HMA
Heads of Medicines Agencies

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Data quality requirements and study design and analysis aspects

Olaf H. Klungel (Utrecht University)

Address relevance when designing 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 analysis**
  - Quality of the data used in the study
- Which aspects of data quality can help provide confidence in the results?
- Data quality dimension *relevance* should be evaluated for each study and addressed in study protocol.

ENCEPP Checklist for Study Protocols (Revision 4)

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☐	☐	☐	
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☐	☐	☐	
3.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☐	☐	☐	
3.4 Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☐	☐	☐	
3.5 Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☐	☐	☐	

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.1 Is the source population described?	☐	☐	☐	
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period	☐	☐	☐	
4.2.2 Age and sex	☐	☐	☐	
4.2.3 Country of origin	☐	☐	☐	
4.2.4 Disease/indication	☐	☐	☐	
4.2.5 Duration of follow-up	☐	☐	☐	

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☐	☐	☐	

Depends on data quality

Study Design choice independent of data source, but driven by research question

ENCEPP Checklist for Study Protocols (Revision 4)

Section 5: Exposure definition and measurement	Yes	No	N/A	Section Number
5.1 Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☐	☐	☐	
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☐	☐	☐	
5.3 Is exposure categorised according to time windows?	☐	☐	☐	
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☐	☐	☐	
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☐	
5.6 Is (are) (an) appropriate comparator(s) identified?	☐	☐	☐	

Section 6: Outcome definition and measurement	Yes	No	N/A	Section Number
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☐	☐	☐	
6.2 Does the protocol describe how the outcomes are defined and measured?	☐	☐	☐	
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☐	☐	☐	
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☐	

Section 7: Bias	Yes	No	N/A	Section Number
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	☐	☐	☐	
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☐	☐	☐	
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☐	☐	☐	

Section 8: Effect measure modification	Yes	No	N/A	Section Number
8.1 Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	☐	☐	☐	

Validity Exposure/Outcome measurement for specific question assessed according to Data Quality Framework

Needs to be considered for each study/question

ENCEPP Checklist for Study Protocols (Revision 4)

Section 10: Analysis plan	Yes	No	N/A	Section Number
10.1 Are the statistical methods and the reason for their choice described?	☐	☐	☐	
10.2 Is study size and/or statistical precision estimated?	☐	☐	☐	
10.3 Are descriptive analyses included?	☐	☐	☐	
10.4 Are stratified analyses included?	☐	☐	☐	
10.5 Does the plan describe methods for analytic control of confounding?	☐	☐	☐	
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☐	
10.7 Does the plan describe methods for handling missing data?	☐	☐	☐	
10.8 Are relevant sensitivity analyses described?	☐	☐	☐	

Section 9: Data sources	Yes	No	N/A	Section Number
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☐	☐	☐	
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☐	☐	☐	
9.1.3 Covariates and other characteristics?	☐	☐	☐	
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☐	☐	☐	
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☐	☐	☐	
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☐	☐	☐	
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☐	☐	☐	
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☐	☐	☐	
9.3.3 Covariates and other characteristics?	☐	☐	☐	
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	☐	

Description of data sources and Data Quality Metrics

These methods can be informed by Data Quality Metrics

Example: Characterising the risk of major bleeding in patients with Non-Valvular Atrial Fibrillation: non-interventional study of patients taking Direct Oral Anticoagulants in the EU (EUPAS16014)

Objective 1:The risk of major bleeding associated with use of DOACs when compared to other oral anticoagulants (OACs) in patients with non-valvular atrial fibrillation (NVAF) overall and in relevant clinical and demographical subgroups in a real-life setting.

Objective 2. The utilization of DOACs in the EU for treatment of NVAF, including the characterization of new DOAC users in NVAF patients.

Objective 3. Prescribers' compliance with recommendations included in sections 4.1, 4.3, 4.4, and 4.5 of the SmPC of each DOAC.

Van den Ham HA, et al. Pharmacoepidemiol Drug Saf 2021;30:1339-52.

Design for objective 1

Design: New User Active Comparator Cohort
Population: NVAF
Exposure: New use of DOAC (and individual DOACs)
Comparator: New use of VKA
Outcome: Major Bleeding (and specific type of bleedings)
Confounders: Risk Factors for outcome
Effect modifiers: Age, Sex
Follow-up time:0.8 – 2.7 yrs

Table 9.1. List of study designs to be conducted in each data source.

	Cohort (objective 1)	Descriptive (objective 2)	Descriptive (objective 3)
Mondriaan		X	X
Danish Registries	X	X	X*
Bavarian		X**	X**
AOK NORDWEST	X	X	
BIFAP	X	X	X
SIDIAP		X	X
CPRD	X	X	X
EGB		X	

Study population – Data quality for Indication assessment

1. A linked diagnosis of NVAF to the first prescriptions of the (D)OAC. If not possible, then:
2. A medical code for NVAF ±3 months around the index date in one of the following files.
    1. GP-record (CPRD, Bifap)
    2. Claims-record (AOK Nordwest)
3. A medical code for NVAF prior to index date + 3 months after the index date in case of Hospital-record (DK)

Exposure assessment – Data quality

- New use of DOAC/VKA based on claims, or prescriptions
  => at least 365 of no use before first prescription)
- Duration of exposure based on :
  1. Prescribed number of tablets and dosage
  2. Median time between prescriptions
  3. When only 1-3 prescriptions available, most frequently occurring estimated prescription duration

Outcome assessment – Data quality

Major bleeding according to definition International Society on Thrombosis and Haemostasis:

- haemorrhagic storke/intracranial bleeding, gastrointestinal bleeding, other extracranial or unclassified bleeding and traumatic intracranial bleeding
- for main analysis all bleeding events / irrespective of admission
- in CPRD additional analysis on hospitalized events only
- in BIFAP validation of GI bleeding and stroke.
- several posthoc sensitivity analysis with different outcome definitions.

Confounder assessment – Data quality

- Assessment at baseline: Sex, weight, BMI, smoking, alcohol use
- Assessment time-dependently: Age, comorbidities, co-medication
- Impact of missing data on BMI, Smoking, Alcohol use by multiple imputation in CPRD

Key messages

- Data quality dimension *relevance* should be evaluated for each study and addressed in study protocol.

- Study Design choice independent of data source, but driven by research question

- Variable definitions and analysis of a study may depend on data quality.

- Data quality framework and metrics can inform fit-for-purpose assessment of data source for specific 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**