acdm

association for **clinical data management**

Highlights from review of EMA reflection paper on use of real-world data in non-interventional studies to generate real-world evidence

Authors: Linda Shostak, Evaldas Lebedys & Silvia Perez Torres

Highlights from review of EMA reflection paper on use of real-world data in non-interventional studies to generate real-world evidence

In March 2025 EMA published their Reflection paper on use of real-world data in non-interventional studies to generate real-world evidence for regulatory purposes. The reflection paper addresses one of the 4 steps in the EMA's roadmap for regulatory guidance on RWD. (See the EMA's Journey towards a roadmap for regulatory guidance on real-world evidence for additional details).

The scope of the paper includes the design, conduct, and analysis of non-interventional studies (NIS) using Real Word Data (RWD). The paper does not address the use of RWD in clinical trials (CT).

The paper includes recommendations for activities that are not the direct or sole responsibility of data managers, such as performing feasibility assessments. The recommendation topics include:

- Legal obligations and regulatory requirements
- Study Design
- Governance and transparency
- Data quality
- Statistical analysis

The ACDM DMEG regulatory considerations reviewed the entire reflection paper and highlighted the following areas of interest and impact for data managers, with specific emphasis on **Section 7 Data quality**.

(Note: Definitions of key terms are provided in the **glossary**)

ACDM DMEG regulatory considerations highlights:

Sections 1 Introduction and 2 Scope

The recommendations relate to both primary and secondary data use unless otherwise specified. EMA notes that recommendations may differ for NIS with descriptive objectives versus NIS with causal objectives. These differences are also specified in the paper.

Section 3 Legal obligations and regulatory requirements

Covers many activities out of the scope of data management; however, data managers may be involved in the selection of RWD sources. In this instance, data managers should note that all legal and regulatory obligations applicable to NIS also apply when utilizing RWD. The paper lists key legal and regulatory references that will likely be applicable when using RWD.

Highlights from review of EMA reflection paper on use of real-world data in non-interventional studies to generate real-world evidence

Section 4 Study Design

When designing the study, we must ensure that the research question drives the selection of the RWD source. EMA recommends the sponsor conduct a feasibility study prior to developing the protocol in order to confirm the RWD will support the research question. While data managers may not be responsible for defining the research question or study design, they will likely participate in the feasibility study and the selection of the data source(s). Data managers should be aware that if the study uses an existing RWD source, the assessment should be conducted in collaboration with the data source holder.

Data managers should also be aware that studies with descriptive objectives and causal objectives may require different feasibility criteria. The paper provides examples of different feasibility considerations for each design type. The paper notes that the target trial emulation (TTE) framework can be employed to design and analyze the NIS in order to reduce or avoid bias.

Section 5 Bias, confounding and effect modification

EMA emphasizes it is important to take steps to reduce bias, confounding, and effect modification when designing the study, selecting the data, and processing the data. Some recommendations that impact data managers include:

- Evaluating the data source to confirm it contains sufficient detail on exposures, outcomes, confounders and eligibility criteria
- Identifying steps for data collection and extraction during the study design phase
- Identifying and evaluating previously performed validations
- Verify the availability of relevant data in the data source(s)
- Minimizing the potential for missing data

Section 6 Governance and Transparency

Data managers should be aware that primary and secondary data use may require specific approvals or consent processes. Data managers may have limited responsibility for ensuring transparency which supports the sharing of study information and the evaluation, interpretation, and reproducibility of the results.

Section 7 Data Quality

Data quality has the most impact on data managers as it provides recommendations and requirements for ensuring the reliability and relevance of the data source(s) in relation to the specific study objectives.

Highlights from review of EMA reflection paper on use of real-world data in non-interventional studies to generate real-world evidence

EMA specifies that the study protocol must address the reliability of the data and its relevance to specific study objectives. The HMA-EMA Data Quality Framework for EU medicines regulation: application to Real-World Data is cited as providing guidance for assessing the RWD quality.

Regarding reliability, EMA notes that the method and results of the reliability evaluation must be documented in the study protocol or a linked document. Information regarding data quality management (data cleaning, extraction, coding, etc.) must also be available. Any validation of the data source(s) should be detailed in the study protocol. The DMEG notes that while the EMA recommendation does not reference risk-based quality management, in our opinion data managers may want to use a risk assessment approach when evaluating reliability.

EMA also points out that due to the nature of RWD it is likely that some data quality issues cannot be resolved and that this will result in some level of uncertainty regarding the data quality. These uncertainties and their possible impact should be documented in the feasibility assessment and study protocol. The DMEG recommends that any issues which were not anticipated and described in the feasibility assessment and study protocol, should be assessed and impact on data quality and reliability of results should be documented and reported to ensure transparency.

In regard to relevance, the EMA reminds us that relevance will be specific to the research question, or study specific. Relevance should be explored as part of the feasibility assessment and discussed in the study protocol. EMA notes that the HARmonized Protocol Template to Enhance Reproducibility can provide a model for presenting this information.

If AI technologies are used to collect or process the RWD, this must be discussed in the protocol including the methods used to assess AI performance, risk of bias, and the potential impact on the study results.

When discussing recommendations for multi-database studies, it is important to note that EMA uses this term to refer to multiple RWD data sources. EMA notes that the use of multiple data sources should not reduce overall quality; in fact, the reliability and relevance of each RWD source should be documented. EMA also emphasizes the importance of using a unique patient identifier to clearly link data from different sources such as genetic data with registry data. The protocol should describe the data elements and methods used to link the data.

EMA does not require use of a specific framework to determine fitness for use. Instead, EMA emphasizes that the sponsor utilizes a framework and makes the results of the evaluation available to regulators. **Section 8 Statistical analyses** focus on statistical concerns, but data managers should make note of Section 8.6 missing data which contains the following recommendations which are like those used in clinical trials:

- Describe the management of missing data in the protocol and Statistical Analysis Plan (SAP)
- Justify assumptions made regarding missing data

Highlights from review of EMA reflection paper on use of real-world data in non-interventional studies to generate real-world evidence

Conclusion

Overall, the scope and recommendations of the EMA paper align with approaches taken by the FDA, Health Canada and the MHRA in recent publications on use of RWD. Key definitions such as primary versus secondary data are consistent across all four regulatory authorities. There is also similar emphasis on evaluating data sources to ensure reliability and relevance of the data, as well as feasibility of use for the intended study objective(s). There is also significant consistency in expectations for transparency, governance, and compliance with appropriate data privacy rules.

Glossary

Causal objectives: a study designed to investigate the effect, causative or preventive, of an exposure in comparison to what has happened to the same individuals under non-exposure or another exposure.

Descriptive objectives: study designed to describe patient characteristics without regards to any causal hypothesis.

NIS: EMA defines a NIS as one that does not fulfill any of the conditions defining a clinical trial in Article 2.2.(2) of Regulation (EU) No 536/2014. Generally, an NIS is an observational study that does not dictate the use of an intervention.

Primary data: data collected from patients, caregivers, healthcare professionals or other persons involved in patient care specifically for the study in question.

Relevance: the extent to which a dataset presents the data elements useful to answer a given research question.

Reliability: the dimension that covers how closely the data reflects what they are directly measuring.

Real-word data (RWD): data that describes patient characteristics, including treatment and outcomes, in routine clinical practice.

Real-world evidence (RWE): evidence derived from the analysis of RWD.

Secondary data: use of existing data for a different purpose than the one for which it is originally collected.