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

Eva-Maria Didden, Rose Ong, Patricia Biedermann, Graham Wetherill
Actelion (now Janssen Pharmaceutical Companies)

Alexander Davydov, Alexandra Orlova, Gregory Klebanov, Michael Kallfelz
Odysseus

The Pulmonary Hypertension Therapeutic Area of

CDISC SDTM

- Introduced and required for regulatory submission of clinical trial data to authorities
- Allows for harmonization of clinical trial and registry data
- Provides not only a data model but is also supported by a controlled terminology
- Meant for organizing raw data, not primarily for analysis purposes

Source data: OPUS and OrPHeUS registries

	OPUS	OrPHeUS
Trial identifier	NCT02126943	NCT03197688
Study location	USA	USA
Study type	Multi-centre, prospective drug registry	Multi-centre, retrospective medical chart review
Observation period	April 2014 – August 2019	October 2013 – March 2017
Latest data cut mapped	August 2019	March 2017
Patients, N	2521	3142
Primary objective	To characterize the safety profile of patients newly treated with macitentan [Opsumit], an endothelin receptor antagonist approved for the treatment of PAH - in routine clinical practice	
Original format	CDISC SDTM*	
Exclusion criteria	• Had an entire date or the year missing from their registry record for important clinical events. • Duplicate patients. • Did not meet the original registry inclusion criteria.	

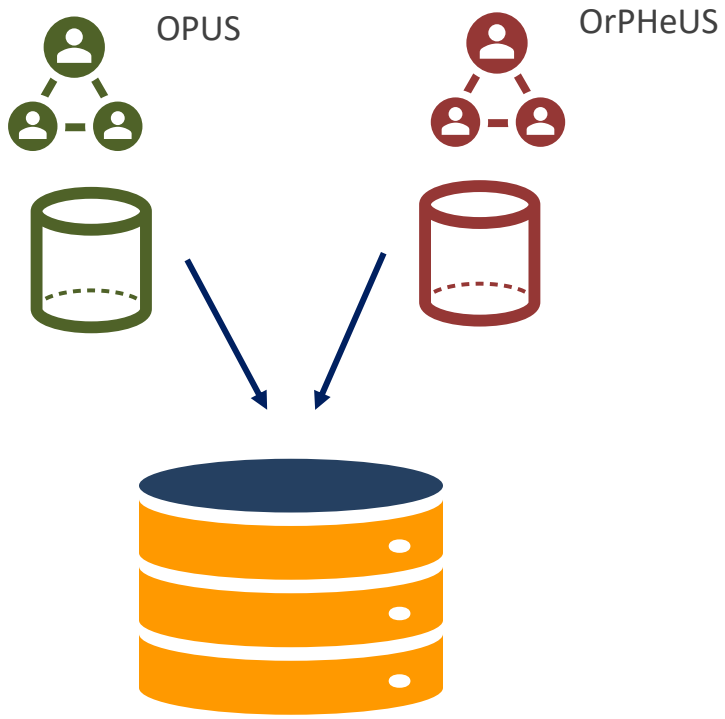
Pulmonary hypertension drug therapy registries [PH]

- OPUS: OPsumit® Users
- OrPHeUS, OPsumit® Historical USers.

Problem statement

Challenge

- Convert rare-disease drug registry data to the OMOP Common Data Model
- Merge 2 Pulmonary Hypertension (PH) data sets



Challenges with SDTM to OMOP conversion

- Meet OMOP CDM constraints
  - Record dates must be complete and consistent (correct/logical sequence)
- Convert and combine information
  - Complex mapping, in particular for adverse events:
    - Extraction of information from multiple sources or source tables
    - Adverse event terms mapping
  - Relationship building, e.g. cause of (adverse) event, (adverse) event, outcome of (adverse) event

Challenges with SDTM to OMOP conversion (2)

- Specific structure of SDTM data
  - Complex and individual Case Report Forms
  - Code systems not widely used in observational health data (e.g. MedDRA, WHODrug)
  - Free text decoding
  - SDTM table content – potential dependencies and inconsistencies
- Uncharted territory
  - No guidelines and conventions for converting SDTM to OMOP

Challenges with SDTM to OMOP conversion (3)

- Access only to “non-curated” registry data (no prior exclusion of patients or records, no prior imputation work)
- Registry data collection may still be ongoing:
  → Case report form may change over time or be amended → Impact on SDTM database content → impact on conversion to OMOP CDM
- Regular feedback sessions between study team and mapping team are crucial to ensure high quality deliverables

Strategies

Methods for imputation rules

- Comparison of the year[, month] of a medical event with calculated 'reference time points':
  - Date of birth / death
  - Opsumit initiation date / exposure end date
  - Date of last available information
  - Latest date of dataset
  - Date of last follow-up visit
- Usage of time points available in the SDTM such as
  - 'ongoing / before at last patient visit'
  - 'ongoing at Opsumit initiation'
  - 'before patient discontinuation’
- For Opsumit, in addition, imputation based on the previous interval of taking a drug
- Extraction of date from free text fields

Examples for imputation rules (1/3):

missing day of Adverse Events [AE]

- If the SDTM time point for study treatment [Opsumit] initiation had the same month and year as the AE date, then the incomplete date was imputed with the day after Opsumit initiation date.
- If the SDTM time point for Opsumit initiation is earlier than incomplete AE date and if the study enrolment date had the same month and year as the AE date, then the incomplete date was imputed with the day of the study enrolment date.
- If the SDTM time point for Opsumit initiation is later than incomplete AE date, then the incomplete date was imputed with the middle of the month.
- If the study enrolment date is earlier than the incomplete adverse AE date, then impute with the first day of the month.

Examples for imputation rules (2/3)

completely missing date of Clinical Management [CM] of an AE

If the start date of a concomitant medication to treat and AE is completely missing in CM table, then the adverse event start date is used for imputation.

Examples for imputation rules (3/3)

death date

- Possible sources:
  - drug safety database – date of an adverse event with a fatal outcome
  - death form – flagged in the demographics table
- Search sequence for possible dates:
  - details from the study CRF → adverse events with a fatal outcome from the study CRF → details from the drug safety database → adverse events with a fatal outcome from the drug safety database → flags in the demographics table
- If date of death is still missing, it is imputed with the date of last available information

Complex ETL rules

- Comprehensive definitions of source concepts for PH WHO groups, etiologies, and sub-etiologies/-groups:
  - 5 different attributes from 2 tables:
    - Medical History [MH]
    - Supplements to Medical History [SUPPMH]

- Extraction of historical context from the medical history free text.

Methods for information rescue

- Keep study name with enrolment consent / date
- Use enrolment date as observation start date
- Build custom concepts
- fact_relationship links
  - Additional characteristic of a medical event (e.g. severity)
  - Outcome of medical event (e.g. hospitalization)
  - Reason for … (e.g. dose change)
  - Adverse event to symptom / treatment linkage
  - Measurements to method linkage

Examples for information rescue (1/2)

Relation to Opsumit

Adverse Events (AE)

AETERM	AESTDTC	AEENDTC	AEREL
LOW HEMATOCRIT AND HEMOGLOBIN	2015-05-06	2015-07-29	POSSIBLY RELATED

AE mapping

Relation to Opsumit mapping (new custom concept)

OMOP CDM MEASUREMENT

measurement_id	measurement_concept_id	measurement_date	value_as_concept_id
1	4151358 - Hematocrit determination	2015-05-06	4267416 - Low
2	4153000 - Hemoglobin level estimation	2015-05-06	4267416 - Low

OMOP CDM OBSERVATION

observation_id	observation_concept_id	observation_date	value_as_concept_id
1	2000000076 - Relationship to Opsumit	2015-05-06	4162850 - Possibly

Link AE and its relation to Opsumit

OMOP CDM FACT_RELATIONSHIP

domain_concept_id_1	fact_id_1	domain_concept_id_2	fact_id_2	relationship_concept_id
1147330 - Measurement	1	1147304 - Observation	1	4165382 -Associated with
1147330 - Measurement	2	1147304 - Observation	1	4165382 -Associated with
1147304 - Observation	1	1147330 - Measurement	1	4165382 -Associated with
1147304 - Observation	1	1147330 - Measurement	2	4165382 -Associated with

Examples for information rescue (2/2)

Measurements to methods

XP

XPTEST	XPDTC	XPMETHOD
Mean Pulmonary Arterial Pressure (mPAP)	2015-02-19	THERMODILUTION METHOD

XP mapping

Method mapping

OMOP CDM MEASUREMENT

measurement_id	measurement_concept_id	measurement_date
3	3028074 - Pulmonary artery Mean blood pressure	2015-02-19

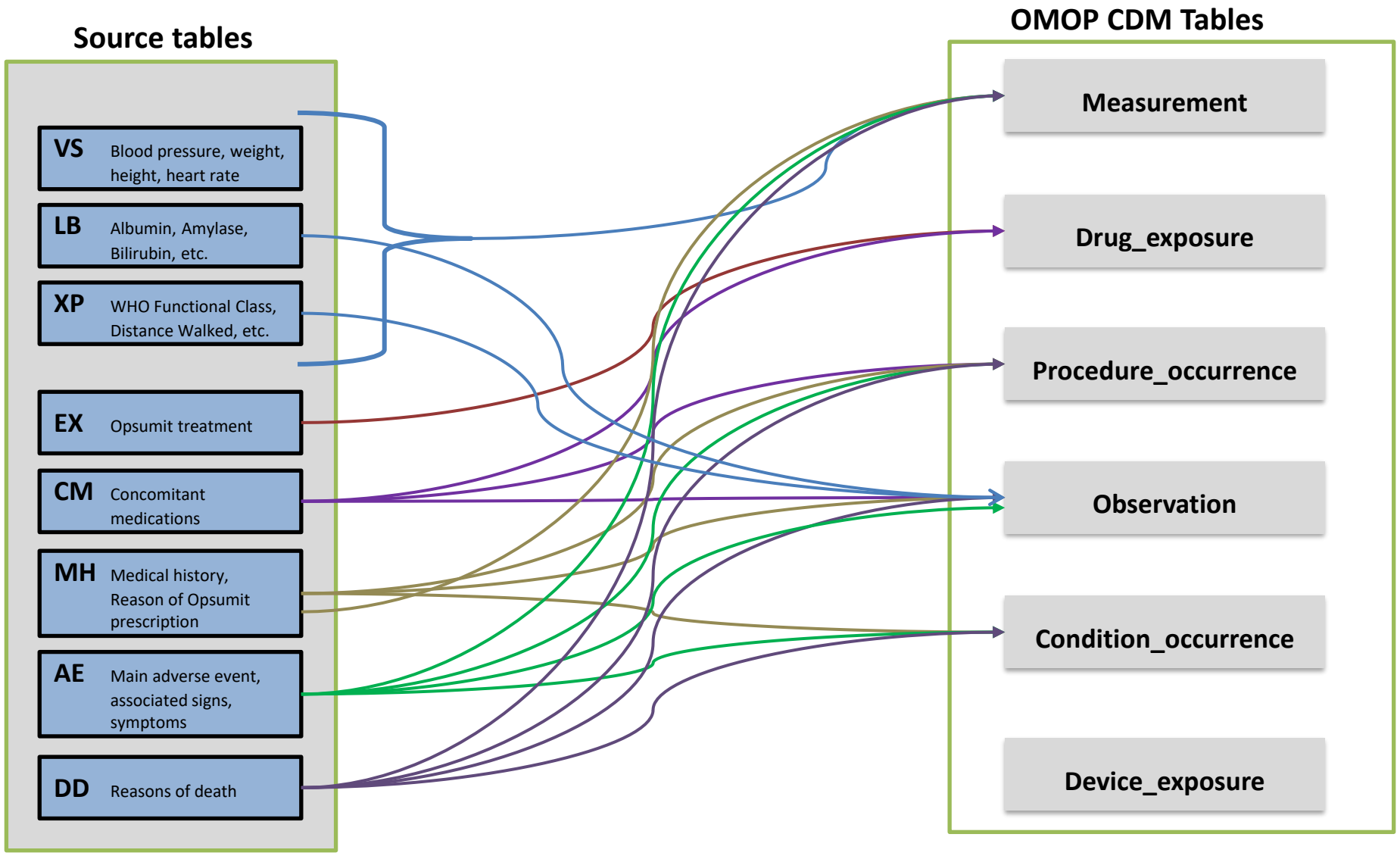
observation_id	observation_concept_id	observation_date
2	4122989 - Thermodilution technique	2015-02-19

Link XP test and its method of measurement

OMOP CDM FACT_RELATIONSHIP

domain_concept_id_1	fact_id_1	domain_concept_id_2	fact_id_2	relationship_concept_id
1147304 - Observation	2	1147330 - Measurement	3	4152892 - Measurement method

SDTM to OMOP data flow



Examples of data flow from „Event“ tables

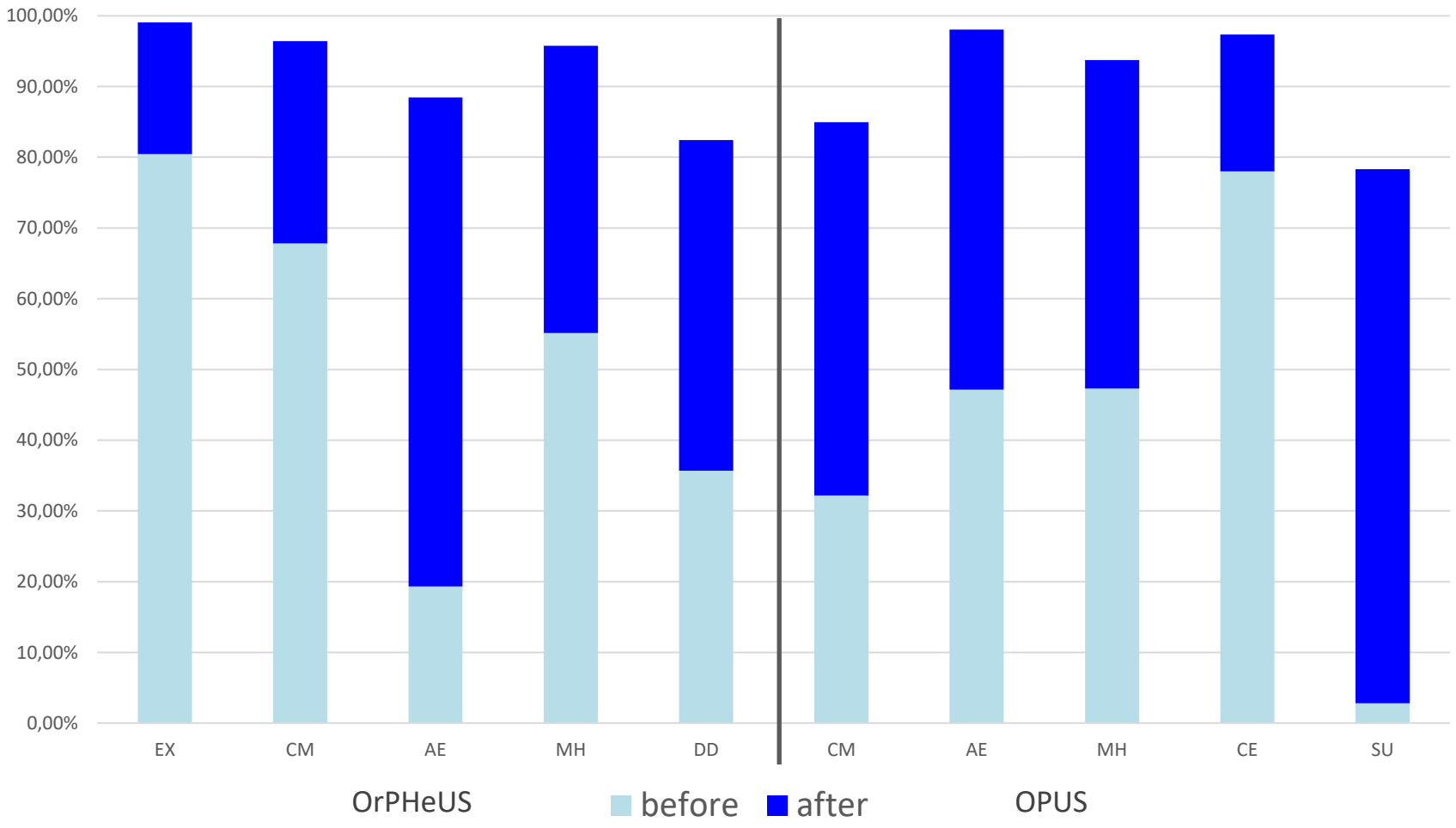
(OPUS Registry)

Results

SDTM to OMOP - Proof of concept

- Facilitates combining Clinical Trial / Registry data with other Real World Data
- Enables use of the Standardized OHDSI Toolset
- Allows entering the OHDSI Network

Data rescue graph



- Selected tables that had a below than 90% coverage

Custom Vocabulary

- MedDRA is recognized as a source vocabulary.
- Concomitant and study medication data (WHODrug and free text in EX table).
- Free text in AE, MH, CM, XP, LB, VS, DD, etc. tables.
- Custom concepts created to reflect the highly-specific terms (PAH subtypes, study drugs, specific tests/assessments/scales).
- Relationships / hierarchy between custom and OMOP vocabulary.