Mapping concepts from the Netherlands Cancer Registry to the OMOP-CDM - experiences and challenges

PRESENTER: **Chiara** Attanasio

INTRODUCTION

The **Netherlands Cancer Registry** (**NCR**) is a population-based structured cancer registry with nationwide coverage since 1989 and **3 million patients** total.

We commenced conversion of the NCR data to the **OMOP-CDM in 2020.** Here we describe our experiences and challenges with the **mapping work** still ongoing as part of an **EHDEN** type two grant.

METHODS

We focus on **two tables** from the NCR:

- The **"Event"** table, which contains clinical events such as diagnostic tests and primary treatment.
- The **"Event Detail"** table, which contains details of entries in the "Event" table and associated values.

Mapping workflow:

1. **Selection** of events and details to map,
2. **Pre-processing** of the source concepts,
3. Multiple **mapping** rounds and reviews,
4. **Post-processing** steps,
5. Final **review** by domain experts,
6. **Implementation**.

RESULTS

The first batch of source concepts from the NCR that we processed within the EHDEN grant were related to the **most frequently occurring diagnostic events**. After the pre-processing step around **350** pre-coordinated source concepts needed to be mapped, accounting for around **10% of all NCR diagnostic concepts**.

Events in the NCR rely heavily on post-coordination. This is not supported in the OMOP-CDM.

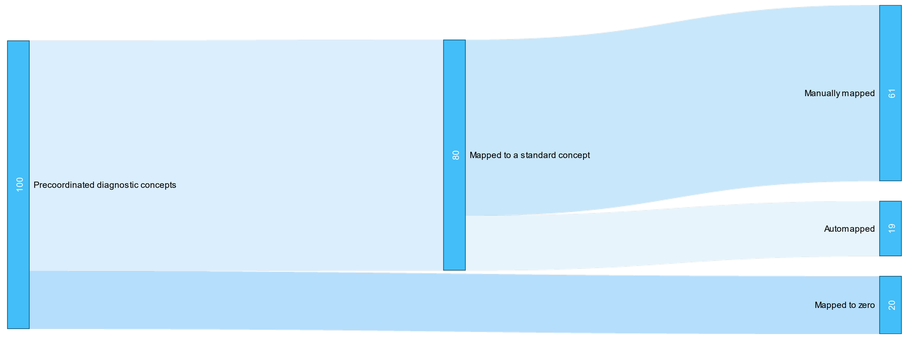

Figure 1: Mapping results for the diagnostic concepts.
Auto mapped means mapped according to the automatic suggestion from the edenceReviewer of edenceHealth.

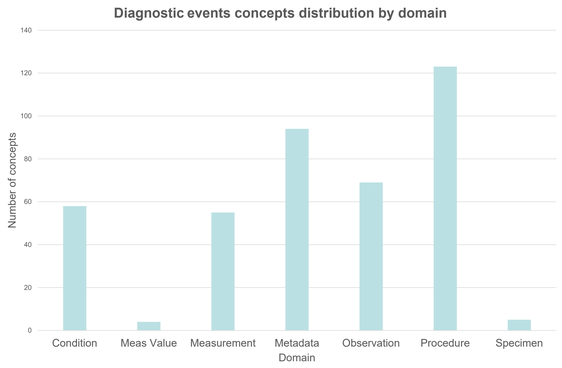

Figure 2: Distribution of the NCR diagnostic events over the OMOP domains.

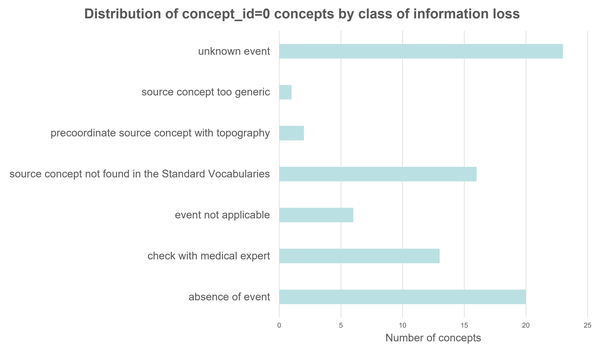

Figure 3: Overview of the reasons why diagnostic events are mapped to zero.

Apart from a few small mapping challenges, the **main roadblock** we faced was linked to the **structure of the "Event" and "Event Detail"** information in the NCR, which did not match the OMOP-CDM. An "Event", "Event Detail", "Value" combination can be mapped in many different ways, and an event may contain more than one detail attached to it.

A big effort went into **identifying all occurring 'mapping' situations** firstly to standardize our approach during the mapping, and secondly to feed this knowledge into the ETL design and implementation step.

CONCLUSIONS

So far, prior the medical review step, we have mapped around **10% of the diagnostic events from the NCR**. If we consider only source concepts that we wanted to add to the CDM, but for which we could not find a suitable standard concept, then **only 5% of these were mapped to 0**.

Within the EHDEN grant effort, we will not be able to map all variables in the NCR, but **we aim to have an interesting data set** to participate in **international studies**, starting by adding primary treatment events.

We have already done so for a **PIONEER** study-a-thon in 2021 and we are currently participating in the **HANA** project on colon cancer treatment effects with South Korea.

Make sure to check the other two posters from IKNL!

Chiara Attanasio[1] (c.attanasio@iknl.nl),
Floor Klijn[1] (f.klijn@iknl.nl),
Jennifer Caffarel[1] (j.caffarel@iknl.nl),
Peter Prinsen[1] (p.prinsen@iknl.nl)
[1]IKNL