

****

**Standardised Assessment Process**

**for Feasibility**

Version 1.0

Main Authors: Lucía Carrasco-Ribelles, Lucía Bellas, Annika Jödicke, Daniel Prieto-Alhambra

Effective date: tbc

Table of content

[1 Summary 3](#_Toc212548107)

[2 Glossary of Terms 3](#_Toc212548108)

[3 Procedures needed before Feasibility Assessment 4](#_Toc212548109)

[4 Methods Feasibility Assessment 5](#_Toc212548110)

[4.1 Feasibility of answering the research question in real-world data 6](#_Toc212548111)

[4.2 Assessment of methods requirements 6](#_Toc212548112)

[5 Resource Feasibility Assessment 6](#_Toc212548113)

[6 Stop/Go before Data Feasibility 6](#_Toc212548114)

[7 Data Feasibility Assessment 7](#_Toc212548115)

[7.1 Exploration 7](#_Toc212548116)

[7.2 Initiation 9](#_Toc212548117)

[7.3 Implementation 9](#_Toc212548118)

[7.4 Execution 11](#_Toc212548119)

[7.5 Dissemination 12](#_Toc212548120)

[8 Feasibility Report and Stop/Go Meeting 12](#_Toc212548121)

[9 Timelines and Responsibilities 13](#_Toc212548122)

[10 References 14](#_Toc212548123)

[Annex I: GREG Feasibility Assessment Report 15](#_Toc212548124)

[Annex II: Phenotype Proposal Form 18](#_Toc212548129)

# Summary

The purpose of this document is to establish a standard process for conducting feasibility assessments within the IHI GREG project in a consistent, transparent, reproducible, and high-quality manner. This process involves different dimensions that need to be evaluated to determine the overall feasibility of a use case, as well as the steps for their evaluation. These dimensions include

1. Methods feasibility
2. Data feasibility
3. Resource feasibility

The feasibility assessment will be embedded in the GREG use case selection process, which comprises the assessment of both relevance and feasibility for prioritisation.

# Glossary of Terms

1. **Research question**: A clear and specific research question framed in PICO (Population, Intervention/Exposure, Comparator, Outcome) or a similar structure.
2. **Use Cases (UC)**: specific, real-world research questions or scenarios that guide the development, testing, and demonstration of federated analytics methods across participating data partners.
3. **Use Case lead/s:** The person(s) who submitted the respective use case and/or the person who will lead the conduction of the Use Case as Principal Investigator.
4. **Landscaping:** Broad screening of potential EHDEN Data Partners, performed by WP6 and the EHDEN Foundation. Landscaping is based on readily available data, typically contained in the EHDEN portal or similar repositories. This process maps the availability of data against a predefined set of clinical concepts.
5. **Feasibility:** Feasibility assessment in GREG will comprise methodological, data, and resource-related aspects. It involves a detailed evaluation of whether the data sources can operationalize the specific study question, including the presence and validity of required variables, population size, temporal coverage, completeness of clinical domains, and fitness-for-purpose to support the planned analyses. The process will be conducted in collaboration of WP4-5-6 and the respective Use Case Lead documented.
6. **Concept set:** A concept set is a standardised, computer-executable list of vocabulary concepts (“code list”) used to define a clinical entity of interest — such as a condition, drug exposure, procedure, or outcome — in a consistent and reproducible way across analyses and Data Partners [2]. It consists of selected concepts from OMOP standard vocabularies (e.g. ICD, SNOMED, ATC) and can include specific codes directly representing the entity and hierarchies of related concepts, allowing broader or more inclusive definitions. Concept sets are designed to be reusable components in various analyses, ensuring that study definitions are consistent, shareable, and transparent.
7. **Phenotype / Cohort:** Set of persons who satisfy one or more inclusion criteria for a duration of time. They are the primary building blocks for executing a research question. These persons are identified within a database based on inclusion and exclusion criteria, using standardized concept sets [2]. Follow-up begins at cohort entry and continues until a defined exit event (e.g. end of treatment, outcome, loss to follow-up, end of data).

# Procedures needed before Feasibility Assessment

A Use Case Prioritisation Task Force (UCPTF) will be formed to formalise and document decisions on the Relevance and Feasibility of the submitted Use Cases. The UCPTF will be formed by 1 member from each of the institutions in the GREG consortium, in addition to key experts in Relevance from WP1-2-3 and in Feasibility from WP4-5-6.

First, for each of the submitted Use Cases, the UCPTF will be tasked with the mission to confirm and document Relevant Use Cases. The process for this is described in a separate document: [Project Portfolio selection proposal - v3.docx](https://mieur.sharepoint.com/:w:/r/sites/GREGEU/_layouts/15/Doc.aspx?sourcedoc=%7BD72C2284-EDEC-440B-9887-1F27D3F45DA8%7D&file=Project%20Portfolio%20selection%20proposal%20-%20v3.docx&action=default&mobileredirect=true). Based on this, the UCPTF will receive the Rankings produced by WP3, containing information on the order of priority of the evaluated use cases based on relevance according to the project tasks, i.e. tasks 4.1-4.3 and 5.1-5.3 as voted by the consortium for wave 1. The UCPTF will consider this ranking and apply it before requesting a Feasibility assessment. In case of ties or disagreement in the ranking, the UCPTF Chair will have an executive vote to break the tie. One of the members of the GREG Ex-com will chair the UCPTF, with a term lasting 1 year per Chair, renewable for a second year where agreeable.

Once a list of prioritized Use Cases is confirmed, the UCPTF will activate the Feasibility assessment, and -with support from the PMO- organize a meeting with WP4-5-6 to discuss Methods and Resource Feasibilities. The criteria for evaluation of these are defined in subsequent sections.

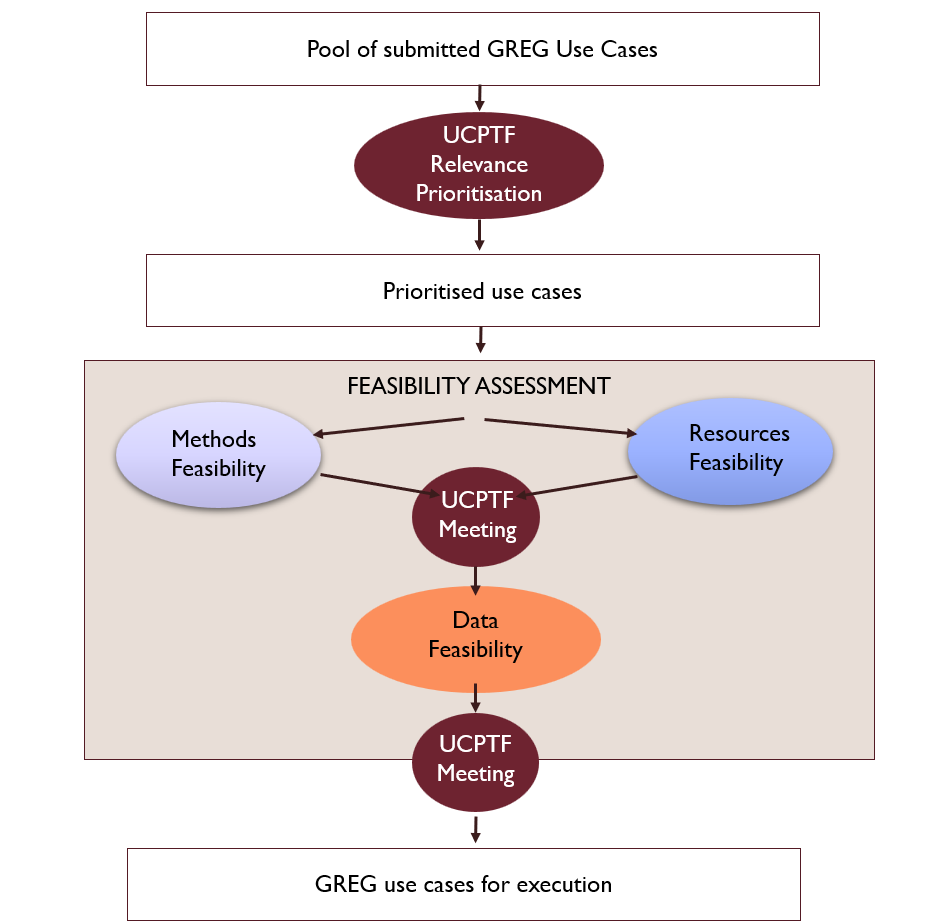


Figure 1. Overview of the use case selection + feasibility assessment procedure within the IHI-GREG project

# Methods Feasibility Assessment

Methods feasibility will be conducted by members of WP4 or WP5, depending on which work package and task the use case was submitted for. The following aspects will be assessed:

1. Feasibility of answering the research question in European real-world data
2. Assessment of methods requirements

The Use Case lead will not be participating in methods feasibility assessment due to a potential conflict of interest but might be contacted in case of need for clarification questions on the proposed methods. Where adaptations of the methods are suggested, these will be described in detail in the Feasibility Report and will be discussed at the UCPTF meeting.

## Feasibility of answering the research question in real-world data

GREG use cases will be focussing on using routinely collected European “real-world data”. In this first step, we will assess if the research question can be addressed using these data, or if the study can only be conducted if data would be collected prospectively, if randomisation was available, or if patients needed to be contacted to collect specific outcomes. In those cases, the study would be deemed not be feasible withing the GREG project lifetime and resources.

## Assessment of methods requirements

Suitability of suggested/required analyses and/or study design

Based on the research question and objectives stated in the use case submission form, nominated experts from WP4 and WP5 will assess if the suggested analytical methods/study type/design are suitable and ensure that these are considered state-of-the-art methods where relevant guidance for the respective analysis/study type exist. This can include, where relevant, study design choices such as suitability of active comparators, follow-up time, methods to minimise confounding or the definition of denominator populations, in line with the proposed methods/objectives in the target tasks (wave 1: task 4.1-4.3 and 5.1-5.3)

Availability of standard analytics

To inform the conduction of the potential use case we assess if/for which analyses/objectives

* 1. Open-source, analytical R packages for the respective analysis on OMOP-mapped data are readily available, i.e. from DARWIN EU®/OHDSI packages or standard R packages. These need to be accessible for all data partner (DP) to allow for federated analyses.
  2. Custom R code needs to be developed, and the complexity of this code. Bespoke code can be used where necessary, but it will imply the need for time and resource for development, testing and execution, far beyond that needed for standard analytics.

# Resource Feasibility Assessment

Staff resource/s

Nominated experts from WP4-5-6 will be involved in the assessment of the resource needed to conduct each prioritised use case. The evaluation of this will include considerations of study complexity (number of objectives, need for bespoke study code or analytics), and number of data partners required. This assessment will be compared to the resource available to the Use Case Lead/s as proposed in their Use Case submission form.

Funding resource/s

Nominated experts from WP4-5-6 will be involved in the assessment of the resource needed to:

* Involve consortium partners as part of the study team (if applicable);
* OR to conduct work on OMOP vocabularies or mapping, or to participate as data partners in the proposed Use Case/s

The former (if necessary) would require in kind contributions from existing consortium partners to participate in the study team, or in-cash contributions to subcontract study activities. The latter would involve the use of ‘contingency funds’ held by Erasmus MC to subcontract external parties (European SMEs or academic institutions) to contribute to the Use Case/s by improving existing OMOP vocabularies, improving existing mapping of available data, or participating as data partners in the proposed Use Case/s.

The evaluation of Funding resource needs will be compared to the available staff, data, and other contributions proposed as part of the Use Case submission by the UC Lead/s.

# Stop/Go before Data Feasibility

A **Preliminary Feasibility meeting** will be held between WP4/5/6 leads and the Use Case leads for the prioritised Use Cases. There, the nominated experts (or deputies) will present on Methods and Resource Feasibility, and ask Use Case leads for any clarification, including on needs and availability of data.

After this, a Preliminary Feasibility decision will be communicated to the UCPTF, who will sign off the decision. Possible outcomes will include:

* Proceed to Data Feasibility (see subsequent sections)
* Wait until Feasibility issues are resolved and resubmit
* The Use Case is unlikely to be feasible within the GREG project lifetime

# Data Feasibility Assessment

This section presents how feasibility of the use case is evaluated from a data perspective, focusing on the availability and suitability of relevant data domains and sources, as well as patient counts. Data feasibility assessment involves the following consecutive phases:

A diagram of a process

AI-generated content may be incorrect.

Figure 2. Data Feasibility Assessment phases

## **Exploration**

This phase involves the exploration of the use case in terms of data feasibility. The data partners shortlisting process will be conducted in a systematic approach, and each step will be documented with the reasoning for transparency. This phase includes two steps, followed by a meeting.

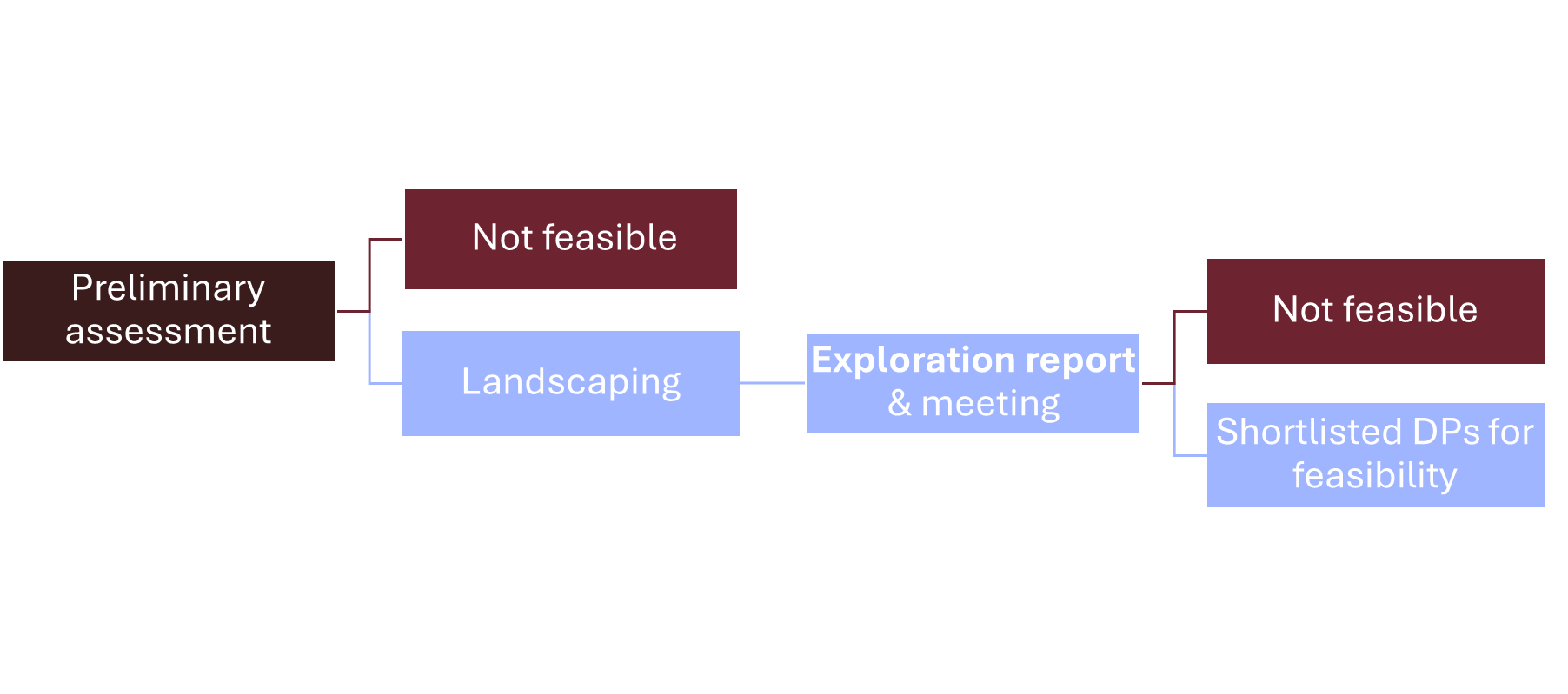


Figure 3. Exploration phase steps and outputs

### Preliminary assessment

WP6 will review the submission forms for the relevant use cases and conduct a preliminary assessment of data feasibility. The following criteria will be evaluated during this step:

1. Healthcare setting suitability. The UC may require, for instance, information on some conditions/procedures that are only/better captured in some settings but not others. Not every data partner covers every setting.
2. Temporal suitability. The UC might need, for instance, information on some specific drug or device that was approved last month, but the latest information available in some data partners might be 6 months ago.
3. Geographical suitability. The UC might need, for instance, information from specific countries that might not be covered by any data partner.
4. Presence of necessary data domains. The UC might need, for instance, information on procedures but not every data partner includes it.

To ensure a systematic and transparent selection process, a flowchart will be produced reporting how many data partners do not meet each criterion. If at least one meets the criteria, landscaping will be performed by the EHDEN-F. Otherwise, written feedback will be provided on the reasons of “non-feasible” assessment, and potential actions if necessary (WP6 + 4/5).

Landscaping for in-house data or data not included in EHDEN data catalogue will be conducted by the local data steward, and will use the criteria laid out in Annex I to produce the output/information needed to populate the Data Feasibility Section in Annex I.

### Landscaping assessment

Landscaping consists on:

1. Based on the use case submission form and with assistance of UC leads, a list of concept IDs (~40 concepts) list in Excel format will be developed to conduct landscaping. Typically, this will include concepts related to the key variables needed for the study, e.g. primary outcome and exposure/s.
2. WP-6 will conduct a landscaping exercise on the list of data partners obtained during the preliminary assessment.This exercise obtains counts of patients who have the previously selected concept IDs recorded, based on the latest information available on the EHDEN portal.

EHDEN-F will produce an **exploration report**, including the flowchart produced in the previous step, the table with patient counts for each concept ID and data partner selected in the preliminary assessment, and a table reporting an overall metric of their data quality, their contractual readiness and timelines. The overall data quality metric reports the percentage of data quality tests passed. Contractual readiness and timelines can be crucial in use cases with tight deadlines, as some data partners in the EHDEN network may face lengthy internal administrative processes—such as signing framework agreements or securing IRB approval—before they can join external studies.

“Exploration meeting”

In a meeting with WP 4/5 and UC leads and WP6, DPs will be shortlisted for feasibility, considering patient counts, minimum sample size requirement, data quality, contractual readiness, and resource prioritisation for the respective UC. From this meeting, WP-6 will include the following output in a final version of the exploration report that will be generated soon after the meeting:

* + List of DPs in order of priority (up to 10) to be invited for Feasibility
  + Documentation of the reasons for shortlisting (e.g. meeting notes)

Feedback mechanism – Relevance

After the preliminary assessment or the “exploration meeting”, if the proposed use case definition is deemed non-feasible, but alternative methods/research aims (e.g. adapted study objectives/design) are available, the next step is to meet with WP2/3 to reassess the relevance of the adapted suggestion and resubmit the use case. If necessary, the Regulatory and HTA Fora may be involved.

## Initiation

This phase mainly covers data partner contracting and will follow the next steps:

1. The EHDEN-F will **contact the top 5 of the selected list of data partners**, sharing the use case and the information that data partners might need to assess their capacity to participate (e.g. required data domains, work timelines, etc.). The identity (public/private) of the UC Leads might be also shared with data partners. Data partners respond regarding their interest and capacity to participate. If a data partner cannot participate in the study, the next one from the list will be selected.
2. WP6 will **contract** theinvited data partners who agreed to participate.
3. UC leads and WP4/5 will share at this point a Phenotype Proposal Form (PPF) (see Annex 2) with WP6. The PPF must include a clinical description, summarising the disease's epidemiology, presenting symptoms, treatments, and potential strengthening or disqualifying factors. If any preliminary **phenotyping** exercise is available in a computable format, they will also share it with WP6. It will be then reviewed from a ‘clinical/epi’ and from a ‘code/computing’ perspective.

## **Implementation**

Implementation covers mainly the phenotype and analytical code development, and include the following steps:

### Phenotype development

Three different might initiate phenotype development.

1. If **an existing computable phenotype** is provided, its feasibility must be evaluated and agreed upon by WP4, WP5, and WP6. This evaluation will consider both structural and functional complexity.
   1. Structural complexity includes 1) the number and type of data domains involved, 2) the number of concepts used, and 3) the presence of cohort intersections or criteria.
   2. Functional evaluation involves 1) the databases on which the provided phenotype has been tested (if any), meaning countries and type of databases (claims, prescriptions); 2) the results or counts obtained from those executions, and 3) the database management systems (DBMS) used during testing.
   3. Additional suitability criteria include 1) date of phenotype generation, 2) vocabulary version used during development (e.g., OMOP version >5.0), 3) relevant publications that reference the phenotype, and 4) prior validation efforts
2. If **a concept set with no other specifications is provided**, it will be run in that format assuming first-ever occurrence of the disease, exposure, outcome, or device.
3. If **no phenotype nor concept set is provided**, WP6 will create a new phenotype

Either if the phenotype is provided and agreed on (A, B) or the phenotype is newly created (C), the output of this phenotyping process will be a Preliminary Phenotype, which use is intended for feasibility analysis exclusively. Preliminary Phenotypes are defined by a concept set—comprising the code list plus logic—or, when appropriate, the concept set plus first-ever occurrence of the disease, exposure, outcome, or device. By default, the most inclusive (best-case) scenario will be assumed unless otherwise specified.

**A maximum of 5 phenotypes will be allowed** per feasibility assessment, so, consequently, some might be prioritised over others. These 5 phenotypes will cover the entry criteria of the cohort as in exposure / outcomes and required covariates to answer the proposed research question. The Use Case lead, WP4-5 and WP6 are responsible for adapting the preliminary phenotypes using OHDSI/DARWIN EU tools.

### Code development

1. A repositoryfor the UC is created in GitHub using the EHDEN-F’s “FeasibilityAssessmentTemplate” **template repository**. This repository will include the necessary directions to install and run the analytical code.
2. WP6/EHDEN-F will **adapt the analytical code and test** it using synthetic data. This analytical code will be based only previously validated/standard EHDEN-F/OHDSI/DARWIN EU tools. This code will only allow stratifications by sex or age, if any.
3. The **final** **preliminary phenotypes (feasibility phenotypes)** will be uploaded to the corresponding folder in the GitHub repository. A final code tests must be performed using these phenotypes. If the feasibility phenotypes were finally different from the preliminary ones, the UC lead would be notified.
4. WP-6 will give **access to the data partners** to the UC GitHub repository. An email notifying this will be sent shortly after.

A diagram of a chemical reaction

AI-generated content may be incorrect.

Figure 4. Phenotyping steps during implementation

## **Execution**

Execution mainly covers the execution of the code by the data partners, the gathering of the results, and the generation of the feasibility report. It follows these steps:

1. The data partners will **execute** the feasibility analytical code. They can report any issue through the corresponding GitHub repository. Adjustment to the analytical code and/or the phenotypes might be required during this stage. If that was the case and changes were made to code/phenotype, data partners and use case leads will be communicated. GitHub will be used as a version control tool for both code and phenotypes.
2. The data partners will **share the results** through the EHDEN-F data transfer zone (DTZ)  or a similar digital environment. EHDEN-F project management office will reach the data partners if the results are not shared in the defined time lapse.
3. The EHDEN-F team will explore the results to identify potential issues. If any issue is identified, the cause will be described. If additional runs of the code could be performed, the phenotyping and/or the analytical code would be updated. Data partners would be notified and kindly asked to run the code again. The UC lead would be notified too.
4. The EHDEN-F team will generate an **execution** **report** that includes:
   1. Counts of people for the target phenotypes stratified by data source/partner
   2. Attrition tables for the target phenotypes stratified by data source/partner
   3. Completeness of required and desired variables, in terms of missingness.
   4. Granularity of required and desired variables, if requested. Some UCs may require having information such as brand or route of administration for drugs, UID for devices, age in days (not years), etc.
   5. Timing for required and desired variables, if requested. Some UCs may require being able to determine onset, duration, recurrence, or resolution of outcomes.

## **Dissemination**

Once the feasibility report has been generated, a **“Execution meeting”** will be held between the EHDEN-F team and the UC lead/s. In this meeting, a final selection of data partners for the subsequent study will be reached and included in the final **data feasibility report**. This final report will include:

1. The **exploration report**, including:
   1. Flowchart of the preliminary assessment
   2. Table with patient counts for the selected concept IDs in the explored data partners
   3. Table with data quality, contractual readiness, and timeliness of the explored data partners
   4. The name/s and description of the shortlisted Data Partners selected for feasibility, and the reasoning for this selection (based on the “Exploration meeting” minutes)
2. The **execution** **report**, including:
   1. Counts of people and attrition table for the target phenotypes stratified by data source/partner
   2. Tables with completeness, granularity, and timing requirements, if requested.
3. **Final selection of Data Partners for the Use Case**, and the reasoning for this selection (based on the “Execution meeting” minutes)

# Feasibility Report and Stop/Go Meeting

Results from feasibility assessment on methods, data, and resources will be comprised in the GREG Feasibility Assessment Form (Annex I), the **“Feasibility Report”**. This will be shared with the UC Lead/s and Steering Committee as part of the Feasibility Report.

A Stop/Go meeting will be held after the Feasibility Report has been shared, where EHDEN-F, WP4-5-6 leads and UC leads meet to decide if the study is going ahead. Ideally, this meeting will be organized for the full wave of relevant / prioritized Use Cases, and all of them will be discussed at the same time. This will facilitate decisions on contracting of data partners and study team/s formation.

# Timelines and Responsibilities

A screenshot of a computer

AI-generated content may be incorrect.

Figure 5. Gantt Chart of the Feasibility Assessment process

# References

1. Dernie F, Corby G, Robinson A, Bezer J, Mercade-Besora N, Griffier R, Verdy G, Leis A, Ramirez-Anguita JM, Mayer MA, Brash JT, Seager S, Parry R, Jodicke A, Duarte-Salles T, Rijnbeek PR, Verhamme K, Pacurariu A, Morales D, Pinheiro L, Prieto-Alhambra D, Prats-Uribe A. Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN EU). Pharmacoepidemiol Drug Saf. 2024 Nov;33(11):e70042. doi: 10.1002/pds.70042.
2. Observational Health Data Sciences and Informatics (OHDSI). (n.d.). The Book of OHDSI. Retrieved from [https://ohdsi.github.io/TheBookOfOhdsi/](https://ohdsi.github.io/TheBookOfOhdsi/)
3. European Medicines Agency. Data Quality Framework for EU Medicines Regulation. European Medicines Agency, https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/data-quality-framework-eu-medicines-regulation\_en.pdf
4. Gatto NM, Campbell UB, Rubinstein E, Jaksa A, Mattox P, Mo J, Reynolds RF. The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework. Clin Pharmacol Ther. 2022 Jan;111(1):122-134. doi: 10.1002/cpt.2466. Epub 2021 Dec 1. PMID: 34716990; PMCID: PMC9299818.
5. Patel D, Guleria S, Titievsky L, Flaherty S, Everage N, Korjagina M, Porkess S, Kou TD, Layton D. Contemporary Practice and Considerations for Real-World Data Source Identification and Feasibility Assessment. Pharmacoepidemiol Drug Saf. 2024 Sep;33(9):e5862. doi: 10.1002/pds.5862.

# Annex I: GREG Feasibility Assessment Report

# ADMINISTRATIVE DETAILS

Use case identifier [from submission form]:  {{ use\_case\_id }}

Use case title:  {{ case\_title }}

Use Case Lead: [{{ case\_lead }}]

Check of completeness of Use Case form conducted by: [{{ completeness }}]

Date of initial feasibility request: [{{ initial }}]

Person(s) responsible for the Feasibility Assessment: [{{ responsible }}]

# ASSESSMENT OF METHODS REQUIREMENTS/FEASIBILITY

*Based on the study type and objectives defined in the use case submission form, this section defines the required methods/statistical analyses for the study.*

Assessment criteria:

|  |  |
| --- | --- |
| Required analyses | *e.g. Incidence/Prevalence; Cohort study* |
| Standard analytics available | *e.g. yes (analytical package); no (cohort study would be using PS matching/weighting)* |
| Guidelines/standard methods available? | *Are there any guidelines/recommendations of state-of-the-art methods that will be used or tested against?* |

RECOMMENDATIONS/PROPOSED METHODS

*Which analytical methods would you recommend to use for this study? What would be their limitations? Is the analytical method (incl. statistical code/packages) available or would it needed to be developed?*

# ASSESSMENT OF RESOURCES/TIMELINES FEASIBILITY

Assessment criteria:

|  |  |
| --- | --- |
| Required FTE | *e.g. %FTE of X staff members working on the study* |
| Data costs | *e.g. X databases* |
| Additional resources? | *Is there a specific contribution (FC/in-kind) available if this use case would be conducted?* |
| Would any extra efford be needed that incures additional costs? | *e.g. mapping of non-OMOP data, mapping medical devices* |
| Timeline | What would be the estimated timeline for completion of the study? |

RECOMMENDATIONS/ESTIMATED RESOURCES

What would be feasible timelines for completion of the study and how many resources would be needed?

# ASSESSMENT OF DATA FEASIBILITY

This section assesses the requirements on data (domain availability, setting, geography, timelines, etc.), relevant feasibility counts, and reports the data partner selection process. Based on these, a recommendation on potential data partners is being made. The start date is reported for reproducibility purposes.

**Start date:** DD/MM/YYYY

Exploration report

This section reports the process for data partner shortlisting for the data feasibility assessment. From the submission form, the following information was extracted regarding the data requirements of this use case:

|  |  |
| --- | --- |
| Healthcare setting suitability | *To be completed with the UC setting needs, e.g., primary care* |
| Temporal suitability | *To be completed with the UC temporal needs, e.g., data from > 2021, data from <2010* |
| Geographical suitability | *To be completed with the UC geographical needs, e.g., data from southern European countries* |
| Presence of necessary data domains | *To be completed with the UC data domains needs, e.g., conditions, drugs, and procedures.* |

After exploring the EHDEN-F data partners specifications, the following flowchart has been produced reporting the number of data partners that do not meet each criterion. This selection process is **systematic** and transparent.

A diagram of a flowchart

AI-generated content may be incorrect.

Figure 6. Flowchart of preliminary assessment of data requirements in EHDEN-F data partners

Landscaping was carried out with the data partners that met all the requirements. As a result, a table reporting the patient counts for the selected concept IDs in these data partners is shown below.

Table 1. Patient counts obtained during Landscaping on the data partners meeting the preliminary assessment criteria.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **DP1** | **DP2** | **DP3** | **DP4** | **DP5** | **DP6** | **DP7** | **DP8** | **DP9** | **DP10** | **…** |
| *Concept I* |  |  |  |  |  |  |  |  |  |  |  |
| *Concept 2* |  |  |  |  |  |  |  |  |  |  |  |
| *Concept 3* |  |  |  |  |  |  |  |  |  |  |  |
| *Concept 4* |  |  |  |  |  |  |  |  |  |  |  |
| *Concept 5* |  |  |  |  |  |  |  |  |  |  |  |
| *Concept 6* |  |  |  |  |  |  |  |  |  |  |  |
| *Concept 7* |  |  |  |  |  |  |  |  |  |  |  |
| *Concept 8* |  |  |  |  |  |  |  |  |  |  |  |
| *Concept 9* |  |  |  |  |  |  |  |  |  |  |  |
| *Concept 10* |  |  |  |  |  |  |  |  |  |  |  |
| *…* |  |  |  |  |  |  |  |  |  |  |  |

The data quality, contractual readiness, and expected timelines of these same data partners are shown in the following table. Data quality is reported as a metric showing the percentage of tests passed, from an exhaustive list of data quality tests. Contractual readiness reports whether the data partner already has a memorandum of understanding or a framework agreement signed with the EHDEN-F. Timelines report the expected timelines for the data partner to participate in a study

Table 2. Data quality, contractual readiness, and expected timelines of the data partners meeting the preliminary assessment criteria.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data partner** | **Data quality** | **Contractual readiness** | **Expected timelines** |
| DP1 |  |  |  |
| DP2 |  |  |  |
| DP3 |  |  |  |
| DP4 |  |  |  |
| DP5 |  |  |  |
| DP6 |  |  |  |
| DP7 |  |  |  |
| DP8 |  |  |  |
| DP9 |  |  |  |
| DP10 |  |  |  |
| … |  |  |  |

Having access to these results, during the “Exploration meeting” the following was decided for each data partner:

Table 3. Comments on the data partners meeting the preliminary assessment criteria during the “Exploration meeting”.

|  |  |
| --- | --- |
| **Data partner** | **Decision** |
| DP1 |  |
| DP2 |  |
| DP3 |  |
| DP4 |  |
| DP5 |  |
| DP6 |  |
| DP7 |  |
| DP8 |  |
| DP9 |  |
| DP10 |  |
| … |  |

Therefore, the list of data partners, sorted by preference, to be invited for the data feasibility assessment is as follows. A number of them were invited according to resource requirements.

1. Data partner to be invited
2. Data partner to be invited
3. Data partner to be invited
4. Data partner to be invited
5. Data partner to be invited
6. Data partner to be invited
7. Data partner to be invited
8. Data partner to be invited
9. Data partner to be invited
10. Data partner to be invited

Execution report

This section reports the process for data partner shortlisting for the study. From the submission form and following discussions with the UC lead, the following information was extracted regarding the data requirements of this use case:

|  |  |
| --- | --- |
| Completeness of required and desired variables | *To be completed with the UC completeness needs, e.g., a maximum of 5% of missingness in dose coverage* |
| Granularity | *To be completed with the UC granularity needs, e.g., route of administration for drugs* |
| Timing for required and desired variables | *To be completed with the UC timing needs, e.g., need to estimate duration of treatment* |

Compliance with these criteria by invited data partners can be seen in the following table:

Table 4. Compliance of data partners selected for feasibility with additional data requirements

|  |  |  |  |
| --- | --- | --- | --- |
| **Data partner** | **Completeness** | **Granularity** | **Timing** |
| DP1 |  |  |  |
| DP2 |  |  |  |
| DP3 |  |  |  |
| DP4 |  |  |  |
| DP5 |  |  |  |
| DP6 |  |  |  |
| DP7 |  |  |  |
| DP8 |  |  |  |
| DP9 |  |  |  |
| DP10 |  |  |  |

Phenotype counts in the form of patient counts for each data partner can be seen in the following table:

Table 5. Patient counts of each phenotype in each data partner. P1 to P5 refer to Phenotype 1 to 5.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data partner** | **P1** | **P2** | **P3** | **P4** | **P5** |
| DP1 |  |  |  |  |  |
| DP2 |  |  |  |  |  |
| DP3 |  |  |  |  |  |
| DP4 |  |  |  |  |  |
| DP5 |  |  |  |  |  |
| DP6 |  |  |  |  |  |
| DP7 |  |  |  |  |  |
| DP8 |  |  |  |  |  |
| DP9 |  |  |  |  |  |
| DP10 |  |  |  |  |  |

The attrition tables obtained from each data partner are included below. The order of application of the inclusion criteria is that of the rows of the table. For each one, the number of records and number of subjects included and excluded are reported.

| **Reason** | **Variable name** | | | |
| --- | --- | --- | --- | --- |
| **number\_records** | **number\_subjects** | **excluded\_records** | **excluded\_subjects** |
| **DP 1; P1** | | | | |
| Qualifying initial records |  |  |  |  |
| Criteria 1 |  |  |  |  |
| Criteria 2 |  |  |  |  |
| **DP1; P2** | | | | |
| Qualifying initial records |  |  |  |  |
| Criteria 1 |  |  |  |  |
| Criteria 2 |  |  |  |  |
| Criteria 3 |  |  |  |  |

Having access to these results, during the “Execution meeting” the following was decided for each data partner:

Table 6. Comments on the data partners meeting the preliminary assessment criteria during the “Execution meeting”.

|  |  |
| --- | --- |
| **Data partner** | **Decision** |
| DP1 |  |
| DP2 |  |
| DP3 |  |
| DP4 |  |
| DP5 |  |
| DP6 |  |
| DP7 |  |
| DP8 |  |
| DP9 |  |
| DP10 |  |

Therefore, the list of data partners, sorted by preference, to be invited for the subsequent study is as follows.

1. Data partner to be invited
2. Data partner to be invited
3. Data partner to be invited
4. Data partner to be invited
5. Data partner to be invited

**OVERALL ASSESSMENT/RECOMMENDATIONS**

[ ] Feasible or [ ] Not feasible

COMMENTS/RECOMMENDATIONS

# Annex II: Phenotype Proposal Form

|  |  |
| --- | --- |
| **Phenotype Proposal Form including Clinical Description** | |
| **Name:** |  |
| **Description** |  |
| **Study ID** |  |
| **UC Lead** |  |
| **Purpose** | **Feasibility / Study** |
| **UC Lead E-mail** |  |
| **Clinical Description** | Overview:  A comprehensive introduction to the clinical phenotype, providing a general understanding of the condition or disease.    Presentation:  Description of the typical signs and symptoms that patients with the clinical phenotype may experience.    Assessment:  Explanation of the diagnostic tests and procedures used to assess and confirm the presence of the clinical phenotype. This may include laboratory tests, imaging studies, and other diagnostic techniques. Criteria and factors considered in establishing a diagnosis of the clinical phenotype, including specific markers, characteristics, or clinical findings.    Plan:  An outline of the treatment and management strategies commonly employed for patients with the clinical phenotype. This may include medications, therapies, surgical interventions, and supportive care measures.    Prognosis:  Factors that influence the prognosis or expected outcome of individuals with the clinical phenotype. This section may identify both positive and negative prognostic indicators. Potential complications or disease progression patterns associated with the clinical phenotype. This section may describe the development of related conditions or transformation into other diseases. |