Data quality in fit-for-purpose assessments

Juan J. Abellan (EMA)

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

26 June 2023

Presenter name and surname, TDA

RWE to address knowledge gaps

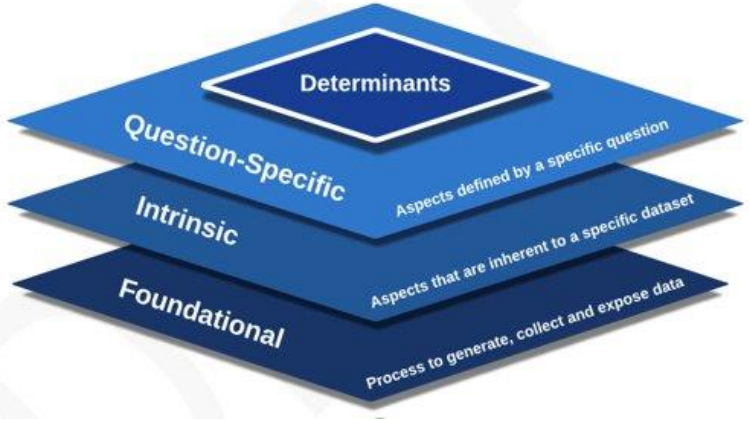
- More evidence-based regulatory decision making requires addressing knowledge gaps
- Knowledge gap -> Research question -> Study
- Which type of study?
  - RCTs preferred tool for efficacy
  - Use of non-interventional studies helpful for research questions on medicines use, safety and effectiveness
- RWD sources are an opportunity, with challenges, for evidence generation

How much can we trust the results of a RWD study?

- Decision makers (end user of study results) may not be familiar with RWD sources included in a study
- How to ensure the RWD sources are fit-for-purpose for a research question?
- Level of trust on the results of a RWD study depends on:
  - Methodological aspects: study design and methods
  - **Quality of the data used in the study**
- Which aspects of data quality can help provide confidence in the results?
- All data quality dimensions are important, but probably two are essential: reliability and relevance

Reliability

Foundational and *Intrinsic* aspects of data quality are revisited through the lens of the research question



- Foundational:
  - What is the primary purpose of data collection in the data source?
  - Are data representative for inferential purposes?
  - Peer-review publications of the data source? Participation in studies?
- Intrinsic:
  - Good values in metrics capturing intrinsic aspects will increase confidence in the data source

Relevance

Description of key data elements available in the data source that justify its inclusion in the study

- Are data needed for my study collected systematically? Data granularity?
  - Data elements needed to identify the population as per inclusion/exclusion criteria, exposure, outcome, adjustment variables
    - Development of computational phenotypes
- Can patients be followed-up retrospectively using records in the data source?
- External validation
  - Are there studies showing good ability of the data source to capture the data relevant for my study (e.g. are patients with a COPD diagnosis code actually COPD patients?)

A framework for study-specific data quality assessment?

- One potential idea inspired in Gatto et al (2022) *The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework* (SPFID)
  - [However, not for ranking comparatively the fitness for purpose of RWD sources]

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, incl. justification where relevant	Suggested extensiveness assessment	Suggested assessment of other quality dimensions	Suggested substantiation by documentation
E.g., Study population	*Inclusion criteria* *Exclusion criteria*	*Data elements required*	*Low/Medium/ High*	*Completeness metrics*	*Precision metrics*	*As relevant*

Illustrative example

What is the mortality rate of severe asthma patients in real world based on EHR?

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element	Suggested extensiveness assessment	Suggested assessment of other quality dimensions	Suggested substantiation by documentation
Study population	Confirmed diagnosis of asthma	Physician diagnosis (SNOMED code or equivalent)	High	% of patients have a diagnosis	% of diagnoses were mapped to SNOMED	Documentation on mapping native coding system to SNOMED
	At least 2 asthma exacerbations in the year prior to index date	Linkage to hospital data	High	% of patients have linked hospital data	Distribution of number of exacerbations is clinically sensible	Documentation on type of linkage to hospital data
	Use of medium-high dose of ICS in the 3 months prior to index date	Medicine prescriptions and dosage collected systematically	Medium	% prescriptions include dose/strength [to ensure it can capture high-dose ICS prescription]	Coherence checks: e.g. patients with no ICS but with a few exacerbations and taking LABA/LAMA	Documentation on how prescriptions are captured in the data source
Key outcomes	Overall Survival	Date of death	High	% of patients with known death have date of death	Statistical checks for reliability of linkage to death registry passed for 100% of patients	Linkage process documentation

Questions to the audience

How can we summarise in a meaningful way the overall fit-for-purpose assessment of a data source for a research question?

Any questions?

Further information

[Insert relevant information sources or contact details as applicable.]

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Telephone +31 (0)88 781 6000
Send us a question Go to www.ema.europa.eu/contact

Follow us on **@EMA_News**