EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

15 April 2024
EMA/CHMP/150527/2024
Committee for Human Medicine Products (CHMP)

Reflection paper on use of real-world data in non-interventional studies to generate real-world evidence

Draft

Draft agreed by Methodology Working Party (MWP)	October 2023
Adopted by CHMP PROM for release for consultation	15 April 2024
Start of public consultation	3 May 2024
End of consultation (deadline for comments)	31 August 2024

Comments should be provided using this EUSurvey form. For any technical issues, please contact the EUSurvey Support.

Keywords	Non-interventional study, real-world data, real-world evidence, feasibility assessment, bias, confounding, data quality

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

Reflection paper on use of real-world data in non-interventional studies to generate real-world evidence

Table of contents

1. Introduction

This reflection paper discusses methodological aspects of non-interventional studies (NIS) using real-world data (RWD) in order to generate real-world evidence (RWE) for regulatory purposes. A NIS is a clinical study that does not fulfil any of the conditions defining a clinical trial (CT) in Article 2.2(2) of Regulation (EU) No 536/2014 on clinical trials on medicinal products for human use, where a CT is defined as a clinical study where (a) the assignment of the subject to a particular therapeutic strategy is decided in advance and does not fall within normal clinical practice of the Member State concerned; (b) the decision to prescribe the investigational medicinal products is taken together with the decision to include the subject in the clinical study; and (c) diagnostic or monitoring procedures in addition to normal clinical practice are applied to the subjects. RWD are data that describe patient characteristics (including treatment utilisation and outcomes) in routine clinical practice. RWE is evidence derived from the analysis of RWD.

CTs are the main source of evidence to evaluate the benefits and risks of medicines in marketing authorisation procedures. As they generally use randomisation, blinding, and a controlled environment, they increase regulators' confidence in the reliability of the evidence submitted. NIS are often used in post-authorisation safety assessment. Their use for assessing medicines efficacy is hindered by methodological limitations. These include absence of randomisation, uncontrolled conditions, non-standardised treatments and uncertainties regarding data quality and completeness.

Healthcare data sources accessible for medicine evaluation have evolved over the last decade. The increasing ability to capture electronic healthcare data and data from registries is now providing new opportunities to use RWD and generate RWE that reflects clinical practice. Examples where NIS using RWD have supported regulatory assessment include:

- To perform post marketing monitoring, investigate safety concerns and evaluate the effectiveness of risk minimisation measures.
- To describe patterns of drug utilisation (e.g. indication, characteristics of drug users, incidence and prevalence of use, doses, duration, and switching patterns).
- To characterise disease epidemiology (incidence, prevalence, risk factors and progression).
- To validate outcome measures, e.g. through a comparison of surrogate and clinical outcomes of disease progression.
- To support the feasibility assessments and the planning of non-interventional post-authorisation safety (PASS), efficacy (PAES) and drug utilisation studies by measuring outcome incidence, treatment exposure, the duration of available follow-up and the sample size effect of different inclusion/exclusion criteria.
- To compare patient characteristics of the study population to those of the clinical practice population in the real-world.
- To understand the clinical context, by describing standards of care, variability in clinical practices and unmet medical needs.

Given the large amount of information that NIS using RWD can generate for regulatory purposes, it is important to understand their limitations as well as how some of these limitations could be overcome or mitigated to increase the reliability of the evidence. This reflection paper is therefore relevant to all stakeholders involved in the planning, conduct and analysis of NIS using RWD to generate RWE for regulatory purposes, including Marketing Authorisation Holders (MAHs) and Applicants, regulatory

authorities, HTA bodies, payers, academia, RWD holders and healthcare professionals' and patients' associations.

2. Scope

The scope of this reflection paper is the design, conduct and analysis of NIS using RWD to generate RWE for regulatory purposes. The use of RWD in the context of CTs, e.g. to provide an external control arm for a single arm trial or to serve as a data source to recruit participants for a CT, is out of scope of this document.

General methodological principles and approaches for the conduct of NIS are described in textbooks and scientific guidelines (1). This reflection paper focuses on methodological principles that are considered critical for the conduct and assessment of NIS using RWD and used for regulatory decision-making throughout a medicine's lifecycle.

A large variety of RWD can be used in NIS, such as data related to healthcare services utilisation, medical claims, prescribing and dispensing of medicinal products, socio-economic and lifestyle data, data from patient registries, data from healthcare professionals' and patients' surveys, data collected with wearable biometric devices and genetic data. A critical aspect when assessing the suitability of RWD for a regulatory purpose is the data quality, including data reliability and relevance as described in Chapter 6, and, depending on the research question, the extent to which RWD truly reflects routine clinical practice. In this context, the data quality frameworks discussed in this document should be considered.

RWD may originate from primary data collection, i.e. data collected specifically for the study in question, or secondary use of existing data sources. In both cases, attention should be paid to the possible selection mechanisms in the data collection, for example the inclusion of specific patients or the collection of specific clinical data. This aspect is important to be addressed in the RWD quality assessment.

This reflection paper makes a distinction between NIS having *descriptive objectives* and NIS having *causal objectives*. This distinction has important implications for the study design. A study with descriptive objectives is designed to describe patient characteristics without regards to any causal hypothesis but it may include a measure of association between the distribution of these characteristics and the categories of other variables. A study with causal objectives is designed to investigate the effect, causative or preventive, of an exposure in comparison to what would have happened to the same individuals under non-exposure or another exposure. In the context of this reflection paper, the exposure is generally a medical treatment, and the outcome of interest is generally a measure of its relative safety or effectiveness. Reference to studies with causal objectives in this document does not imply an expectation that specific analytical methods will be used.

3. Legal obligations and regulatory requirements

The relevance of including NIS using RWD in a regulatory procedure to support safety and/or efficacy should be discussed between the MAHs and Applicants and the regulators at an early stage during the development phase of the product. NIS using RWD may be proposed to fill knowledge gaps, but its relevance depends on the intended regulatory purpose within the context of a specific application. The regulatory assessment does not mandate a specific study design but requires that the evidence generated is sufficiently reliable to support the regulatory objective. The relevance of a NIS using RWD to generate RWE for a specific application can therefore only be determined on a case-by-case basis.

The legal obligations and regulatory requirements applicable to NIS should be followed. The following documents are particularly relevant in the context of this reflection paper.

Legal obligations

- Regulation (EU) No 536/2014 on clinical trials on medicinal products for human use.
- Regulation (EC) No 726/2004 laying down Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency.
- Commission Implementing Regulation (EU) No 520/2012 on the performance of pharmacovigilance activities.
- Directive 2001/83/EC on the Community code relating to medicinal products for human use.
- Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

Regulatory requirements

- Guideline on good pharmacovigilance practices. Module VIII -Post-authorisation safety studies.
- Guideline on good pharmacovigilance practices: Module XVI – Risk minimisation measures: selection of tools and effectiveness indicators.
- Scientific guidance on post-authorisation efficacy studies and Post-authorisation efficacy studies: Questions and Answers.
- Guideline on registry-based studies.
- Good Practice Guide for the use of the Metadata Catalogue of Real-World Data Sources V 1.0
- European Medicines Agency (EMA) and Heads of Medicines Agencies (HMA) Data Quality Framework for EU medicines regulation.
- ICH E9 (R1). Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical principles for Clinical Trial. 2019.

4. Study design

4.1. General Considerations

The design of the NIS should be primarily driven by the need to obtain reliable evidence regarding the research question. It is the MAH's and Applicant's responsibility to justify that the use of RWD is appropriate and feasible to meet the pre-defined study objectives. Methodological standards for NIS and use of RWD, such as those described in the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology (1), should be applied. The ENCePP Checklist for Study Protocols (2) should be included as an Annex to the protocol of non-interventional PASS submitted to a regulatory authority, as a check that important study design components have been addressed.

For any type of NIS, it is essential that the research question is expressed with sufficient detail and attention to the regulatory question targeted. The specific aim of the study forms the basis for the selection of data source(s), study design, and analysis approach. It also allows the critical discussion regarding the feasibility of the study to meet regulatory objectives.

4.2. Feasibility assessment

A feasibility assessment is recommended prior to writing the study protocol to guide its development and facilitate early discussions with regulatory authorities. It is a preparatory document for the protocol, and it does not replace the submission of the protocol and statistical analysis. It should include:

- an evaluation of the reliability and relevance of the proposed RWD source(s) in order to meet the study objective (see Chapter 6); this discussion should address, as appropriate, the feasibility of the planned study design based on the data source, the choice of the study population (inclusion/exclusion criteria) and the availability of data on exposures, endpoints, and covariates;
- a discussion of how the proposed inclusion and exclusion criteria will impact sample size and whether the available sample size may provide sufficient precision of key estimates, overall and in subgroups (e.g. country-specific estimates in a descriptive drug utilisation study).
- as applicable, data on the incidence of study outcomes in the source population for the study, as they will inform on precision but also on anticipated timelines for the study to meet the regulatory objective;
- a discussion on how the above evaluation may impact the milestones for the study;
- as applicable, options to increase feasibility.

When an existing RWD source is proposed to be used, the feasibility assessment should be performed in collaboration with the data source holder to ensure data availability and set-up realistic timelines for the completion of the study. If relevant, exploratory analyses should be conducted to document the study feasibility.

The format of the feasibility analysis recommended in the Guideline on registry-based studies can be considered and adapted.

4.3. Studies with descriptive objectives

Studies with descriptive objectives may be conducted for different purposes, including risk assessment or prognosis, diagnosis, and service evaluation. They generally aim to observe and accurately measure patient characteristics at a single time point or over time.

Depending on the research question, it may be essential that the study population is representative of the real-world target population. When some of the characteristics studied may be influenced by the setting in which they are observed, the study should pay attention to the conditions that may influence the results of the study and this information should be addressed in the study design and analysis to help understand their impact, e.g. through sensitivity analyses. The aspects to be considered may include:

- The healthcare setting where the RWD has been collected, e.g. primary care, specialist care, hospital care, disease registries, claims data, longitudinal drug prescription, dispensing or other drug utilisation data, which may result in selection mechanisms that can influence the study feasibility and results.
- The healthcare system of the country(-ies) where the RWD are collected, which may influence the availability and accessibility of exposure and outcome data, for example data related to specialist care in hospitals, and the possible duplication of data if patients may consult different general practitioners.

- Regional differences in clinical practice and healthcare systems management, e.g. diagnostic criteria, prescribing practices, prescribing formularies, coding practices or reimbursement policies.
- The specificities of coding terminologies for medicinal product exposure and clinical events, use of a common data model and how data quality is assessed and managed (e.g. data quality metrics, data quality controls, misclassification and missingness, benchmarking).

4.4. Studies with causal objectives

The causal interpretation of any treatment effect requires a comparator in order to isolate and quantify the effect. The aim of the study design in studies with causal objectives should be to achieve valid comparisons between exposure groups by dealing with the risk of selection bias, information bias and confounding.

The target trial emulation (TTE) framework should be considered as a strategy that uses existing tools and methods to formalise the design and analysis of NIS using RWD with causal objectives (3-5). The first step of this framework is to envision the key elements of a hypothetical (target) trial that would answer the research question, including its target population, eligibility criteria, assignment procedure, treatment conditions, outcome, causal contrasts (i.e. the estimand) and analysis plan. The second step is to design a NIS as close as possible to the hypothetical trial using epidemiological methods.

The TTE framework is considered useful for the following reasons:

- It provides a structured and coherent framework for the design of NIS with a causal objective, with similarities with CTs in terms of terminology, definition of the estimand and analytical approaches.
- It helps the investigators to consider potential bias and adequate methods to address them.
- Given the need to explicitly describe the design elements needed to emulate the CT, it provides a high level of transparency on the study design, the assumptions needed to emulate the trial and the definition of causal effects; this level of transparency may facilitate the evaluation and the replication of the study.
- The detailed and structured definition of inclusion/exclusion criteria and allocation of time periods defined by study entry and estimated start of treatment have been shown to reduce bias, such as the prevalent user bias (6) and the immortal-time bias (7).

Although the TTE framework can improve the internal validity of the NIS, the lack of randomisation and blinding still requires attention to the prevention and/or control of selection bias, information bias and confounding described in Chapter 4.5.

To increase the coherence between definitions of exposures, endpoints and intercurrent events, the estimand framework described in the ICH E9 (R1) Addendum on Estimands and Sensitivity Analysis in Clinical Trials should be considered in the design of the hypothetical trial, such as the attributes of the estimand, intercurrent events and strategies to manage ICEs. The main statistical analysis may also be aligned with the estimand framework, e.g. concerning the approach to missing data handling and sensitivity analyses.

4.5. Bias and confounding

The non-experimental nature of NIS may lead to bias that distorts the measure of association due to processes of selection (selection process in the overall dataset or systematic differences in the

selection and follow-up of study groups), misclassification (differences in the classification of individuals as regards their exposures, outcomes or covariates, including time-related classification and measurement errors in continuous variables), and confounding (difference in underlying disease risk between the treatment groups). Confounders may be unknown or inadequately measured. These sources of bias should be clearly identified at the design stage. They are not easily controlled in the analysis and the design should attempt to minimise their impact on the results.

4.5.1. Selection bias

For many of the data sources used, there are potential selection mechanisms that need to be addressed not only in the feasibility assessment but also later in the design of the study. Selection bias is difficult and often impossible to address in the analysis stage. At the design stage, the following steps are therefore recommended.

- The selection of specific data sources over others considered in the feasibility assessment should be justified on scientific grounds to prevent the risk of introducing bias.
- Any inclusion and exclusion criteria should be adequately defined and justified in the protocol with a description of the diagnostic and/or procedural codes and of any algorithm used to include patients in the analysis. This description should also address the completeness and possible misclassification of the data used to define the inclusion and exclusion criteria. If misclassification may be present, the possible impact on the study results should be addressed. It is generally recommended to use wide inclusion criteria when the data are extracted from the original RWD source, if applicable, and apply stricter inclusion/exclusion criteria in the analysis stage. This allows for stratified analyses and sensitivity analyses that can inform the interpretation of results.
- The definition of inclusion and exclusion criteria for the study population should consider implicit selection criteria resulting from the method used to identify the study population and define exposure categories, as these selection criteria may not be balanced between these categories. For example, differences between health seeking behaviours in vaccinated and non-vaccinated persons identified through the healthcare system may be related to their socio-economic status, their probability of being vaccinated and their probability of presenting with the outcome.
- Depending on the research question, a new (incident) drug user design should be considered instead of including both prevalent and incident drug users. Prevalent drug users are patients. already taking the study treatment or a similar treatment before the start of follow-up. Including such patients can cause selection bias because patients who experience the outcome of interest early during treatment become underrepresented (i.e. depletion of susceptibles). Bias can also arise if exposure and confounding factors are time-dependent or if variables impacting medicine prescription at study entry (e.g. disease severity) are influenced by use of previous medicine that is part of the study exposure. The use of prevalent new user cohorts can also be used in some situations, e.g. to allow inclusion of initiators of the new drug who were previously on an older comparator (8).
- Comparisons of study populations from different RWD sources or different time periods may introduce bias. The variables influencing the inclusion of individuals in these study populations may vary across time and settings and may not be known or measured, hence they may affect the exposure status and/or the study outcome. This design therefore needs to be justified and the likelihood of bias and confounding should be recognised.

4.5.2. Information bias

Information bias, or misclassification bias, may arise when key study variables (exposure, outcome, or confounders) are inaccurately measured or classified. Misclassification can arise at many different steps of data collection and extraction: diagnosis, coding, recording, data transformation, data aggregation, summarisation, and analysis. The following is recommended:

- The different steps of data collection or extraction applied in the RWD source should be identified at the stage of study design. Ideally, these steps should be verified in order to evaluate if the data source(s) contain(s) enough details on exposures (e.g. dose, duration, time period and indication) and outcomes (e.g. diagnostic code, disease severity and date of occurrence) to correctly classify the patients.
- Any validation study previously performed should be identified and evaluated. A new validation study may be proposed as part of the feasibility analysis.
- Misclassification is traditionally categorised as differential or non-differential. It is often stated that differential misclassification can lead to biases in any direction, whilst non-differential misclassification typically drives the association between the exposure and the outcome towards the null value. This is, however, not always the case. From a regulatory perspective, non-differential misclassification is often presented as being preferable for superiority efficacy studies as it is conservative. At the design stage this reasoning should be avoided, as the assumption that misclassification will be non-differential is difficult to verify or requires additional analyses that may not be done. In studies with causal objectives, the impact of misclassification also largely depends on study objectives that may co-exist in a same study, e.g. non-inferiority or superiority, efficacy, or safety. It is therefore more important to identify potential misclassification of exposures, outcomes, and relevant covariates in order to minimise and measure it during the study if possible.

4.5.3. Time-related bias

Time-related bias, including immortal time bias (10), is a misclassification bias that may occur in cohort studies where the exposure status may change over time and the allocation of time periods of observations to the non-exposed/exposed person-time is incorrect. It is therefore recommended:

- to define at the design stage the timing of study entry, start of treatment and changes of patient exposure or outcome status, and plan the data collection or extraction of important dates
- to consider the appropriate assumptions concerning risk windows and ensure that the available length of follow-up is sufficient to include the relevant risk windows
- to plan sensitivity analyses to evaluate the impact of these assumptions
- to include in the study protocol graphical representations of the study design and study diagrams (11)
- to consider the target trial emulation framework to help mitigate misclassification bias (see Section 4.4).

4.5.4. Confounding

Confounding should be addressed already at the design stage. The following is recommended:

- Potential confounders (risk factors for the outcome of interest) should be identified from disease knowledge or previous studies to plan the data collection or extraction for the variables to be accounted for; some confounders such as age, sex, socio-economic status, and geographic location are common to many studies. Any potential confounders should be identified irrespective of availability of measured confounders in the available RWD. It is particularly important to identify potentially important unmeasured confounders.

- The analytical methods to address potential confounding should be pre-specified in the protocol or analysis plan.

- For studies with causal objectives, use of an active comparator should be preferred to a non-pharmacological comparator or a non-user comparator to increase the similarity in measured patient characteristics between treatment groups and reduce potential confounding by indication, disease severity or unmeasured variables(9). Its use is optimal in the context of the new user design whereby comparison is between patients with the same indication initiating different treatments. Depending on the study objectives, a justification should be provided if the use of an active comparator was considered not appropriate or not feasible.

- For studies with causal objectives, inclusion of a control exposure and/or control outcome should be considered to help in the interpretation and appraisal of results. A negative control exposure (exposure known not to be causally associated with the outcome of interest) or negative control outcome (outcome known not to be causally associated with the exposure of interest) can help assess the presence of residual confounding by revealing an association where none is expected. A positive control exposure (exposure known to be causally associated with the outcome of interest) or positive control outcome (outcome known to be causally associated with the exposure of interest) can help identify confounding towards the null value by revealing a reduced or null association where a clear departure from the null is expected. The choices between a negative and a positive control, and between a control exposure and a control outcome depend primarily on the study objective, taking also into account the available data and the clinical context. Both negative and positive controls could be used.

4.6. Effect modification

- In studies with causal objectives, any potential effect modification of concern should be addressed at the design phase in order to verify the availability of relevant data in the data sources planned to be used, ensure that the sample size will be sufficient to allow appropriate analyses to characterise effect modification, and pre-specify the analytical approach in the statistical analysis plan.

- In case the study population characteristics may influence the treatment effect estimates and affect the generalisability of the study results to a defined target population, a comparison should be made between the cohort's key characteristics and those of the target population using other information (e.g. published research or national statistics). The effects should be measured across relevant characteristics through sensitivity analyses. In some studies, such as registry-based studies, it may be possible to compare the characteristics of patients included and excluded in the study, such as age, sex, socio-economic status, disease severity and medication use.

5. Governance and transparency

5.1. Governance

- The study governance should follow the principles described in the ENCePP Code of Conduct (12), and for vaccines the ADVANCE Code of Conduct (13).
- Knowledge of the governance applied to the RWD used in a study (data source, raw data, and results) helps understand any restrictions related to the conditions to access and publish the data, including constraints that may affect the availability of some data. Governance principles applied to the data source should therefore be adequately described and disclosed to regulators.

5.2. Transparency

Transparency is essential to share study information and support the evaluation, interpretation, and reproducibility of results. The main tools for the transparency of NIS are the EMA-HMA Catalogue of data sources (described in the Good Practice Guide for the use of the Metadata Catalogue of Real-World Data Sources) and the EMA-HMA Catalogue of Non-interventional Studies (14).

MAHs, Applicants and concerned stakeholders should:

- register in the EMA-HMA Catalogue of Non-interventional Studies all NIS performed by the MAH or Applicant, together with the study protocol and study report;
- register in the EMA-HMA Catalogue of data sources the data sources used in NIS submitted to regulators; if the data source is already registered, it would be appropriate to update this information if the last update was performed more than 12 months ago; this provision may be included in the contractual agreement between the MAH or Applicant and the data owner, as relevant;
- make publicly available the codes used for the creation of the analytical data set and the programming code for the statistical analyses.

6. Data quality

RWD are often used in NIS without the ability to influence the way they were collected, coded, and recorded. For this reason, it is essential to adequately characterise the dataset and assess the quality of the RWD that will be used for a specific study and present it in the feasibility analysis. At least two key dimensions of data quality should be addressed in the study protocol: reliability, which is the property of the data irrespective of their use in any specific study, and relevance, which should be evaluated in relation to specific study objectives and data needs (15).

6.1. Reliability

Reliability determines whether data represent the intended underlying medical concepts and are complete, trustworthy, and credible. Different dimensions of reliability can be evaluated and documented by using a data quality framework (see Chapter 6.5). The following is recommended:

- Adequate information on results of the evaluation of reliability should be made available in the study protocol as an Annex or a linked document. Reference can be made to the description of the RWD source, its data elements and the quantitative information included in the HMA-EMA Catalogue of data sources.

- Information regarding the standard data quality management applied to the RWD source should also be made available, such as the steps of data cleaning, extraction and transformation, the data quality checks applied to detect logical inconsistencies and erroneous, missing or out-of-range values, and the remedial actions that are taken at the level of the data source.
- Any validation study performed on the data source should be highlighted in the study protocol.
- It is expected that some data quality issues inherent to the data source deemed appropriate for the study objective are difficult or impossible to resolve, and therefore that uncertainties on some data quality aspects will remain. These uncertainties and their possible impact on the study results should be clearly identified in the feasibility analysis and the study protocol.

6.2. Relevance

Relevance determines, for a given research question, whether key data elements (exposure, outcomes, covariates) are available, the size of the population is adequate, the population is representative and of sufficient coverage of the target population for the study objective, and the study design is appropriate to fully answer the research question. The following is recommended:

- Relevance is study-specific and should be discussed in the feasibility analysis and the study protocol. The HARPER protocol template (16) provides a format for the presentation of information on the study design, exposure, outcome, and covariates, and it should be considered as a complement to the applicable protocol formats.
- Depending on the research question, summary statistics of important variables (e.g., age, sex, disease severity, medication use) may be presented in comparison to RWD from the literature or other sources in order to understand the value of the RWD source to fulfil the study objectives.
- The evaluation of the relevance of the proposed RWD source and study design should lead to a justification in the feasibility analysis and study protocol of why they are deemed fit-for-purpose to answer the research question.

6.3. Multi-database studies

The number of RWD sources available for secondary use is increasing, and this increase is associated with an increasing number of multi-database studies (17). If several data sources are used in a study, the following is recommended:

- The number of data sources and the associated increase in sample size should not reduce the quality requirements, and information on the reliability and relevance of each of the RWD source should be presented.
- Whilst consistency of conclusions across different sources may provide stronger confidence in the evidence, any heterogeneity between results generated from different RWD sources based on a same protocol is a source of uncertainties and needs to be addressed in the final study report. Qualitative and quantitative assessments of the sources of heterogeneity should be provided to better understand potential differences in the results across data sources (see Section 7.6).

6.4. Data linkage

Data sources may need to be linked to combine individual-level RWD on exposure, outcomes, and covariates from different sources, such as genetic data, mother-child data or data from different registries using a unique patient identifier. For these studies, the study protocol should describe:

- The data elements used to link the data.
- The linkage methodology, including a description of how the performance of the matching will be measured in case of probabilistic matching; the impact of an imperfect matching on the study results needs to be evaluated and discussed in relation to possible bias.

6.5. Data quality frameworks

Several data quality frameworks provide a set of characteristics determining the fitness-for-use of data (18). While they differ by the number of dimensions and the names given to these dimensions, they do not diverge substantially. It is therefore recommended:

- To use a data quality framework appropriate to the data source, such as the tool developed by the Observational Health Data Science and Informatics (OHDSI) network for databases conforming to the Observational Medical Outcomes Partnership (OMOP) common data model (19) and make the results available to regulators
- To develop expertise for the implementation, analysis, and interpretation of at least one data quality framework
- To follow the HMA-EMA Data Quality Framework for EU medicines regulation document (20).

7. Statistical analyses

The statistical analyses should be performed according to a pre-defined statistical analysis plan and this plan should be developed before the preparation of the analysis dataset. The following aspects may deserve additional attention.

7.1. Hypothesis testing, estimation, and precision

Hypothesis testing is often the focus of confirmatory RCTs as it serves to decide whether an effect has been demonstrated at a pre-specified level. For NIS, use of estimates quantifying the magnitude of the effect and of confidence intervals describing the precision of these estimates, both overall and in important subgroups, is essential to support decision making derived from the data (21).

For NIS based on large RWD sources, statistical analyses may produce statistically significant results and narrow confidence intervals that may not be clinically relevant and may be subject to other types of bias. An integrated evaluation is therefore essential for any conclusion based on results, i.e. to estimate the magnitude of the effect and its clinical relevance, to provide an appropriate description of the precision of estimates with confidence intervals, and to supplement this estimation with an assessment of the impact of selection bias, misclassification bias and confounding.

7.2. Time-dependent analyses

Longitudinal follow-up in cohort studies require accurate accounting of time periods in each exposure category from study entry until end of follow-up.

- A time-dependent analysis should be planned when appropriate to the research question in the cohort studies where events occur at different time points to account for time-dependent exposure and confounding.
- Treatment switches is a common problem in longitudinal pharmacoepidemiological studies. The methods used to handle treatment switch is dependent of the specific outcome studied. Methods to handle treatment switches should be pre-planned.

7.3. Stratified analyses

Stratified analyses may provide further insights into the results and can fulfil several objectives: to provide results of the analysis in sub-groups of interest as part of the research question, to measure an effect estimate within relevant categories of a confounder, and to perform analyses in sub-groups defined by potential effect modifiers.

- Stratified analyses should be pre-specified and may be planned as sensitivity analyses.
- In multi-database studies, stratified analyses may also help assess the robustness of the results across important subgroups/datasets; they should be performed by country and setting (e.g. primary care or hospital setting) in addition, for example, to age, gender and disease severity.
- Additional stratified analyses may be requested by regulatory assessors. If formulated at an early stage of the discussions with MAHs and Applicants, they should be addressed with the research question or, at the latest, in the development of the study protocol or statistical analysis plan.

7.4. Sensitivity analyses

Sensitivity analyses examine the potential impact of biases, of choices made in the study design or of assumptions made in the analysis.

- Sensitivity analyses should be pre-specified in the statistical analysis plan.
- In studies with causal objectives, the sensitivity analyses should consider the recommendations of the ICH E9(R1) Addendum on Estimands and Sensitivity Analysis in Clinical Trials.
- Sensitivity analyses may also address uncertainties related to assumptions in the main analysis. They should be discussed and aligned with regulators at an early stage if they affect the study design and the choice and definition of study variables.

7.5. Missing data

Missing data can lead to bias and confounding, and the following is recommended:

- to describe the management of missing data in the study protocol and the statistical analysis plan;
- to provide a thorough justification for the assumptions made regarding missing data and the appropriateness of the method chosen to handle them in the analysis;
- to consider sensitivity analyses to missing data assumptions made in the main analysis to understand their impact on the results;

- to follow the recommendations of the ICH E9(R1) Addendum on Estimands and Sensitivity Analysis in Clinical Trials.

7.6. Heterogeneity

In multi-database studies, different estimates may be found even when the same protocol is applied across all data sources (22). Many factors may explain heterogeneity in addition to random variation: different healthcare systems and reimbursement policies for healthcare, different populations covered by the data source, different practices for data collection, coding, and recording. It is recommended:

- to anticipate differences between data sources and study populations in the study protocol in light of the study objectives and to describe how they will be managed
- to discuss in the statistical analysis plan quantitative management of heterogeneity through appropriate methods for evidence synthesis (23)
- to consider use of forest plots and other visualisation methods of the results (point estimates and 95% confidence intervals) to help evaluate the results
- if Bayesian models for evidence synthesis are used, to present graphs of the full posterior distribution of key model parameters as well as relevant summaries of these posteriors (e.g. mean and/or median, 95% credible intervals and relevant posterior probability statements).

8. References

1. ENCePP. ENCePP Guide on Methodological Standards in Pharmacoepidemiology. (Revision 11), 2023.
2. ENCePP. ENCePP Checklist for Study Protocols (Revision 4), 2018.
3. ENCePP. Target trial emulation. ENCePP Guide on Methodological Standards in Pharmacoepidemiology. Chapter 4.4.2. Target trial emulation.
4. Hernán, M. A., & Robins, J. M. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8),758–64.
5. Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC). Chapters 3.6 (The target trial) and 22 (Target trial emulation).
6. Hernan MA, Alonso A, Logan R et al. Observational studies analysed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology 2008;19(6):766-79.
7. Dickerman BA, García-Albéniz X, Logan RW et al. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med. 2019;25(10):1601-6.
8. Webster-Clark M, Mavros P, Garry EM, et al. Alternative analytic and matching approaches for the prevalent new-user design: A simulation study. Pharmacoepidemiol Drug Saf. 2022;31(7):796-803.
9. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. Curr Epidemiol Rep. 2015;2(4):221-8.
10. Suissa S. Immortal time bias in pharmaco-epidemiology. Am J Epidemiol. 2008;167(4):492-9.
11. Schneeweiss S, Rassen JA, Brown JS et al. Graphical Depiction of Longitudinal Study Designs in Health Care Databases. Ann Intern Med. 2019;170(6), 398–406.
12. ENCePP. The ENCePP Code of Conduct. Rev. 4, 2018
13. Kurz X. et al. The ADVANCE Code of Conduct for collaborative vaccine studies. Vaccine 2017;35(15):1844-55.
14. EMA. The EMA/HMA catalogue of non-interventional studies (EU PAS Register). (2023)

15. FDA. Draft guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (2021)

16. Wang S. et al. HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR task force. Pharmacol Drug Saf. 2022.

17. ENCePP. Chapter 9.2.Models of studies using multiple data sources. ENCePP Guide on Methodological Standards in Pharmacoepidemiology (Revision 11), 2023.

18. ENCePP. Chapter 13.2 Data Quality Frameworks. ENCePP Guide on Methodological Research Standards in Pharmacoepidemiology, (Revision 11), 2023.

19. OHDSI. Data Quality Dashboard.

20. HMA-EMA. Data Quality Framework for EU medicines regulation [to be finalised]

21. Wasserstein RL, Lazar NA. The ASA Statement on *p*-Values: Context, Process, and Purpose. The American Statistician 2016;70(2):129-33

22. Russek M, Quinten C, de Jong VMT et al. Assessing heterogeneity of electronic health-care databases: A case study of background incidence rates of venous thromboembolism. Pharmacoepidemiol Drug Saf. 2023 ; 32(9):1032-48.

23. Franklin JM, Pawar A, Martin D et al. Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. Clin Pharmacol Ther. 2020;107(4), 817–26.