REPORT January 2022

Generating Evidence from Real-World Data in Health Technology Assessment

Methodological guideline

Generalitat de Catalunya

Salut/ Agència de Qualitat i Avaluació Sanitàries de Catalunya

The Agency for Health Quality and Assessment of Catalonia (AQuAS) is a public body affiliated to the Catalan Health Ministry. Its mission is to generate relevant data and knowledge to guide decision-making in the area of public health and to promote the sustainability of the Catalan Health Care System. AQuAS is a founding member of the International Network of Agencies of Health Technology Assessment (INAHTA) and the International School on Research Impact Assessment (ISRIA). It is a corporative member of Health Technology Assessment International (HTAi), of the CIBER group (Networked Biomedical Research Centres) in Epidemiology and Public Health, of REDISSEC (the Spanish Research Network on Health Care in Chronic Diseases), and the Research Network on Chronicity, Primary Health Care and Health promotion (RICAPPS) and it is also an associated unit of the INGENIO research centre at CSIC-UPV. In 2019, AQuAS was awarded the Josep Trueta medal for services to healthcare by the Catalan government.

This document should be cited as follows: Vivanco-Hidalgo RM, Blanco-Silvente L. Generation of Evidence with Real-World Data in Health Technology Assessment. Methodological guideline. Barcelona: Agency for Health Quality and Assessment of Catalonia. Ministry of Health. Government of Catalonia; 2022.

For further information regarding this document, please contact:
Agency for Health Quality and Assessment of Catalonia.
Roc Boronat, 81-95 (second floor). 08005 Barcelona
Tel.: 93 551 3888 | Fax: 93 551 7510 | http://aquas.gencat.cat

Published by: Agency for Health Quality and Assessment of Catalonia
First edition: Barcelona, January 2022
Correction: Communication Area

Generating Evidence with Real-World Data in Health Technology Assessment. Methodological guideline

Authors

Rosa Maria Vivanco-Hidalgo
Health Technology Assessment Area, Agency for Health Quality and Assessment of Catalonia (AQuAS)

Lidia Blanco-Silvestre
Health Technology Assessment Area, Agency for Health Quality and Assessment of Catalonia (AQuAS)

Partners

Xabier Garcia-Albéniz
RTI Health Solutions

Acknowledgements

Ramon Roman
Programme for Data Analytics for Research and Innovation (PADRIS), Agency for Health Quality and Assessment of Catalonia (AQuAS)

Àgata Carreño
Health Technology Assessment Area, Agency for Health Quality and Assessment of Catalonia (AQuAS)

Conflict of interest statement

The authors declare that they have no conflicts of interest in relation to this document.

Contents

Summary

In certain situations, the process of Health Technology Assessment (HTA) is obliged to conclude that the evidence available is insufficient or of poor quality. After identifying the knowledge gaps, HTA reports provide recommendations on how to generate the evidence necessary to address the problem at hand.

The functions of the Agency for Health Quality and Assessment of Catalonia (AQuAS) include HTA and the management of access to the reuse of health data in order to generate new knowledge that is able to improve the health of the public.

Therefore, observational studies are needed to fill in the knowledge gaps detected. To do so, they should follow standardized processes that promote the development of studies of excellence.

The present document, for internal use, aims to guide the development of studies of this kind at AQuAS through the application of protocols, checklists and templates.

Introduction

Context

Health Technology Assessment (HTA) is a multidisciplinary process that uses specific methods to determine the value of a particular health technology at different points in its life cycle (i.e., pre-market entry, during its approval, post-market entry, and during its disinvestment). Therefore, the aim of HTA is to provide information on decision-making in order to promote fairness, efficiency and excellence in the health system.

In certain situations, HTA is obliged to conclude that the evidence available is insufficient or of low quality. After identifying the knowledge gaps, HTA reports provide recommendations on how to generate the evidence necessary to address the problem at hand.

As described in its statutes, the mission of the Agency for Health Quality and Assessment of Catalonia (AQuAS) is to generate data and knowledge able to improve the quality, safety and sustainability of the healthcare system in Catalonia. This knowledge facilitates decision-making not only for the general public, but also for healthcare professionals, managers and the bodies responsible for health planning. AQuAS also aims to promote the involvement of health professionals in the system and to foster their commitment to achieving common goals in the area of healthcare excellence.

The functions of AQuAS include HTA and the management of access to health data in order to generate knowledge able to improve the health of the general public. AQuAS is thus in an ideal position to help the healthcare system in Catalonia to manage uncertainty in the implementation of new technologies, via the design and execution of observational studies.

Today, several HTA and regulatory agencies have shown an interest in using observational studies with real-world data in their decision-making processes, along with clinical studies, to assess the effectiveness of technologies. That is, they aim to generate knowledge with real-world data.

However, concerns have been voiced about these studies, as the evidence they provide may be of low quality (or low reliability). Therefore, observational studies using real-world data are needed in order to fill in the knowledge gaps detected. To do so, they should follow standardized processes that promote the development of studies of excellence.

Real-world data (RWD) are the observational or administrative data that provide information on routine healthcare delivery and on the current state of health of the target population.

Real World Evidence (RWE) is the evidence generated by the analysis of RWD.

This Methodological guideline has been written as part of the internal policy of the AQuAS Assessment Area for 2021, under the strategic line 12 (LE12) “Data analysis in all assessment processes”. The present document is for internal use. It is a guide for the development of real-world data studies carried out at AQuAS and it standardizes the methodology to be followed when carrying out observational studies.

Objectives

This document presents a methodological guideline for the execution of studies with real-world data, led or co-led by AQuAS, along with researchers from the Catalan health system.

Methodology

A bibliography search was carried out to identify the documentation provided by organizations involved in decision-making in healthcare at European level. These documents offered recommendations on best practices for the generation or use of RWE.

Decision-making organizations of this kind usually publish recommendation documents directly on their websites rather than presenting their work for peer review. For this reason, a grey literature search was conducted targeting the websites of these key organizations rather than a systematic review of peer-reviewed literature, although a selection of peer-reviewed journal articles was also consulted.

In addition, an epidemiologist with extensive experience in the field (XGA) supported the development of the guideline through online sessions attended by AQuAS members LBS and RMVH.

Based on the review of the documents, the researchers developed a protocol, two templates, and a checklist. Finally, an internal flowchart was also created which detailed the process for writing reports with the key agents that would be involved. This flowchart was presented at a general meeting on 29 October 2021 and was approved by the AQuAS management.

NOTE: Further updates are expected, due to the huge volume of new proposals for including RWE in HTA, and the participation of AQuAS in international forums / working groups to keep up to date with the developments in the field.

How to conduct an observational study with real-world data in health technology assessment at AQuAS

The steps involved in an HTA observational study are described below.

Requirements for developing further observational studies with RWD in HTA

The execution of an observational study using real-world data should comply with the programme specified in Appendix 1.

Before conducting an observational study with RWD, the following requirements must be met.

NOTE: If any of the questions are answered “no”, the study will not be conducted.

1. Are any knowledge gaps identified in the available evidence during the technology assessment? [*Yes or No*]	
2. Is the research question (PICO) explicitly defined? [*Yes or No*]	
3. Is the collection of additional data feasible, in terms of time constraints, type of study, population and costs? [*Yes or No*]	
4. Is the study necessary, bearing in mind the possible existence of other similar studies in progress? [*Choose one of the following options*].	**a.** Yes, because no other similar studies are planned or in progress. **b.** Yes, because even if there are similar studies planned or in progress, this study will provide additional value. **c.** No, because the studies already planned or in progress will provide enough information to fill the knowledge gap.
5. Will the study provide additional information relevant to technology assessment and decision-making? [*Yes or No*]	

Those requirements are adapted from *Criteria to select and prioritize health technologies for additional evidence generation*. *Work Package 7*, EUnetHTA. July 2012.

Protocol synopsis

Before developing the final protocol in the template, the researchers or the person requesting the report/study must provide a protocol synopsis of it (see Appendix 2). This information will be evaluated by the AQuAS Data-Assessment Area in order to determine the feasibility of the study. It will also be considered whether the available data sources are suitable for answering the research question and objectives (see Appendix 3 for a checklist to assess the availability and the quality of data).

Study protocol

Once the requirements have been met and the feasibility of the protocol has been assessed, the research question must be determined. The recommended templates are presented below with descriptions of the different sections. This protocol template, as well as the checklist, has been created on the basis of the following documents:

- European Medicines Agency. Guidance for the format and content of the protocol of no interventional postauthorisation safety studies: EMA/623947/2012.
- Health Technology Assessment International. Real-world evidence in the context of Health Technology assessment processes. Global Policy Forum HTAi; 2019.
- Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, et al. STaRT-RWE: structured template for Planning and reporting on the implementation of real world evidence studies. BMJ 2021; 372:m4856.

NOTE: This template is standard and it is intended to assist in the development of a protocol for a study using real-world data, of whatever design (descriptive, predictive or causal inference). That is, there will be sections of the protocol that will not apply, for example, to descriptive studies. In this case, the required response is N/A (not applicable). This will also apply to the protocol checklist (Appendix 4). Version 1 (17/09/2021)

Protocol

Study using real-world data to generate quality evidence for Health Technology Assessment

Protocol title:
Protocol version:
Date:
Requester:
AQuAS strategic line:

Protocol title:	
AQuAS code:	
Protocol registration:	

1. DEVELOPMENT TEAM / WORKING GROUP

ROLE	NAME	AFFILIATION	CONTACT
Senior Manager (SM)		**AQuAS**	
Operations Manager (OM)		**AQuAS**	
Methodological supervision			

2. PARTNERS

ROLE	NAME	AFFILIATION	CONTACT

3. SUMMARY

Title	
Protocol version (date)	
AQuAS code	
Type of study	
Senior Manager (SM)	
Rationale and background	

Research questions and objectives	
Study design	
Population	
Variables	
Data source	
Sample size	
Data analysis	
Milestones	

4. VERSION HISTORY

DATE	PROTOCOL VERSION	RATIONALE FOR CHANGES / AMENDMENTS

5. MILESTONES

[This section should present the planning of the study's expected milestones]

MILESTONES	EXPECTED DATE
Protocol registration [if applicable]	
Start of data extraction	
End of data extraction	
Mid-term report(s) [if applicable]	

MILESTONES	EXPECTED DATE
Final report	

6. BACKGROUND AND RATIONALE

[Describe the context and the reason for conducting the study]

7. RESEARCH QUESTIONS AND OBJECTIVES

[Include research questions here]

7.1 Main objective

[Clarify whether the aim of the study is to DESCRIBE, PREDICT or INFER CAUSALITY. This point is important, as it will determine the study design as well as the methodological approach. For further details, see this reference]

7.2 Secondary objective(s)

[Clarify whether the aim of the study is to DESCRIBE, PREDICT or INFER CAUSALITY. This point is important, as it will determine the study's design as well as the methodological approach.]

8. METHODOLOGY

[A table summarizing the methodology can be found in Appendix 5]

8.1 Study design

[The information regarding methodology that does not come under the other headings of the protocol can be included here]

For descriptive studies:

[Specify the characteristic(s) or event(s) (e.g., the age and gender of the patients or mortality among those receiving care)].

[Specify the statistical estimator used for the description of each of the characteristic(s) or event(s) (e.g. means, minimum and maximum values for continuous variables, or proportions for categorical variables)].

For predictive studies:

[Specify the characteristics that are to be predicted and which predictors will be assessed (e.g., whether the aim is to predict the probability of mortality using age and gender in patients receiving care as predictors)].

[Briefly specify which analytical methods are considered for the prediction of the characteristic(s) (e.g. risk difference, logistic regression. or others)].

For causal inference studies (*target trial emulation*):

[Specify a) the overall study design and rationale (in the event of using the target trial emulation framework, in most cases it will be a cohort study), b) the comparison groups, c) the primary and secondary endpoints and the main effect measures, and d) causal contrast (if the aim is to consider the target trial as "intention to treat" or as "per protocol"). For further details on the target trial emulation design, the following references are recommended:

- https://pubmed.ncbi.nlm.nih.gov/34153099/,
- https://pubmed.ncbi.nlm.nih.gov/34089045/,
- https://pubmed.ncbi.nlm.nih.gov/33528595/,
- https://pubmed.ncbi.nlm.nih.gov/28411091/,
- https://pubmed.ncbi.nlm.nih.gov/27669524/].

If a target trial is implemented, the following points must be described:

8.1.1 *Index trial*. Is an index trial available? This is a clinical study that has already been conducted, with either published or unpublished results, addressing the same or a similar research question as this target trial emulation protocol. The target trial is a hypothetical study that may or may not be based on an index trial. If it is based on an index trial, this reference is recommended; otherwise this reference can be consulted.

8.1.2 *Time of exposure*. State whether the exposure of the target study will be a one-off procedure (e.g., a type of surgery, or a diagnostic procedure) or whether it will continue over time (e.g., screening strategies, periodical follow-up or medium- to long-term medication. This reference is recommended in the case of one-off exposure, and this reference in the case of sustained exposure.

8.1.3 *Time zero* or *baseline*. State whether the *time zero* or the *baseline* in the target trial emulation will be the same for all fields to be compared, or whether there will be several (this may be the case if the eligibility criteria are met more than once over time). In this case, describe how the different *time zeros* will be included in the target trial emulation. On the emulation of studies that address different *time zeros* the following references are recommended:

- https://pubmed.ncbi.nlm.nih.gov/31591592/,
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4832051/).

8.1.4 *Characterization of baseline variables*. Specify when the baseline variables will be characterized in the target trial emulation. If the study emulates different studies (see section 8.1.3 *Time zero* or *baseline*), the baseline variables should be characterized at the start of each study or cohort.

8.1.5 *Exposure strategies*. Describe in detail the exposure strategies of the target trial. These strategies should be specified and aligned with the procedures that are carried out in clinical practice. For example, in the case of a pharmacological treatment, the scenarios in which treatment should be stopped due to toxicity should be specified. By way of example, the following table displays a well-defined exposure strategy:

Poorly defined strategies	Properly defined strategies
1. The patient should take statins	**1.** The patient should start any statin treatment at time zero and continue the treatment during the follow-up until the development of a contraindication (e.g., liver impairment or myopathy).
2. The patient should not take statins	**2.** The patient should not start any statin treatment during the follow-up until the development of an indication (LDL cholesterol > 5 mmol/L) When a contraindication or an indication occurs during follow-up, the patients and clinical staff should decide whether to start, stop or change the treatment. Participants should have a primary care visit at least once every four years.
3. Annual screening mammography	**3.** Women continue to have an annual screening mammography (with a grace period of 3 months, meaning that they may have the mammography in the 12^{th}, 13^{th} or 14^{th} month after the last one. They are excused from further screenings after a diagnosis of breast cancer.
4. No annual screening mamm	**4.** Women have no further screening after baseline. In both strategies, a diagnostic screening mammography is allowed when clinically indicated.

8.1.6. *Grace periods*. If it is planned to use grace periods in the target trial, this should be specified. Grace periods are used to allow flexibility in the exposure strategies so that the target trial emulation accurately represents those performed in the actual practice. For example, in clinical practice it may be that cancer treatment is not initiated at the same time as the cancer diagnosis; there may be a delay until the treatment is initiated, for administrative, diagnostic, or other reasons. Grace periods allow a proper allocation of early events that take place, and avoid selection biases such as immortal time bias. For further details on grace periods, the following references are recommended:

- https://pubmed.ncbi.nlm.nih.gov/25794605/,
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4832051/).

8.1.7. *Patients' classification according to study strategies.* If patients use more than one exposure strategy during part of their follow-up (to use grace periods or because their exposure strategies are maintained over time, with common exposure periods), explain how this situation will be addressed. In the context of a target trial emulation, clones can be used. For this application, this reference is recommended.

8.1.8. *Causal contrast.* Describe the causal contrast that is to be analysed. For some examples of causal contrasts in a target trial design, see:

- Intention-to-treat effect: In a target trial, this effect estimates the fact of being assigned to an experimental branch compared to the control branch when complete follow-up of the procedure is performed. In a target trial, this effect requires adjustment for loss to follow-up. The observational counterpart of the intention-to-treat effect implemented in target trial emulation compares the effect of initiating exposure in the experimental branch and the control branch when complete follow-up is performed. This effect requires adjustment for baseline confounding (since there is no randomization) and for any losses to follow-up.
- Per-protocol effect: In a target trial, this effect estimates the fact of receiving the treatment in the experimental branch compared to in the control branch, with complete follow-up and adherence to the treatment. In a target trial, this effect requires adjustment for baseline confounding, for time-varying confounding and for any losses to follow-up. The observational counterpart of the per-protocol effect implemented in target trial emulation estimates the same effect as in the target trial, requiring the same adjustment.

Taking all these concepts into account, complete the following table. For further examples, the following references are recommended.

- https://pubmed.ncbi.nlm.nih.gov/29112066/,
- https://pubmed.ncbi.nlm.nih.gov/32753444/,
- https://pubmed.ncbi.nlm.nih.gov/28822996/.

COMPONENT	TARGET TRIAL	TARGET TRIAL EMULATION
Eligibility criteria		
Therapeutic strategies or procedures		
Treatment assignment		

COMPONENT	TARGET TRIAL	TARGET TRIAL EMULATION
Outcomes		
Follow-up		
Causal contrast		
Statistical analysis		

8.2 Setting

Study population and representativeness of the sample

[Specify the types of population, the setting, and whether the whole population is included or only a sample. Indicate whether convenience or random sampling is used. If the study does not include the whole population, describe the type of sample and the source of the data. Specify the geographical area of the data].

Time period

[The start and the end of data collection]

Selection criteria

- **Inclusion criteria:**
- **Exclusion criteria:**

8.3 Variables

Exposure

[Describe the groups that are compared in the study. If the target trial emulation framework is used, there will always be an exposure which would be the equivalent of the intervention in a randomized clinical study].

In this section, point 8.1.5 (Exposure strategies) can be used as a reference or it can be expanded.

Outcome measures

[Define the primary and secondary variables]

Confounders

[Identify potential confounders and whether they are available in the database]

Effect modifiers

[Identify possible modifiers and describe their influence on the exposure/intervention]

8.4 Data source

[Describe the database/record from which the study data will be obtained. State whether there has been a linkage between the different databases/records. To assess data quality, indicate whether the data related to the exposures or outcomes have been previously validated, and whether previous studies exist for reference purposes. If not, describe how this will be addressed in the study. In both cases, indicate whether expert panels have been involved. This information should be checked with the AQuAS data area].

8.5 Sample size

The sample size will be the size available in the data source. The aim of this section is to define the accuracy of the effect estimators (https://www.jclinepi.com/article/S0895-4356(21)00273-0/fulltext). Describe the approximate sample size (e.g., in the form of a range). Ideally, prior sample estimates should have been obtained (i.e., how many people have made use of a particular technology, e.g., how many people have undergone TAVI). Based on the approximate sample size, the following calculations can be made:

1. An estimation of the 95% confidence interval (95%CI), taking into account the following specifications:

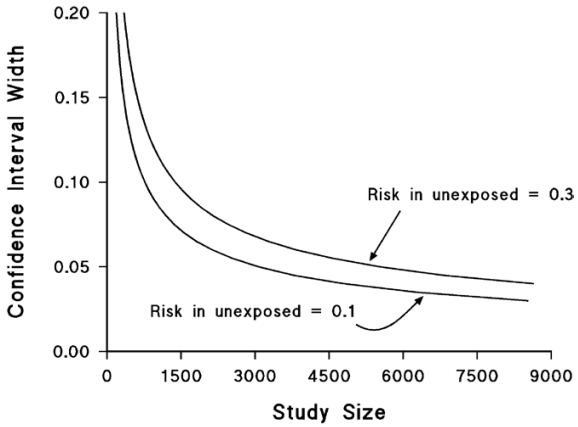
    a. Effect estimator (e.g., risk difference, difference in incidence rate, risk ratio, etc.)

    b. Risk or incidence in the control group

    c. Risk or incidence in the exposure group or, alternatively, an effect (e.g., risk difference)

    d. Expected distribution of patients in each branch (e.g., 1:1, 1:2 or others)

With these specifications the range of the CI can be estimated, and the expected accuracy of the study can be predicted. At this point, describe the expected accuracy of the data, assessing both the statistical power and the accuracy. A chart similar to the one below should be included (Source: Rothman et al. 2018. Epidemiology; 29(5), 599-603, https://pubmed.ncbi.nlm.nih.gov/29912015/).



2. An estimation of the probability that the upper bound of the CI will exceed the *a priori* threshold, using the following specifications:

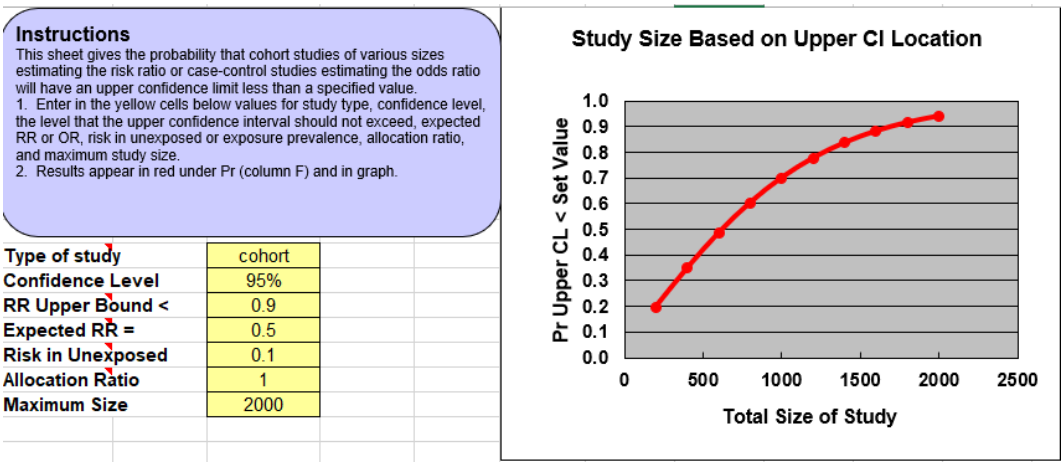
   a. Threshold of the upper bound of the CI (e.g., in a study of efficacy, the aim is to detect the minimum efficacy by setting an upper bound)

   b. Expected effect of the intervention

   c. Risk in unexposed people

   d. Distribution of patients in each branch (e.g., 1:1, 1:2 or others)

This estimation can be performed with Rothman's Episheet tool, under the tab "*Study Size*" (http://krothman.hostbyet2.com/Episheet.xls). For example, in the case of a cohort study design aiming to assess the efficacy of an intervention, and in which 1,000 to 2,000 eligible individuals are expected, one could assume that the expected efficacy is to halve the risk of the event, with an upper limit of 95% CI of < 0.9. If the risk in unexposed people is 0.01, and patients are distributed 1:1 between the intervention and control branches, Episheet can calculate the probability that the upper limit of the 95% CI is < 0.9. Therefore, from 1,500 patients, a 95% CI with the desired range will probably be obtained.



In both cases (1 and 2), the variance may increase when statistically adjusting for confounding factors.

8.6 Data management

[Describe AQuAS's procedures for data management, coding, anonymization, the software used, and so on]

8.7 Data analysis

Creation of the cohorts

[In the case of target trial emulation, define *time zero*. The *time zero* occurs when a) the eligibility is defined, b) the exposure is assigned, and c) the follow-up has begun. Include a figure/sketch to illustrate how the cohort(s) is/are created. Consider including a hypothetical example of observed exposure, cohort generation and the strategy for group assignment].

As described in 8.1.3 *Time zero* or baseline, the way in which *time zero* will be used for cohort creation should be defined. In this section, it is recommended to introduce a chart for the creation of cohorts, following this reference.

Assignment to exposure and follow-up

[Describe how a patient is assigned to the exposure group or the comparator group]

In this section, it is important to describe the follow-up according to the causal contrast being studied. For example, for intention-to-treat or per-protocol, it may be necessary to implement artificial censoring, which will determine the type of follow-up.

In this section, expand on the concepts outlined in 8.1.6 Grace periods, 8.1.7 Classification of patients according to study strategies, and 8.1.8 Causal contrast.

Outcome measures

[Describe how the effect will be measured for each of the outcome variables. Measures of effect should be expressed at an absolute level (e.g., risk difference)]

Setting up the analysis

[Describe how the analysis will be adjusted to ensure interchangeability. After adjustment for confounders, patients should be interchangeable between groups. Define the type of adjustment and explain the rationale for it (propensity score and others)].

	Model 1	Model 2	…
Confounding adjustment method: - Baseline confounding - Time-varying confounding - Losses to follow-up Artificial censoring			

Missing data

[Describe the impact of missing values and how it will be adjusted for the analysis]

	Variable affected	Variable affected	Variable affected	...
Specify the method based on the results of the feasibility study (e.g., imputation of missing values)				

Sensitivity analysis

[Describe the sensitivity analyses: What is the parameter that varies, and why? If possible, identify the role that it will play in the conclusions]

	Variable	Rationale	Strengths (*)	Limitations (*)
Sensitivity analysis 1				
Sensitivity analysis 2				

(*) In terms of the main analysis.

8.8 Quality control

[Describe the procedure used for the validation of data quality. See Appendix 3 for a Checklist for data quality/availability assessment by AQuAS]

8.9 Methodological limitations

[Describe the limitations related to any aspect of the methodological approach: sample selection and representativeness, sample size, confounding factors, adjustment of the analysis, among others]

9. DATA PROTECTION

[A study based on RWD should not present a risk to patients. Describe the procedures applied to ensure data confidentiality/anonymization, the role of the ethics committee, and so on]

10. TRANSPARENCY AND REPLICABILITY

[To ensure transparency and replicability, it is recommended to register protocols, e.g., on websites such as ClinicalTrials.gov, or even to publish them in peer-reviewed journals. Once done, specify the protocol registration number]

11. DISSEMINATION PLAN AND COMMUNICATION OF RESULTS

[Include the specific dissemination plan for the study and the communication of the results in the application. Include the dissemination of results through AQuAS's communication channels such as the website and social media]

12. REFERENCES

[Include the specific dissemination plan for the study and the communication of the results in the application. Include the dissemination of results through AQuAS's communication channels such as the website and social media]

Appendices

Appendix 1. Internal circuit

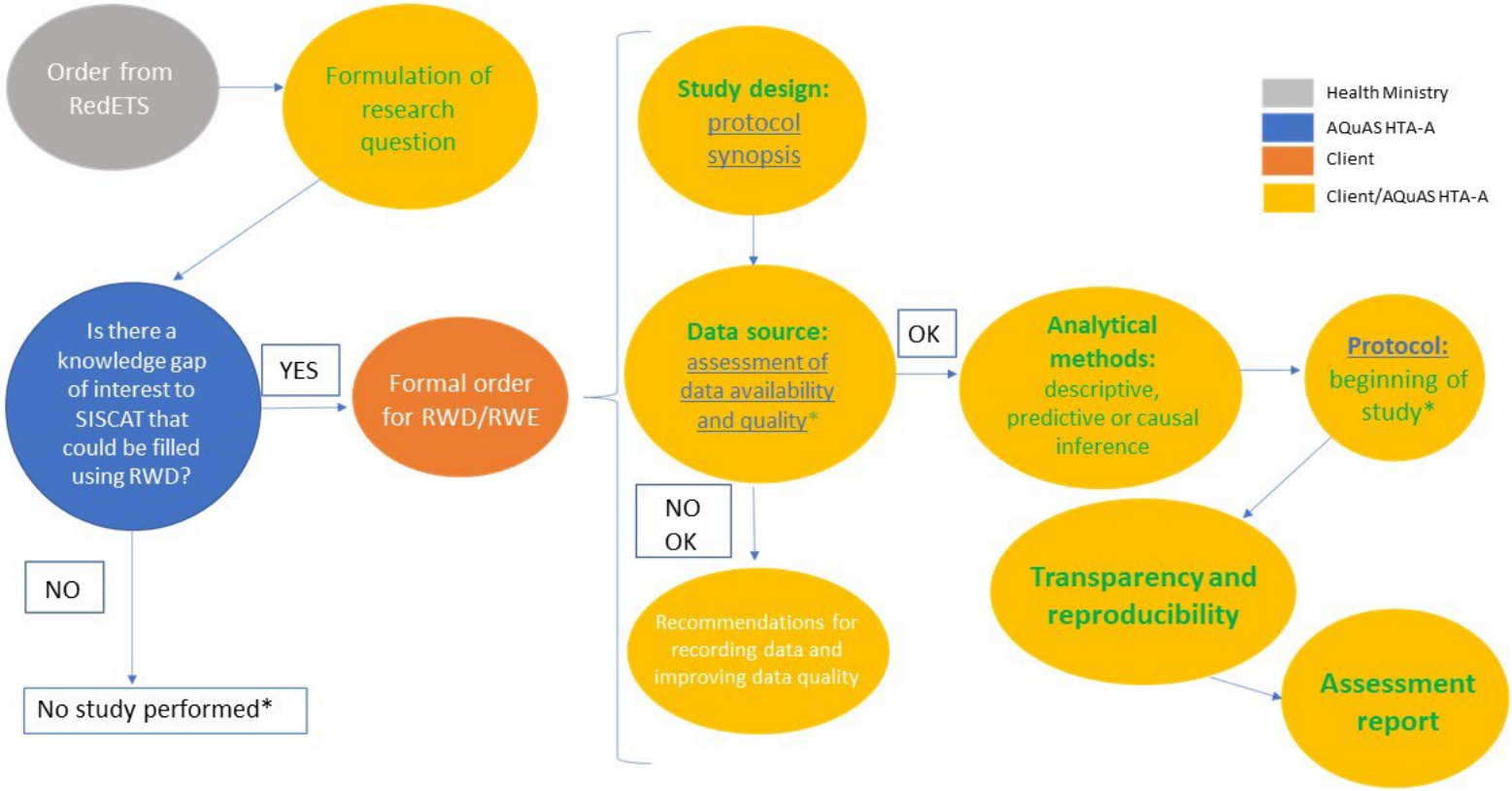
Appendix 2. Protocol synopsis

Appendix 3. Checklist for data quality/availability assessment

Appendix 4. Protocol checklist

Appendix 5. Summary of the methodology

APPENDIX 1. Internal circuit



*A feasibility study will be performed of the data at this stage if there is some uncertainty regarding the quality of the data collection
*The PIs will be appointed (clients advised by AQuAS HTA-A)

If studies cannot be carried out because the point at issue cannot be resolved using RWD, the Health Technology Assessment Area (HTA-A) can present a proposal with another appropriate study design. It may be found that the research question can be answered by consulting studies already published and available in scientific literature. In this case, a systematic review is suggested, as this is the standard methodology used in HTA.

APPENDIX 2. Protocol synopsis

Instructions: To tick a box, place the cursor to the left of the box, click the left button, select "properties" and then select "checked".

Information regarding the study cohort

How is the study population defined? List the eligibility criteria and the time period in which they will need to be assessed:

Criteria	Time window

1a. In what kind of setting is the information used to define the study population recorded? (Tick all that may apply):

- ☐ Primary care centre
- ☐ Hospital outpatient clinics
- ☐ In-patients
- ☐ Accident and emergency department
- ☐ Outpatient prescription
- ☐ Inpatient prescription
- ☐ Other: _____

1b. What information is needed to identify eligibility criteria?

- ☐ Diagnostic codes
- ☐ Procedure or therapeutic codes
- ☐ Laboratory results
- ☐ Other: _____

1c. Is validation of any of the eligibility criteria required?

- ☐ No
- ☐ Yes (specify criteria):

_____

Information on exposure

Define the exposure of interest (e.g., receiving a liver transplant, taking statins according to the recommended therapeutic programme, or a biannual digital screening mammography for breast cancer). Please specify whether this is a one-off exposure (i.e., the first example above) or whether it is sustained over time (i.e., the second and third examples).

2a. In what kind of setting is the information collected to identify exposure? (Tick all that may apply):

- ☐ Primary care centre
- ☐ Hospital outpatient clinics
- ☐ In-patients
- ☐ Accident and emergency department
- ☐ Outpatient prescription ☐ Inpatient prescription
- ☐ Other: _____

2b. If the exposure is medication or treatment, is the differentiation between "prescription" and "administration" relevant?

☐ Yes ☐ No ☐ Other: _____

2c. If the exposure is sustained over time, please list the variables or events that may determine patients' adherence to the intervention (e.g., a patient who develops liver failure is likely to stop taking statins, a patient diagnosed with lung cancer may decide to stop breast cancer screening, etc.):

Outcomes

Please define the outcome of interest (e.g. cancer diagnosis, mortality).

3a. In what setting is the information collected to identify the outcome? (Tick all that may apply):

- ☐ Primary care centre
- ☐ Hospital outpatient clinics
- ☐ In-patients
- ☐ Accident and emergency department
- ☐ Outpatient prescription
- ☐ Inpatient prescription
- ☐ Other: _____

3b. What information is needed to identify the outcome?

- ☐ Diagnostic codes
- ☐ Procedure or therapeutic codes
- ☐ Laboratory results
- ☐ Other: _____

3c. Is validation of the outcome required?

- ☐ No
- ☐ Yes

Information on other aspects of a cohort study

Please list the potentially confounding baseline variables and the reference on which they are based, if available.

Variable	Reference

How long is the planned follow-up period?

During the follow-up period, have there been any important changes in the diagnosis, treatment or care setting of the study? (e.g., does it include follow-up during the COVID-19 pandemic?)

APPENDIX 3. Checklist for data quality/availability assessment

This checklist aims to assess the quality/availability of data provided by AQuAS to answer questions referring to HTA studies (a feasibility study once the protocol synopsis is available).

Description of the data provided by AQuAS

This description should include the following information:

Context of data generation
• In what setting are the data collected? Public, private, outpatient, inpatient, geographical area, calendar.
• What are the reasons why people are in the database? Is there a profile of people who are not in the database? (e.g., destitute, unregistered persons, foreigners using only private healthcare)?
• What are the reasons for collecting the data? Are there incentives to collect data of a particular kind?
• Who collects the data? (e.g., medical staff, nursing staff, administrative staff, or others)
• How are the data collected?
• How long should data be collected? Have there been any changes in the process over time? (e.g., changes in the code dictionary)
• How many people are included in the database?
• What is the time span between the collection of the data to their becoming available for research? How often is the database updated?
• What is the time span from the implementation of a new health intervention to its being recorded? (e.g., if a new radiopharmaceutical is approved for diagnosis or treatment, how long does it take for it to obtain a specific code)?
• Can additional information be collected from the professionals providing the data or from the patients? (for example, if it is known that weight is an important variable for a certain study and the data are insufficient, could they be completed?)
• Can the professionals providing the data be contacted in order to validate the variables? (e.g., if the study outcome is a diagnosis of breast cancer, can pathology records or medical records be accessed to confirm this?)

Context of data generation
• Describe the type of studies that have been conducted so far using the database, and provide references.
• Can the database be cross-checked with other data sources?
• What is the procedure for accessing the data? Who is the contact person? Are there any restrictions on access? Are there any restrictions on displaying the results?
• Is review by an ethics committee required?

Data content
• What fields are collected and what code dictionary is used?
• Is free text available?
• Is any kind of data validation available? (e.g., logical filters)
• Is the information on the number of missing values available? Do they differ according to the type of variable? Has the amount been quantified? Is it known why the values are missing? (e.g., does the loss occur randomly or is it influenced by a variable?)
• How has the COVID-19 pandemic affected data collection?
• Have any other external events affected data collection? (e.g., administrative or hospital restructuring)
• Are the data transformed in any way or are they provided in raw form?
• Is it possible to follow a person's diagnostic or therapeutic history through different healthcare providers? (e.g., is a person who goes to the Primary Care Centre and then goes to the emergency department counted as a duplicate?)
• What information is missing from the data? (e.g., prescriptions and therapies not covered or unavailable in the public health system)

APPENDIX 4. Protocol checklist

Protocol checklist

Real World Data - Real World Evidence Study

Protocol title:
Protocol version:
Date:
Requester:
AQuAS strategic line:
AQuAS code:
Protocol registration:

1. DEVELOPMENT TEAM / WORKING GROUP

1.1. Does the study have an SM assigned?	
1.2. Does the study have a member assigned for methodological supervision?	
1.3. Is the information required from each member of the study team specified?	

2. PARTNERS

2.1. Is the requester included in the study as a partner member?	
2.2. Are partners with experience in the field of study included?	
2.3. Is the information required from each partner specified?	

3. SUMMARY

3.1. Are all sections of the abstract duly completed?	

4. VERSION HISTORY

4.1. Is the version history of the protocol up to date? [if applicable]	

5. MILESTONES

5.1. Have the expected due dates for each milestone been met?	

6. BACKGROUND AND RATIONALE

6.1. Are the health problem and target population described?	
6.2. Are the characteristics of the health technology or procedure described?	
6.3. Is the reason for the study request identified and explained?	
6.4. Are the origin and the requester of the study/report specified?	
6.5. Is the scope of the study/report and its potential users specified?	

6.6. Does the protocol comply with the request for the study/report?	

7. RESEARCH QUESTIONS AND OBJECTIVES

7.1. Has the researcher or person requesting the study/report completed the protocol synopsis to assess its feasibility (Appendix 2)?	
7.1. Do the research questions include the target population, the health technology or procedure, the comparator, and the outcome?	
7.2. Does the PICO research question match the request of the study/report?	
7.3. Does the primary objective specify whether the study is intended to describe, predict or infer causality?	
7.4. Do the secondary objectives specify whether the study is intended to describe, predict or infer causality?	

8. METHODOLOGY

8.1. Is the type of study design specified (descriptive, predictive or causal inference)?	
8.2. Is an explanation given for the type of design selected?	
8.3. Is the procedure for data quality control specified?	
8.4. Is/are the control group(s) specified?	
8.5. Are the primary and secondary endpoints detailed?	
8.6. Are the main measures of effect specified?	
8.7. In the case of target study emulation based on causal inference, is it specified whether the approach will be 'intention-to-treat' or 'per-protocol'?	
8.8. Is the study population described?	
8.9. Is it specified whether the entire study population will be considered or just a sample? In the case of a sample, is it stated whether the sampling is convenience or random?	
8.10. Are the start and the end dates of data collection defined?	
8.11. Are the inclusion criteria clearly and concisely specified?	

8.12. Are the exclusion criteria clearly and concisely defined?	
8.13. Is the representativeness of the study population and the data source specified?	
8.14. Is the exposure clearly and concisely specified? Are the study groups properly defined?	
8.15. Are the primary variable and the secondary variables properly defined as outcome measures?	
8.16. Are potential confounders specified?	
8.17. Is it specified whether potential confounders are recorded in the database?	
8.18. Are possible effect modifiers identified?	
8.19. Is the influence of possible effect modifiers on the procedure/exposure described?	
8.20. Is the database/register used to extract the study data described in detail?	
8.21. In the case of a link to other databases/registers, is it described?	
8.22. Is it specified whether the data have been pre-validated? If so, are the validation studies referenced?	

8.23. If the data have not been previously validated, is the manner in which they will be addressed in the study described in detail?	
8.24. Is it specified whether experts were involved in the validation of the data?	
8.25. In the case of a descriptive or predictive study, is the expected power of the data described?	
8.26. In the case of a causal inference study, is the expected accuracy of the data described?	
8.27. In the database/register used to extract the data for the study, Is the management and handling of the data (coding, anonymization, etc.) described in detail?	
8.28. Is the data analysis specified according to the design? (i.e., descriptive, predictive or causal inference)?	
8.29. In the case of a target trial emulation, is the procedure for allocating a patient to the exposure or comparator group described?	
8.30. Is there a description of how the effect will be measured for each of the outcome variables (both primary and secondary)?	
8.31. In the case of a causal inference study, is the method of adjustment for confounding factors described?	
8.32. Is the method for imputing missing values described?	

8.33. Are sensitivity analyses indicated for each of the outcome variables?	
8.34. In terms of the main analysis, are the rationale, strengths and limitations described for each sensitivity analysis?	
8.35. Are methodological limitations related to methodological aspects (sample selection and representativeness, sample size, confounders, adjustment of the analysis, etc.) described in detail?	

9. DATA PROTECTION

9.1. Is the handling of data confidentiality and the involvement of the ethics committee described?	

10. TRANSPARENCY AND REPLICABILITY

10.1 Is the study protocol recorded in a database such as www.clinicaltrials.gov or others?	

11. PLAN FOR THE DISSEMINATION AND COMMUNICATION OF RESULTS

11.1. Is the expected dissemination of the results specified (communication of the results to the requester, communications to congresses, scientific publications, or others)?	
11.2. Is the dissemination of the results through AQuAS's communication channels envisaged?	

12. REFERENCES

12.1. Is there a list of references to the studies/documents cited throughout the protocol? Are the studies cited correctly?	

APPENDIX 5. Summary of the methodology

Protocol title	
Background and rationale	
Research question	The study question will be defined according to the PICOTS scheme Patients- Procedure-Comparator- Outcome - Time (Time horizon) - Framework/scenario (setting)
Objectives (primary and secondary)	Clarify whether you intend to DESCRIBE, PREDICT OR INFER CAUSALITY
Methodology	
Study design	Descriptive, predictive, causal inference Figure with design, if applicable
Setting	Study population and representativeness (type, care setting (hospital, primary care, emergency, etc.), geographical area Time period: start and end of data collection Eligible cohort entry period (recruitment period) Inclusion and selection criteria
Variables	Define exposure, outcome measures, confounders, effect modifiers, washout window (if time is required until the assessment of a new event/outcome) Description of the index date (day 0) and definition

	With regard to the predefined covariates: • Characteristics (definitions and how they will be assessed) • Variable type • Assessment window • Site of care (inpatient, outpatient, casualty, etc.) • Type of code used to assess the diagnosis Source of algorithms for defining covariates (if necessary)
Data source(s)	Definition and type(s) of the data source(s) used Date to be extracted Criteria for data extraction (data quality, linkage) Data conversion (if required to perform the extraction) Source of algorithm for defining selection criteria Highlight all key points of the data extraction, the criteria used and the assessment times
Sample size	Estimation of CI range and threshold probability
Data management	Type of code used Anonymization process Software
Data analysis	Creation of cohorts: definition of *time zero* Exposure assignment and follow-up Definition of the measurement of the effect Definition of the model(s)

	Adjusting the analysis: -Propensity score matching -Propensity score weighting -Propensity score stratification (stratum definition) -Other (specify details) Loss management: Patients selected, excluded, remaining, and assessable Flowchart of patient selection and exclusion from the main analysis Missing data: Methods for assessing missing data Sensitivity analysis
Quality control	Description of the procedure for the validation of data quality

Generalitat de Catalunya

Salut/ Agència de Qualitat i Avaluació Sanitàries de Catalunya