

Pathways for advanced transformation of CDISC SDTM data sets into OMOP CDM

PRESENTER: Michael Kallfelz

INTRO:

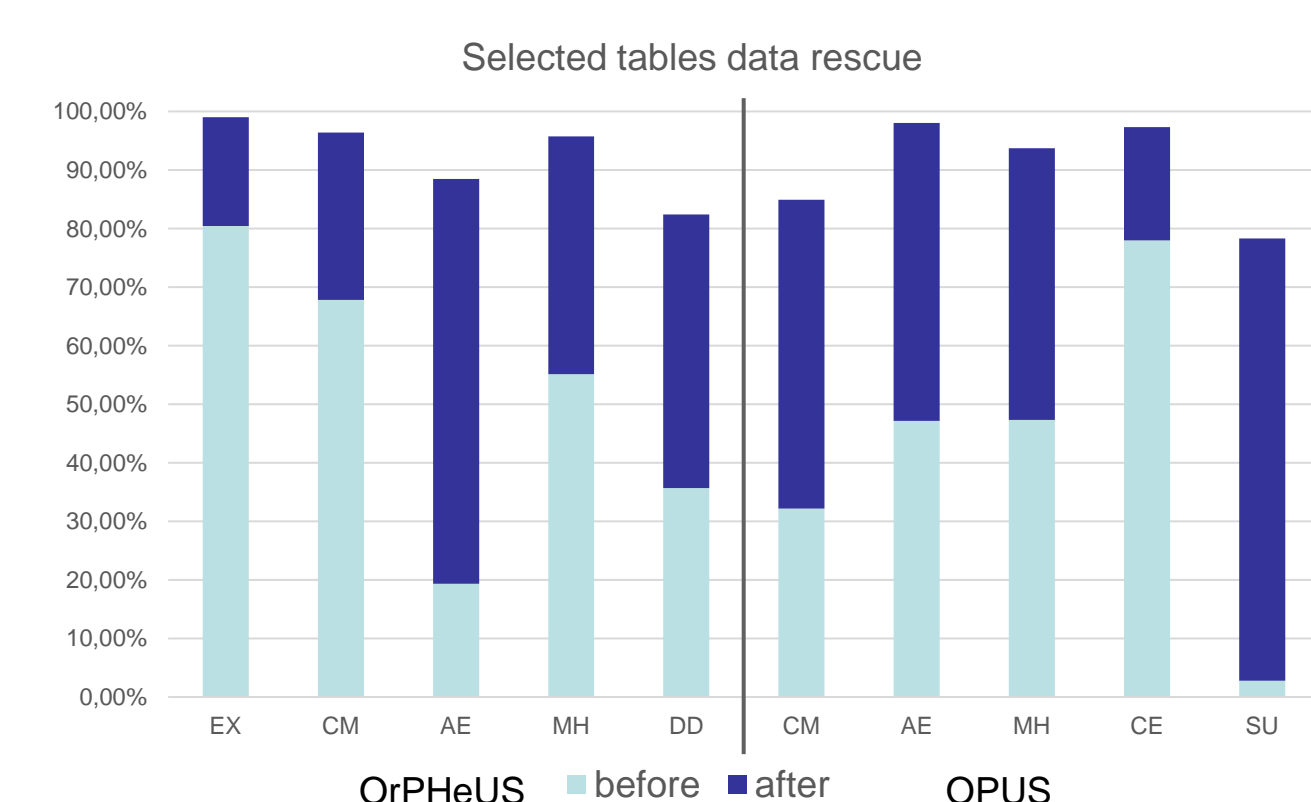
- Data model disparity between CDISC SDTM and OMOP CDM can lead to considerable data loss following a **conventional mapping approach**. It can be particularly problematic in complex diseases with rare subgroups and multiple etiologies, such as Pulmonary Hypertension [PH]. However, a large percentage of otherwise lost data can be rescued by **fine tuning the conversion process** according to the OMOP constraints and restituting missing parts by **imputation rules**.

METHODS

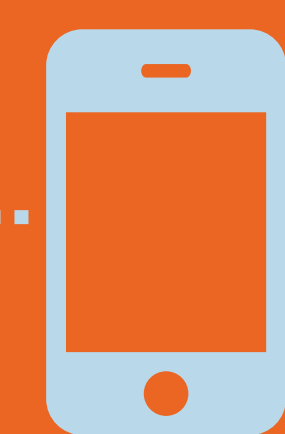
- Initial conversion of CDISC SDTM data lead to further investigation of possible process improvements.
- PH registry data from OPUS and OrPHeUS were taken as a basis to develop and improve disease- and SDTM-specific conversion logic.
- Quantification of data rescue and validity through comparison of source and target tables
- Imputation:**
Impute complete and consistent dates from incomplete sources
Specific Conversion:
Sophisticated conversion into specific concepts and use of fact-relationships.

RESULTS

dark blue shows percentage of rescued values for selected source tables



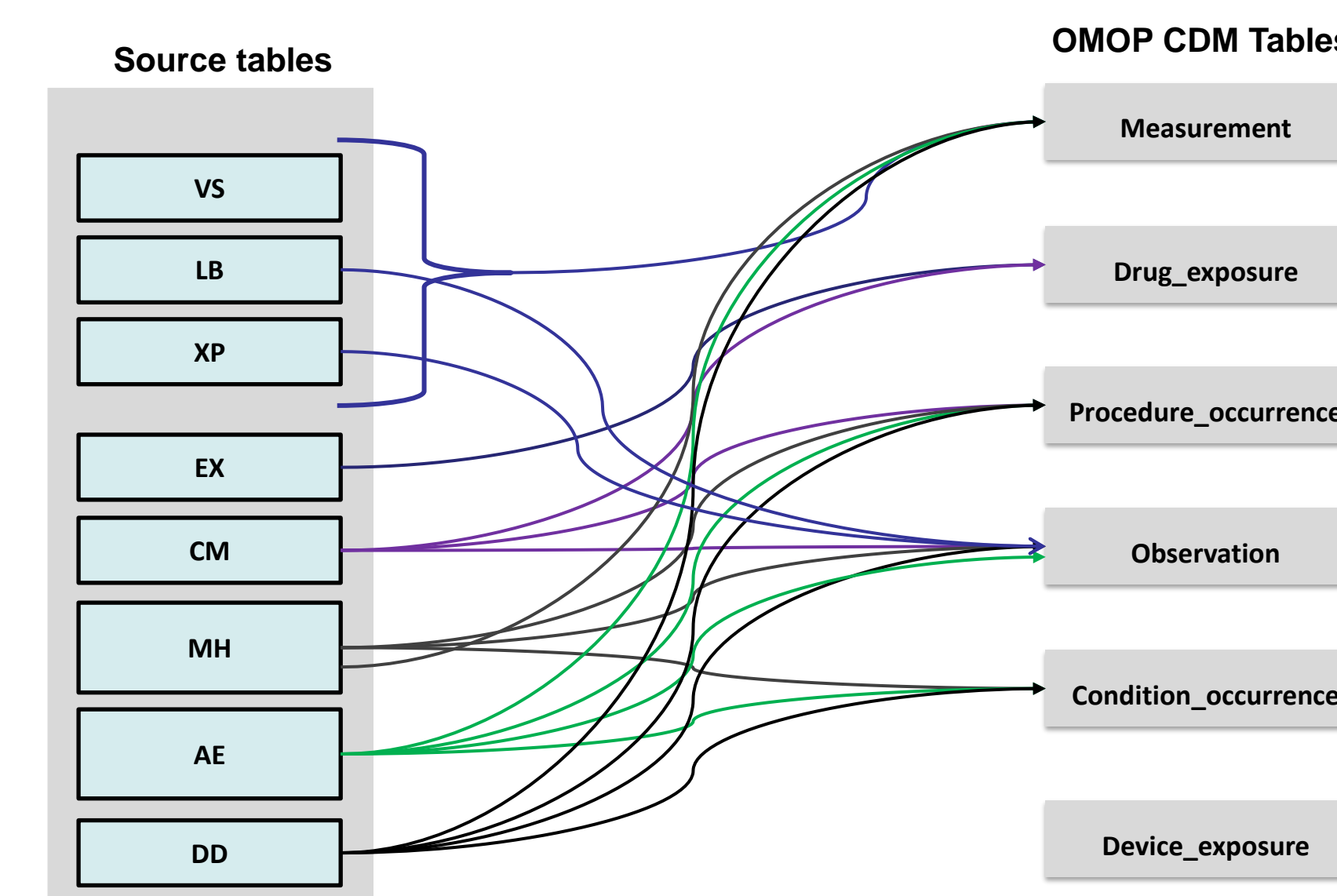
Reduce data loss in SDTM conversion, using date imputation rules and SDTM specific information rescuing.



Take a picture to access the showcase repository or follow [this link](#)

Findings

- OMOP CDM can become more accessible and richer than original SDTM data through imputation and information derivation from multiple sources
- Strategies to deal with missing standard concepts in rare diseases
 - Custom concepts, embedded in ontologies and hierarchies
 - Association of multiple concepts



- Most wanted source vocabularies:
 - MedDRA (mapped)
 - WHODrug (OMOP'ed and mapped)
 - DrugBank (OMOP'ed)

AUTHORS: Eva-Maria Didden, Patricia Biedermann, Rose Ong, Graham Wetherill – Actelion (now Janssen) Alexander Davydov, Alexandra Orlova, Gregory Klebanov, Michael Kallfelz – Odysseus

