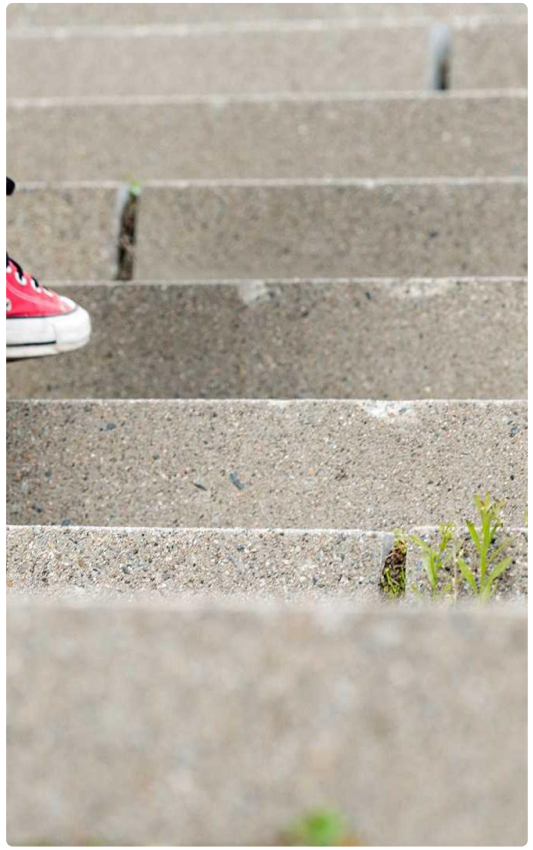
A Data Scientist's perspective and lessons learned

Iiro Toppila, M.Sc. (Tech)

Medaffcon Oy

Data Analysis Lead

Agenda



1. Medaffcon and me in 30 seconds
2. Lessons learned: Data mapping
3. ~~Lessons learned: Readiness and collaboration~~

Focus Areas of Medaffcon

1

Market Access

2

Real World Evidence

3

Medical Affairs

Iiro Toppila

Data Analysis Lead
M.Sc. (Tech.)

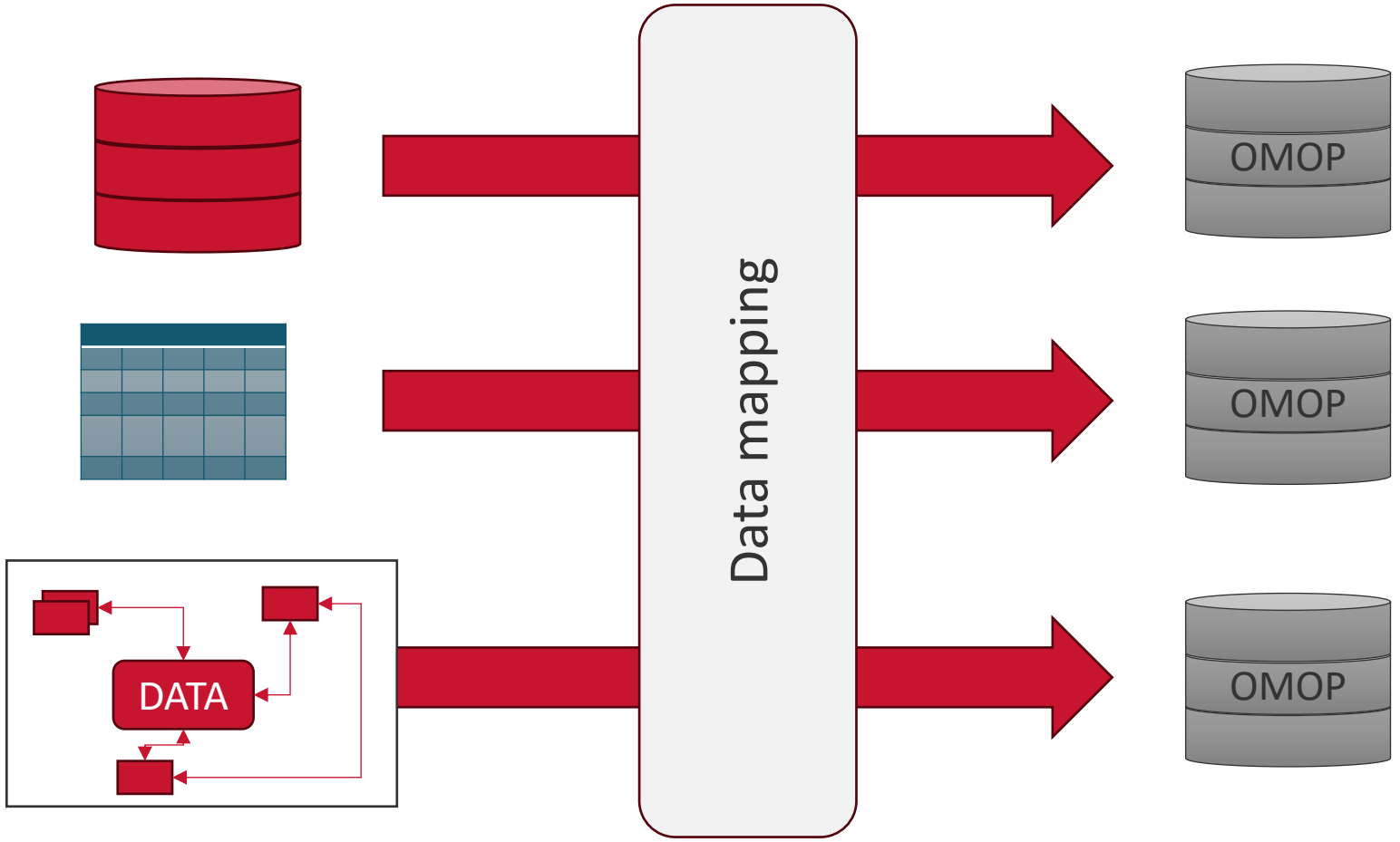
At Medaffcon since 2017

20+ Peer-reviewed publication on
RWE studies by Medaffcon

No conflict of interest
nor
Interest in conflicts

Data mapping

Important step – data mapping to OMOP



- Retain all relevant data and avoid relevant information loss in data transformation
- Personal opinion: Requires insight and experience from primary data source, and the possible use cases of mapped data
- Any resulting analyses from mapped OMOP data are as good as the mapping

Case example

- Patient group: CVD Secondary prevention patients
  - i.e. patients, that have had major CVD event
- Main treatment goal is to prevent any consecutive MACE events (via interventions)
- One of the most important outcome measurements in epidemiological setting (on top of survival) is the recurrence rate and time to recurrence.
  - But the recurring event might not be straight forward to identify from the data…

Let’s consider ”dummy” case example of epidemiological study, where we want to assess the event rate of major CVD event recurrency using register or hospital data (in OMOP setting naturally)

Previous work


Check for updates

European Journal of Preventive Cardiology

ESC European Society of Cardiology

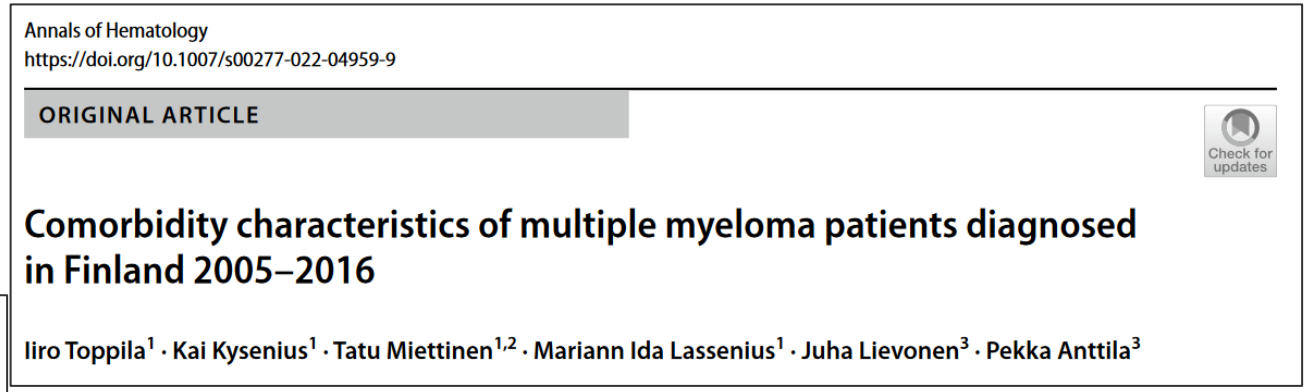
Full research paper

Cardiovascular event rates increase after each recurrence and associate with poor statin adherence

Mariann I Lassenius[1], Iiro Toppila[1], Susanne Bergius[2], Julia Perttilä[2], KE Juhani Airaksinen[3] and Mikko Pietilä[3]

European Journal of Preventive Cardiology
0(00) 1–10
© The European Society of Cardiology 2020
Article reuse guidelines:
sagepub.com/journals-permissior
DOI: 10.1177/204748732090433
journals.sagepub.com/home/cpr

SAGE

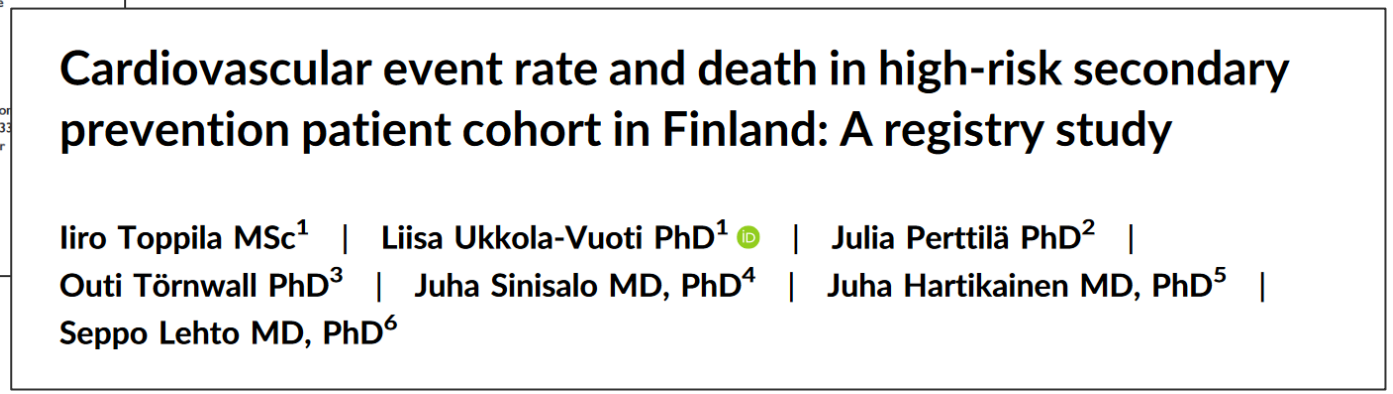

Annals of Hematology
https://doi.org/10.1007/s00277-022-04959-9

ORIGINAL ARTICLE

Check for updates

Comorbidity characteristics of multiple myeloma patients diagnosed in Finland 2005–2016

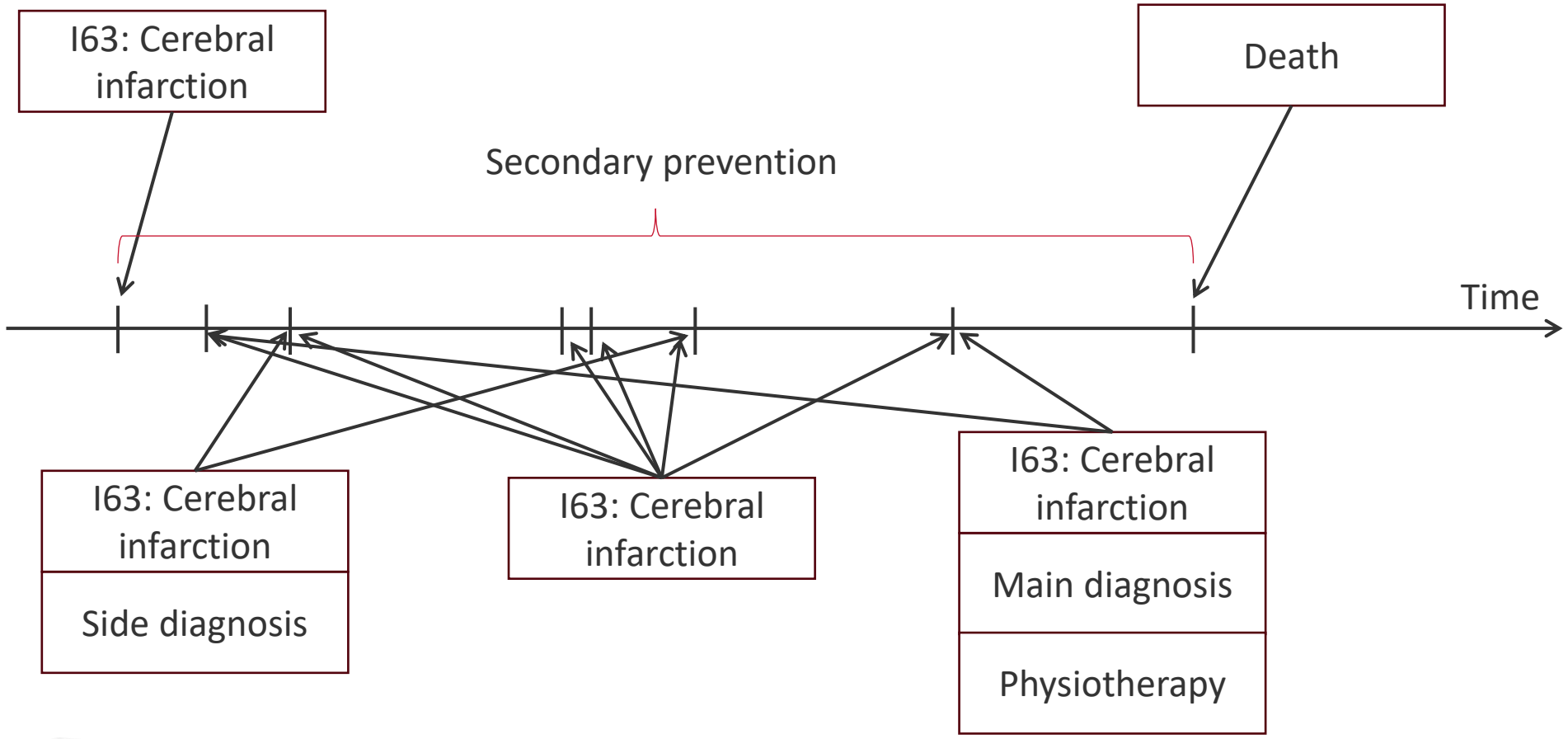
Iiro Toppila[1] · Kai Kysenius[1] · Tatu Miettinen[1,2] · Mariann Ida Lassenius[1] · Juha Lievonen[3] · Pekka Anttila[3]


Cardiovascular event rate and death in high-risk secondary prevention patient cohort in Finland: A registry study

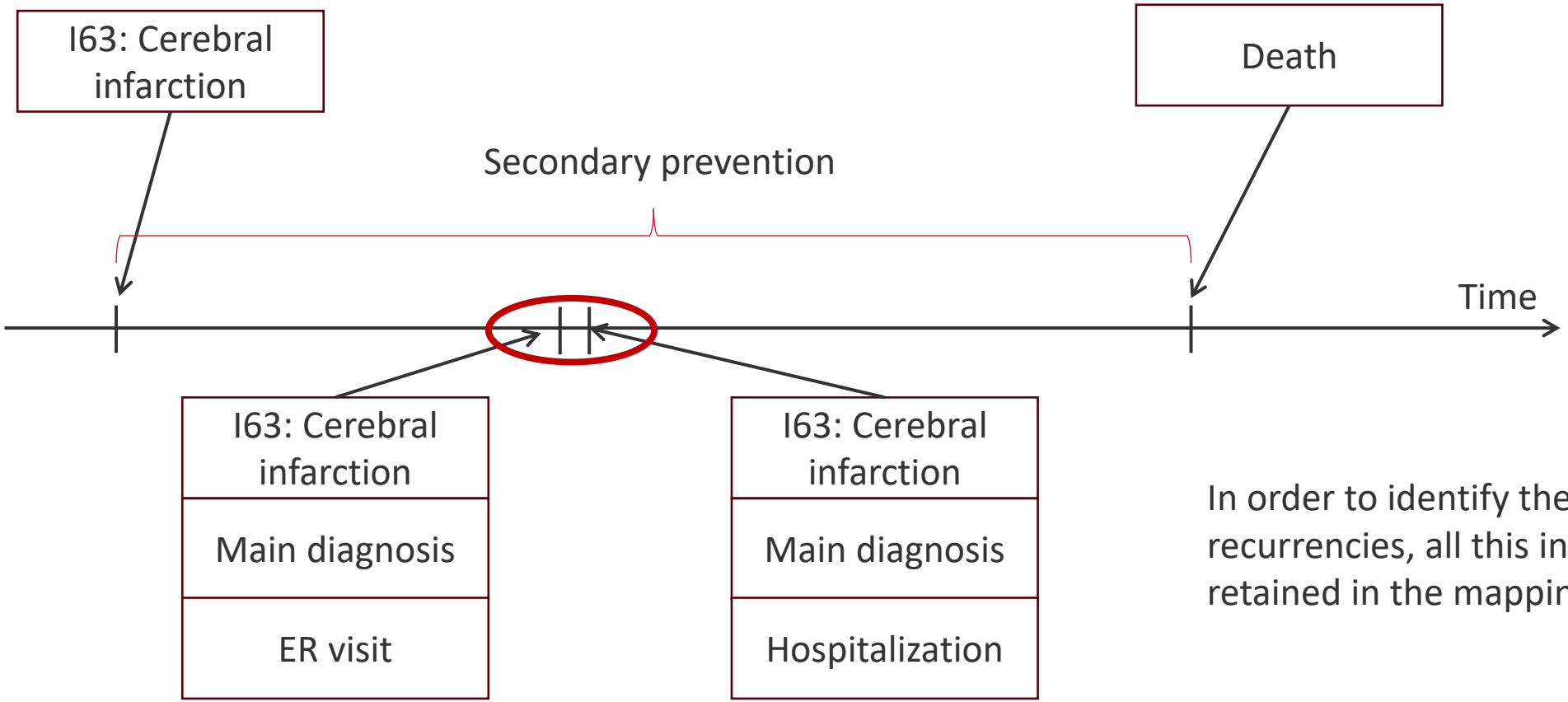
Iiro Toppila MSc[1] | Liisa Ukkola-Vuoti PhD[1] | Julia Perttilä PhD[2] | Outi Törnwall PhD[3] | Juha Sinisalo MD, PhD[4] | Juha Hartikainen MD, PhD[5] | Seppo Lehto MD, PhD[6]

- Non-OMOP setup; varying definitions and ways to identify MACE and recurrent CVD events
- Primary source hospital data-lakes or national registers

Dummy example – Stroke patient

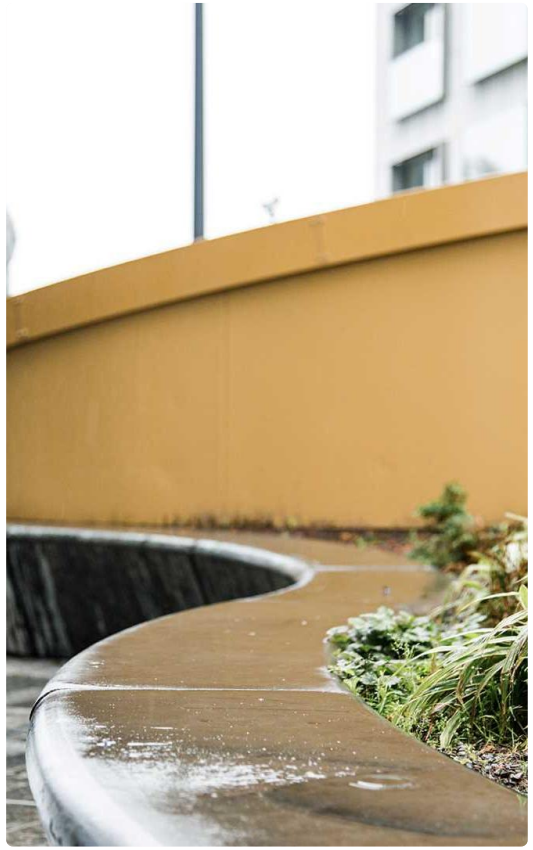


Dummy example – Stroke patient



In order to identify the actual recurrencies, all this information must be retained in the mapping!

Take home message & final thoughts



- When mapping the data, try to retain all possible (relevant) info
  - Who knows what is relevant?
  - Poor mapping will result into low quality data which will further produce unreliable results
  - Invest into Sr. Personnel whom understand the primary data, but also the later use cases and how to analyze the data
- When using the OMOP data, understand the primary source and that it is fit for purpose!
  - National register, hospital data, cohort study, and quality register are completely different, even when OMOP'ed
  - Clear documentation of mapping process, and even open-source scripts?

Contacts

Iiro Toppila
Data Analysis Lead
+358 44 314 15 97
iiro.toppila@medaffcon.fi

News letter subscription