European Federation of Pharmaceutical Industries and Associations

Assessing the EMA data quality framework (DQF) dimensions using REQUeST: a decentralized registry use case






Pamela Dobay, PhD | Meritxell Sabidó, PhD

Objectives

EFPIA acknowledges there are important learnings and challenges to review with EMA relevant to measuring and characterising data quality (DQ) in the context of RWE generation. In this presentation, we will:

- Share industry experience applying EMA DQF principles to a network of decentralized registries in the context of a PASS
- Discuss the strengths and limitations of using the REQuEST tool to facilitate assessments and documentation of DQ before commencing a study
- Capture key learnings and potential improvements in DQ assessments and documentation to align processes with principles in the DQ framework

Case study: evaluation of fitness-for-purpose of data from a network of decentralized registries with indirect data access for PASS

	Description
Rationale	• High representativity • Long follow-up • Potential to collect data that can be harmonized systematically • Potential for large-scale, standardized analysis
Composition	Several national and international registries with decentralized governance
Electronic data capture (EDC) system	• Multiple EDCs designed to meet individual registry objectives • No common data model for all PASS as of June 2023
Data flow and access	• Analysis performed at individual registries • Aggregate data provided to third party (i.e., analyses is not federated)

The consortium completed **REQuEST** as part of EMA PASS qualification. This was prior to the EMA DQF draft publication. The case study is based on **a revision of REQUeST** completed by the consortium in 2023.

REQUeST was designed by EUnetHTA to support HTA bodies and other actors in evaluating registry data, and is included in the EMA guideline on registry-based studies

Note: *other tools to assess data quality are available and are currently being evaluated by the EFPIA IEGU DQ workstream*

Methodological Information	Essential standards (data quality and protection)	Additional requirements (good practice questions)
• Type of registry • Use for registry-based studies and previous publications • Geographical and organizational setting • Duration • Size • Inclusion and exclusion criteria • Follow-up • Confounders	• Registry aims and methodology • Governance • Informed consent • Data dictionary • Minimum data set • Standard definitions, terminology and specifications • Data collection • Quality assurance • Data cleaning • Missing data • Financing / sustainability • Protection, security and safeguards	• Interoperability and readiness for data linkage • Data sources • Ethics

*REQUEST TOOL AND ITS VISION PAPER – EUNETHTA: GUIDELINE ON REGISTRY-BASED STUDIES (EUROPA.EU)

How can we use REQUeST to assess EMA DQF* dimensions and metrics?

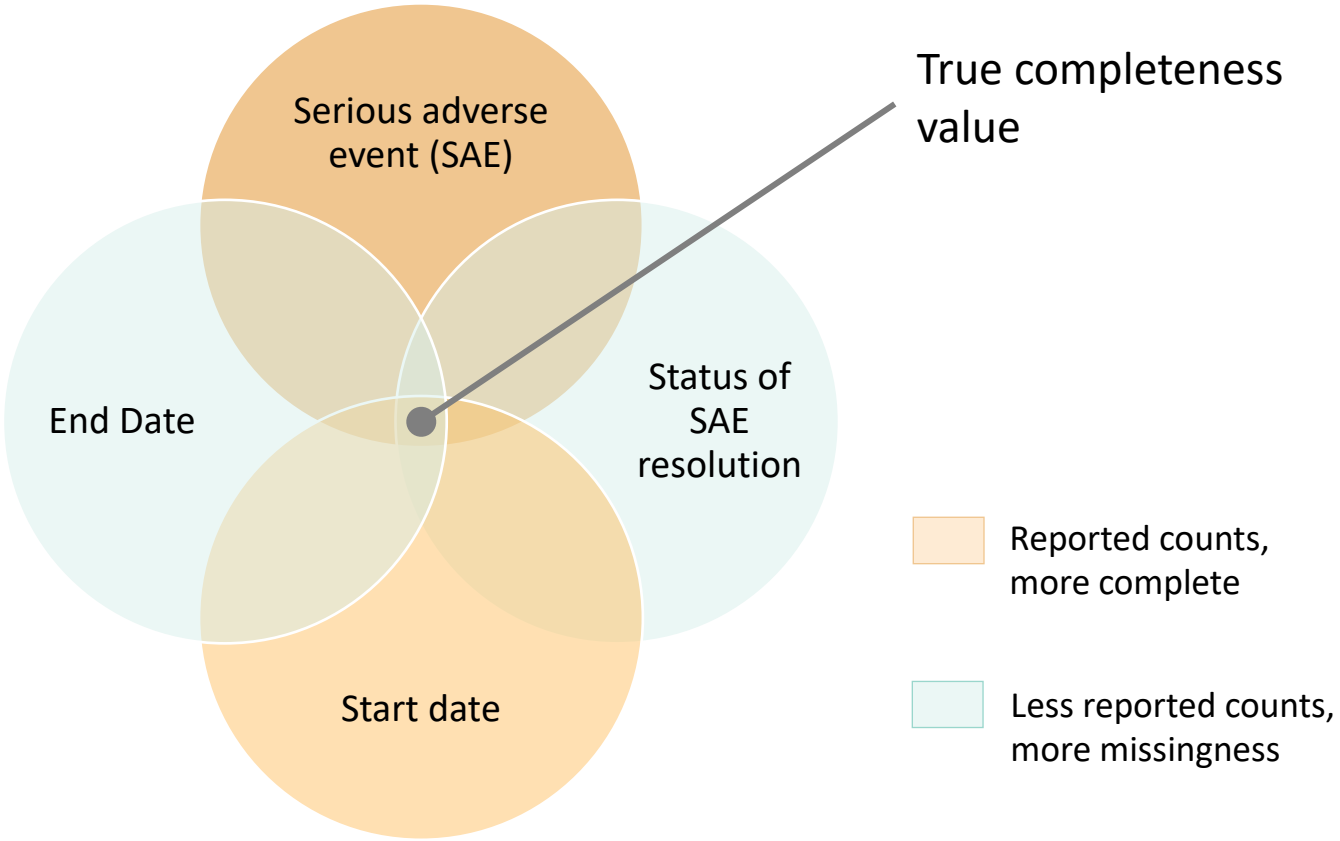
Note: *Relevance is not covered, as it is study specific*

DQ dimensions and metrics (Sub dimensions)	Information used in assessment	Assessment outcome
Reliability (Precision, Accuracy, Plausibility)	• Data dictionaries or variable lists • Methodology for data capture • Quality assurance and use of data quality indicators (DQIs)	+ Elements of reliability can be assessed through REQUeST elements (data dictionary, data capture processes) *- Reliability can only be fully assessed with analyses*
Extensiveness (Completeness, Coverage)	Registry-reported coverage and completeness for each data element in a core PASS** protocol	*- Registry-reported coverage often has limitations* *-Completeness can only be fully assessed with analysis* ***(Example 1)***
Coherence (Format, Structural, Semantic, Uniqueness, Conformance)	• List of core data elements for PASS • Data dictionaries or variable lists • Methodology for data capture and correction	+ Format and structural coherence can be assessed from data dictionary *- Semantic coherence can only be fully assessed with analyses*
Timeliness (Currency, Lateness)	Methodology on timeliness of data capture, export, and correction	+ Currency and lateness can be assessed through REQUeST elements

*DATA-QUALITY-FRAMEWORK-EU-MEDICINES-REGULATION_EN.PDF (EUROPA.EU) ** FOR EMA QUALIFICATION

When evaluating composite outcomes such as safety, completeness should be reported based on counts of patients that have all constituent data element available

Example 1



Using REQUEsT to assess DQF dimensions is complex in decentralized registries with indirect access to patient-level data elements

Learnings on applying DQF principles to a decentralized registry

- **Learning 1:** Registries that meet or implement processes to enhance DQ according to the EMA DQF are valuable sources of RWD to support regulatory decision-making
    - Assessing and documenting DQ helps shape roadmaps to identify remediable DQ gaps and opportunities to improve operations in a registry
    - Registries could benefit further from the EMA DQF by having an avenue (outside of qualification) to consult the EMA on their DQ

Learnings from using REQUeST

- **Learning 1:** The efforts to sufficiently assess DQ using the EMA DQ framework are substantial for industry and registry holders
- **Learning 2:** While useful and detailed to assess the foundational determinants of DQ, REQuEST needs to be complemented with analyses to address intrinsic determinants
- **Learning 3:** For decentralized registries, the level of DQ required for individual registries depends on regulatory purpose of data, and is influenced by the data governance model

We offer suggestions to the EMA on how to improve DQ assessments and documentation to implement the DQF

The recommendations are geared towards more efficient DQ determination and enabling operational sustainability for all stakeholders (regulators, registries, industry)

Suggestion 1

Continue leading dialogues to:

- Set clear expectations for DQ assessments, processes, and documentation
- Facilitate discussions that will ultimately make quality assessments more efficient and aligned between all stakeholders
- Maximize the utility of registry data for regulatory purposes

Suggestion 2

Consider guidance on minimum information that should be readily available to improve efficiency and transparency of DQ assessments:

- Data dictionaries, common data models (CDM), and data quality indicators (DQIs)
- Where there is a CDM, relevant mappings and CDM testing routines
- DQ documentation (e.g., in DARWIN)

Suggestion 3

Lead the co-development and piloting of tools aligned with the DQF, together with all stakeholders. Initial suggestions are:

- Self-explanatory checklist to operationalize EMA DQF
- Generalizable DQIs for evaluating risk / benefits of using specific data sources
    - E.g.: Loss to follow-up, representativeness, outliers
- Guidance on acceptable thresholds per DQI

In Summary

We highlighted important considerations and recommendations for registry-based RWE generation.

Some DQ principles can be extrapolated to other sources of RWD such as claims and electronic health records, while others are not, highlighting the need for data-source specific DQ considerations.

We look forward to future engagements with EMA on DQ standards and tools relevant to sources of RWD in Europe.

On behalf of EFPIA, thank you for your attention.