HMA
Heads of Medicines Agencies

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Session 4: Data Quality metrics for Real-World Data

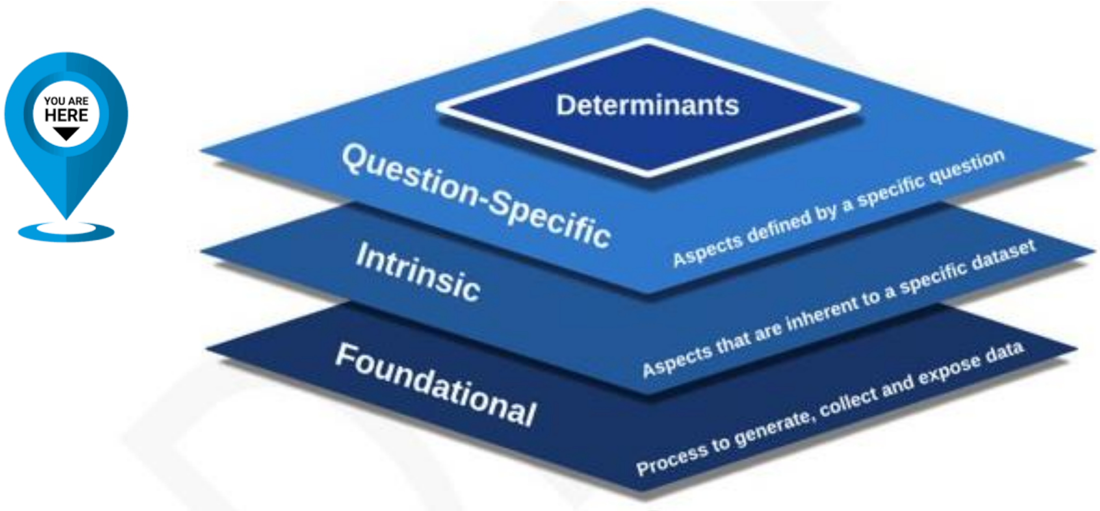
Multi-stakeholder workshop on Real World Data (RWD) quality and experience in use of Real World Evidence (RWE) for regulatory decision-making – 26/06/2023

Presented by Katerina-Christina Deli (EMA)
Moderated by Jesper Kjær (DKMA)

Proposed data quality metrics for RWD and challenges

Presented by Katerina-Christina Deli (EMA)

In this session, we focus on intrinsic determinants of quality



What quality aspects can be assessed or observed about a dataset, independently from the way this was generated

"Metrics"

Metrics for Real-world Data

- The general DQF introduced the concept of "intrinsic determinants" as all quality aspects that can be observed in a dataset without neither knowing how this dataset originated, nor its intended use
- In the context of RWD, "intrinsic determinants" correspond to metrics applied to datasets
- Metrics are relevant for DQ in general but have a specific significance in RWD, as often the quality of secondary use of data cannot be controlled at source, but can only be assessed at a second stage

Metrics in RW-DQF

- The aim of this document is to provide a framework able to best describe the data quality of a data source, on the basis of:
  - The measured dimensions
  - The corresponding best-fit metric (considering also the complexity required to implement)
- Examples for relevant metrics are included
- The metrics presented in the RW-DQF are not meant to be exhaustive nor normative, as different extensive lists of metrics and checks have been proposed
  - The objective of this framework is to help aligning different metrics and to identify and address gaps in current practices
- **Metrics measure a "quality level" – but they do not characterise the "adequacy" of a given quality level to address a question**
  - This point is addressed by the question-specific determinants

Data quality dimensions the metrics aim to measure

RELIABILITY	EXTENSIVENESS	COHERENCE	TIMELINESS	RELEVANCE*
Is the data correct? Is it representing what is meant to represent?	How much data is there?	Can data be analysed as a whole?	Is data available at the right time?	Is this the kind of data I need?

*Relevance should be considered as the suitability to a research question

Categorisation of metrics

OBJECTIVE DATASET ASSESSMENT

Metrics regarding the dataset structure, for which no additional knowledge or information is required

PLAUSIBILITY CHECKS

Metrics assessing likelihood of data being accurate based on general knowledge about clinical data

CONFORMANCE CHECKS

Metrics assessing conformity to external standards dictating data structure or format

COMPARISON TO OTHER DATASETS

Metrics based on a comparison with an external dataset acting as a 'reference'

CHECKS ON DATASET DESCRIPTORS

Metrics on additional ('meta') data that may come with a dataset

A systematic view on RWD metrics

	RELIABILITY	EXTENSIVENESS	COHERENCE	TIMELINESS
OBJECTIVE DATASET ASSESSMENT	E.g. number of decimal points	E.g. completeness	E.g. potential duplicates	E.g. most recently recorded timestamps
PLAUSIBILITY CHECKS	E.g. values are within clinical range			
CONFORMANCE CHECKS			E.g. conformance to external standards	
COMPARISON TO OTHER DATASETS	E.g. similar densities of data values	E.g. similarity of patient characteristics		
CHECKS ON DATASET DESCRIPTORS	E.g. values derived from imputation	E.g. use of explicit negation	E.g. conformance to allowable ranges	

7 Proposed data quality metrics for RWD and challenges

Example metrics

Dimension	Metric category	Example metrics	Example
Reliability	Plausibility checks	% of records where logical constraints between values agree with expectations	X% of records of pregnancy were attributed to females
	Checks on dataset descriptors	% of variables/datasets that are based on imputation or derivation	End of treatment date is derived for X% of patients from treatment start date and treatment cycle length
Extensiveness	Comparison to other datasets	Deviation score between patient characteristics in data source and reference data source	Cohort of patients with chronic kidney disease in France has significantly older age and higher severity classification compared to national French population from national claims database
Coherence	Conformance check	For relevant variables, % of patient records where the precision of values is fitting a target standard	X% of records have HbA1c levels reported with one decimal digit
Timeliness	Objective dataset assessment	Average time of updates in a database	Timestamps indicate patient records are updated on average every 3 months (after hospital visit)

Maturity model – intrinsic determinants (metrics)

- **Maturity level 0:** Metrics may have to be estimated and self-reported by the data owner with approximate knowledge of general data trends ('qualitative assessment')
- **Maturity level 1:** Owner performs sampling and spot checking of records, with documented results ('quantitative assessment' for this and further maturity levels)
- **Maturity level 2:** Comprehensive test cases, results, including summary statistics and scores, available through a metadata catalogue for decision making of fitness-for-purpose
- **Maturity level 3:** Fully automated testing in data conformant with CDM, results and summary statistics auto disseminated to catalogue systems
- **Maturity level 4:** DQ edit checks available in capture system during data collection, correction propagated through the generation system

Points to consider (1/2)

- **When to apply metrics.** Metrics may be applied at different points in time of the lifecycle for RWD: they can serve for internal error detection and subsequent correction of data, and/or they can provide an intermediary or final quality control
- **Metrics may change for a subset of data of interest.** Some metrics may change (and need to be re-assessed) when a specific subset of a dataset of interest is identified (e.g., precision of age may change if only a paediatric population is considered)
- **Responsibility for generating metrics** Throughout an evidence generation process, not all actors have visibility on all data (some may only have access to downstream aggregated data). Different actors are responsible to maintain DQ assessment for the part they can control, in order to maintain an overall chain of evidence

Points to consider (2/2)

- **Data Source Catalogue to capture data quality metrics.** Integration of a comprehensive data quality module within the data source catalogue would provide the linkage with the studies run on a data source and provide transparency
- **Metrics benefit from automation and standardisation.** The use of standard CDM and automated set of tests increases the feasibility and utility of metrics
- **Large number of tests can be difficult to interpret.** The experience with automated testing showed that it can be challenging to interpret a large number of data quality checks
- **Maturity level could be useful in guiding DQ assessment**. The maturity level could help orientate the data user on the overall level of data quality of a given data source

Proposed discussion points

- Automated metrics can provide an overwhelming amount of data. How can these be summarised in a synthetic way?
  - Are counts of tests fail/pass useful?
  - Should detailed values for key variables be reported?
- How much is it really feasible to have "gold standards" or general reference datasets to base quality assessment on? Is there a role for "silver standards" (e.g., comparisons among related datasets)?
- Should metrics be included that cannot be derived from a simple dataset (e.g., percentage imputed values). If so, how feasible is it at scale?