APPLICATION OF THE EMA DQF FOR FEASIBILITY ASSESSMENT OF 10 USE CASES IN THE TARGET EU PROJECT

ROC19: TARGET-EU

COMPARATIVE EFFECTIVENESS AND SAFETY STUDIES USING THE TARGET TRIAL EMULATION AND ESTIMAND FRAMEWORKS

SPECIFIC CONTRACT 04 IMPLEMENTING FWC EMA/2020/46/TDA/L5.06

03 November 2025

PhD MD Mònica Sabaté

Monica.sabate@vallhebron.cat

Fundació Hospital Universitari Vall d'Hebron - Institut de Recerca (VHIR)

EU PE&PV RESEARCH NETWORK

INDEX

The presentation has been produced in part by making use of the results under the specific contract no.04 FWC EMA/2020/46/TDA/ L5.06 managed by Universiteit Utrecht, in its role as coordinator of the EU PE&PV Research Network consortium.
This presentation expresses the opinion of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties.
This study has been registered in the HMA-EMA Catalogue of Real-World Data under the EU PAS numbers: EUPAS1000000539, EUPAS1000000791

PURPOSE:

To characterise which **European RWD sources are suitable** to be used in the context of such studies as well as key aspects to pay attention to in **feasibility assessment of RWD sources** (including databases or registers) for emulating a hypothetical target trial for each of the 10 selected cases.

BASIS:

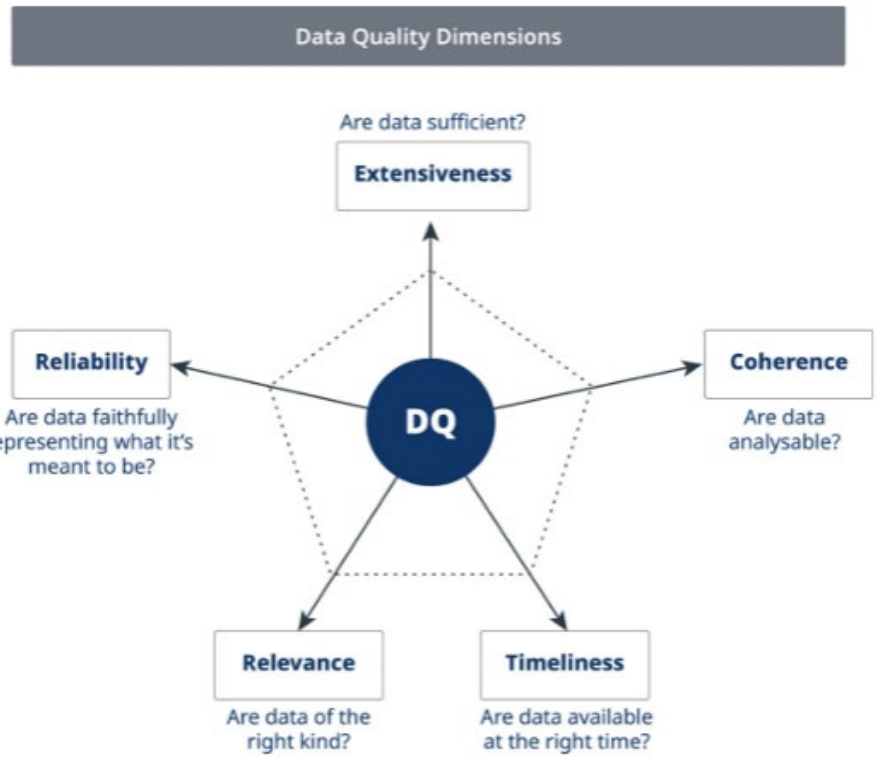
A process of **3 steps** (plus a pre-screening and a final assessment) was established to select fit-for-use European RWD sources for the **10 case** studies. This process was mainly inspired by the EMA data quality framework for medicines regulation applied to RWD.

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

4 November 2024
EMA/503781/2024
Committee for Medicinal Products for Human Use (CHMP)

Data Quality Framework for EU medicines regulation: application to Real-World Data

THE EMA DATA QUALITY FRAMEWORK



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Data source characterization checklist

- ☐ Rationale and scope for the RWD source creation
- ☐ Data collection process
- ☐ Selection of data sources and their onboarding
- ☐ Data management infrastructure
- ☐ Data management and governance
- ☐ Data manipulation steps
- ☐ Data augmentation steps
- ☐ Known quality issues and independent QA assessment
- ☐ RWD source data representation
- ☐ RWD source declared SLAs
- ☐ RWD source licensing and restrictions
- ☐ Feedback

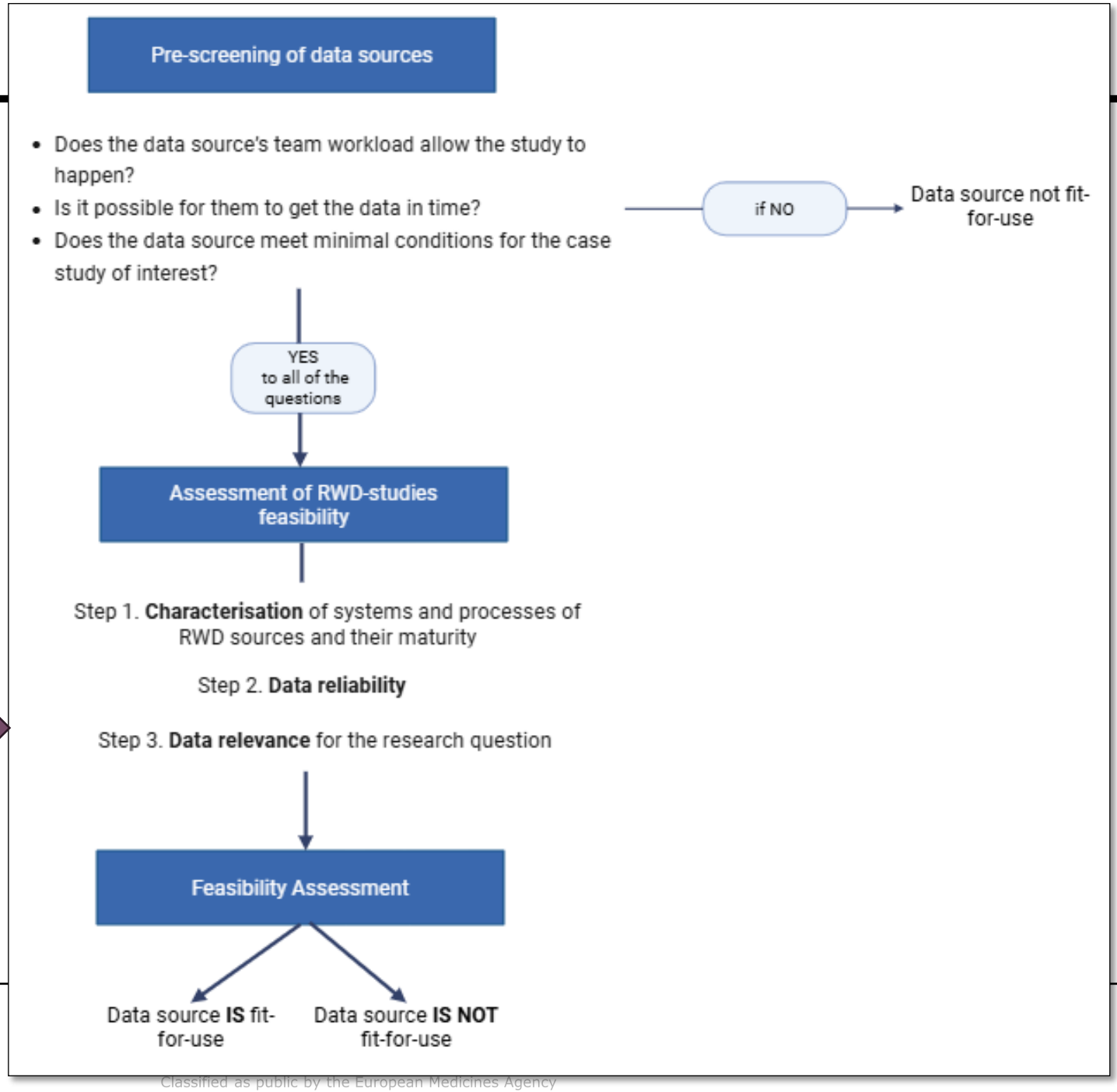
Figure 2 - RWD source characterisation checklist overview. SLA – service level agreement.

https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/data-quality-framework-eu-medicines-regulation_en.pdf
https://www.ema.europa.eu/en/documents/other/draft-data-quality-framework-eu-medicines-regulation-application-real-world-data_en.pdf

THE FEASIBILITY ASSESSMENT (FA) WORKFLOW

The FA workflow

Based on the Hypothetical target trial protocol



EU PE&PV RESEARCH NETWORK

IMPLEMENTING THE FEASIBILITY ASSESSMENT TO A CASE-STUDY

Case Study	Exposure	Comparator	Indication	Population	Outcome	Study type	Original study design	Data sources
1	SARSCoV-2 mRNA vaccine (BNT162b2)	No vaccination	NA	Adult/General population	COVID-19 infection	PAES	RCT	VID, CPRD
2	nivolumab plus ipilimumab combined with two cycles of chemotherapy	pembrolizumab combined with two cycles of chemotherapy	non-small-cell lung cancer	Adult/General population	Death due to any cause	PAES	RCT	NCR
3	Dapagliflozin	Placebo	Type II Diabetes Mellitus	Population with indication at high risk of atherosclerotic cardiovascular disease	MACE (cardiovascular death, myocardial infarction or stroke)	PAES	RCT	CPRD, BIFAP
4	Rivaroxaban	other oral anticoagulants	Atrial fibrillation	Elderly	Safety	PASS	Cohort	DNR, SIDIAP
5	Vilanterol/fluticasonfuroaat	Inhaled corticosteroids (ICS)/Long acting beta agonists (LABA)	Asthma	Adolescents	Pneumonia	PASS	Cohort	CPRD, Finnish registers
6	Sacubitril/valsartan	Angiotensine converting enzyme (ACE) inhibitors	Heart failure (HF)	Adult/General population	angioedema and other specific safety events	PASS	Chort	CPRD, PHARMO
7	Valproate (paternal exposure)	no valproate (paternal exposure)	Epilepsy/Bipolar disorder	Pregant women	Pregnancy outcomes/harmful risk to offspring	PASS	Cohort	VID
8	Nirsevimab	No immunization	prevention of lower respiratory tract disease caused by RSV	All infants	RSV-lower respiratory tract infection, RSV related hospitalization	PAES	Cohort	PEDIANET
9	Tolvaptan	Placebo	Autosomal Dominant Polycystic Kidney Disease	Adults >16y	hepatotoxicity, Basal cell carcinoma and Galucoma	PASS	RCT	CPRD
10	CapOx chemotherapy (capecitabine + oxaplatin) in combination with bevacizumab	CapOx chemotherapy (capecitabine + oxaplatin)	Metastatic colon cancer	Adult/General population	Overall survival and progression free survival	PAES	RCT	NCR

Feasibility Assessment: Step 1 – DATA SOURCE CHARACTERISATION

Item	Sub-item	Description
Data base identification	Country	the Netherlands
	Data Access Provider	Netherlands Comprehensive Cancer Organisation (IKNL)
	Organisation type	Quality institute for oncological and palliative research and practice. National, regional, or municipal public founding
Rationale and scope for the RWD source creation	Primary purpose for which data are collected	The main goal of the Netherlands Comprehensive Cancer Organisation (IKNL) is to reduce the impact of cancer, from the personal to the societal level. With the Netherlands Cancer Registry (NCR) as its core activity, IKNL enables health care professionals, researchers, policy makers and others to reflect on cancer and on palliative care.
	What triggers a record in the database	Event triggering registration of a person in the data source: having performed a biopsy through PALGA (the national pathology database) and having a cancer diagnosis in LBZ
Data collection or recording process	Key data elements captured (are they always recorded, are they optional, is there a planned coverage over time, ...)	Disease information, rare diseases, prescriptions of medicines, indication for use, procedures, clinical measurements, patinet-reported outcomes, unique identifier persons, diagnostic code, medicinal product information, quality of life measurements, sociodemographic information (age, gender). These are items grouped by: patient, tumor and treatment. [...] The day of death is known by linkage to CBS. TNM recorded
Data management and governance	Measures to prevent data alterations by unauthorised parties (cybersecurity)	IKNL has an IT department that is responsible for cyber security. There are also Information Security Officers that monitor this.
	Auditing and DQ improvement procedures in place	IKNL is NEN-7510 certified. Quarterly internal audits are performed, as well as regular external audits. There is also a working group responsible for DQ. They perform checks on the data. Researchers can also signal potential DQ issues.
The RWD source representation	Description of data model or models used (OMOP, FHIR, ...)	OMOP, ETL completed. IKNL uses its own data model for the NCR. [...] The data in the OMOP-CDM is updated a few times per year.

Feasibility Assessment: Step 2 – DATA RELIABILITY

Dimension	Sub-dimension	Metrics	Description
Timeliness	Currency	How often is the database updated (i.e., frequency of updates)	NCR updates are daily. However, data is registered 6-12 months after diagnosis so there is a lag there. Vital status is indeed checked once per year.
		The time gap between the latest available data and date when data is delivered to user (i.e., how up-to-date data are when it reach the user)	1 to 2 years, as data managers only have access to EHR once per patient to capture the primary treatment plan
		The time elapsed from when a user requests the data to when they actually receive it	~2 months
		❑ Median time (years) between first and last available records for unique individuals	0.7 years
Extensiveness	Coverage	Percentage of a target population present in a database	>95% coverage of the total population in The Netherlands. >= 18 y Population size: 3,677,269
	Completness	% of subjects in the data who had a prescription/dispensing with a recorded code for the medicine	99.27% of registered chemotherapies have an ATC code.
Reliability	Precision	Exposures codes precision level, including medicines and vaccines (e.g., active principle, therapeutic group, …)	Active principle (ATC level 5 codes)
		Precision of date of birth (e.g., day, month, year)	Day, month, year; but this is generally not shared in a data request, instead age at diagnosis is shared, for example
		Precision of date of death (e.g., day, month, year)	Day, month, year
	Traceability	Provenance of event and exposures records	EMR. Death is from CBS
Coherence	Semantic coherence	For EVENTS, codelists/data dictionaries being employed according to external standards	Indication: ICD-O; Procedures vocabulary: own vocabulary; Diagnosis/medical event vocabulary: ICD-O Stage: TNM
		For EXPOSURES, codelists/data dictionaries being employed according to external standards	Prescription: ATC level 5, own vocabulary

Feasibility Assessment: Step 3 – DATA RELEVANCE

Design elements	Operationalization of definitions	Data elements for valid capture of variables	Criticality of the quality of the element	Extensiveness assessment (if applicable)	Reliability assessment (if applicable)
Study population	Inclusion criteria				
	Histologically confirmed mCRC diagnosis in the last year prior to randomization	Pathology results	High	100% of individuals have available information	Once year all cancer dx are reviewed to identify cancer patients that did not have a biopsy and pathology finding.
	Age > or = 18y	Date of birth	High	100% of individuals have available information	As the format is not known, precision can not be evaluated. Low impact in the study?
	ECOG < or = 1	ECOG score	High	Recorded 15% missing	ECOG is dependent on the eye of the beholder.
	[...]				
	Life expectancy longer than 3 months	Pathology results	High	100% of individuals have available information	Once year all cancer dx are reviewed to identify cancer patients that did not have a biopsy and pathology finding.
	No prior systemic therapy for mCRC or previous treatment with oxaliplatin or bevacizumab	Medication code Date of prescription/dispensing	High	Only first line treatment	
	Adequate hematologic/ clotting, hepatic and renal function	Laboratory tests	High	Unknown missingness	
	Exclusion criteria				
	Pregnant or breastfeeding women	Pregnancy/breastfeeding status	High	Not registered for colon cancer patients	
Treatment/exposure	Bevacizumab (7.5 mg/kg IV, on day 1 of a 3-week cycle) + Capecitabine-Oxaliplatin regimen (IV/3wk).	Medication code Date of prescription/dispensing	High	Prescription first line treatment, dose not registered	

Feasibility Assessment: Step 3 – DATA RELEVANCE

Design elements	Operationalization of definitions	Data elements for valid capture of variables	Criticality of the quality of the element	Extensiveness assessment (if applicable)	Reliability assessment (if applicable)
Comparator group (if applicable)	Capecitabine-Oxaliplatin regimen (IV/3wk).	Medication code Date of prescription/dispensing	High	Prescription first line treatment, dose not registered	
Key endpoint(s)	Progression Free survival (PFS)	Date of treatment initiation Date of progression (imaging) Date of death	High	A date of death is recorded for 100% of individuals who are known to have died	Vital status checked once per year. As the date of death is registered it will be possible to calculate. PFS is not directly provided, although an algorithm using prognostic markers has been used in this database to predict PFS, being included in published papers.
Intercurrent events	Treatment discontinuation	Medication code Treatment end date	Low	100% of individuals have available information	
	[...]				
	Treatment switch	Medication code Date of prescription/dispensing Date of discontinuation Treatment duration	Low	Only first line treatment	As only first line treatment is recorded it won't be posible to diferentiate descontinuation than switch
	Local treatment	Date of procedure Procedure code	Low	Procedures available, cancer related surgery might be picked if a specific code is available	

EU PE&PV
RESEARCH NETWORK

Feasibility Assessment: FINAL FEASIBILITY ASSESSMENT

Case study	**10 (Capecitabine with Oxaliplatin (CapOx) plus Bevacizumab versus CapOx in patients with Metastatic Colorectal Cancer)**
RWD source	**NCR**
Sample size estimation form the hypotetical trial protocol	With an approximate estimated sample size of 440 individuals (based on a 1:1 ratio between treatment arms, comparing CAPOX plus bevacizumab versus CAPOX alone), and considering that the Netherlands Cancer Registry (NCR) recorded 22,192 patients aged ≥70 years with metastatic colon cancer between 2005 and 2020—of whom 23% received targeted therapy—the target sample size is anticipated to be reached.
Feasibility assessment (yes/yes, with limitations/no)	Yes, with limitations on a design element
Rationale for the feasibility assessmernt	Elements with high criticality are available and fairly reliable, with reservations regarding a design element endpoint. The time elapsed from when a user requests the data to when they actually receive it is 2 months. Data recency is ~12 months before extraction, reasonably enough for the research question. Sample size is achievable.
Limitations identified during the feasibility assessment and categorisation	·Potentially major: ○ Progression free survival (key endpoint) is not directly provided, although an algorithm using prognostic markers has been used in this database to predict PFS. ○ ECOG is 15% missing. ○ The median lenght of follow-up per patient is approximately 9 months. ·Minor: ○ Some cancer patients do not have a biopsy and pathology, but might be picked by diagnostic code. ○ Only prescription of first line of treatment is available, but cancer stage changes mean a new first treatment line is started. ○ Data is registered 6-12 months after diagnosis so there is a lag. ○ Imaging information to assess progression-free survival is not available, only death is captured. ○ Procedure codes are available, but cancer-related surgery might only be picked if a specific code is available.
Description of potential impact of the identified limitations on the study results	Although PFS is not directly available, a previously developed algorithm using prognostic markers has been applied in this database to estimate PFS. Missing ECOG data may prevent us from including certain subjects. Although the median follow-up time in the NCR is 9 months, this includes patients with all types of cancer with different survival durations. However, this variation is likely non-differential, meaning it is not expected to bias the results in favour of or against any particular cancer group. If the patients included in the study have a longer survival time, the registry will allow for the follow-up required by protocol.

GENERAL LEARNINGS AND CONSIDERATIONS

The tables:

- Novelty: they have been created for this project and were unfamiliar to some researchers, this made them difficult to navigate in some instances.

The content:

- Public information alone often lacked depth.
- Direct contact with DEAPs was crucial for revising feasibility tables and provides insights beyond public information.
- Required onboarding sessions and follow-ups due to complexity and time demands.

Roles:

- Step 3 required coordination with case-study leads for the protocols being ready.

THANK YOU FOR YOUR ATTENTION

UU, UMCU, LSHTM, VAC4EU, AEMPS, UNIVR, FISABIO, VHIR, IDIAPJGol, ARS, Santeon, PHARMO, SoSeTe/Pedianet, DUTh, UEF, RSU, NIPH, UCPH, UGENT, Teamit

PhD MD Mònica Sabaté
Monica.sabate@vallhebron.cat