IMI 116026 HARMONY; IMI 945406 HARMONY PLUS - HARMONY Big Data Platform

First published: 01/02/2024

Last updated: 17/10/2024

Data source **Human** **Cancer registry** **Other**

Registration with healthcare system

Administrative details

Administrative details

Data source ID

1111137

Data source acronym

HARMONY Big Data Platform

Data holder

HARMONY Alliance

Data source type

Cancer registry

Other

Registration with healthcare system

Data source type, other

clinical trial and real-world datasets on haematologic malignancies (HMs) from pharmaceutical companies, hospitals, cooperative working groups, registries, and interventional and non-interventional trials

---

Main financial support

Funding from public-private partnership

Care setting

Other

Primary care – specialist level (e.g. paediatricians)

Data source qualification

If the data source has successfully undergone a formal qualification process (e.g., from the EMA, ISO or other certifications), this should be described.

Yes

Description of the qualification

CHMP Qualification of novel methodologies for medicine development.

EMA/SA/0000122783

Data source website

https://www.harmony-alliance.eu

Contact details

HARMONY Coordination Office harmonyoffice@ibsal.es

Main

harmonyoffice@ibsal.es

Data source regions and languages

Data source countries

Austria

Belgium

Brazil

Canada

Denmark

France

Germany

Greece

Italy

Kuwait

Netherlands

Poland

Qatar

Singapore

Spain

Sweden

Switzerland

Taiwan

United Kingdom

United Kingdom (Northern Ireland)

Data source languages

English

Data source establishment

Data source established

15/06/2017

Data source time span

First collection: 01/09/2018

The date when data started to be collected or extracted.

Publications

Data source publications

Szócska M, Jayawardana S, Smand C, et al. Big data for better outcomes: supporting health care system transformation in Europe. Eurohealth. 2017;23(1):1-5. http://eprints.lse.ac.uk/69798/

Malcikova J, Tausch E, Rossi D, et al. ERIC recommendations for TP53 mutation analysis in chronic lymphocytic leukemia - Update on methodological approaches and results interpretation. Leukemia. 2018;32(5):1070-1080. https://doi.org/10.1038/s41375-017-0007-7

do Valle ÍF, Menichetti G, Simonetti G, et al. Network integration of multi-tumour omics data suggests novel targeting strategies. Nature Communications. 2018;9(1):4514. https://doi.org/10.1038/s41467-018-06992-7

Curti N, Giampieri E, Levi G, et al. DNetPRO: A network approach for low-dimensional signatures from high-throughput data. Scientific Reports. 2022;12(1):22253. https://doi.org/10.1038/s41598-022-25549-9

de Waal EPB. Hematology and Big Data, Enabling Better and Faster Treatment. Hemasphere. 2017;1(1):e17. https://doi.org/10.1097/hs9.0000000000000017

Data elements collected

The data source contains the following information

Disease information

Does the data source collect information with a focus on a specific disease? This might be a patient registry or other similar initiatives.

Yes

Disease details

Acute myeloid leukaemia

Chronic lymphocytic leukaemia

Chronic myeloid leukaemia

Polycythaemia vera

Myelofibrosis

Disease details (other)

Malignant hematological disorders (HMs): acute lymphoblastic leukaemia (ALL), myelodysplastic syndromes (MDS), multiple myeloma (MM), Waldenström macroglobulinemia (WM), non-Hodgkin lymphoma (NHL), Hodgkin lymphoma (HL), myeloproliferative neoplasms (MPN), essential thrombocythemia (ET), myelofibrosis (MF), paediatric HMs and other rare blood cancers

Rare diseases

Are rare diseases captured? In the European Union a rare disease is one that affects no more than 5 people in 10,000.

Yes

Pregnancy and/or neonates

Does the data source collect information on pregnant women and/or neonatal subpopulation (under 28 days of age)?

No

Hospital admission and/or discharge

No

ICU admission

Is information on intensive care unit admission available?

No

Cause of death

Captured

Cause of death vocabulary

Other

SNOMED

Cause of death vocabulary, other

HARMONY generated

Prescriptions of medicines

Not Captured

Dispensing of medicines

Not Captured

Advanced therapy medicinal products (ATMP)

Is information on advanced therapy medicinal products included? A medicinal product for human use that is either a gene therapy medicinal product, a somatic cell therapy product or a tissue engineered products as defined in Regulation (EC) No 1394/2007 [Reg (EC) No 1394/2007 Art 1(1)].

No

Contraception

Is information on the use of any type of contraception (oral, injectable, devices etc.) available?

No

Indication for use

Does the data source capture information on the therapeutic indication for the use of medicinal products?

Not Captured

Medical devices

Is information on medicinal devices (e.g., pens, syringes, inhalers) available?

No

Administration of vaccines

No

Procedures

Does the data source capture information on procedures (e.g., diagnostic tests, therapeutic, surgical interventions)?

Captured

Procedures vocabulary

Other

SNOMED

Procedures vocabulary, other

HCPCS, HemOnc, HARMONY generated

Healthcare provider

Is information on the person providing healthcare (e.g., physician, pharmacist, specialist) available? The healthcare provider refers to individual health professionals or a health facility organisation licensed to provide health care diagnosis and treatment services including medication, surgery and medical devices.

Yes

Clinical measurements

Is information on clinical measurements (e.g., BMI, blood pressure, height) available?

Yes

Genetic data

Are data related to genotyping, genome sequencing available?

Captured

Genetic data vocabulary

HGNC

Other

Genetic data vocabulary, other

LOINC, CPT4, OMOP Extension, OMOP Genomic, HARMONY generated

Biomarker data

Does the data source capture biomarker information? The term “biomarker” refers to a broad subcategory of medical signs (objective indications of medical state observed from outside the patient), which can be measured accurately and reproducibly. For example, haematological assays, infectious disease markers or metabolomic biomarkers.

Captured

Patient-reported outcomes

Is information on patient-reported outcomes (e.g., quality of life) available?

Yes

Patient-generated data

Is patient-generated information (e.g., from wearable devices) available?

No

Units of healthcare utilisation

Are units of healthcare utilisation (e.g., number of visits to GP per year, number of hospital days) available or can they be derived? Units of healthcare utilisation refer to the quantification of the use of services for the purpose of preventing or curing health problems.

Yes

Unique identifier for persons

Are patients uniquely identified in the data source?

Yes

Diagnostic codes

Captured

Diagnosis / medical event vocabulary

Other

SNOMED

Diagnosis / medical event vocabulary, other

ICDO3, LOINC, HARMONY generated

Medicinal product information

Captured

Medicinal product information collected

Active ingredient(s)

Brand name

Dosage regime

Route of administration

Medicinal product vocabulary

Other

RxNorm

If 'other,' what vocabulary is used?

SNOMED, ATC, NDFRT, Nebraska Lexicon, HemOnc, HARMONY generated

Quality of life measurements

Not Captured

Lifestyle factors

Not Captured

Sociodemographic information

Captured

Sociodemographic information collected

Age

Gender

Quantitative descriptors

Population Qualitative Data

Population age groups

Paediatric Population (< 18 years)

Preterm newborn infants (0 – 27 days)

Term newborn infants (0 – 27 days)

Infants and toddlers (28 days – 23 months)

Children (2 to < 12 years)

Adolescents (12 to < 18 years)

Adults (18 to < 46 years)

Adults (46 to < 65 years)

Elderly (≥ 65 years)

Adults (65 to < 75 years)

Adults (75 to < 85 years)

Adults (85 years and over)

Estimated percentage of the population covered by the data source in the catchment area

30%

Population

Population size

45769

Active population size

25868

Population by age group

Age group	Population size	Active population size

Paediatric Population (< 18 years)	8406	7228
Infants and toddlers (28 days – 23 months)	421	310
Children (2 to < 12 years)	6098	5512
Adolescents (12 to < 18 years)	1887	1406
Adults (18 to < 46 years)	5402	3193
Adults (46 to < 65 years)	13546	7467
Elderly (≥ 65 years)	16974	7751
Adults (65 to < 75 years)	9483	4782
Adults (75 to < 85 years)	6194	2552
Adults (85 years and over)	1297	417

Median observation time

Median time (years) between first and last available records for unique individuals captured in the data source

4.50

Median time (years) between first and last available records for unique active individuals (alive and currently registered) capt

6.00

Data flows and management

Access and validation

Governance details

Documents or webpages that describe the overall governance of the data source and processes and procedures for data capture and management, data quality check and validation results (governing data access or utilisation for research purposes).

116026_HARMONY_D1.11_Data Quality Supervision Committee_v1.6

English (266.43 KB - PDF)

View document

116026_HARMONY_D4.10 Definition of user access requirements_vf

English (563.57 KB - PDF)

View document

HARMONY_Platform_Data_Flow_v1

English (1.04 MB - PDF)

View document

Biospecimen access

Are biospecimens available in the data source (e.g., tissue samples)?

No

Access to subject details

Can individual patients/practitioners/practices included in the data source be contacted?

Yes

Description of data collection

The data provider prepares the “Data Dictionary” that describes the data to be submitted, listing and defining all fields with all possible values. Common data dictionaries exist in some HMs, to facilitate this process. The data providers then prepare the data itself, ensuring all fields are fully defined, and subjecting it to a first stage of pseudonymisation. The pseudonymized data are submitted through a Trusted Third Party (TTP), following required file names and formats. The TTP takes the data through another pseudonymisation process before uploading to the HARMONY Platform. Then, a Data Quality Supervision Committee (DQSC) verify and check the quality of the data from a clinical and technical point of view to guarantee compliance with different quality levels. After pseudonymisation and quality control, the data are standardised to a common model to address the heterogeneity of incoming data from different sources: OMOP CDM and custom concepts created by HARMONY when they are absent in OMOP. The approach consists of transforming the information from the origin databases into a common representation in terms of vocabularies, terminologies, or coding schemes. If necessary, the data provider and clinicians will validate the harmonisation process to ensure completeness and trustworthiness of the data. After harmonisation is complete, the data are deposited in the HARMONY Platform.

Custom HARMONY OMOP concepts are subject to validadation from OHDSI relevant workgroups (https://www.ohdsi.org/workgroups/)

Currently the HARMONY Platform is mainly based on secondary used of anonymized data and not includes EHR. Therefore, there is not an specific way to creating a record (pregunta 34), to register a person (pregunta 35) o to de-register a person (pregunta 36) and it is not linked to another data sources (pregunta 37). Pilot projects to include EHR information are foreseen in Q4 2023.

Data source linkage

Linkage

Is the data source described created by the linkage of other data sources (prelinked data source) and/or can the data source be linked to other data source on an ad-hoc basis?

No

Data management specifications that apply for the data source

Data source refresh

Yearly

Informed consent for use of data for research

Other

Possibility of data validation

Can validity of the data in the data source be verified (e.g., access to original medical charts)?

Yes

Data source preservation

Are records preserved in the data source indefinitely?

No

Data source preservation length (years)

10 years

Approval for publication

Is an approval needed for publishing the results of a study using the data source?

Yes

Informed consent, other

Every Data Provider confirms at data sharing contract signature that they complies with all applicable laws, rules, regulations, ordinances, and directives, including laws on the protection of personal data and that clinical trial protocols, informed consents, and ethics committee approvals for the Contributed Data are in place and do not specifically prohibit the pseudonymization or anonymization and the secondary use of the pseudonymized or anonymized data.

Data source last refresh

31/12/2022

Common Data Model (CDM) mapping

CDM mapping

Has the data source been converted (ETL-ed) to a common data model?

Yes

CDM Mappings

CDM name

OMOP

CDM website

https://www.ohdsi.org/Data-standardization/

Data source ETL CDM version

5.3.2001

Data source ETL frequency

0,03 months

Data source ETL status

In progress