STARTER
STarting an Adult Rare Tumour European Registry

European Reference Network
for rare or low prevalence complex diseases
Network Adult Cancers (ERN EURACAN)

REGISTRY OF THE EUROPEAN REFERENCE NETWORK ON RARE ADULT SOLID CANCERS "EURACAN"

Study protocol February 27, 2025,

version 8

BACKGROUND

Despite the rarity of each of the 198 identified rare cancers, collectively they represent 24% of all new cancer cases diagnosed in the EU28 each year (*Gatta G et al, Lancet Oncol 2017*). Five-year relative survival is worse for rare cancers (47%) than for common cancers (65%), and differences in survival exist across European countries, suggesting the existence of inequalities in healthcare (*Gatta G et al, EJSO 2019*).

There is a wide consensus that pathological diagnosis and primary treatment of rare cancers in particular should be centralized to reference centres and/or collaborative health networks with multidisciplinary, highly specific expertise *(Blay J-Y et al, Ann Oncol 2017; Ray-Coquard I et al, Eur J Cancer 2017; Blay J-Y et al, Ann Oncol 2019).* In addition, clinical and translational research needs a high level of centralization and international collaboration. The problem of rare cancers is by definition the low number of cases. This limiting factor can be overcome only through large collaborations. New methodologies for clinical studies may help (*Casali PG et al, Ann Oncol 2015*), but they are nonetheless more likely to be adopted within networks specializing in rare cancers.

Rare cancers are covered by widespread cancer registration in the EU. As of 2019, there are nearly 200 active population-based cancer registries (CRs) in Europe. However, the quality of a CR inevitably depends on the local healthcare environment and the available sources of information. Quality of care is relevant to quality of CRs. For example, inappropriate pathological diagnoses will result in misclassification. Rare cancers are particularly exposed to discrepancies in quality of care, with some (e.g. sarcomas) being particularly affected compared to others (e.g. squamous cell carcinomas of the head and neck). There tends to be a lack of clinically relevant data, e.g. on detection, staging, treatment and treatment effects, across CRs *(Siesling et al, Eur J Cancer 2014).* One way to obtain clinically relevant information is to perform ad hoc observational studies (i.e. high-resolution studies) but these are costly and time consuming, too often lack detailed information, and are unable to provide information in real-time.

In its Rare Cancer Agenda 2030, JARC recommends developing clinical registries on rare cancers within networks specializing in rare cancers, to prospectively collect clinical information on the entire patient journey, with a view to increasing knowledge on rare cancers and supporting clinical research.

A clear window of opportunity has been provided by the creation of European Reference Networks (ERNs). ERNs are virtual networks of selected institutions targeting rare conditions. The purpose of ERNs is to provide multidisciplinary second opinions for the management of complex cases, disseminate knowledge on rare cancers, produce clinical practice guidelines, and foster research by setting up shared registries. Twenty-four

ERNs were launched in March 2017, four dedicated to or involving rare cancers: the ERN on rare adult solid cancers (ERN EURACAN), the ERN on haematological diseases (EuroBloodNet), the ERN on paediatric cancer (ERN PaedCan), and the ERN on genetic tumour risk syndromes (ERN GENTURIS).

AIMS

We aim to set up a clinical registry for the ERN EURACAN (https://www.euracan.eu/research-and-registries/registries). EURACAN focuses on 10 of the 12 families of rare cancers (accounting for 75% of all rare cancers), each corresponding to a EURACAN "domain": G1, sarcomas; G2, rare neoplasms of the female genital organs and placenta; G3, rare genitourinary cancers; G4, neuroendocrine tumours; G5, rare gastrointestinal cancers; G6, endocrine cancers; G7, rare head and neck cancers; G8, rare thoracic cancers; G9, rare skin/eye melanoma; G10, central nervous system tumours. Paediatric and haematological cancers, the two remaining families of rare cancers, are covered by ERN PaedCan and EuroBloodNet, respectively.

The EURACAN Registry will cover all EURACAN domains. Its objectives will be:

1. to help describe the natural history of rare adult solid cancers (how the rare cancer develops, progress, possible association with other diseases, etc.);
2. to evaluate factors that influence prognosis (e.g. mortality, survival, progression-free survival) and treatment response;
3. to assess treatment effectiveness (systemic, radiotherapy, surgery, target therapy, immunotherapy and possible combinations);
4. to measure indicators of quality of care (diagnostic and staging procedures, treatment strategies, follow-up etc.).

The collected information can be used to improve our ability to prevent, diagnose and treat all aspects of rare adult solid cancers, with the ultimate goal of improving survival and quality of life for patients with rare cancers.

Furthermore, the registry aims **to collect information, when available, on the storage of biological samples** at the facilities of participating healthcare providers (HCPs). **Sharing biological samples is not required.** The goal is to gather information about the availability of such samples, not to collect them, as the registry is focused on data collection. In the event of future studies on the biology of rare solid cancers, this information will help identify which HCPs already have biological data that could potentially be contributed.
Finally, the registry will also serve as a tool to steer EURACAN towards achieving its objectives.

REGISTRY DESIGN

The Registry is designed to **prospectively and retrospectively collect clinical** data derived from diagnostic tests and treatments performed by the HCP as part of patient management.

The data collected for the Registry will not entail further examinations or admissions to the HCP and/or additional appointments to those normally provided. Medications, procedures, visits and check-ups are prescribed according to standard clinical practice. The only intervention will be to collect health status data, already present in the clinical file.
In other words, it will be an observational, real-world registry.

- **Patient inclusion criteria**

1. Adult patients (aged ≥18 years) diagnosed with any of the rare solid adult cancers included in EURACAN. The diagnosis should be performed or verified by the expert centre entering the patient information in the registry.
2. Malignant cancers; the inclusion of cancers of uncertain malignant potential will be domain- specific (e.g. G1, sarcoma domain will include sarcoma with malignant/3 and uncertain malignant potential/1).
3. Patients entering the HCP at any clinical phase of the disease (diagnosis, treatment of primary cancer, treatment of recurrence, treatment of M+ etc.). The HCP will collect information on the entire course of the disease regardless of when it started to manage the patient. The HCP can decide, based on its resources, the number of patients on whom it can collect data and/or type of patients (e.g. only those treated by the HCP from diagnosis onwards, only relapsed/metastatic ones, etc.). The HCP must clarify to the coordinator which type of patients he includes in the Registry
4. The starting year for patient enrollment depends on when the HCP joins the registry. **Newly participating HCPs should include patients diagnosed in the same year they join.** In addition, they may include patients who are still in follow-up at the HCP and were diagnosed within the three years preceding the year in which the HCP joins the registry.

- **Data collection and storage**

The registry will exploit data available from:

- national or regional registries/databases (DBs) dedicated to the rare adult solid cancer of interest;
- HCP registries/DBs;
- ad hoc data collection by HCPs.

If a European registry/DB (e.g. the ENETS Registry) is already available for any of the rare adult solid cancers of interest, formal collaboration will be discussed and finalised.

If a registry or DB exists at national level or at the HCP level, information will be extracted according to the standard format agreed for the EURACAN registry, after checking semantic interoperability. Extracted data will be entered in the DB developed ad hoc for the EURACAN registry.

In the absence of a registry or DB, data will be manually entered in the EURACAN registry's e-CRF.

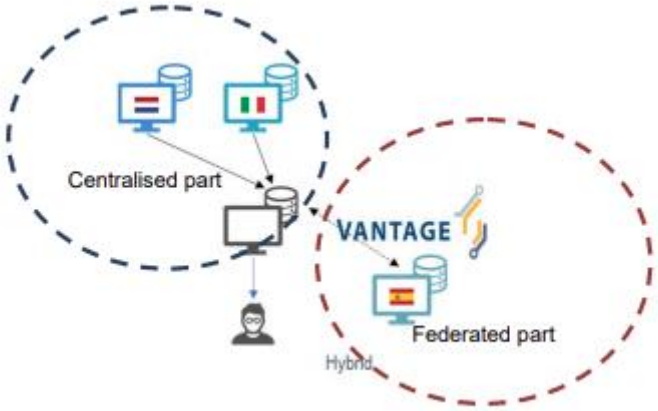
- **Registry architecture**

The registry utilizes a hybrid architecture, where data can either be shared in pseudonymized form through the registry coordinator's e-CRF in a centralized model or stored locally at the participating HCP in a federated model (*Fig.1*).

In the centralized registry, all data collected via the coordinator's CRF are stored in a single central DB managed by the Registry Coordinators.

In the federated model, data are kept at its source thus, no copies of datasets are generated and/or shared with third parties. Data are collected in a pseudo anonymised form locally using a e-CRF shared by the coordinator but locally installed. The Personal Health Train (PHT) enables data from multiple organizations to be analysed without identifiable data leaving the organization. Vantage6 is the open-source implementation of the PHT (https://distributedlearning.ai/). Vantage6 uses the mathematical principle of "federated learning", typically applied to horizontally partitioned data (i.e. organizations provide data from different patient cohorts, but with similar characteristics/items). Federated learning is based on the mathematical principle of splitting computations into (a) parts at the station (local HCP or registry) and (b) a central part (https://distributedlearning.ai/).

Figure 1. Registry hybrid architecture



A EURACAN registry helpdesk will be set up to support: installation of the different software at the local level; data integration solutions (e.g. data transfer from an existing DB or data warehouse etc.); and data collection. Each HCP and/or registry should identify an IT person responsible for software installation and related issues (e.g. server, security licences etc.).

- **Data quality checks**

Data analysis is performed only when all or most of the errors have been corrected. Data quality checks aim to assess whether data values are present, valid and believable in terms of:

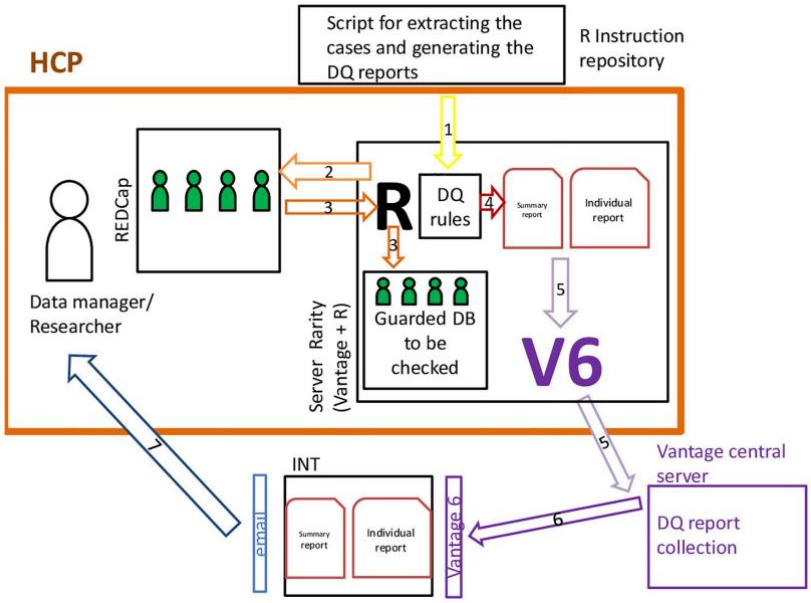
- Validity: the data should be in the correct format (e.g. allowed only some values, data formats)
- Plausibility: the data should be acceptable (eg, internal consistency; temporal and atemporal)
- Completeness: the data should not have missing values or miss data records

Validity and plausibility checks are embedded in the e-CRF in the form of alerts and errors during the data input. Additional checks on completeness and data distribution are performed by the coordinator either on the centralized DB (centralized model) or on the local DB leveraging the federated infrastructure via R script locally implemented. It is worth stressing that in the federated model, the coordinator launches data quality checks as described below, but he/she never accesses individual record data. The coordinator gets back only the results of the data quality checks that are locally performed.

What happens at the local level. The R script, including the checks, is downloaded locally from an online instruction repository (Fig 2, step 1). The R script extracts from the Research Electronic Data Capture

(REDCap) (the IT solution used for the registry CRF) all the completed cases, stores a copy of the DB in the **dedicated local server** and runs the checks locally (Fig 2, step 2 and 3). The results of these checks are summarized in two reports: a summary and an individual report (Fig 2, step 4). The summary report collects: the number of patients with a complete and valid information on the core data elements; some relevant distribution (e.g. treatment by stage for each cancer site included in the registry, the proportion of cases with missing information for each variable). The individual report includes a summary of error for each patient, together with a specification of the variables with the errors to be corrected. Thanks to the Vantage6 software, **the two reports (not the data)** reach the registry coordination team (INT) to be monitored and discussed with each data provider (Fig 2 step 5, 6 and 7).

Figure 2. Data quality steps in the federated model



After the corrections made by the data providers, all checks will be re-run and quality reports (coming from the centralized DB and from the different federated DB) reviewed by INT. Interaction with data providers will be reiterated until sufficient data quality is achieved. The DB with sufficient data quality will be saved and used for federated learning analysis. These checks will be performed annually and will ensure high data quality within the federated and centralized DB. The degree of data completeness will be summarized and available for the researchers and consumers of the analysis of the registry.

- **Data analyses**

Statistical analyses will be performed based on a study protocol. Queries will be developed, in collaboration with clinical experts, to interrogate the EURACAN DB to generate the descriptive statistics and relevant information needed to plan the statistical analyses envisaged by the study protocol.
Registry data will support observational studies based on secondary use of available data.

Data collected for the EURACAN registry could also be used for global collaborative studies performed inside and outside the EU.

- **Data to be collected**

Core data elements will include the "Set of common data elements for Rare Disease Registration" developed by JRC (https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements) to address specificities of rare cancers as compared to rare diseases.

Following the EURACAN registry objectives, data are collected on patient characteristics, exposure and outcomes. Patient characteristics will consist of descriptive patient data, such as information on patient demographics, including race, lifestyle, medical history, health status, etc. Exposure data will focus on the disease, devices, procedures, treatments or services of interest. Outcome data will describe patient outcomes (e.g. survival, progression, progression-free survival, death, etc.). In addition, data on potential confounders will also be collected.

- **Healthcare Providers included in the registry**

The aim of the registry is to collect data on all patients treated or followed up for a rare adult solid tumour at the HCPs of the ERN EURACAN, at non-EURACAN expert HCPs, or at national and/or regional clinical networks in European and non-European countries. The list of EURACAN centres is available on the EURACAN website, the list of HCPs contributing to the registry is also available and updated regularly on the website.

- **Ethics**

INT is the coordinator of the EURACAN registry as well as a data provider. At the INT, and at each HCP involved, responsible investigators will ensure that the EURACAN registry is implemented in compliance with this protocol, following the instructions and procedures described herein. Each HCP is a controller and will identify a data processor.

Patient protection

Personal data will be recorded and stored in pseudonymised format. In the federated model this solution will enable data sharing among multiple organizations without identifiable data leaving the HCP.

All parties involved in registry development will maintain strict confidentiality to ensure that neither personal privacy nor the privacy of the families of patients participating in the registries is violated. Data will be processed exclusively by authorized personnel participating in development of the EURACAN registry (a data processor will be identified at each HCP involved). Access to computer systems and the premises where they are kept will be controlled by appropriate security measures which comply with privacy regulation requirements. The processing of the personal data of patients taking part in the EURACAN registry, and specifically in relation to consent-related data, will comply with local privacy legislation and the General Data Protection Regulation 2016/679 (GDPR) of the European Union.

The registry protocol will be submitted to the ethics committees (EC) of the HCPs involved. Furthermore, the ECs of the HCP involved will authorize in advance any research carried out using EURACAN registry data.

GOVERNANCE

To ensure appropriate access and use of the data, the registry will be supervised by a Steering Committee (SC) composed of:

- EURACAN domain and sub-domain leaders;
- Transversal Task Force (TTF) leaders;
- European-Patient Advocacy Group leaders (e-PAGs) from each domain;
- Representatives of national registries contributing data to the EURACAN registry;
- The EURACAN registry coordinator (Fondazione IRCCS Istituto Nazionale dei Tumori).

The SC will include all domain leaders and e-PAGs (1 domain leader and 1 e-PAG per domain). One alternate should be identified per domain leader and ePAG. Additional expertise may be brought into the Steering Committee as required.

The functions of the SC are:

- To launch, plan and supervise studies and publications based on registry data;
- To define and review the rules for accessing EURACAN registry data;
- To ensure adherence to the publication policy and guidelines for accessing registry data;
- To plan and endorse modifications of the registry structure (e.g. pathological classification or staging changes, etc.);
- To deliberate on applications for enrolment into the EURACAN registry, from non EURACAN centres or networks linked to it;
- To promote the use of registry data also for international collaborative studies (e.g. with US, Rare Cancers Asia, etc.), once the registry is fully functioning. In other words, to ensure the EURACAN registry is **FAIR** (findable, accessible, interoperable, reusable);
- To review the EURACAN registry governance
- To identify financial support opportunities to maintain the registry.

The SC is chaired by the Registry Coordinator (Annalisa Trama, Fondazione IRCCS Istituto Nazionale dei Tumori, INT) assisted by a Co-coordinator (Jean-Yves Blay, Centre Léon Bérard, as EURACAN coordinator).

The SC is supported in its activities by a dedicated scientific secretariat based at INT (contact: EURACANregistry@istitutotumori.mi.it – Tel. 02 2390 3564).

REGISTRY INTEGRATIONS

To support the development of registries within ERNs, the European Commission decided to set up a European Platform on Rare Disease Registration (EU RD Platform) and to develop specific standards for the interoperability of rare disease registries, drawn up by the Commission's Joint Research Centre (JRC). The EU RD Platform makes rare disease registry data searchable and findable, thus increasing the visibility of each registry. This is ensured by the European RD Registry Infrastructure (ERDRI), which supports the activities of existing and the creation of new registries.

The EURACAN registry will be fully compliant with ERDRI requirements, to ensure its interoperability with other rare disease registries, and will be enrolled in ERDRI. The EURACAN registry will therefore be registered in the EU RD Platform.

In addition, synergies with other ERN registries will be fostered through the ERN Registries Task Force. Opportunities for partnerships with population-based CRs will also be explored. Finally, the EURACAN registry will be interoperable with the Clinical Patient Management System (CPMS), i.e. the tool developed by the Commission to provide multidisciplinary second opinions for complex case management (i.e. teleconsultations across different expert centres).