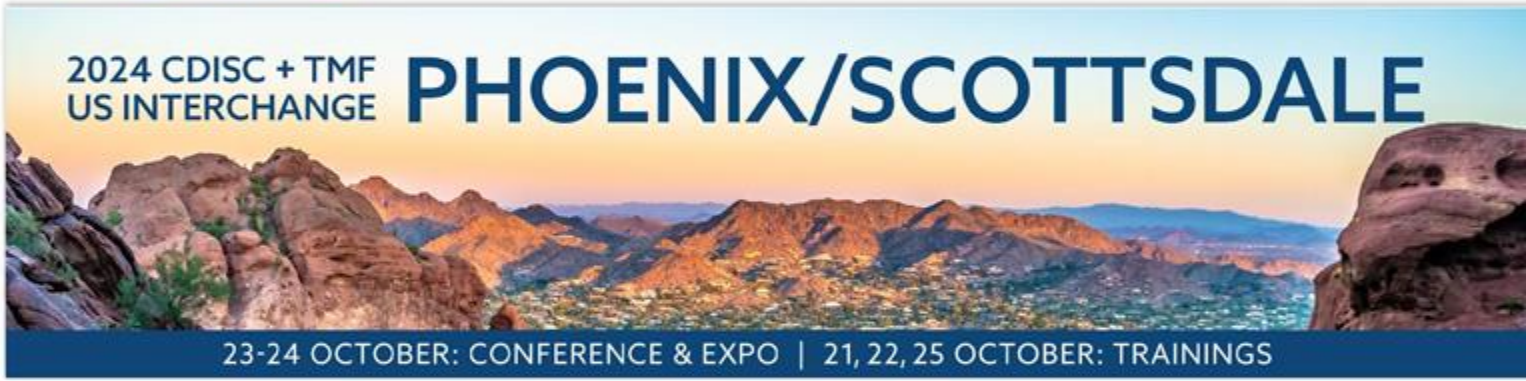


RWD Lineage Initiative

A traceability standard to trust SDTM generated from RWD for regulatory decision-making

Meet the Speakers

Tasha Nagamine

Title: CTO

Organization: Droice Labs

Tasha is an entrepreneur and seasoned technologist with 10+ years of experience in AI, research, and tech strategy. She brings deep expertise in RWD to build data-driven products that have processed over 100 million patient lives. Tasha received her BS in physics from Brown University and left a PhD in AI/deep learning at Columbia University to start Droice.

Anita Umesh, Ph.D.

Title: Biomedical Data Standards Specialist

Organization: Roche/Genentech

Anita is a member of Roche/Genentech's Data Standards and Governance Group. Originally trained as a molecular biologist/biochemist with research experience in cardiovascular science and clinical informatics experience in oncology, she contributes to the Data Tabulation efforts to develop modeling strategies of complex oncology and molecular data into SDTM. She has volunteered on a number of CDISC groups since 2016.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *The author(s) have no real or apparent conflicts of interest to report.*

Agenda

1. Introducing the RWD Lineage Initiative
2. Real-World Data & Data Reliability
3. Lineage & Traceability for Reliable RWE
4. CDISC Initiative: RWD Lineage

Introducing the RWD Lineage Initiative

CDISC standard for lineage to provide the reliability required by FDA to use RWE as primary evidence

Introducing RWD Lineage

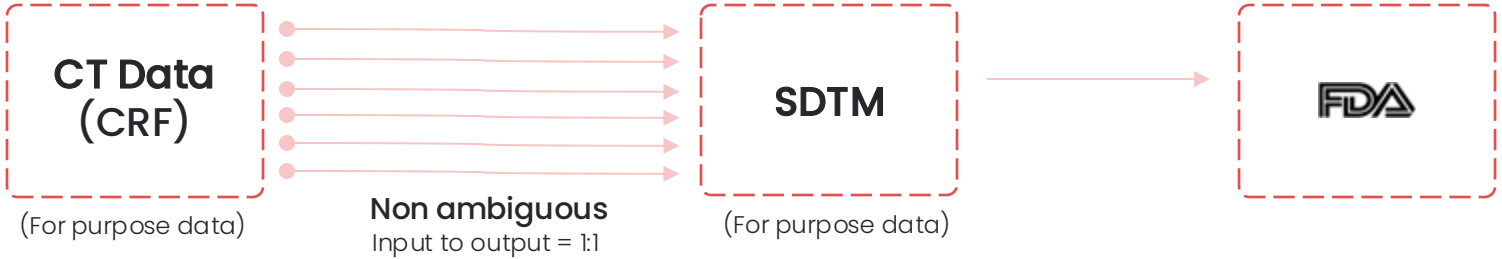
Project Goal

Create a CDISC data exchange standard for lineage metadata that is supplied along with RWD-derived SDTM, which provides the data reliability required by FDA to use RWE as primary evidence.

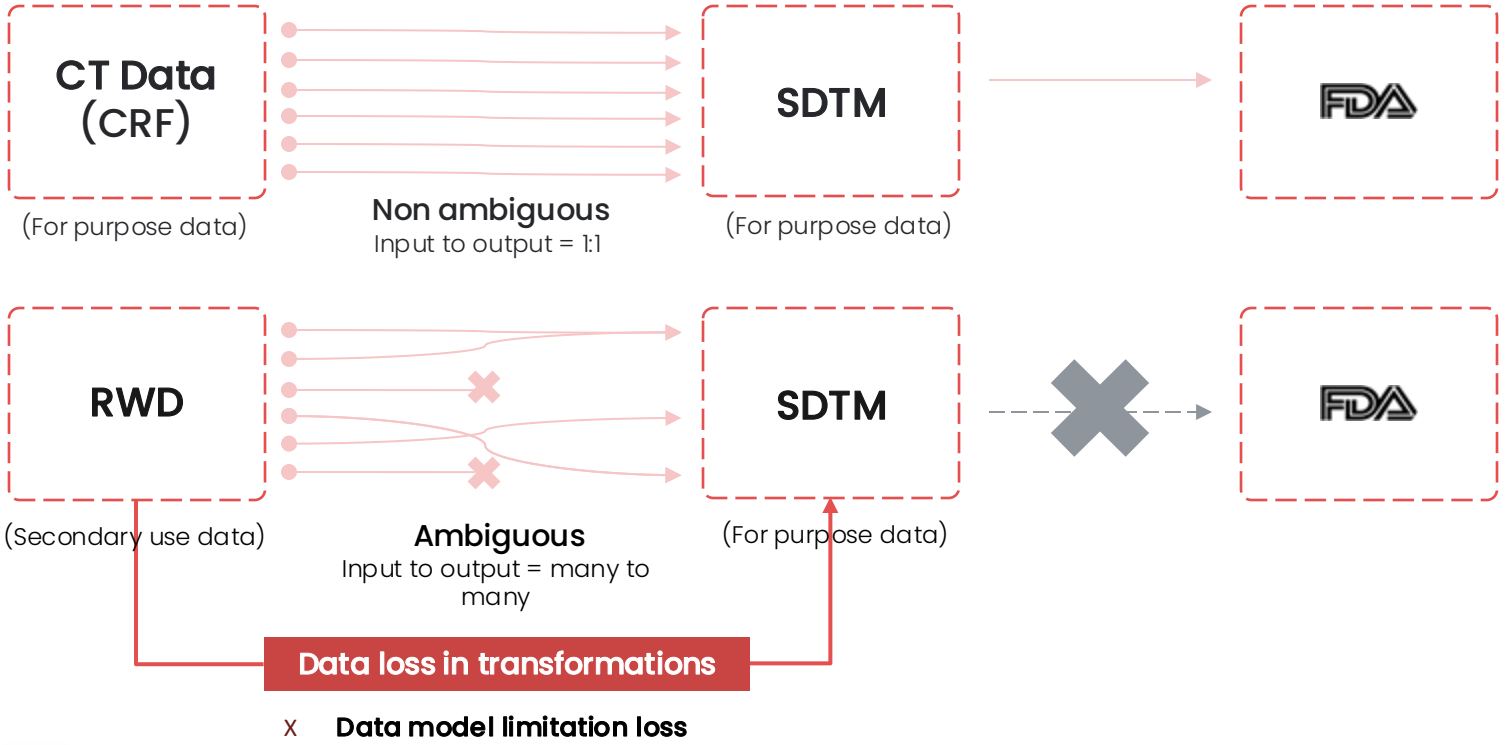
Real-World Data & Data Reliability

Real-world data (RWD) presents both FDA & sponsors with unique challenges in data quality & reliability

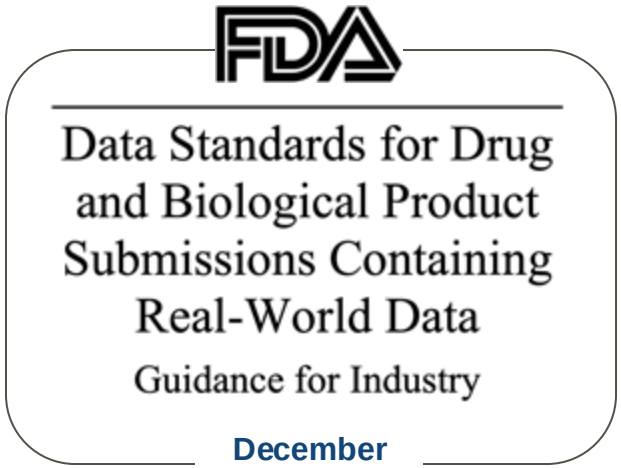
Reliability in RWD: Challenges



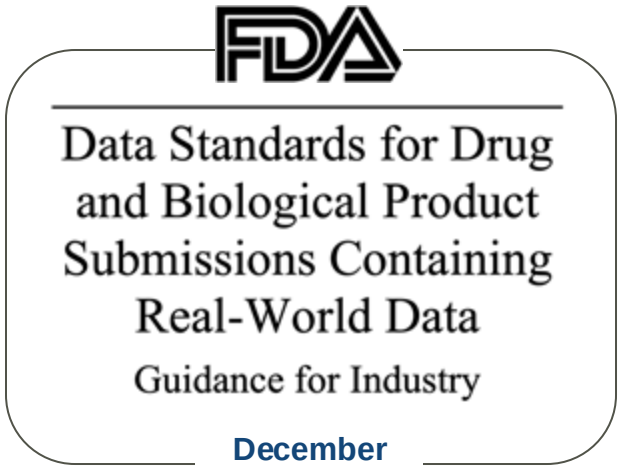
Reliability in RWD: Challenges



FDA: SDTM is the Standard for RWD/RWE



FDA: SDTM is the Standard for RWD/RWE



"Currently, and absent a waiver, sponsors submitting clinical and nonclinical study data (including those derived from RWD sources) in submissions subject to section 745A(a) of the FD&C Act are **required to use the formats described in the Study Data Guidance and the supported study data standards listed in the Catalog**."

FDA: No Lowered Quality Standards for RWD/RWE

FDA: No Lowered Quality Standards for RWD/RWE

"For all study designs, it is important to ensure the **reliability**… of the data used to help support a regulatory decision. For the purposes of this guidance, the term **reliability includes accuracy, completeness, and traceability**."

FDA: Need to Interrogate Source RWD

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

FDA: Need to Interrogate Source RWD

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

FDA

Use of Electronic Health Record Data in Clinical Investigations

Guidance for Industry

July 2018

"FDA must have access to records and may inspect and copy all records pertaining to a clinical investigation in accordance with 21 CFR 312.62, 312.68, 812.140, and 812.145. **All relevant information in the EHR pertaining to the clinical investigation must be made available to FDA for review upon request** (21 CFR 312.62(b), 312.68, 812.140(a), and 812.145).20 This information should be made available and viewable to FDA as original records in the EHR or as certified copies."

FDA: Need to Interrogate Source RWD

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

FDA

Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry

August 2023

"Sponsors must ensure that they are able to submit patient-level data for any RWD that have been analyzed as part of the clinical study included in a marketing application when required under 21 CFR 314.50 and 601.2… If certain RWD are owned and controlled by other entities, sponsors should have agreements in place with those entities to ensure that relevant **patient-level data can be provided to FDA and that source data[15] necessary to verify the RWD are made available for inspection as applicable**."

[15]For the purposes of this guidance, source data include all information in original records and certified copies of original records of clinical findings, observations, or other activities in a clinical investigation used for reconstructing and evaluating the investigation.

cdisc

https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug

FDA: Need to Interrogate Source RWD

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

FDA

Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry

July 2024

"...sponsors should evaluate the completeness, accuracy, and plausibility of the data, including **verifying data against its original source** (e.g., discharge notes, pathology reports, registry records)…"

"...subject matter experts' **review of medical records** (including structured and unstructured data) may be a preferred reference standard for validation of clinical events identified by diagnosis codes or automated algorithms…"

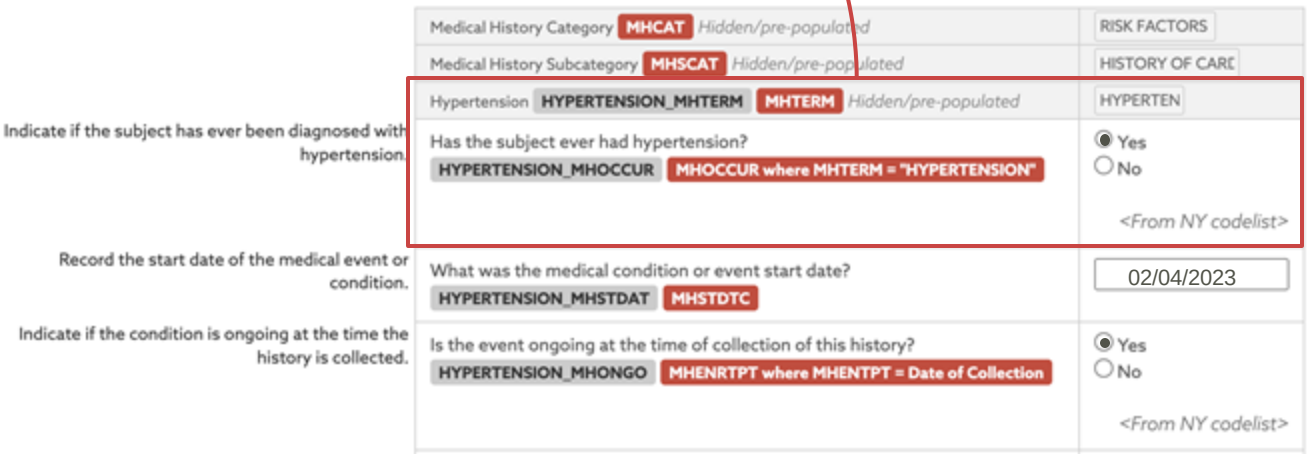
Current State: Traceability for RWD-Derived SDTM

SDTM

MH

…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

aCRF



Current State: Traceability for RWD-Derived SDTM

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

Traditional lineage

- Mapping specs (with or without rationale as comments)
- Define.xml (e.g., value-level metadata, explanation of oddities within the SDTM dataset, whether variables and variable values were derived vs. assigned in protocol vs. collected)
- Identifiers within SDTM dataset and variables such as --XFN that can be used to point back to source data within the EDC or image
- cSDRG providing a description of where source data ended up in SDTM
- Programs/code (not submitted to HAs for SDTM)

Individual data points are not directly traceable back to source

Current State: Traceability for RWD-Derived SDTM

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

Traditional lineage

SPECS				
Source table	Source column	Target domain	Target field	Description
PT_DX	ICD10	MH	MHTERM	Map value to MHTERM
VITALS	VITAL, VALUE	MH	MHTERM	MHTERM = "HYPERTENSION" and MHOCCUR = "Y" if VITAL = BP and VALUE (systolic) > 120 or VALUE (diastolic) > 100
…	…	…	…	…

MAPPINGS	
VALUE	MHTERM
I10	HYPERTENSION
I11.0	HYPERTENSION
…	…

Individual data points are not directly traceable back to source

Current State: Traceability for RWD-Derived SDTM

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

Source RWD

PT_DX		
PT_ID	ICD10	TERM
21962	I10	Essential hypertension
21962	N18.3	Chronic kidney disease, stage 3
…	…	…

VITALS		
Patno	Vital	Value
19251	BP	150/110
19251	BMI	28.3
…	…	…

NOTES	
PT_ID	TEXT
19251	Medical history: Type 2 DM on insulin, CHF, HTN, CKD3, …
19251	Discharge summary: …
…	…

N > 10,000 (EHR)

Individual data points are not directly traceable back to source

Lineage & Traceability for Reliable RWE

Lineage that enables traceability of individual data points from SDTM to source RWD can enable reliability for RWD

Reliability: Atomic Lineage for Individual Data Points

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

Reliability: Atomic Lineage for Individual Data Points

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

Source RWD

PT_DX		
PT_ID	ICD10	TERM
21962	I10	Essential hypertension
21962	N18.3	Chronic kidney disease, stage 3
…	…	…

Reliability: Atomic Lineage for Individual Data Points

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

Source RWD

PT_DX		
PT_ID	ICD10	TERM
21962	I10	Essential hypertension
21962	N18.3	Chronic kidney disease, stage 3
…	…	…

VITALS		
Patno	Vital	Value
19251	BP	150/110
19251	BMI	28.3
…	…	…

NOTES	
PT_ID	TEXT
19251	Medical history: Type 2 DM on insulin, CHF, HTN, CKD3, …
19251	Discharge summary: …
…	…

Reliability: Atomic Lineage for Individual Data Points

SDTM

MH

…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

Source RWD

PT_DX

PT_ID	ICD10	TERM
21962	I10	Essential hypertension
21962	N18.3	Chronic kidney disease, stage 3
…	…	…

VITALS

Patno	Vital	Value
19251	BP	150/110
19251	BMI	28.3
…	…	…

NOTES

PT_ID	TEXT
19251	Medical history: Type 2 DM on insulin, CHF, HTN, CKD3, …
19251	Discharge summary: …
…	…

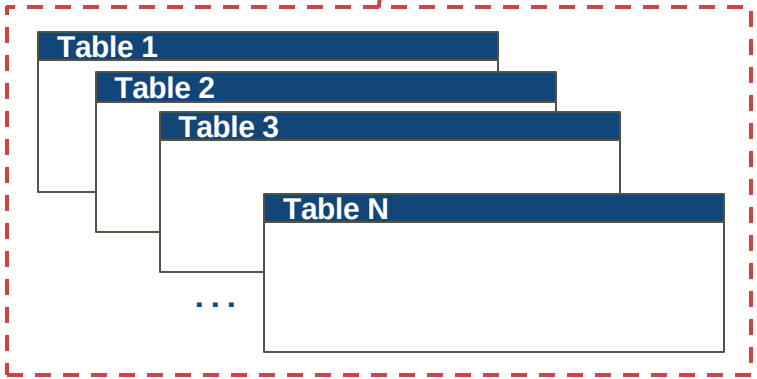
No hypertension for USUBJID = 003 in these tables…

Reliability: Atomic Lineage for Individual Data Points

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

Source RWD



N > 10,000 (EHR)

…or any other table in source RWD

Reliability: Atomic Lineage for Individual Data Points

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

True positive

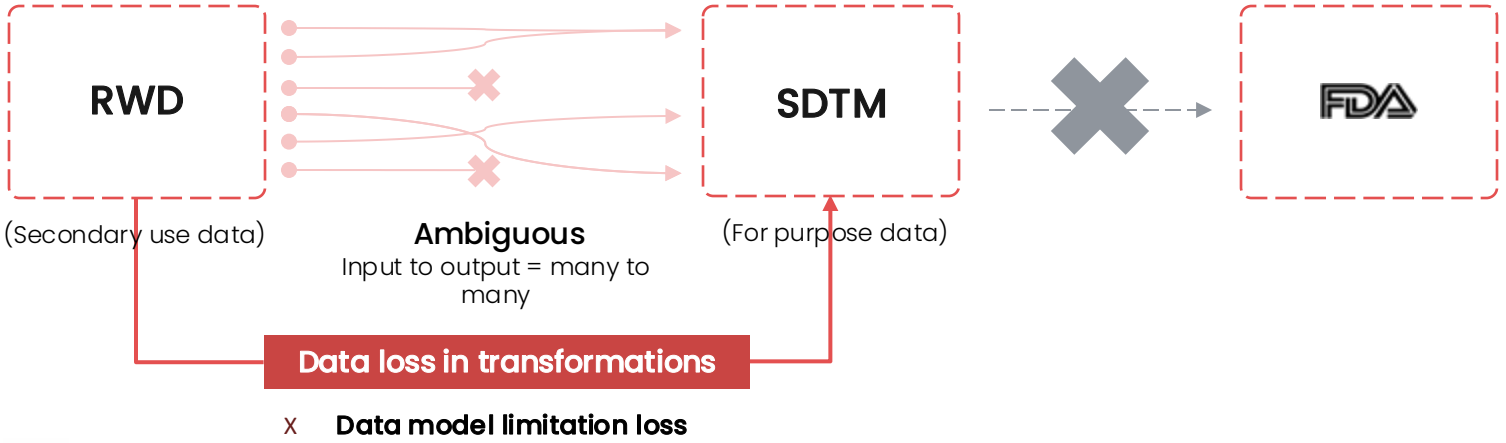
True positive

False negative

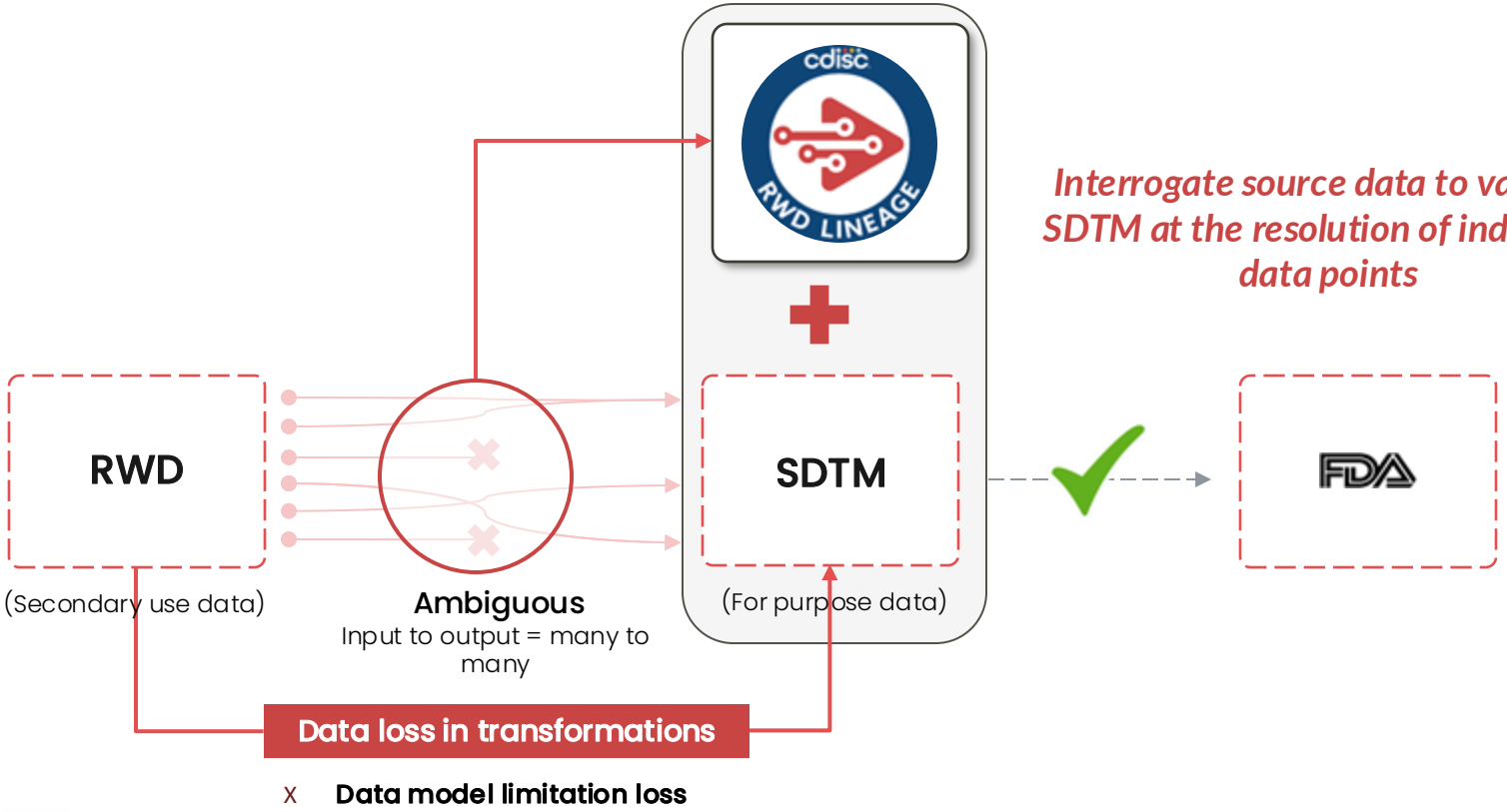
Interrogate source data to validate the accuracy of individual data points

Atomic lineage provides traceability of individual data points, including coordinates and values, from target SDTM to original source data

Reliability in RWD: Challenges



RWD LINEAGE for Reliable RWE



Interrogate source data to validate SDTM at the resolution of individual data points

CDISC Initiative: RWD Lineage

An initiative to create a data exchange standard for lineage metadata that is supplied along with RWD-derived SDTM, which provides the data reliability required by FDA to use RWE as primary evidence

Introducing RWD Lineage

Motivation

- To generate reliable RWD in SDTM for regulatory use, additional information is needed to audit source data and quantify the information loss and performance of data transformations.
  - Traceability + Quantitative Quality

Initial definition:

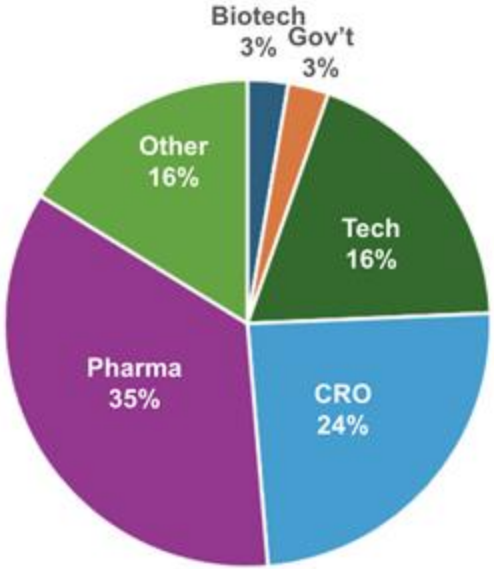
- RWD Lineage will be a standardized and comprehensive representation of data lineage for each source patient data element that specifies either 1) the location of the element in the output SDTM dataset (Positive Lineage), or 2) that the element was not used in the output analysis dataset (Negative Lineage).

Standards development:

- RWD Lineage and quality will be represented in a CDISC standard metadata model.
- This model will be a machine-readable data exchange standard

Introducing RWD Lineage Team

- Team kickoff meeting: July 30, 2024
- 47 team members registered
- Participation from 37 unique organizations
- Six meetings to date
- Meets every other Tuesday on 11am Eastern (next meeting 5th Nov)
- https://wiki.cdisc.org/display/RWDLIN/RWD+Lineage



RWD Lineage Use Cases

Use cases collected from team members with a focus on submission to regulators for decision-making

- Pragmatic trials
- External control arms (ECA)
- Prospectively collected RWD (via EDC)
- Validation studies
- Natural history studies

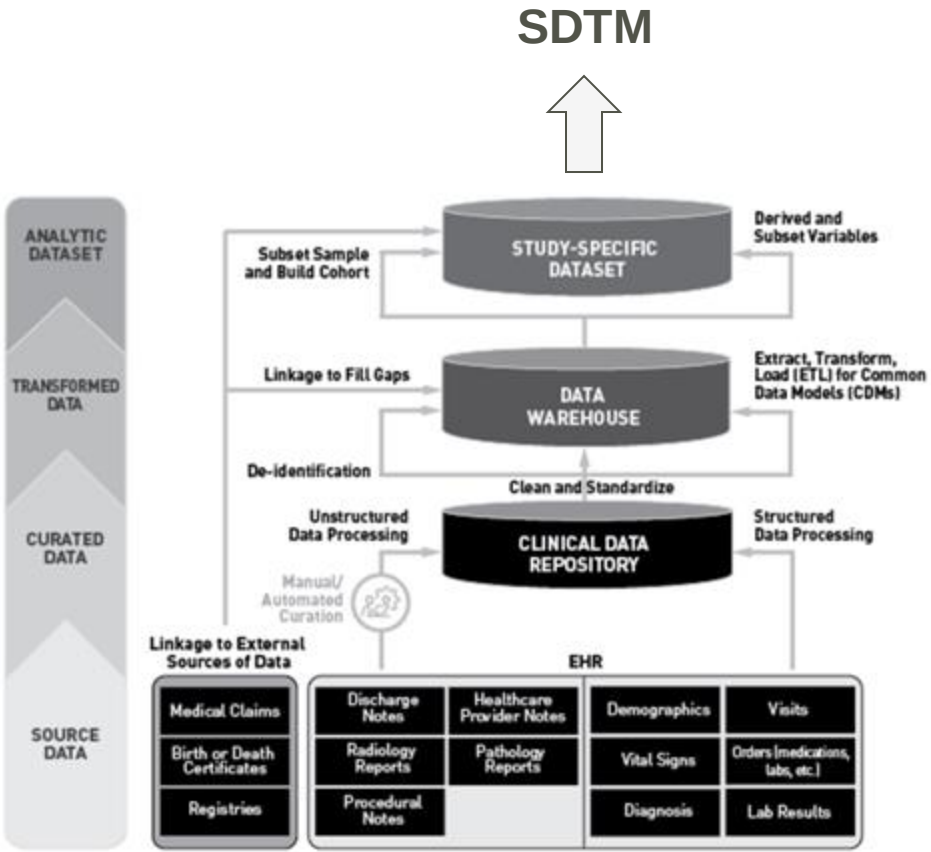
https://wiki.cdisc.org/display/RWDLIN/Use+Cases

Project Scope: Initial Requirements

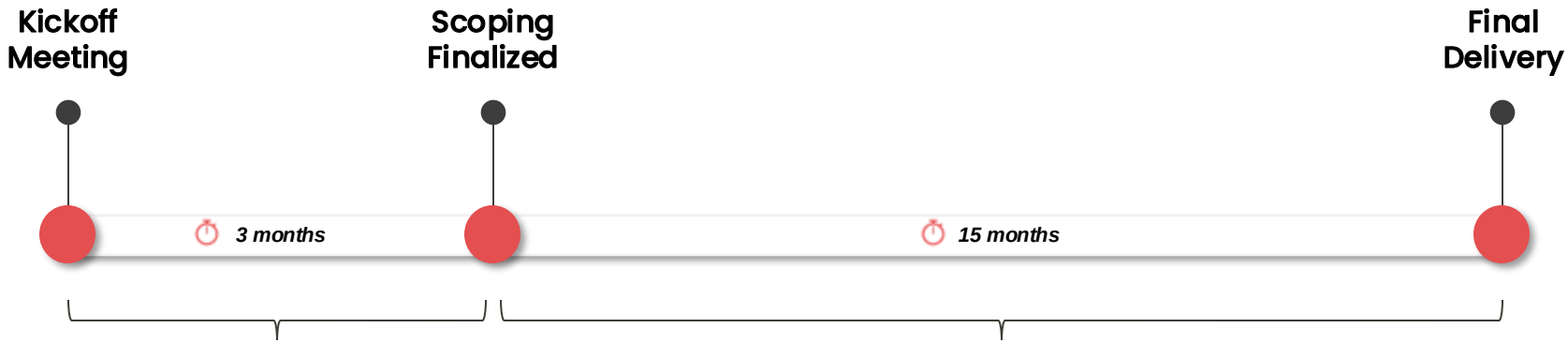
- Enable Priority 1 dimensions of use cases
- Meet requirements in FDA guidance
- Compatible with current submission standards
- Low barrier for adoption and use by sponsors and regulators
  - e.g., start with current define.xml and expanding define.xml
- Support varied data models to accommodate variety of RWD

Project Scope: Proposed Phases

- **Phase I: Reliability**
  SDTM to source data that has been selected for use
- **Phase II: Relevance**
  Fit-for-purpose assessment (selected data back to raw source)
- **Phase III: Implementation**



Project Timelines



- Decide use cases
- Define timelines

- Implement use cases to inform the lineage
- Package lineage into a standard

Read More…

- FDA RWD/RWE Guidance for Industry
  - Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products
  - Real-World Data: Assessing Registries To Support Regulatory Decision-Making for Drug and Biological Products
  - Use of Electronic Health Record Data in Clinical Investigations
  - Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products
  - Data Standards for Drug and Biological Product Submissions Containing Real-World Data
- PhUSE-US 2024
  - Transforming RWD for Regulatory Submissions: How to Use SDTM for RWD

Thank You!

Join us as a volunteer!

cdisc