cdisc

2023
EUROPE
INTERCHANGE
COPENHAGEN | 26-27 APRIL

Challenges and opportunities for CDISC at the intersection of observational and interventional research: ECRAID-Base experience with CDISC implementation

Meet the Speaker

Ankur Krishnan

Title: Data and Project Manager

Organization: Heidelberg Institute of Global Health

Ankur is the Data and Project Manager at Heidelberg Institute of Global Health (HIGH) for the ECRAID-Base project. He led the design and implementation of the ECRAID-Base FAIR-by-design strategy, which includes: i) CDISC-compliant eCRFs for observational studies within ECRAID-Base, ii) cross-walks between CDISC – OMOP CDM (through EHDEN project) and CDISC – HL7 FHIR (through Horizon Standardization Booster project), iii) machine-actionable study- and site-level metadata schemas and, iv) machine-actionable Data Management Plan (DMP).

Ankur has over 5 years of experience in data standards, harmonization, collection and management within the translational and evidence-based clinical research landscape. And, 7 years of experience in project management.

He holds Master's degrees in Forensic Science (2011 – 2013: NICFS, New Delhi, India) and Applied Biotechnology (2016 – 2018: McGill University, Montreal, Canada).

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 ECRAID-Base project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 965313*

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

Agenda

1. Introduction
2. CDISC-compliant CRF development:
   A 'central library' approach
3. Challenges and Opportunities
4. Considerations and Recommendations

Introduction

ECRAID-Base

- **Ecraid** stands for the **European Clinical Research Alliance on Infectious Diseases** (https://www.ecraid.eu/)
- Our purpose is to **reduce the impact of infectious diseases** on individual and population health in Europe.
- ECRAID-Base addresses ID **outbreak preparedness and response** through generating rigorous evidence to **improve the diagnosis, prevention and treatment of infections and to better respond to infectious disease threats**.
- This is facilitated by a **European multidisciplinary clinical research network and innovative research approaches**.
- ECRAID-Base aims to provide a 'warm base' for clinical trials (contracts, trained team, data capture system) and sentinel sites for (re)emerging IDs through **five perpetual observational studies**

Perpetual observational studies (POS)

- A POS is a **prospective, observational clinical study that perpetually enrolls patients**, collecting a set of demographics, clinical characteristics, and outcomes, mostly available through routine care, as described in a core protocol.

- POSs in ECRAID-Base are **multicentre studies** designed **to address key clinical research gaps, including variations in clinical practices, incidence of ID syndromes, and associated risk factors.**

- Each POS creates a **clinical research backbone**, ready to concurrently or sequentially **embed studies (observational, experimental, investigator-initiated, or commercial)** and efficiently advance the evidence base for infectious diseases management.

- **Enhancing intervention studies** - Cohort participants as the external control (historical control) arm for intervention studies, Pre-trial, longitudinal data on participants and RWD to better the definition of clinical endpoints (clinical and microbiological cure) in trials

Please see - Hassoun-Kheir N, van Werkhoven CH, Dunning J, et al. Perpetual observational studies: new strategies to support efficient implementation of observational studies and randomized trials in infectious diseases. Clin Microbiol Infect. 2022;28(12):1528-1532. doi:10.1016/j.cmi.2022.07.024

POS	ID syndrome	Healthcare setting	N anticipated sites	N anticipated patients, per year	Data types	Status
POS-ICU-VAP	Ventilator-associated pneumonia	Intensive care unit	40	4000	EHR, laboratory and microbiological data	10 sites 396 participants
POS–cUTI	Complicated urinary tract infection	Hospital	40	3000	EHR, laboratory and microbiological data	14 sites 286 participants
POS-ER-ARI	Community-acquired acute respiratory tract infection	Secondary Care	40	4000	EHR, laboratory and microbiological data	Live (14 April 2023)
POS-PC-ARI	Community-acquired acute respiratory tract infection	Primary care	50-100	2000	EHR, laboratory and microbiological data	Not live yet
POS-ER-Disease X	(Re)emerging ID syndromes amongst immunocompromised patients	Hospital	5-8	400	EHR, laboratory, microbiological and OMICs data	Not live yet

How CDISC helps ECRAID-Base deliver on its mission

Compliance	• Submission ready core dataset that can be expanded for intervention studies that respond to (re)emerging IDs • Historical controls, external control arms to facilitate RCT • RWD to better the definition of clinical endpoints in trials
Interoperability	• Between POSs • Across related ID studies, including antimicrobial resistance (AMR)
Utility	• Real-time AMR surveillance by EPI-Net • FAIR-by-design data to enhance reuse • CDISC-compliant, prospectively-harmonized eCRFs

CDISC-compliant, FAIR-by-design CRF development: A 'central library' approach

‘Central library’ approach - CRF development

Developed list of domains/CRFs based on the existing CDISC domains (e.g., Demographics, Vital Signs, etc.) and commonly used CRFs in observational studies on ID

Study teams indicated the domains/CRFs relevant for their study to establish a set of ‘common domains/CRFs’ across studies.

Any domain/CRF indicated by more than one study team was included in the ‘central library’

The common domains/CRFs were populated with variables and controlled terminology using -

- CDISC resources (IGs, eCRF portal, TAUGs, etc.)
- ISARIC COVID-19 CRFs
- CRFs of existing studies on the disease area (led by the study teams)
- Input from study teams, LAB-Net and EPI-Net.

Alignment between common domains/CRFs in ‘central library’ and CDISC domains

Common form	CDISC domain
Informed consent	DS domain
Screening - Inclusion/Exclusion criteria	DS domain
Admission details	HO domain
Demographics	DM, SC and RP domain
Vital signs	VS domain
Predictive scores and scales	QRS domain
Comorbidities and Risk factors	MH domain
Prior and Concomitant Medications (including vaccination history)	CM domain
Signs and symptoms	CE domain

Common form	CDISC domain
Laboratory tests	LB domain
Radiological tests	PR and FA domains
Microbiological Identification and Susceptibility testing	MB and MS domains
Treatment	CM and PR domains
Events and complications	CE, AE and FA domains
Microbiological cure	MB and MS domains
Outcomes	DS, DD, HO and FA domains
Withdrawal	DS domain
Exposures and Recent Travel	ER domain

‘Central library’ approach - CRF development

Master lists and Controlled Terminologies

Created master lists of controlled terminologies with study teams, for -

- Diseases and health conditions (MedDRA/ICD-10)
- Specimen types
- List of Pathogens provided by LAB-Net (mapped to CDISC CT/NCBI taxonomic IDs)
- Resistance Mechanisms/Profiles
- Antimicrobial agents (ATC)

POS-specific eCRF development

- Study team chose the data items and categories/controlled terminologies in each common form and master list, that need to be collected for their study
- Any new study specific data items and forms were mapped to CDASH/SDTM
- Where a variable could not be mapped to CDASH/SDTM, we created Non Standard Variables (NSV)/supplemental qualifier.

POS name	% of variables mapped to CDASH/SDTM	% of NSVs	% of variables for internal data management *
POS-ICU-VAP	81	15	4
POS–cUTI	85	12	3
POS-ER-ARI	88	7	5
POS-PC-ARI	68	30	2
POS-ER-Disease X	83	14	3

- Based on the total number of variables, variables mapped to CDASH and NSVs created, for each study
- POS-PC-ARI highly preferred to use CRFs from their previous point-prevalence audit survey studies. Did not use ‘central library’ approach. Study CRFs aligned to ‘central library’ and mapped to CDASH/SDTM, as best possible
- **POS-ER-Disease X will export laboratory and microbiological data directly from the sites** (i.e., not captured via eCRF). Align and map each sites data export format to ‘central library’ and CDASH/SDTM

Challenges and Opportunities

CDISC resources & support

We used CDASH and SDTM IGs, CDISC library, CDISC wiki, CDISC Knowledge base, CDISC eCRF portal, CDISC TAUGs and existing studies/use cases such as ISARIC COVID-19 CRFs

- **Internal support** (at the beginning) - Steve Canham, ECRIN
- **External support** (towards the end) - Gary Walker, CDISC
- Subsequent to our mapping work, we found the **CDISC RWD webinar** and **JMIR Medical Informatics article** helpful and encouraging

Challenges -

- At the beginning, identifying and **understanding how to navigate and use the resources was difficult**
- **IGs are ~ 450-page documents**
- **CDISC library and wiki are great for targeted searches** i.e., one has some prior (at least basic) understanding of the standards
- **Access to CDISC implementation support/expert**
  - Cost-prohibitive to have CDISC support/expert on retainer in government-funded research projects

Understanding CDASH and SDTM for CRF design and development

Important considerations during CRF development -

- Study-, disease area- and setting-specific data collection needs and practices
- Minimizing the burden of data collection
- Aligning with end-users needs and objectives such as EPI-Net, STAT-Net
- CDASH/SDTM IGs and conformance rules

Finding the balance between -

Pushing the study team to change the format of their variable to fit CDASH/SDTM IG and rules	- Bending conformance rule(s) - Creating a new Non-standard variable (NSV)/supplemental qualifier

At times, **understanding SDTM was important when designing CRFs** because how data are tabulated in SDTM affects –

- Descriptive metadata collection (e.g., anatomical location where temperature was taken)
- How variables should be partitioned (or not)

Specimen types such as Endotracheal aspirate, Bronchoalveolar Lavage and Suprapubic Aspirate

- **From a CDASH/SDTM perspective**, better to collect them as Specimen material type, Anatomical location and Method variables, separately, in the CRF
- **From a clinicians/researcher's perspective**, these are commonly collected specimens. Splitting them in CRF would not make sense and likely, increase the burden of data collection
- Best approach to request them as new terms in CDISC CT?

Sample type	MBSPEC (SPECTYPE Codelist)	MBLOC (LOC Codelist)	MBMETHOD (METHOD Codelist)
Endotracheal aspirate/fluid	Endotracheal Aspirate/Fluid	Trachea	Not found
Bronchoalveolar lavage	Lavage Fluid	Lower Respiratory System OR MBLOC = Multiple; MBLOC1 = Bronchus; MBLOC2 = Alveolus	Not found
Suprapubic Aspirate	??	Suprapubic Region	Not found

CDISC Controlled Terminology and other external terminologies

Potential variations across geographical areas, cultures and laboratory practices

- **RACE and ETHNIC codelists**
  - Not extensible
  - Controlled terminologies in codelists are US-centric
  - CDISC CT could not be used for our EU-based studies
- **MBTEST, MSTEST and METHOD codelists**
  - Extensible
  - Some of the tests and methods used by laboratories in our network were not available in the codelists or could not be mapped completely (loss of granularity)

Difficult to know **when to extend CDISC versus when to bring in a different terminology**

- Effect of this on compliance is not clear
- What happens if we need to extend an inextensible CDISC CT codelist?

Potential effect on semantic interoperability of -

- Extending CDISC CT (codelist)
- Using external CTs
- Sponsor-defined codelists

CDASH/SDTM mapping - an observational study design perspective

Mapping ‘Treatment’ administered as part of routine care

- Exposure (EX) domain for trials
- Treatment mapped to CM domain in our studies
- CM domain includes both medication being taken at the time of admission (prophylaxis, maintenance therapy, etc.) and administered for treatment

Understanding time in CDISC

- Some timing related SDTM variables/concepts may not be available or align with observational study design
- How **‘Actual and Relative Time Assumptions’ - SDTMIG** and **‘Mapping Relative Times from Collection to Submissions’ - CDASHIG**, relate to sponsor-defined study reference period difficult to understand
- More guidance on use of RELMIDS, --EVLINT, etc. and terms like "DURING", "DURING/AFTER", "COINCIDENT“, etc. from an observational/cohort study perspective

CDASH/SDTM mapping - an observational study design perspective

Signs and Symptoms, Onset date, Diagnosis, Recurrence and Reinfection in CDISC -

- In our studies -
  - **At Day 0/VAP screening** - Signs and Symptoms, Date of Onset and (Working) Diagnosis
  - **At Day 14** - Follow-up assessment of Signs and Symptoms, Primary Diagnosis
  - **At Day 30** - Reinfection and Recurrence

- As per CDASHIG and TAUGs, multiple approaches possible -
  - **CE and FA(CE) domain/dataset**
  - MH and FA(MH) domain/dataset; using MHEVDTYP
  - MH, SM (MIDS = "DIAGNOSIS") and FA(MH) domains/dataset

- How do we map Signs and Symptoms, Date of Onset, Date of Diagnosis (Working and Final), Date of Reinfection and Recurrence within the CE domain?

- **--EVDTYP (using EVDTTYP codelist) for CE domain** or something similar might be needed

New Term Requests and Non-Standard Variables (NSV)/Supplemental Qualifiers (SQ)

- We had to –
  - Extend the CDISC CT codelists
  - Create some Non-Standard Variables (NSVs)

- These additions **come at a cost for interoperability** and **may affect regulatory compliance**

- TAUGs (TB, COVID-19 and Influenza), Vaccine Administration mapping document and ISARIC COVID-19 CRFs were good resources

- How do we **notify and request creation of new NSV**? How do we **track requested NSVs**?

- Are **all NSVs incorporated into CDISC standards?** Sometimes, it might be too late by the time a NSV is officially incorporated.

- Slightly difficult to understand the process of requesting a new term or a clinical scale/scores/questionnaires (and CDISC's decision making process behind approving new term)
  - What if we had to extend an inextensible codelist?

How do we measure & validate CDISC conformance?

Our study data will be **converted to SDTM, only when needed**.

We were lucky to have **Gary Walker (Implementation expert, CDISC)** to help us review new variables for conformance

How do we **measure CDISC conformance during CRF development?**

Jentoft et al., Journal of the Society for Clinical Data Management 2(3), 2023

- Ongoing SDTM conversion throughout the trial
- Provided SAS data exports every two weeks to an external CDISC consultant, to create and update the SDTMs
- Consultant exported the SDTM output to the Pinnacle 21 data review validator
- Data or structural issues resolved

Not ideal (maybe, not even possible from a time and resources perspective) in our project

How do we maintain CDISC conformance & interoperability?

Sample type	During CRF development - SDTMIG v3.3 (SDTM v1.7)			SDTMIG v3.4 (SDTM v2.0)		
	MBSPEC (SPECTYPE Codelist)	MBLOC (LOC Codelist)	MBMETHOD (METHOD Codelist)	BSSPEC (SPECTYPE Codelist)	BELOC (LOC Codelist)	BECLMETH (CLMETH Codelist)
Endotracheal aspirate/fluid	Endotracheal Fluid	Trachea	Not found	Endotracheal Fluid	Trachea	Aspiration
Bronchoalveolar lavage	Lavage Fluid	Lower Respiratory System OR MBLOC = Multiple; MBLOC1 = Bronchus; MBLOC2 = Alveolus	Not found	Lavage Fluid	Lower Respiratory System OR MBLOC = Multiple; MBLOC1 = Bronchus; MBLOC2 = Alveolus	Lavage
Suprapubic Aspiration	??	Suprapubic Region	Not found	??	Suprapubic Region	Aspiration

How are these differences reconciled across versions of CDISC? How do you know the level of interoperability between CDISC versions?

Considerations and Recommendations

General challenges

- Deciding on an ontology
- Making judicious use of standards' flexibility
- Determining a budget for applying CDISC or any ontology
- Convincing the PI/funder that the budget for applying CDISC of any ontology is warranted
- Building & keeping talent
- Linking stakeholders (Researchers, Funding bodies, Regulatory bodies, and Standard Developing Organizations) to share responsibility and take action on multiple levels (e.g., funding mandates, training, improved tooling for standards implementation)
- Need for crosswalks between CDISC, OMOP-CDM and HL7-FHIR and other standards (fractured FAIR landscape)

Considerations

- CDISC, through CDASH, enables FAIR-by-design CRF development approaches
- CDISC data standards are very well suited for the intersection of observational research/RWD and interventional research
    - Enables regulatory body submission compliance and FAIR principles
    - Highly important in Infectious Diseases and Antimicrobial Resistance research landscape
    - Can enable better outbreak preparedness and response
- CDISC has a different stakeholder landscape than some other commonly used standards and the focus on fit-for-regulatory submission may be more important than fit-for-cross-study interoperability
- CDISC is perhaps too rigid at times (observational study data might not be as conformant) and too flexible at others (study data might not be as interoperable)

Recommendations for CDISC

- **Improved tooling and simplified guidance** to help orient new-users and to understand and navigate of existing CDISC resources
- **Cross-walks and standardization across standards and terminologies** used in healthcare research, to enable, encourage and enhance structural and semantic interoperability
- Some **mechanism to assess conformance** (perhaps, NLP-based) – ideally as CRF is being developed.
- Some **guidance/checks/SOPs that can be incorporated into study's data validation plans and quality checks to ensure a level of conformance**
- Guidance on **how compatible different versions of CDISC standards are with one another** to understand when to move to the next version
  - Some level of **automation for (re)mapping terms to new domains** when a new, more appropriate domain is established in a new version of CDISC

Recommendations for CDISC

- **CDISC Portal** -
  - Researchers implementing CDISC can register
  - Provide details such as data source(s), research area, etc.
  - Submit keywords and highlight domains to tell CDISC what they are working on
  - Produces targeted updates if there are new domains added, updates to relevant domains, etc.
  - Researchers can notify of sponsor-defined categories, sub-categories and codelists

- **Accessible NSV/SQ registry:** A tool/portal to -
  - Notify and request creation of new NSV (study-/disease area-/healthcare setting-specific)
  - Access existing, approved NSVs
  - **Integrate into CDISC library**
  - Guidance on CDISC's decision making process around new variables

Recommendations for CDISC

- **CDISC trained and certified SMEs**
  - Academic research projects can sub-contract (internal or external funding initiatives)
  - Support protocol and CRF development, conformance checks, etc.

- **More dialogue between CDISC and researchers** to help CDISC react to researcher and other stakeholder needs through:

  - Extending the CDISC RWD Connect Initiative
  - CDISC RWD Working group
  - Forum hosted on CDISC website
  - Freely available webinars
  - More guidance tailored to observational studies
  - Published articles and white papers

Acknowledgment

ECRAID-Base

- Steven Canham
- Lauren Maxwell
- Frank Leus
- Lisanne Vincent
- Ellen van Deuren
- Roxanne Schaaks
- Thomas Jaenisch

CDISC

- Gary Walker
- John Owen
- David Evans

Thank You!