Real-World Data Management Systems

Development of **standardized** and **centralized** data repositories and management systems for **Real World Evidence**

enabling multi-center projects on clinical research studies

**Web-based applications**

Provide data availability and security in a user friendly way
Simplicity, data reliability, easiness in use and flexibility as necessary conditions

*User – friendly, adjusted to user requirements*

---

**Real-world evidence (RWE)** biomedical research is an increasingly important component of INAB | CERTH activities
RWD management systems

OBJECTIVES

A. collect and transform clinically relevant RWD into evidence

B. correlate data in order to provide accurate information about the diagnosis, prognostic assessment and management

C. facilitate translational and clinical research

Organize and analyze clinically relevant data from the daily practice, by gathering homogenized high-quality datasets, thus improving the quality and delivery of medical care
RWD management systems

APPROACH

1. Standardization & Harmonization
   Agreement on common policies and procedures with standards-based approaches, paving the way for standardized registration of RWD

2. Simplicity & Flexibility
   Design in expandable modules allowing the rapid introduction of additional data categories and extension of data domain when and if needed

3. Data Availability & Usability
   Unified access to clinically relevant structured data while fulfilling the data quality assurance requirements
RWD management systems

AIM

Combine information into a data integration framework

- Data integrity, data accuracy and quality assessment
- Data correlations and statistical analysis
- Data analytics, Data mining and knowledge discovery
- User-friendly integration of ontologies, terminologies and standards
- Data protection, security and availability
RWD management systems

DATA PROTECTION

**GDPR compliance**
- Data are collected solely for specific and legitimate purposes
- Data are adequate and related to the purpose of collection
- Data are processed and treated lawfully and fairly in a transparent manner

**Data anonymization procedures**
- Pseudonymized identifier for each patient
- The correspondence between original and anonymized records will be stored only at the local level.
- Fully anonymized data used for statistical analysis. Aggregated results are reported.

**ISO 27001:2013**
for Information Security Management System

**ISO 22301:2019**
for Business Continuity Management System

**Appropriate technical and organizational measures**
- data confidentiality, availability and integrity
- data protection and security
RWD management systems

DATA MODEL

Describe all the relevant information to form an accurate and complete representation of patient diagnosis and disease course
RWD management systems

**DATA MODEL**

- Patient
  - Diagnosis setting
  - Follow-up of disease course

- Assessment
  - Investigations
  - Lab results
  - Molecular data
    - Low-Throughput
    - High-Throughput
      - summaries, references

- Treatment options
  - Patient status
  - Disease phase
  - Observations / Events / Related diseases

- Patient and disease characteristics
  - Demographic Data
  - History
  - Family history
  - Social history
  - Medical history
  - Comorbidities
  - Response to treatment
  - Adverse events
  - Interpretations
  - Mutation analysis
DATA REQUIREMENTS

Detailed description of data

<table>
<thead>
<tr>
<th>Field</th>
<th>Data type</th>
<th>Data format</th>
<th>Unit</th>
<th>Value constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td>list</td>
<td>Predefined</td>
<td></td>
<td>CLL,MBL,SLL</td>
</tr>
<tr>
<td>Date of diagnosis</td>
<td>date</td>
<td>DD/MM/YYYY</td>
<td></td>
<td>&gt;1980</td>
</tr>
<tr>
<td>Rai stage at diagnosis</td>
<td>list</td>
<td>Predefined</td>
<td></td>
<td>0,I,II,III,IV</td>
</tr>
<tr>
<td>Binet stage at diagnosis</td>
<td>list</td>
<td>Predefined</td>
<td></td>
<td>A,B,C</td>
</tr>
<tr>
<td>Comorbidities at diagnosis</td>
<td>list</td>
<td>Predefined</td>
<td></td>
<td>YES,NO</td>
</tr>
<tr>
<td>Blood count date</td>
<td>date</td>
<td>DD/MM/YYYY</td>
<td></td>
<td>less than or equal to current date</td>
</tr>
<tr>
<td>Hb</td>
<td>decimal</td>
<td>#0.0-100.0</td>
<td>g/dL</td>
<td>0.0-100.0</td>
</tr>
<tr>
<td>WBC (x10^9/l)</td>
<td>decimal</td>
<td>#0.00-1000.00</td>
<td>(x10^9/l)</td>
<td>0.00-1000.00</td>
</tr>
<tr>
<td>Treatment status</td>
<td>list</td>
<td>Predefined</td>
<td></td>
<td>Treated,Untreated</td>
</tr>
<tr>
<td>Response to first treatment</td>
<td>list</td>
<td>Predefined</td>
<td></td>
<td>CR,CRI,Nodular PR,PR,PR with lymphocytosis,Progressive,Stable</td>
</tr>
<tr>
<td>Date of last follow-up</td>
<td>date</td>
<td>DD/MM/YYYY</td>
<td></td>
<td>less than or equal to current date</td>
</tr>
<tr>
<td>Survival status</td>
<td>list</td>
<td>Predefined</td>
<td></td>
<td>Alive,Dead</td>
</tr>
</tbody>
</table>

- Detailed list of data categories, entities, terminologies
- Definition of data types, data format and allowed values
- Configuration of data relationships and rules
- Consistency and integrity constraints

User scenarios and workflows
RWD management systems

DATA REQUIREMENTS

— Consistency and integrity constraints

**Required** information

<table>
<thead>
<tr>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID</td>
</tr>
<tr>
<td>Diagnosis</td>
</tr>
<tr>
<td>Date of diagnosis</td>
</tr>
<tr>
<td>Rai stage at diagnosis</td>
</tr>
<tr>
<td>Binet stage at diagnosis</td>
</tr>
<tr>
<td>Treatment status</td>
</tr>
</tbody>
</table>

**Data redundancy** control

*Examples:*
- The Patient id must uniquely characterize one patient
- A patient has only 1 RAI stage at diagnosis record
- A patient cannot have 2 visits with the same date

**Depended / Essential** information

*Examples:*
- If patient is reported DEAD, then Date of death is a required field
- If patient is reported Treated, then Start date of treatment and treatment type are required fields

**Additional constraints and rules for the comparison of related fields**

*Examples:*
- Start date of treatment cannot be before Date of diagnosis
- Treatment-related data are not applicable for Untreated patients
- A patient with enlarged lymph nodes at time of diagnosis cannot be reported with RAI stage: 0
Center - based access

Project - based access

Lab-based access

Controlled access to data of the registration center

✓ Role-based access
Role assignment: Restricted or Extended Access Privileges
RWD management systems

USER MANAGEMENT

User role examples of **center-based** access

**Principal investigator of center (PI)**
Overall access, review registration and management procedures

**Data administrator (DA)**
Overall read & write privileges

**Data moderator (DM)**
Restricted read & write permissions to selected data categories

**Read - only user**
Restricted read permissions to selected features

Role-assignment example

**Cytogenetics-lab User**
Access only to cytogenetic data

User role example of **project-based** access

**General PI**
Read-only access to data from all centers in a project
RWD management systems

SYSTEM DESIGN

Users

User authentication system

User authorization

Define access privileges

Web-Based Application

Database

Tools

Data management tools
- Data registration
- Data update
- Data import
- Data validation

Data retrieval tools
- Search
- Filter
- Preview
- Download

Data analysis tools
- Visualization and Export modules
DATA COLLECTION METHODS

Real-time registration systems

A. Online web forms for prospective data registration

✓ Organized and Dynamically adjusted according to selected criteria
RWD management systems

DATA COLLECTION METHODS

Real-time registration systems

B. Retrospective data registration tools

* specifically designed template registration files according to requirements

* adjusted to local necessities and project-specific data collection

✓ Protected sheets, including data validation
  - warnings, notes, validation rules and drop down lists

Data upload, validation and import tool

facilitate data collection in a retrospective way

integrate multi-originated data from different sources into a central repository
RWD management systems

DATA COLLECTION METHODS

Data Validation and Integration

- **translation** of terms / **conversion** of data
- **validation** of data formats
- detection of **data redundancy**
- additional constraints and rules for the **comparison of related fields**

Retrospective data registration

Data validation procedure → Data validation reports

Data validation report

Upload of Amsterdam dataset

Dataset status
- Center: Amsterdam
- Number of uploaded cases: 43
- Number of validated cases: 11 / 43 (number of cases with errors: 32)
- Error: 2 (error occurs if required)
- Total number of errors: 56

Data validation report
- Mean data (35 rows):
  - 15 / 45 processed rows with data have been validated
  - Number of rows with errors in sheet: 28
  - There are warnings (4) to review in this sheet
- Subsequent lines of treatment (3 rows):
  - 3 / 4 processed rows with data have been validated
  - Number of rows with errors in sheet: 5
  - There are warnings (2) to review in this sheet
  - Reference exists in the sheet for 4 / 45 (8.89%) cases

Download data validation report in a tab-delimited text file.
RWD management systems

Data Validation and Integration

- **translation** of terms / **conversion** of data
- **validation** of data formats
- detection of **data redundancy**
- additional constraints and rules for the **comparison of related fields**

Retrospective data registration

Data validation procedure  ➔  Data validation reports

Re-upload  ➔  Corrected dataset  ➔
RWD management systems

DATA COLLECTION METHODS

Data Validation and Integration

- **translation** of terms / **conversion** of data
- **validation** of data formats
- detection of **data redundancy**
- additional constraints and rules for the **comparison of related fields**

**Retrospective data registration**

Data validation procedure ➔ Data validation reports

Data curation

Data organization

Data homogenization

Data import

Central Database
RWD management systems

DATA RETRIEVAL AND ANALYSIS

1. data selection
   - **Query tools** supporting dynamic definition of selection filters.

2. data retrieval
   - Including data visualization modules

3. data export options
RWD management systems

DATA ANALYTICS

A web-based application for online statistical analysis

✓ Project-based configuration
RWD management systems

DATA ANALYTICS

A web-based application for online statistical analysis

Data correlations
RWD management systems

ERIC LL database

Collection of prospective and retrospective clinical and biological data from patients with Chronic Lymphocytic Leukemia at the time of diagnosis and follow-up on a project basis

A project of ERIC, the European Research Initiative on CLL, a Scientific Working Group (SWG) of the European Hematology Association (EHA) aimed at improved management of CLL through collaborative research

http://www.ericll.org/

Supervised by INAB
a large-scale initiative aimed at addressing the outstanding basic, translational and clinical research questions in CLL

Current status

Cases: 20463
Participating Centers: 91
Countries: 31

Challenges and Solutions for Collecting and Analyzing Real World Data: The ERIC CLL Database as an Illustrative Example

Anastasia Chatzidimitriou, E Vu Minga, Thomas Chatzikonstantinou, Caro Moreno, Konstantinos Stamatopoulos, Paolo Ghia, on behalf of ERIC, the European Research Initiative on CLL.

Challenges of gathering high-quality real-world data

Collection and analysis of real-world data (RWD) can prove both effective and efficient for advancing precision medicine and improving the quality and delivery of medical care, provided these come along with data quality.

The amount of biomedical data continuously improves due to technological advances, thus raising the necessity for designing and developing standardized approaches and methodologies to be implemented in clinical practice.

Data acquisition is usually a process distributed among different health professionals potentially leading to data quality problems across domains, such as data redundancy, heterogeneity (e.g., different data formats and inconsistencies), e.g., due to diagnoses after the data of treatment, mainly resulting from lack of standardization and data correlation processes. Such problems are particularly pertinent in the case of multi-institutional efforts, whose multidimensional and multi-engaged data are collected. Furthermore, the rapid increase of data complexity captured during patient care, especially data produced by the application of novel methodologies, can generate sequencing, poses challenges that cannot be addressed with standard computational approaches.

Thus, there is an imperative to impose real-world evidence generation by optimizing the integration of the heterogeneous information through automated and thorough quality control and correlation mechanisms and, support and harmonize with multibioinformatics. This will provide suitable and standardized access to valid, accurate and comparable data. Practical and feasible tools are required, capable of providing context in use, flexibility and simplicity, in order to facilitate the fast entry procedures and encourage the registration and organization of clinically relevant data from the daily practice. (4)

Towards the development of a unified data management framework

Harmonization of heterogeneous data is a prerequisite for gathering homogeneous high-quality datasets and solving the many forms of methodological and medical information. A common approach that can be adapted to local and protocol-specific requirements, will inevitably facilitate biological tran...
RWD management systems

THE ERIC CLL DATABASE

Online platform for project data management and analysis

ERIC CLL database

Registry of clinical and biological data of patients with Chronic Lymphocytic Leukemia
- A project of ERIC, the European Research Initiative on CLL -
www.ericll.org

Welcome to the ERIC data management system
for the collection of prospective and retrospective data on a project basis
- You should be an authorized user to login.
RWD management systems

THE ERIC CLL DATABASE

Online platform for project data management and analysis

Project-based data collection

---

**Treatment Sequencing**

Real-World Evidence on Therapeutic Strategies and Treatment Sequencing in Patients with Chronic Lymphocytic Leukemia: An International Study of ERIC, the European Research Initiative on CLL

---

Current registration status:

- **80** Participating Centers
- **28** Countries
- **291** Uploaded Cases
- **258** Verified Cases

---

- ![Detailed Project Status](#)
- ![View List of Participating Centers](#)
- ![Add New Participating Center](#)
- ![Upload a Received File for Data Validation](#)
- ![Upload and Save a Final/ Curated Dataset](#)
- ![Review Final Datasets](#)
RWD management systems

THE ERIC CLL DATABASE

Online platform for project data management and analysis

Upload tool

Upload form

Template Name
ERIC_CLL-db_TreatSeq

Participating Center
Amsterdam - Amsterdam Medical Center, Hematology, (Amsterdam, NL)

File
AmsterdamERIC_CLL-db_TreatSeq.xlsx

Data validation report

Dataset status
Center code: Amsterdam
Number of uploaded cases: 43
Number of validated cases: 32 (number of cases with errors: 11)

Data validation report

Main data (44 rows)
- 2 rows with errors (28)
- 1 row with warnings (5)

Subsequent lines of treatment (8 rows)
- 1 row with errors (5)
- 1 row with warnings (5)
- Reference exists in this sheet for 7 rows (5/799) cases
RWD management systems

THE ERIC CLL DATABASE

Online platform for project data management and analysis

Project overview

Data overview

Center: Amsterdam

- Number of cases: 43
- Median age at diagnosis: 68 years
- Median survival time: 60 months
- Median time to treatment: 22 months

Gender distribution:
- Male
- Female

Treatment status:
- Treated
- Untreated

Survival status:
- Alive
- Dead

Total lines of treatment percentage

No of patients diagnosed per year

Developed and hosted by Institute of Applied Biosciences (INAB) | Centre for Research and Technology Hellas (CERTH)