Collaborator Showcase Poster/Software Demo Session #3
4:15 pm - 5:00 pm	Collaborator Showcase Lightning Talk Session #2
5:00 pm - 6:00 pm	Titan Awards, Wednesday Closing Activity
6:00 pm - 6:15 pm	Group Photo
6:15 pm - onward	Free Time

Agenda • Thursday, Oct. 9

Time (ET)	Meetings
7:00 am - 8:00 am	Lite Breakfast, Exhibits
8:00 am - 10:00 am	Session 1 of Workgroup Activities Featuring: Africa Chapter, APAC Chapter, Medical Imaging, GIS - Geographic Information System, HADES Hackathon, Oncology, Common Data Model, ATLAS/WebAPI, Phenotype Development and Evaluation, Dentistry, and Latin America
10:00 am - 10:30 am	Break, Exhibits
10:30 am - 12:30 pm	Session 2 of Workgroup Activities Featuring: Perinatal and Reproductive Health, Industry, Natural Language Processing, GIS - Geographic Information System, HADES Hackathon, Oncology, Common Data Model, ATLAS/WebAPI, Phenotype Development and Evaluation, Early-Stage Researchers, and Vocabularies
12:30 pm - 1:30 pm	Buffet Lunch and Exhibits
1:30 pm - 3:30 pm	Session 3 of Workgroup Activities Featuring: Surgery and Perioperative Medicine, Rare Diseases, Medical Devices, Psychiatry, HADES Hackathon, Health Equity, Evidence Network Data Partners, Eyecare and Vision Research, Women of OHDSI, CDM Survey
3:30 pm - 5:00 pm	Workgroup Summary

ohdsi.org/ohdsi2025

Global Symposium: Oct. 7-9

ohdsi.org/ohdsi2025

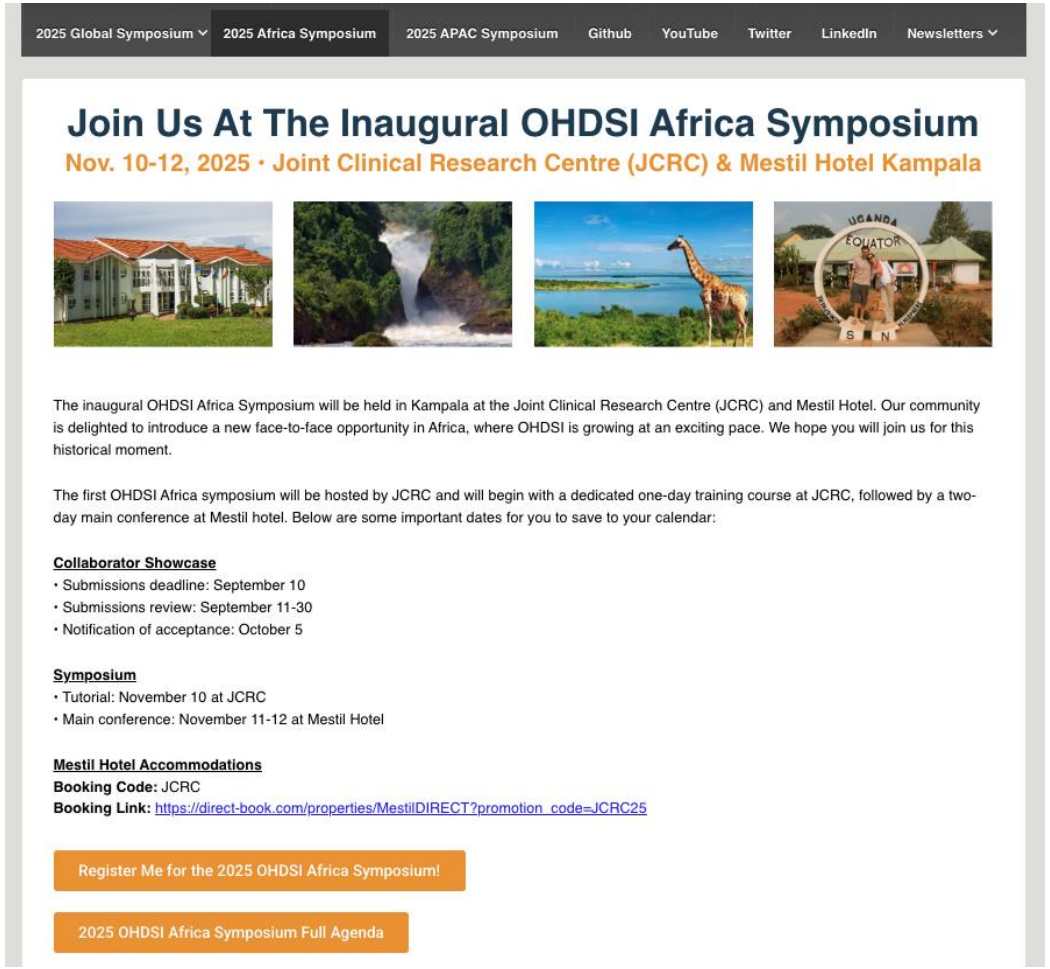
Titan Award Nominations Are Open

The Titan Awards have been handed out annually since 2018 to recognize OHDSI collaborators (or collaborating institutions) for their contributions towards OHDSI's mission.

Nominations for the 2025 Titan Awards are now open. **Please complete your nominations by our Sept. 9 (8 pm ET) deadline!**

ohdsi.org/titan-awards

Africa Symposium: Nov. 10-12



ohdsi.org/africa2025

APAC Symposium: Dec. 6-7

The 2025 OHDSI APAC Symposium will be held Dec. 6-7 in Shanghai, China at the Shanghai Jiao Tong University. It will feature a 1-day tutorial and a 1-day main conference. Here are some important dates for you to save to your calendar:

Collaborator Showcase

- Submissions deadline: September 7
- Submissions review: September 8 – October 9
- Notification of acceptance: October 17

ohdsi.org/apac2025

SURVEY DATA AND THE OMOP CDM: LANDSCAPE ASSESSMENT

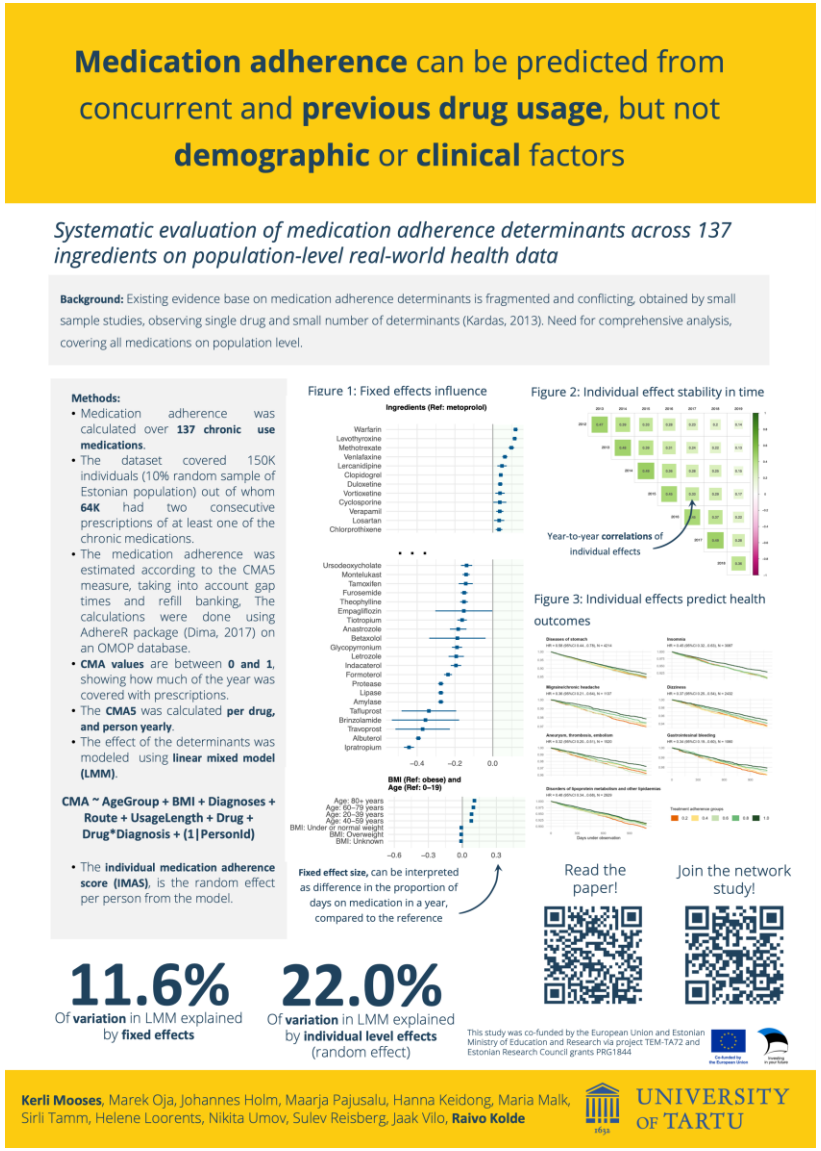
- https://forms.gle/f18ufspAFT3jSYrk6
- **Open through August 31, 2025**

#OHDSISocialShowcase This Week

Monday

Systematic evaluation of medication adherence determinants across 137 ingredients on population-level real-world health data

(Kerli Mooses, Marek Oja, Maria Malk, Helene Loorents, Maarja Pajusalu, Nikita Umov, Sirli Tamm, Johannes Holm, Hanna Keidong, Taavi Tillmann, Sulev Reisberg, Jaak Vilo, Raivo Kolde)

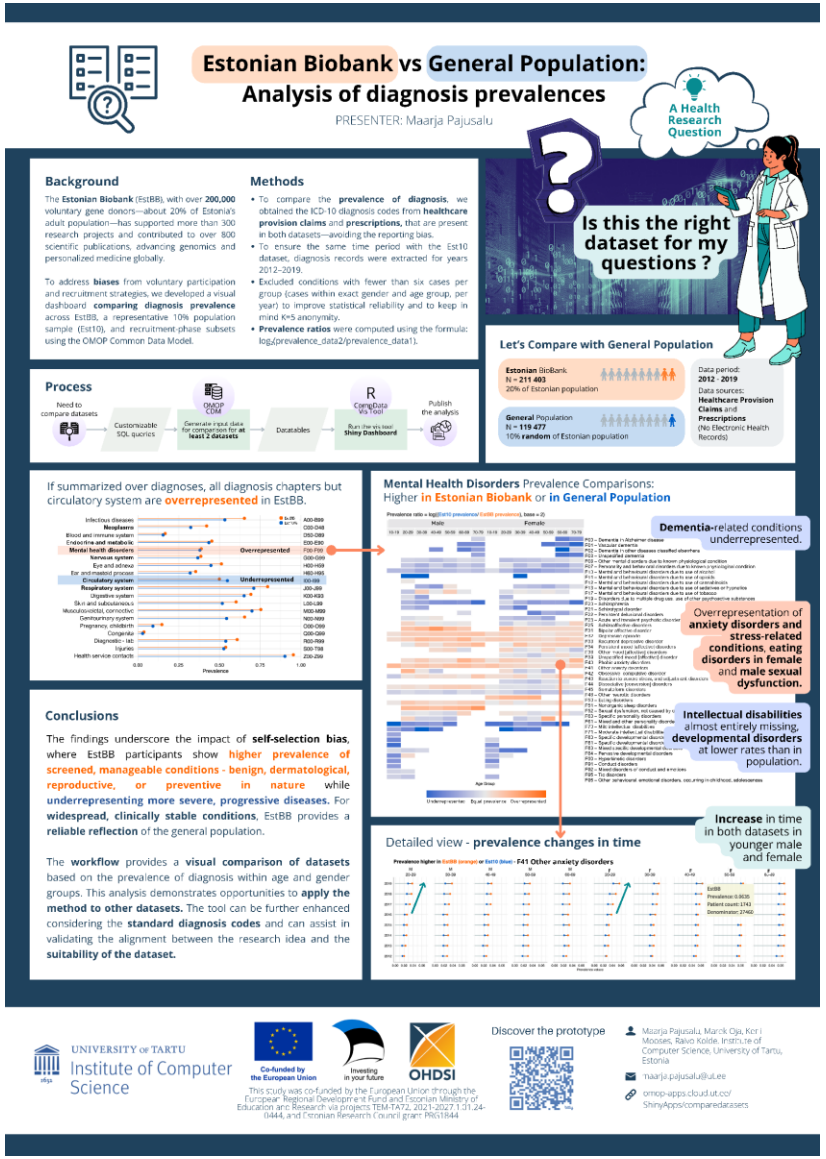


#OHDSISocialShowcase This Week

Tuesday

Estonian Biobank vs General Population: Analysis of diagnosis prevalences

(Maarja Pajusalu, Marek Oja, Raivo Kolde)



#OHDSISocialShowcase This Week

Wednesday

Impact of prior observation requirements on denominator populations

(Berta Raventós, Martí Català, Guillaume Verdy, Romain Griffier, Angela Leis, Juan Manuel Ramirez, Miguel-Angel Mayer, James Brash, Akram Mendez, Timothy Howcroft, Vishnu V Chandrabalan, Marek Oja, Raivo Kolde, Edward Burn)

Excluding individuals based on prior observation requirements can affect the age of the included population

Impact of prior observation requirements on denominator populations

Background: Prior observation requirements are commonly used as an exclusion criterion to define study populations. However, there is no clear guideline on the amount of prior observation that should be considered.
Aim: To examine the impact of choosing different amounts of prior observation and how they influence the identification of denominator populations.

Methods:

- We identified individuals in observation as of 01/01/2019 using different observation requirements (0, 30, 90, 180, 365, 730, 1095 days).
- Included and excluded patients were characterized for comparison.
- Analyses were performed using the "IncidencePrevalence" R package.

Data sources (*):

- CDWBordeaux (hospital; France)
- IMASIS (hospital; Spain)
- Lancashire (hospital; UK)
- IQVIA-DA (primary care; Germany)
- MAITT (primary and secondary care; Estonia)

(*) All defined observation periods based on healthcare encounters, except for MAITT (observation started at data collection start or date of birth).

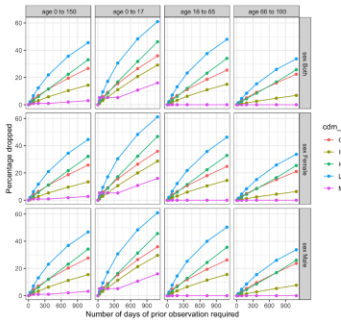


Figure: Patients dropped (%) using different prior observation criteria

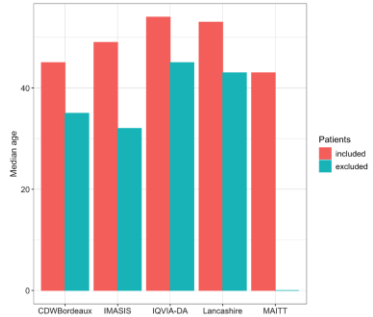


Figure: Median age of patients using a 365-days prior observation requirement.

Results:

- A greater number of patients were excluded as prior observation requirements increased. This particularly affected younger age groups, which resulted in an older population being included.
- In MAITT, only individuals aged <1 year were excluded due to the logic used to define observability.
- Careful consideration should be given when choosing prior observation requirements, taking into account the research question and database-specific factors.

Berta Raventós,[1] Martí Català,[2] Guillaume Verdy,[3] Romain Griffier,[3] Angela Leis,[4] Juan Manuel Ramirez,[4] Miguel-Angel Mayer,[4] James Brash,[5] Akram Mendez,[5] Timothy Howcroft,[6] Vishnu V Chandrabalan,[6] Marek Oja,[7] Raivo Kolde,[7] Edward Burn[2]

[1] Erasmus Medical Center, The Netherlands; [2] University of Oxford, UK; [3] University Hospital of Bordeaux, France; [4] Hospital del Mar/Hospital del Mar Research Institute (IMIM), Spain; [5] IQVIA, Real World Solutions, UK; [6] Lancashire Teaching Hospitals NHS Foundation Trust, UK; [7] University of Tartu, Estonia.

#OHDSISocialShowcase This Week

Thursday

Data cleaning and imputation approach for a real-world prescription database and its effect on medication adherence calculations

(Maria Malk, Kerli Mooses, Marek Oja, Johannes Holm, Hanna Keidong, Nikita Umov, Sirli Tamm, Sulev Reisberg, Jaak Vilo, Raivo Kolde)

With carefully designed imputation pipeline, it is possible to meaningfully **improve the quality of prescription datasets**

Data cleaning and imputation approach for a real-world prescription database and its effect on medication adherence calculations

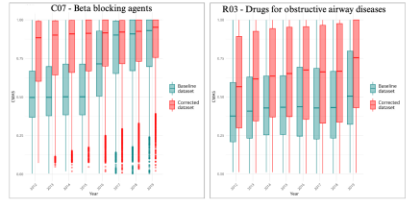
Background: For accurate drug usage statistics and drug adherence calculations, we need to have an accurate days' supply value for each prescription. However, this information is not always provided. Neither is daily dosing to calculate days' supply. Therefore, methods need to be applied to acquire aforementioned values.

Results:

Table 1. Days' supply imputation summary by drug form

MAIN GROUP	DATASET INFORMATION	METHOD OF FINDING DAYS' SUPPLY	PRESCRIPTIONS WITHIN MAIN GROUP
Tablets, capsules, suppositories	Daily dose is provided	Calculated using given daily dose	2 823 867 (45.75%)
	Treatment course is provided	Treatment course was used	864 091 (13.99%)
	Daily dose and treatment course is not provided	Calculated using imputed daily dose value	2 485 182 (40.25%)
	Days' supply could not be calculated nor imputed	-	1 445 (0.02%)
Semisolid dosage forms	Treatment course is provided	Treatment course was used	119 748 (37.55%)
	Treatment course is not provided	Imputed as 30 days* per package	199 154 (62.45%)
Medicinal nail polish	Treatment course is provided	Treatment course was used	3 088 (41.33%)
	Treatment course is not provided	Imputed as 210 days or 180 days*	4231 (58.67%)
Ear drops	Ear drops - treatment course is provided	Treatment course was used	5 661 (57.13%)
	Ear drops - treatment course is not provided	Imputed as 7 days* per prescription	4 252 (42.89%)
Eye drops	Eye drops in single dose containers - daily dose is provided	Calculated using given daily dose	6 724 (2.35%)
	Eye drops - treatment course is provided	Treatment course was used	62 332 (21.81%)
	Eye drops - treatment course is not provided	Imputed as 30 days** per package	216 740 (72.36%)
	Days' supply could not be calculated nor imputed	-	43 (0.02%)
Oral drops	Daily dose is provided	Calculated using given daily dose	52 085 (53.07%)
	Treatment course is provided	Treatment course was used	33 055 (33.67%)
	Daily dose and treatment course is not provided	Imputed as 30 days* per prescription	12 983 (13.20%)
	Days' supply could not be calculated nor imputed	-	120 (0.12%)
Inhalation medication	Daily dose is provided	Calculated using given daily dose	78 541 (57.50%)
	Treatment course is provided	Treatment course was used	19 885 (14.56%)
	Treatment course is not provided	Imputed as 60 days* per prescription	27 137 (27.19%)
	Days' supply could not be calculated nor imputed	-	1 825 (0.75%)
Nasal sprays	Anti-fungal nasal spray - treatment course is not provided	Imputed as 7 days* per prescription	7 758 (8.72%)
	Daily dose is provided	Calculated using given daily dose	26 361 (29.70%)
	Treatment course is provided	Treatment course was used	40 551 (45.69%)
	Daily dose and treatment course is not provided	Imputed as 30 days* per prescription	14 094 (15.88%)
Syrups	Antibiotic syrups/solutions - treatment course is not provided	Imputed as 14 days* per prescription	29 130 (34.31%)
	Daily dose is provided	Calculated using given daily dose	37 251 (43.87%)
	Treatment course is provided	Treatment course was used	2 094 (3.64%)
	Treatment course is not provided	Imputed as 30 days** per package	15 436 (18.18%)
Transdermal patch	Hormonal patch - treatment course is not provided	Impute as 7 days per patch plus 7 days*	13 131 (81.18%)
	Analgesic patch - treatment course is not provided	Imputed as 3 days* per patch	1 127 (7.01%)
	Treatment course is provided	Treatment course was used	599 (5.94%)
	Treatment course is not provided	Impute as 7 days* per patch	865 (5.38%)
Intravaginal contraceptives	-	Imputed as 30 days* per item in package	21 165
Implants and vaccines	-	Imputed as 1 day	41 919
Insulin injections	Treatment course is provided	Treatment course was used	2 603 (3.14%)
	Treatment course is not provided	Imputed as 60 days* per prescription	80 286 (96.89%)
Other	Treatment course is provided	Treatment course was used	49 (nn)
	Days' supply was not be calculated nor imputed		126 895

* based on (SPC) issued by Estonian Agency of Medicines
** medication not shelf-stable for more than 30 days after opening



Example of single dose medications (tablets, capsules, suppositories)

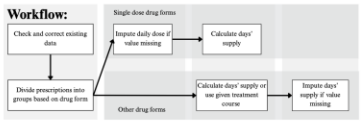
Example of non single dose medications

Figure 1. Adherence calculated by year by medication groups

Method:

Data: 2012 to 2019 of randomly selected 10% of Estonian population, which has been transferred to OMOP CDM

Workflow:



Validation: continuous multiple interval measures of medication availability measure (CMA)

Key takeaways:

- We developed and implemented a multi-step data cleaning and imputation approach to address missing or incomplete information in prescription records
- When imputing missing data in prescriptions, the use of SPCs, domain knowledge, and information from similar prescriptions proves to be most useful

Maria Malk, Kerli Mooses, Marek Oja, Johannes Holm, Hanna Keidong, Nikita Umov, Sirli Tamm, Sulev Reisberg, Jaak Vilo, Raivo Kolde

maria.malk@ut.ee

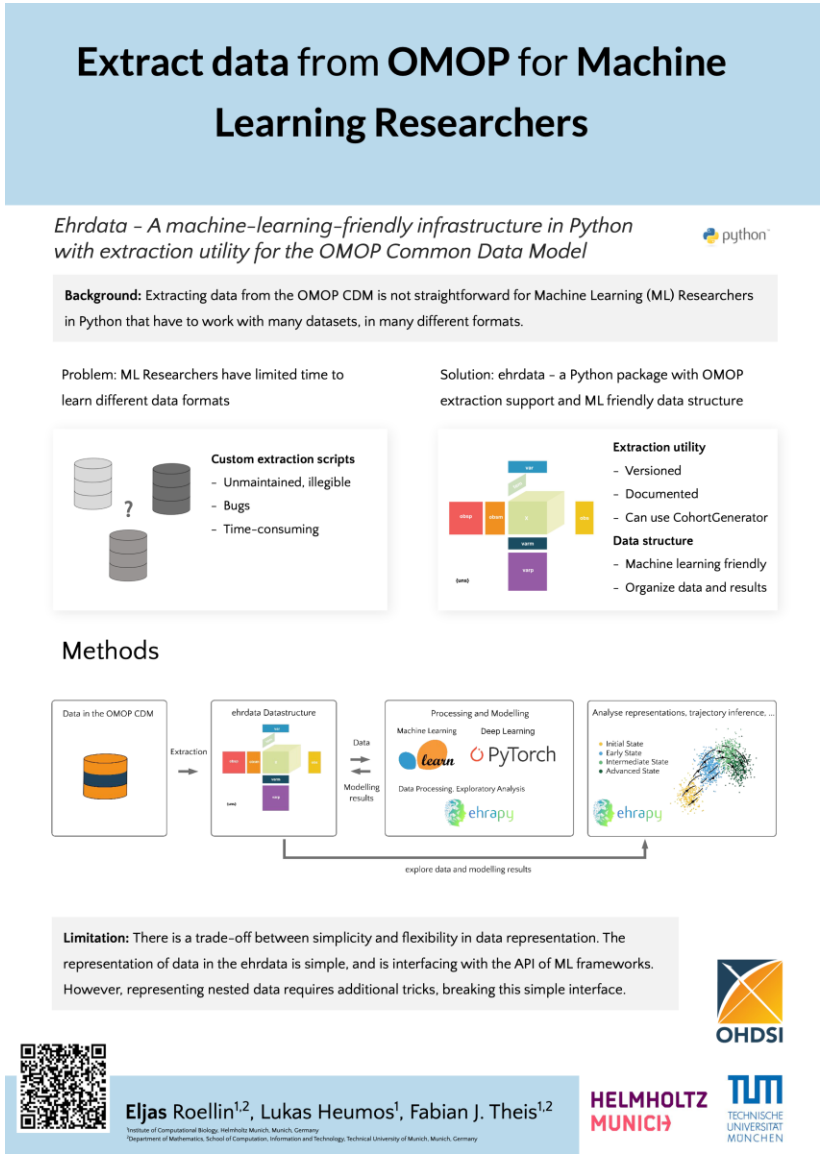
This study was co-funded by the European Union and Estonian Ministry of Education and Research via project TEM-TA72 and Estonian Research Council grant PRG1044

#OHDSISocialShowcase This Week

Friday

Ehrdata - a machine-learning-friendly infrastructure in Python with extraction utility for the OMOP Common Data Mode

(Eljas Roellin, Lukas Heumos, Fabian J. Theis)



Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

The weekly OHDSI community call is held every Tuesday at 11 am ET.

Everybody is invited!

Links are sent out weekly and available at:
ohdsi.org/community-calls-2025