Design of a framework to detect temporal clinical event trajectories from health data standardized to the OMOP CDM

PRESENTER: Sulev Reisberg

INTRO:

- Temporal disease sequences (trajectories) can characterize the dataset and describe disease progressions within the population
- However, the number of disease trajectory studies is small due to:
    1. lack of syntactic and semantic interoperability of observational health data
    2. no common principles for that kind of study
- While the first issue is effectively tackled by the OHDSI community by developing the OMOP Common Data model, the second issue has remained a challenge

AIM:

- propose **a standardized framework for detecting** the most prominent **temporal clinical event trajectories** in the observational health dataset
- **test** the framework and package on electronic health records from Estonia and the Netherlands and compare the results with previous findings in the Danish population

The framework is implemented as an open source **R package**. The package will be freely available on GitHub after the publication of the manuscript.
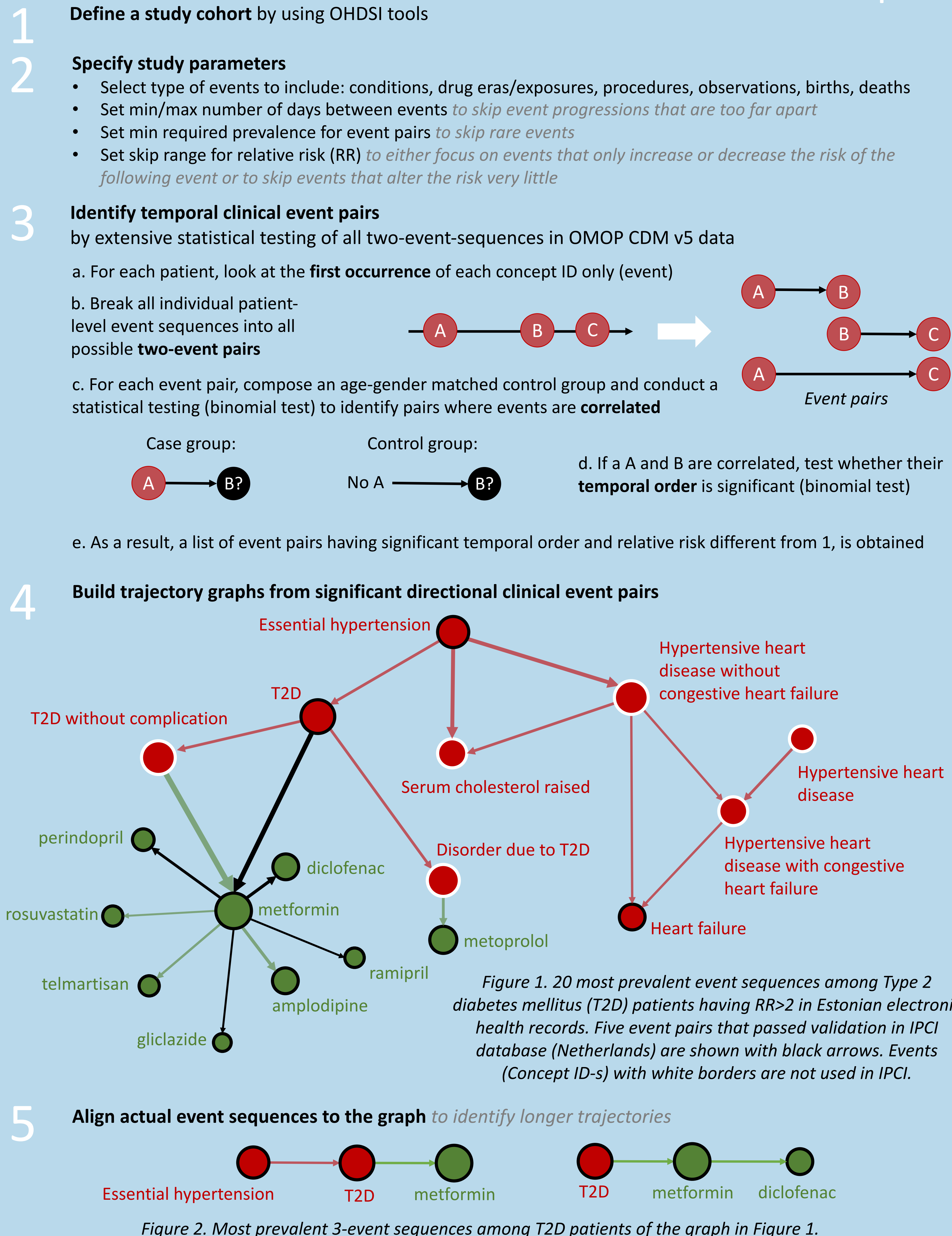
Framework description

1

Define a study cohort by using OHDSI tools

2

Specify study parameters

- Select type of events to include: conditions, drug eras/exposures, procedures, observations, births, deaths
- Set min/max number of days between events *to skip event progressions that are too far apart*
- Set min required prevalence for event pairs *to skip rare events*
- Set skip range for relative risk (RR) *to either focus on events that only increase or decrease the risk of the following event or to skip events that alter the risk very little*

3

Identify temporal clinical event pairs
by extensive statistical testing of all two-event-sequences in OMOP CDM v5 data

a. For each patient, look at the **first occurrence** of each concept ID only (event)

b. Break all individual patient-level event sequences into all possible **two-event pairs**

c. For each event pair, compose an age-gender matched control group and conduct a statistical testing (binomial test) to identify pairs where events are **correlated**

Event pairs

Case group:

Control group:

d. If a A and B are correlated, test whether their **temporal order** is significant (binomial test)

e. As a result, a list of event pairs having significant temporal order and relative risk different from 1, is obtained

4

Build trajectory graphs from significant directional clinical event pairs

Figure 1. 20 most prevalent event sequences among Type 2 diabetes mellitus (T2D) patients having RR>2 in Estonian electronic health records. Five event pairs that passed validation in IPCI database (Netherlands) are shown with black arrows. Events (Concept ID-s) with white borders are not used in IPCI.

5

Align actual event sequences to the graph *to identify longer trajectories*

Figure 2. Most prevalent 3-event sequences among T2D patients of the graph in Figure 1.

RESULTS
IN ESTONIA VS. DENMARK:

- In 10% of a random sample of Estonian electronic health records (n=147K patients, 8 years), we validated 7733 most prominent temporal event pairs in the Danish population having RR either <=0.8 or >=1.2 (Siggaard et al., n=7M patients, 25 years)
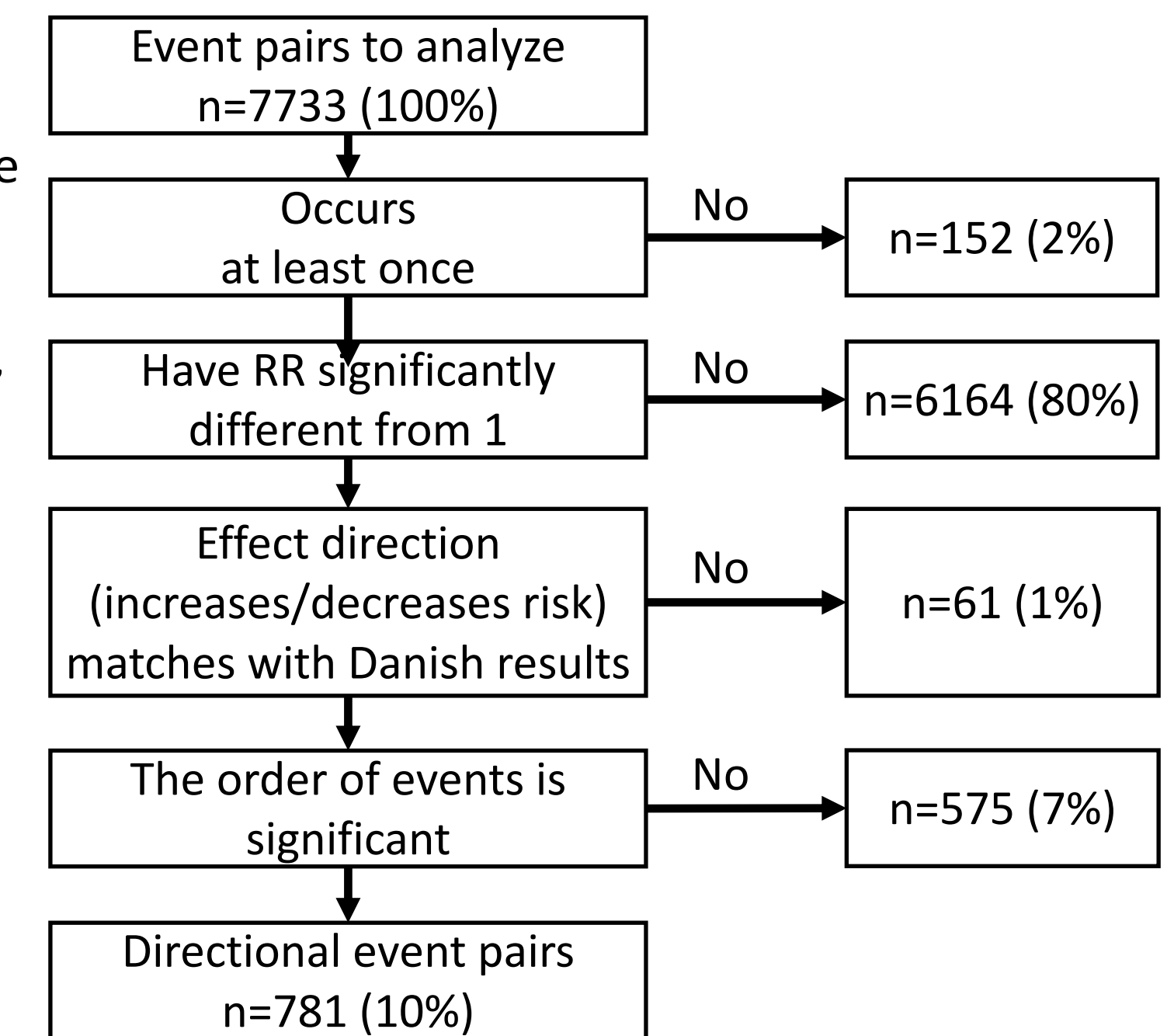- We confirmed RR<>1 and direction of 781 pairs (10%)

Figure 3. Attrition diagram, showing the number of event pairs after various stages in the validation analysis

RESULTS IN ESTONIA VS. NETHERLAND (IPCI):

- In Estonian data, we identified 22 directional event pairs having RR>2 and occurring on at least 5% of Type 2 Diabetes patients
- Out of these,
    - 5 passed the validation in Netherlands' data (IPCI database, n=2.5M) (Figure 1)
    - Concept ID-s used in 14 pairs are not used in IPCI

CONCLUSION:

- The proposed framework identifies and visualizes significant clinical event progression patterns in health data standardized to the OMOP CDM. The open-access R package, the first of its kind, allows researchers to run the same framework on their OMOP-formatted health data and compare results across databases to allow for the identification of clinical event associations
- Using different Concept ID-s for the same underlying event in different OMOP databases makes the cross-dataset comparison of event trajectories challenging
- Before moving to investigate longer global trajectories, a global consensus on the simplest trajectories - pairs - need to be established first

Kadri Künnapuu, Solomon Ioannou, Kadri Ligi, Raivo Kolde, Sven Laur, Jaak Vilo, Peter Rijnbeek, Sulev Reisberg