iPAAC
INNOVATIVE PARTNERSHIP
FOR ACTION AGAINST CANCER

Co-funded by
the Health Programme
of the European Union

Results of the iPAAC Work Package7 pilots: enriching cancer registry data

Author(s):	Lead author: Annalisa Trama, Silvia Francisci, Milena Sant Co-authors: Alice Bernasconi, Simone Bonfarnuzzo, Laura Botta, Roberto Lillini, Anna Gigli, Stefano Guzzinati, Tania Lopez, Sandra Mallone, Daniela Pierannunzio, Andrea Tavilla
Version:	2.0
Date:	21. 12. 2021

Contents

This report arises from the Innovative Partnership for Action Against Cancer Joint Action, which has received funding from the European Union through the Consumers, Health, Agriculture and Food Executive Agency of the European Commission, in the framework of the Health Programme 2014-2020. The content of this report represents the views of the author/s only and is his/her/their sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use that may be made of the information it contains. The authors are not responsible for any further and future use of the report by third parties and third-party translations.

Abbreviations

CRs	Cancer registries
EPAAC	European Partnership for Action Against Cancer
CANCON	EU Cancer Control Joint Action
iPAAC	Innovative Partnership for Action Against Cancer
EUROCARE	EUROpean CAncer REgistry based study on survival and care of cancer patients
CI5	Cancer Incidence in Five Continents
CONCORD	Global surveillance of Cancer survival
ENCR	European Network of Cancer Registries
IARC	International Agency for Research on Cancer
AYA	Adolescent and Young Adult
ICDO3	International Classification of Disease for Oncology, 3rd edition
ATC	Anatomic Terapeutical Chemical classification
HDR	Hospital Discharge Records
DB	Database
CT	Computed Tomography
ECOG	Eastern Cooperative Oncology Group
ICD9-CM	International Classification of Diseases, 9th revision - Clinical Modification
ICD10-CM	International Classification of Diseases,10th revision - Clinical Modification
GDPR	General Data Protection Regulation

Executive summary

Pilot studies of iPAAC Work Package 7 on "Cancer Information and registries" aimed at assessing the feasibility of enriching cancer registries (CRs) data with information from administrative datasets to define quality of care (pilot 7.2), costs of care (pilot 7.3) and long-term comorbidities in cancer survivors (pilot 7.4; with a focus on adolescents and young adults). All pilots were based on individual patients' records linked by the CRs to the different data sources necessary to achieve the pilot-specific objectives. The CRs used a unique anonymous identification code for each patient allowing the tracing of each patient in all data sources provided. Several countries contributing to these pilots (Belgium, Norway, Italy, Poland, Slovenia and Basque Country in Spain). Study protocols including objectives, necessary data, indicators and analysis plan were developed per each pilot. The protocols were tailored to the specific context of participating countries, based on a careful review and harmonisation of the country-specific available data sources. Analyses were performed according to the identified common procedures, either by each CR (de-centralised application) or by the pilot coordinator (centralised application).

The results of the pilots showed the feasibility of using administrative dataset to add relevant clinical and economic information to the cancer registries data. Thus, the results confirm the key role that CRs should play in clinical and translational research fields, as well as cancer planning and monitoring.

Efforts at EU level should be continued to emphasize to national policy makers the need and urgency of CRs development and to provide political and financial support. The mandate of CRs should be strengthened and widened by broadening their scope and the range of data they collect. CRs should increasingly become key players in the EU health data space as they can be a 'corner stone' to develop more comprehensive cancer information systems, joining both clinical and public health components.

1 Rationale and objectives

1.1 Rationale

Cancer registries (CRs) have been expanding in Europe since the early 1900s. Over the past three decades, cancer registration has become an important element of the EU strategy against cancer, promoted as part of the framework of the European Action against Cancer Programme (1985–2008), the European Partnership for Action Against Cancer (2009–2014), the EU Cancer Control Joint Action and the Innovative Partnership for Action Against Cancer (2018-2021).

CRs have the staff, logistics and methodology to store, analyze, interpret cancer data and to control the quality of their data. Furthermore, different experiences (CI5; EUROCARE; CONCORD; high resolution studies) have generated evidence supporting the importance of CRs and the high quality of their data.

A minimum set of data to be collected by CRs was proposed by the International Agency for Research on Cancer and the European Network of Cancer Registries. However, **additional data should be collected, to better contribute to cancer research and planning.** An opportunity to enrich the CR dataset is provided by the availability of administrative datasets**.** Thus, European CRs already use a wide range of sources, including pathology reports, medical and discharge records, radiotherapy wards, death certificates. Additional sources such as hematologic labs, palliative care / hospice records, public or private hospital practice records, primary care records, and health insurance records are used in varying ways by some CRs.

Against this background, iPAAC Work Package 7 (WP7) pilot studies aimed to assess the feasibility to enrich CRs data linking with heath and administrative data sources to derive key indicators useful for cancer care and management. Three different pilots were carried out to provide information on:

- patterns and quality of care (Task 7.2);
- costs of care (Task 7.3);
- long-term outcomes in Adolescent and Young Adult (AYA) cancer survivors (Task 7.4). AYA are those aged 15-39 years at first cancer diagnosis. Survivors are those alive at least 5 years after cancer diagnosis.

1.2 Specific objectives

Task 7.2 evaluated the feasibility of linking individual patient's data included in the participating population-based CRs with administrative and health data, in order to:

1. describe the complete pathway of cancer patients from diagnosis to rehabilitation or terminal care, including the use of health care resources at the end of life;
2. assess the adherence of the administered treatments to standard clinical guidelines.

Additionally, Task 7.2 explored the feasibility of investigating:

- pathological events (e.g., relapse, unwanted effects of anticancer treatments) occurring during the disease course;
- co-morbidities and socio-economic status.

Task 7.3 assessed the feasibility of extending to other European countries the procedures developed in the framework of the EPICOST Italian project to produce indicators on the direct costs related to diagnosis, treatment and follow-up of cancer patients.

Specific objectives of this task were:

- to identify for each participating country a prevalence cohort stratified by phase of care according to the study design;

- to assess the availability of the relevant information (cost of procedures and treatments related to cancer patient's health care starting from diagnosis to possible recovery or death) at individual level;
- to assess the possibility for the participating country to link this information in order to reconstruct patterns of care and corresponding costs of the prevalence cohort;
- to select for each data source/country a list of codes of cancer-related procedures;
- to calculate cost indicators according to the same methodology used in EPICOST and described in the methods section.

The objectives of Task 7.4 were:

- to estimate the burden of late effects;

- to compare late effect indicators across the different countries.

The late effects included both clinical and socio-economic outcomes:

- Clinical outcomes
  - Multiple subsequent malignant neoplasms;
  - Hospitalizations (as a proxy of comorbidities) eg. infectious and parasitic diseases, endocrine nutritional and metabolic diseases, and immunity disorders, diseases of the blood and blood-forming organs, diseases of the nervous system and sense organs;
  - Mortality;
  - Infertility and/or complications of pregnancy, childbirth, and puerperium.
- Socio-economic outcomes
  - Attained education;
  - Marital status;
  - Utilisation of social/financial benefit (for unemployment, for disability);
  - Occupation and unemployment.

2 Key contextual features

Availability of population based Cancer Registry and other health administrative databases was a prerequisite. A survey on available electronic data sources for individual linkage was conducted among the European cancer registries involved in iPAAC WP7 (Task 7.1). Overall, 27 population-based CRs from 14 different countries replied to the questionnaire. Health administrative data sources available for linkage to European cancer registries are not homogeneously accessible and used, additionally the quality, completeness and extent of these data sources differ across countries and within regions of a same country. This situation reflects heterogeneous health care systems, data owners, legal frameworks and socio-economic conditions. A significant proportion of registries, however, incorporates these data sources in their routine activity and for research purposes. Part of these sources are sufficiently standardised in terms of coding classification and data structure and can be considered valid for deriving comparable indicators on cancer care.

Results of the survey to European CRs developed in Task 1 of WP 7 are available at the iPAAC website – Work Package 7(https://www.ipaac.eu/en/work-packages/wp7/).

Furthermore, for data integration, the level of legal enforcement of CRs plays a key role and varies widely across countries and health systems. Clear rules and/or procedures enabling CRs to link other health administrative databases are an essential requisite.

Finally, it is advisable to have a good understanding of the information included in administrative databases and their quality problems. If CRs do not have such knowledge, collaboration with administrative data owner should be ensured.

3 Methods

All pilots proposed in the context of the WP7 were based on individual records linked by the CRs to the different data sources necessary to achieve the pilot-specific objectives.
The CRs used a unique anonymous identification code for each patient allowing the tracing of each patient in all data sources provided.
Pilots' study protocols including case selection criteria, index cancers, data sources, core indicators and variables needed to derive them, were developed for each pilot (Tasks 2-4). They are available at the iPAAC website – Work Package 7(https://www.ipaac.eu/en/work-packages/wp7/).

Two alternative types of study design were used: cross-sectional and longitudinal.

A cross-sectional prevalence-based study design was shared by Task 7.2 and Task 7.3. The study cohort includes patients diagnosed with the index tumors during all years of activity of the CRs and still alive at the prevalence index date (prevalence cohort). The prevalence date is the most recent one on which the CRs database has been updated; an entire year of follow-up (ascertainment of life status) after the prevalence index year must be available for the whole cohort.

Invasive, primary, malignant neoplasms of colon (International Classification of Diseases for Oncology, 3rd revision – ICDO3 topography C18), rectum (ICDO3 topography C19-20), pancreas (ICDO3 topography C25), and skin melanoma (ICDO3 topography C44, morphology 8720-8790) diagnosed in adult (aged ≥15 years) patients were eligible for inclusion in Task 7.2 and Task 7.3 (these are called index tumors).

Cross-sectional study design is the most suited for Task 7.2 and Task 7.3 purposes, as it allows to derive patterns of care and cost indicators using the most updated information available, in distinct phases of the disease course by considering events occurring 2 years before and 1 year after the prevalence index date (Figure 1).

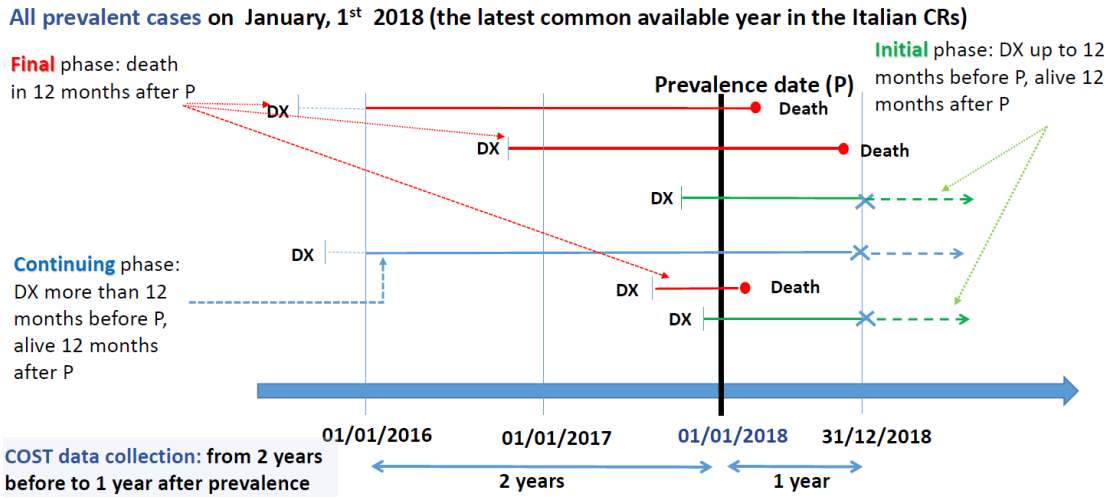


Figure 1 - Example of 7.2 and 7.3 pilots' study design with prevalence index date on January 2018, 1st

Concerning Task 7.2, this cross-sectional study design allows to obtain indicators of diagnostic and therapeutic procedures along the disease course, including terminal care and to study the adherence

with clinical guidelines. It allows capturing the effects of technological progress and the introduction of new cancer therapies potentially affecting outcomes.

Regarding Task 7.3, this prevalence-based cross-sectional study design presents the following advantages:

- it produces an updated snapshot of total costs delivered to cancer patients in a given calendar year;
- it allows to capture the effects of technological progress and the introduction of targeted cancer therapies particularly relevant for costs at the disease onset;
- it is commonly used in understanding the overall impact of disease on health plan budgets, in monitoring resources used by patients with a similar cancer and in planning appropriate future resources;
- cost data are directly collected at constant inflation rate and consequently there is no need of adjustments with temporal price indices to transform nominal into real values.

An incidence based longitudinal study design was used for the task 7.4

In this case, CRs identify incident cases aged 15-39 at first cancer diagnosis in each year of incidence covered by the registry. Cancer survivors are defined as those alive 5 years after the first cancer diagnosis. Minimum selection criteria for CRs were:

- 15 years of follow-up (because of the long period of time - median time 15 years - that needs to elapse to observe the late effects) and at least 5 years of follow-up for the latest incidence year (to identify AYA survivors);
- availability of mortality information for the incident cases and for the cancer free population covered by the CRs;
- cancer and not-cancer related hospitalisation data for the incident cases and for the cancer free population covered by the CRs. Incident cases have to be linked with: all their multiple primary malignancies and other available sources necessary to study the late effects.

Let's take the example of the Figure 2 below.

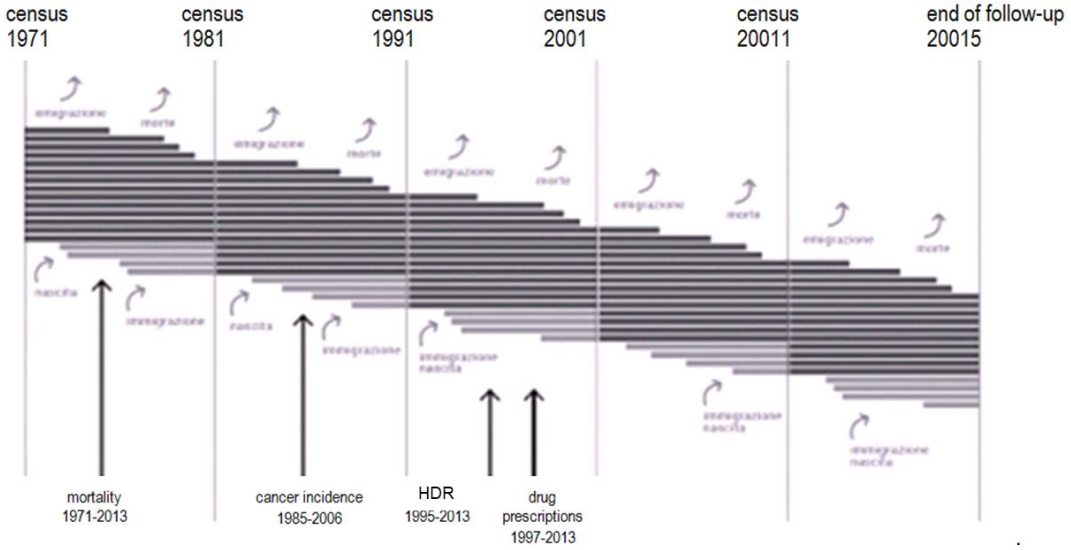


Figure 2: example of data sources available for linkage with individual cancer registry data with respective time frame

In the example there are 5 fixed census cohorts starting from 1971 to 2011, on the other hand mortality information is linked to the census cohorts from 1971, cancer incidence is linked from 1985 to 2006, hospital discharge records (HDR) are linked from 1995 to 2013, drug prescriptions are linked from 1997-2013.
In this case, all AYA patients diagnosed with cancer from 1985 to 2006 are part of the cohort. Socio-economic information related to each AYA cancer case will be obtained linking the various censuses as follows:

- AYA diagnosed with cancer between 1985 and 1990 are linked to 1991, 2001 and 2011 censuses
- AYA diagnosed with cancer between 1991 and 2000 are linked to 2001 and 2011 censuses
- AYA diagnosed with cancer between 2001 and 2006 are linked to 2011 census.

The linkage between AYA cancer cases and census was carried out only for those alive at the census date. In case data of several censuses are connected at individual level, the socio-economic effects will be defined by investigating inter-censuses changes in educational level, marital status, occupation, type of residence.

For each AYA patient diagnosed with cancer from 1985 to 2006 health outcomes (e.g. mortality, hospitalization) were obtained linking the relevant data source from the first year of availability as follows:
- from 1985 to 2013 linkage with cause of death for the entire cohort of AYA diagnosed with cancer from 1985 to 2006;
- from 1995 to 2013 linkage with the HDR for AYA diagnosed with cancer from 1990 to 2006;
- from 2000 to 2013 linkage with drug prescriptions for AYA diagnosed with cancer from 1995 to 2006.

These details are to be defined for each Cancer Registry according to the data source and the time frame available. An implementation example can be found in Bernasconi A, et al, 2020.

An alternative approach has been used in Norway where we designed a matched cohort study design with population-based incidence cohort matched with a cohort of general population. In particular, the Norwegian Cancer Registry (NCR) identified AYA with a first cancer diagnosis, recorded between January 1, 1991 and December 31, 2010 and who are alive 5-years after the diagnosis. For each AYA with cancer, 5 general-population members were randomly selected from a pool of individuals who are alive and free of cancer on the date of the matched person's cancer diagnosis and before 39 years of age, as recorded in the NCR (the index date), matched on birth year, sex and county of residence. All matched controls were alive 5 year after first diagnosis of the patient.

AYA cancer patients and general-population members were followed-up from the cancer diagnosis/index date until 31 December 2019, date of death or date of emigration whichever came first.

3.1 Study indicators

3.1.1 Task 7.2

Cancer specific study indicators on the entire pathway of cancer patients were identified by literature review, collaboration with oncologists and profited by the experience carried out with the High Resolution cancer registries based studies.
The following indicators of standard care in the diagnostic, continuing or end of life disease phase were considered for each tumour:

colorectal cancer [Labianca R et al, 2013; Van Cutsem E et al, 2014; Glynne-Jones R et al, 2017]

- Percentage of screen detected (by organised or opportunistic screening) colorectal cancer patients;
- Percentage of colon cancer patients diagnosed by endoscopy who received a complete colonoscopy;
- Percentage of stage III resected colon cancer patients treated with adjuvant chemotherapy [if feasible, only R0 resections included] (in common with WP10);
- Percentage of resected colorectal cancer patients died within 30/90 days from surgery (in common with WP10).
- Percentage of stage III resected rectal cancer patients treated with neo-adjuvant radiotherapy [if feasible, only R0 resections included] (in common with WP10);
- Percentage of metastatic colorectal cancer patients treated with biological drugs (molecular targeted, monoclonal antibodies).

skin melanoma [Dummer R et al, 2015]

- Percentage of stage IV skin melanomas receiving mutation testing;
- Percentage of skin melanomas with information on the maximum thickness in millimetres (Breslow);
- Percentage of skin melanoma patients with clinically negative nodes (cN0) and tumour thickness of >1 mm receiving sentinel lymph node biopsy;
- Percentage of metastatic skin melanoma patients treated with immunotherapy.

pancreatic cancer [Ducreux M et al, 2015]

- Percentage of pancreatic cancer patients receiving CT scan at diagnosis; Percentage of resectable pancreatic cancer patient treated with curative surgery;
- Percentage of metastatic pancreatic cancer patients with information on ECOG performance status;
- Percentage of resected pancreatic cancer patients died within 30/90 days from surgery (in common with WP10).

Furthermore, for **all tumour** under study:

- Type of hospital (oncological, general hospital, oncological department within general hospital) where patients received the main treatments.

Indicators of quality of care at the end of life [Barbera et al, 2015], taking into account that in this phase high hospitalisation or anticancer drugs use are considered indicators of inappropriate care:

- A new hospital admission in the last 30 days of life;
- An intensive care unit (ICU) admission in the last 30 days of life;
- Chemotherapy use in the last 2 weeks of life;
- Percentage of dead cancer patients with information on place of death;
- Enrolment in specialist palliative care program.

3.1.2 Task 7.3

Prevalence by phase of care

Three phases of care are defined as follows: initial (12 months after cancer diagnosis); continuing (time elapsed between initial and final); final (last 12 months before death due to cancer). In a cross-sectional study design phases of care are mutually exclusive.

On prevalence date, each patient of the prevalence cohort belongs to only one phase of care, depending on the interval between prevalence date and diagnosis date and on the occurrence of death for cancer during the following year. More details about phases of care definition are in: Gigli A, 2021, and about methods to estimate complete prevalence by phase of care are in Gigli A, 2021.

Cost by phase of care

Each patient of the prevalence cohort contributes to the study with a 12-month time interval and is linked to the available administrative/health care databases in order to trace every event of interest during the follow up time.
In order to take into consideration only those events that are related to the index tumours, a list of events (specific for each index tumour) is utilised for each database. These lists were created by expert oncologists in the framework of the EPICOST Italian project and referred to diagnoses, interventions and procedures coded according to the ICD9-CM classification for Hospital Discharges DB and for outpatient services DB, and to the Anatomical Therapeutic Chemical Classification System (ATC code) for drug prescriptions DB. The lists are available within the EPICOST study for colon, rectum and breast cancers only. More details about the methodology and the lists for breast cancer are in Busco S, 2021.
Costs are expressed in Euros and are defined as the direct expenditure related to diagnosis, treatment, follow-up and end-of-life care provided to cancer patients.
We identified homogeneous groups of patients according to clinical and demographic variables affecting the patterns of care: age, stage at diagnosis (for patients in the initial phase only), and we computed costs as simple averages over patients belonging to the same homogeneous group.
The following indicators were considered for each of the three phases of care:

- Patient monthly average cost, Ci: all costs sustained on average for a patient in month i, obtained by dividing costs sustained for all patients in month (i) by the corresponding number of person-months. A cost profile is a series of 36 patient monthly average costs Ci over the three phases of care;
- Patient annual average cost, CA: all costs sustained on average for a patient in a year, obtained by summing up patient monthly average costs, i.e. CA = Σ Ci 12 i=1;
- Total annual cost: all costs sustained in 12 months for all patients, obtained by multiplying the patient annual average cost (CA) by the total number of patients.

These costs are computed by phase of care and/or by type of health care service.

Patterns of care indicators
In order to better describe and interpret results on costs in the initial phase of care, a list of patterns of care indicators is computed, the list is specific for each index tumor considered.
Here an example of indicators computed by age at prevalence and stage at diagnosis, applicable to colon cancer:

- percentage of patients receiving at least one surgery treatment;
- percentage of patients receiving at least one chemotherapy over all patients in initial phase of care;
- time occurring between surgery and chemotherapy.

3.1.3 Task 7.4

The following indicators were computed:
− standardized incidence ratio (SIR) of cancer and other chronic diseases;
− standardized mortality ratio (SMR);
− risk ratio of attaining a low education level, of not getting married, of being financially dependent, of being unemployed.

All indicators are calculated using as reference the cancer free population covered by the CRs. The SIR was calculated considering the cancer incidence in the non-AYA population.

3.2 Information and data sources required

All tasks are based on Cancer Registry data, to which specific additional data were linked.
For each cancer case, the CRs provided all variables included in the 2015 ENCR-JRC Call for Data study protocol [ENCR-JRC Call for Data 2015; EUROCARE-6 protocol 2015].

3.2.1 Task 7.2

The following table shows selected indicators for colon (CC), rectal (RC) and pancreatic (PC) cancers, skin melanoma (SM), and for investigating patients' end of life (EoL) in all patients, with the type of information required to derive them.

Topic	Indicator	Required information
CC	Percentage of screen detected *(by organised or opportunistic screening)* CC	Screen detection and date, type of screening
RC	Percentage of screen detected *(by organised or opportunistic screening)* RC	
CC	Percentage of CC patients diagnosed by endoscopy with biopsy	Stage at diagnosis, diagnostic procedure and date
RC	Percentage of RC patients diagnosed by endoscopy with biopsy	
SM	Percentage of stage IV SM receiving mutation testing	Stage, mutation testing procedure
PC	Percentage of PC patients receiving CT scan at diagnosis	Diagnostic procedure and date
SM	Percentage of SM patients with clinically negative nodes (cN0) and tumour thickness ~~of~~ >1 mm receiving sentinel lymph node biopsy	Pathological and clinical stage, thickness / Breslow (the latter could be indirectly reconstructed by pT); sentinel node biopsy
CC	Percentage of stage III resected CC patients treated with adjuvant chemotherapy	Stage, All treatments (type or category of drugs) and their dates *(important for distinguishing between pre or post-operative, for primary cancer or relapse/progression)*
CC	Percentage of metastatic CC patients treated with targeted therapy or monoclonal antibodies	
RC	Percentage of stage III resected RC patients treated with neo-adjuvant radiotherapy	Type of radiotherapic treatments
RC	Percentage of metastatic RC patients treated with targeted therapy	Stage, treatment (type or category of drugs)
SM	Percentage of metastatic SM patients treated with immunotherapy	Stage; treatment (type or category of drugs)
PC	Percentage of resectable PC patient treated with curative surgery	Stage; treatment
EoL	Chemotherapy use in the last 2 weeks of life	Dates (dd,mm,yy); treatment (type or category of drugs)
SM	Percentage of SM with information on the maximum thickness in millimetres (Breslow)	Stage at diagnosis with thickness / Breslow (it could be indirectly reconstructed by pT)
PC	Percentage of metastatic PC patients with information on ECOG performance status	Stage, ECOG
CC	Percentage of resected CC patients died within 30 days from surgery	Dates of death (dd,mm,yy), life status, date of surgery
RC	Percentage of resected RC patients died within 30 days from surgery	
PC	Percentage of resected PC patients died within 30 days from surgery	

EoL	Percentage of dead cancer patients with information on place of death	Place of death
EoL	A hospital admission in the last 30 days of life	All hospital admissions and their dates
EoL	An intensive care unit (ICU) admission in the last 30 days of life	Unit of hospital admission (e.g., emergency room, etc.)
EoL	Enrolment in a specialist palliative care programme	It includes information on setting where palliative care (Pcare) was administered, e.g.: • home specialised Pcare service • hospice admission • residential hospice admission • hospital Pcare consultation • inpatient hospital Pcare admission • outpatient Pcare clinic

3.2.2 Task 7.3

Cancer Registry data
All variables included in the 2015 ENCR-JRC Call for Data study protocol [EUROCARE-6, 2015] are required for Task 7.3.

In particular:
- stage at diagnosis is requested only for prevalent cases diagnosis up to 12 months before the prevalence date;
- all multiple primaries occurred before the prevalence date.

In case of prevalent cases with multiple primaries, patients with index tumours to be included in the study cohort were defined centrally (by Task 7.3 coordination group) by selecting: a) patients with index tumours diagnosed as most recently as possible; and b) other primaries (any cancer type) that occurred 5 or more years before the index tumour diagnosis date.

Each cancer case was linked to all the available administrative/health care data sources and to the mortality file. This individual linkage was performed by each CR

Administrative/health care data
Number and contents of administrative/health care data sources considered for the linkage might vary according to the country health care data system. The aim is to include as much information as possible to estimate the total amount of health care expenditures directly related to diagnosis, treatment and monitoring of the prevalence cohort during the study period. However, at least cancer and not-cancer related hospitalisation data were required.

From administrative/health care data sources, the following information were analysed:
- Type of procedure (diagnostic procedures, treatments, outpatient procedures and visits) classified according to the ICD9-CM (or ICD10-CM), pharmaceutical prescriptions classified according to the ATC code;
- Date of procedure;
- Quantity of procedure;
- Cost per unit of procedure (in Euros);

— Total cost of procedure (in Euros);

Any other additional variable is data source-specific.
As an example, in the case of the Hospital Discharge database in Italy: regimen (with or without overnight stay in hospital), number of days of stay, multiple diagnostic codes (main diagnosis, secondary diagnoses up to five), multiple treatment codes (main treatment, secondary treatments up to ten), DRG code.
Table below includes an example of data from administrative/health care data sources used for EPICOST in Italy.

Data source	Variables
Hospital admission/discharge file (HA)	-Patient ID (the same one used in the CR database) -Demographic variables (sex, place of residence, date of birth, civil status, education level) -Type of admission (ordinary, day hospital) -Dates of admission at the hospital and of discharge; -Diagnosis (principal diagnosis + the other secondary up to five); -Diagnostic and intervention procedures (principal intervention + the other secondary interventions up to ten); -Dates of diagnostic and intervention procedures; -Discharge modality (patient death, ordinary discharge, transfer to other unit same hospital, transfer to other hospital); - DRG code -Total claim (in Euros)
Outpatient Services database (OPS)	-Patient ID (the same one used in the CR database) -Dates of service; -Code of Diagnostic and intervention procedure; -Description of Diagnostic and intervention procedure; -Date of Diagnostic and intervention procedure; -Branch of the procedure (numerical code corresponding to homogeneous groups of interventions: diagnostic, visits, radiotherapy, genetic tests, …) -Quantity: number of diagnostic or intervention procedure; -Tariff: unitary cost per single diagnostic or intervention procedure; -Total claim (in Euros): Quantity X Tariff, when quantity is >1
Drug Prescriptions database (DP) / Hospital Drugs (HD) record track	-Patient ID (the same one used in the CR database) -Dates of pharmaceutical prescription; -ATC code; -AIC code; -Quantity: number of doses indicated in the prescription;

	-Tariff: unitary cost per single dose of drug; -Total claim (in Euros) corresponding to the total cost of the prescription (quantity X Tariff)

The datasets (one for each health care source) included a record per each patient of the selected prevalence cohort and per procedure (i.e multiple procedures for the same patient correspond to multiple records).

Mortality data
As regards the mortality file, the following variables were required:
- Patient-ID (the same one used in the CR database sent for the pilot);
- Date of death;
- Cause of death.

3.2.3 Task 7.4

The list of data and corresponding data source required to derive each specific indicator on AYA survivors are provided in the following Table.

Indicators	Data required	Data source
SIR of multiple cancers	multiple primary malignancies of AYA cancer survivors and of the not-AYA cancer population	Cancer registry database
SIR of chronic diseases	Information on cardiovascular, respiratory, endocrine diseases on infertility and pregnancy problems for the AYA cancer survivors and for the cancer-free population	Clinical discharge records: hospitalisation for a specific disease proxy of the disease occurrance (standardised hospitalisation ratio).
SMR	Mortality (including cause specific mortality) for AYA cancer survivors and for the cancer-free population	Mortality register
RR of low education	Information on the attained education for AYA cancer survivors and for the cancer free population	Census. Education ISCED classification
RR of not being married	Information on marital status for AYA cancer survivors and for the cancer-free population	Census. Marital status classified as: married, not married, divorced, widowed,single
RR of being unemployed or being in an unskilled working class	Information on occupational conditions and occupational class for AYA cancer survivors and for the cancer-free population	Census. Occupational conditions classified as Employed, Unemployed/looking for first occupation, Housewife, Student, Retired; occupational class classified as Bourgeoisie, Middle class, Self-employed with employees, Self-employed without employees, Skilled working class, Unskilled working class.
RR of using social/financials benefit	Information on the use of social/financial benefit status for AYA cancer survivors and for the cancer-free population	In Italy INAIL/INPS (collaboration in progress). In other countries to be defined.

4 Implementation steps

The implementation steps were the same in each pilot and are listed in the following.

1) Formal steps: Ethical Committee and assessment of data/datasets compliance to the GDPR.

2) Operational steps:
- feasibility assessment and finalization of the study protocol based on registry-specific data availability;
- identification of patient's cohort (prevalent or incident cases);
- selection of relevant administrative health database;

- assessment of administrative database quality and completeness;
- individual record linkage to available data sources;
- adaptation of calculating algorithms to the country specific data;
- calculation of target indicators;
- outcomes validation via comparison with available literature or with precedent studies, e.g. in the case of High Resolution studies;
- discussion with relevant stakeholders.

3) Data analyses were performed according to two possible options:

- CRs performed the analyses by themselves. In this case, a common plan of analyses was discussed and shared with the pilot leaders, (*de-centralised approach*);CRs sent the data to each pilot leader. In this case, the analyses were performed centrally, (*centralized approach*).

These choices were defined considering that the recent implementation of the GDPR (and, in some cases, its different interpretation by individual countries) has increased the administrative burden linked with the access health information sources and to sharing them with research groups outside the Cancer Registry (data sharing agreements). Some of the health information sources are highly complex in terms of variables and classification (i.e. health insurance claims) and it is therefore impossible for the data to be transmitted outside the cancer registry.

5 Results of pilot implementation

5.1 Task 7.2 - Quality of care

In Italy, the pilot study on quality of care was carried out by centralising the CRs data for analyses:

- Information from 5 Italian CRs with different cancer sites (**Friuli-Venezia Giulia** with colon-rectum; **Napoli ASL3 South**; **Palermo**; **Trapani**; **Veneto**) were received along with the datasets coming from health administrative sources (e.g., hospital discharge records, hospital and pharmacy drug prescriptions, integrated home care, etc.);
- A deterministic record linkage among the different datasets of every single CR, was done by unique identifying code.

The following Table shows the Cancer Registries, tumour sites and number of cases included in the pilot 7.2.

	Friuli-Venezia Giulia	*Veneto*	*ASL3 Napoli South*	*Palermo*	*Trapani*
Tumor site	COLON	MELANOMA	COLON, MELANOMA	COLON	MELANOMA
Number of prevalent cases	**8609**	**2143**	**3155, 1439**	**2616**	**650**
Date of prevalence	31.12.2013	01.01.2016	31.12.2016	01.01.2011	31.12.2013
Date of follow-up	01.01.2014 - 31.12.2018	08.01.2016 – 22.01.2018	01.01.2017 – 04.11.2019	01.01.2011 – 07.07.2015	04.08.2003 – 31.12.2019

The following two tables report the distribution (as percentages on total prevalent cases included in the study) of the selected clinical indicators for colon cancer and cutaneous melanoma in three Italian cancer registries.

Selected indicators of care for colon cancer along the disease course	*Friuli – Venezia Giulia*	Napoli	Palermo
Prevalent cases	***8609***	***3155***	***2616***
Screen detected (by organised or opportunistic screening)	Not Available	Not Available	Not available
Diagnosis by endoscopy with biopsy	64.2%	37.3%	53.2%
Stage III resected colon cancer patients receiving adjuvant chemotherapy	7.4%	50.0%	57.0
Resected CC patients died within 30 days from surgery	1.5%	0.1%	1.8%
Metastatic patients treated with targeted therapy/biological drugs	4.0%	14.7%	0.0%
Chemotherapy use in the last 2 weeks of life	0.3%[a]	5.5%	31.0%
A hospital admission in the last 30 days of life	45.2%	19.5%	13.3%
An intensive care unit (ICU) admission in the last 30 days of life	5.2%	Not available	Not available
Enrollment in a specialist palliative care programme	5.9% (25.7% on the deceased cases)	Not available	Not available
Dead patients with information on place of death	Not Available	Not Available	Not available

Selected indicators of care for cutaneous melanoma along the disease course	Veneto	Napoli	Trapani
Prevalent cases	***2143***	***1439***	***650***
Stage IV melanomas receiving mutation testing	5.6%	0.0%	0.0%
Availability of information on the maximum thickness in millimetres (Breslow)	15.2%	64.6%	65.1%
Melanoma patients with tumour thickness of >1 mm receiving sentinel lymph node biopsy	7.5%	21.5%	14.0%
Metastatic skin melanoma patients treated with immunotherapy	10.0%	0.0%	0.0%
Chemotherapy use in the last 2 weeks of life	0.7%	10.3%	4.1%
A hospital admission in the last 30 days of life	18.6%	22.4%	19.8%
An intensive care unit (ICU) admission in the last 30 days of life	31.3%	Not available	Not available
Enrollment in a specialist palliative care programme	0.09% (1.5% on deceased cases)	Not available	Not available
Dead patients with information on place of death	31.3%	Not available	Not available

Results on skin melanoma obtained by combining Cancer Registry information and administrative health data in Veneto were presented at the GRELL annual meeting in 2021 (Lillini R et al.).

The approach used in the Pilot 7.2 was able to provide population-based indicators for patterns of care related to diagnosis, treatment, end of life.

Using international classification systems (ATC, ICD9-CM) to identify health procedures and treatments ensured comparability of methods and results.

Integration of CRs records with multiple data sources allowed to derive indicators that are not directly available in the databases (e.g., enrollment in palliative care program).

Although the pilot was carried out in a single country, the methodology can be exploited for international studies.

However, several LIMITS should be overcame in order to generalize the results and replicate studies using the same methodology, eg:

- Scarce standardisation of variables present in the additional databases;
- Different availability of databases additional to those commonly used by CR;
- Incompleteness of databases related to cancer care (e.g., hospital pharmaceutics).

5.2 Task 7.3 - Direct costs of care

Cost indicators represent a key information for policy makers in order to better allocate resources needed for cancer care. The approach proposed in Task 7.3 is based on the EPICOST model (Francisci et al, 2020) allowing to estimate the amount of resources allocated to people living with a cancer diagnosis in three phases of care, reflecting clinical patterns: initial (diagnosis and first line treatment), monitoring (follow up) and final phase. It has been successfully implemented in Italy to estimate direct costs related to breast cancer in the female population and rectal cancer. Here population-based cohorts of prevalent cases were identified according to the study protocol and linked at individual level with administrative data sources to derive information on patterns of care and associated costs. A centralized approach has been used.
This approach is replicable in all countries/regions where a cancer registry is present and linkable at individual level with other data sources reporting costs information.
In some European countries, the feasibility of extending the model of cost analysis and the procedures developed in the EPICOST Italian project has been assessed. According to the GDPR, a decentralized approach has been used.
Attempts to adjust the Epicost approach to national-based data have been implemented in Belgium, Norway, Poland, Croatia, and Spain: in the latter two countries the adjustment was unsuccessful, as data on individual costs were not available.

The approach used in Italy was modified to comply with the Belgium and Norway Cancer Registries rules.

As concerning Belgium, cancer patients can be linked to their health insurance information only in the time period from year before the diagnosis year up to 5 years after the diagnosis year. Therefore, Belgian study cohort includes adult patients (aged 15+) diagnosed with index tumors over a 5-year period before the prevalence date and still alive at prevalence date (prevalence cohort), definition of continuing and final phases of care have been adapted consequently.

As concerning Norway, exact dates of diagnosis and birth are not available, the distance in days between the date of prevalence and date of birth/diagnosis is provided instead. Moreover, exact date of prevalence is unknown, it is provided as a six months interval including the exact unknown date. As a consequence, prevalence by phase of care and costs indicators refer to the year across the six months interval including the prevalence date rather than being centered to the prevalence date.

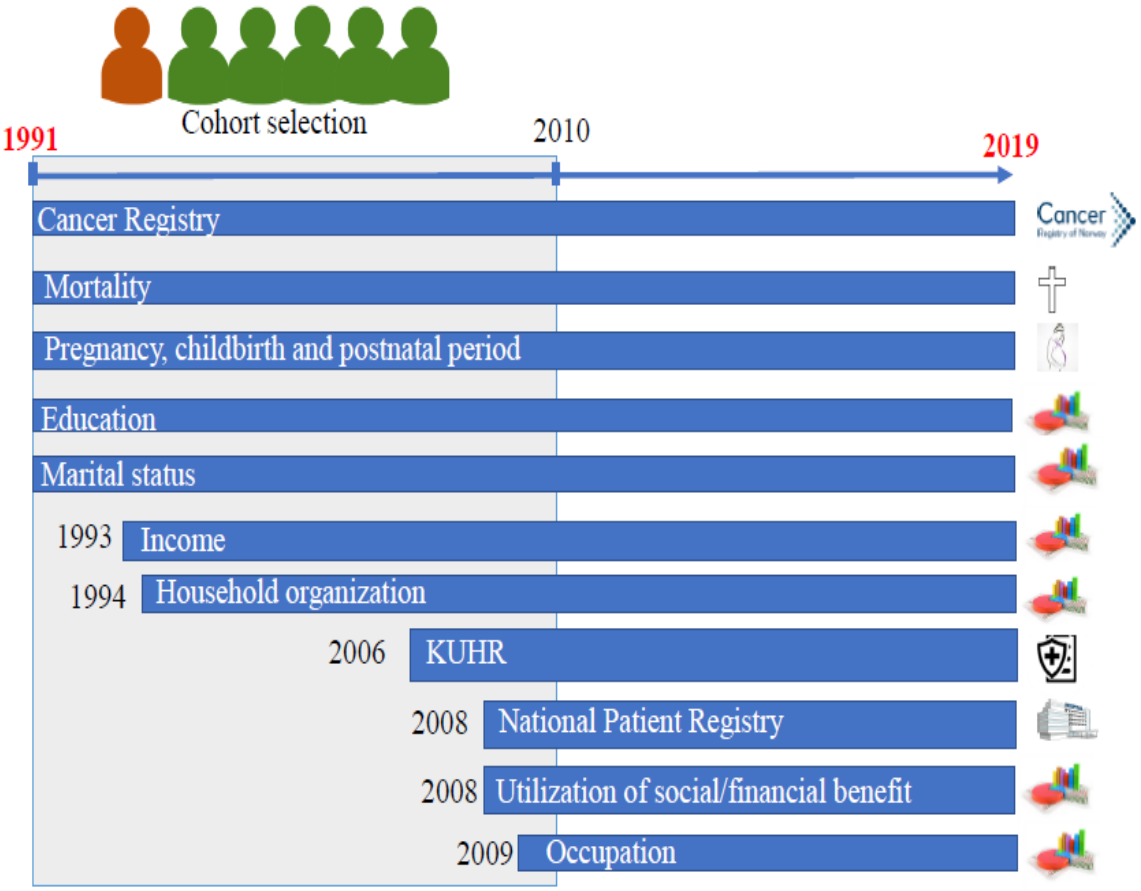
For more details on the Italian case, please refer to Gigli 2021, Francisci 2020.

5.3 Task 7.4 Long-term outcomes of AYA survivors

In Italy, we developed the first AYA cancer survivors cohort. We showed that exploiting already available data sources, it is possible, with a limited effort, to study late effects occurring in cancer survivors. Our analyses, have highlighted that AYA cancer survivors face persistent risks for a broad range of diseases (i.e. subsequent primary cancers, several comorbidities) and excess mortality in comparison with the general population up to 25 years after cancer diagnosis. AYA cancer survivors face many life transitions in terms of education, employment, social relations, and family formation. Late effects could thus have far more physical and social consequences for AYA than for older adults. Our findings, underscore the need for strict evidence-based and personalised follow-up plans for survivors, to prevent chronic cancer-induced conditions and minimise the burden of follow-up examinations.

More details are reported in Trama A, 2021. and in Trama A et al (in press)

We modified the approach used in Italy to comply with the Norway Cancer Registries (NCR) rules. In Norway, it is currently under development the cohort of AYA cancer survival matched with the general population. The following Figure present the different data sources selected in Norway



We used the approach used in Italy to define the excess risk of subsequent primary neoplasms (SMN) in AYA cancer survivors in Basque Country. The Italian cohort included 67,692 AYA cancer survivors diagnosed in 1976-2013 (median follow-up=8 years). The Spanish (Basque Country) cohort included 9,100 AYA cancer survivors diagnosed in 1986-2014 (median follow-up=13 years). First primary tumour distribution in AYA survivors was similar: breast cancers and lymphomas followed by melanomas and testicular germ cell tumours were the most common cancers in both countries. However, thyroid cancer was more common in Italy compared to Spain. In both countries AYA survivors had a 60% excess risk

of developing any SMNs (SIR=1.6); the highest risk was observed for survivors of digestive tract tumors (SIR=2.1 Italy, SIR=2.2 Spain) and lymphoma. The only differences between Spain and Italy was observed for lymphomas (SIR=2.5 Italy vs SIR=1.7 Spain). This difference is partially explained by the number of subsequent thyroid cancers which was high in Italy. We showed that AYA cancer survivors are at heightened risk of SMNs, regardless of their primary tumor. The excess risk is similar in Italy and Spain most likely because they share similar risk factors. However, observed differences may be in part attributable to thyroid cancer overdiagnosis. AYA cancers are rare, collaborative studies are important to strengthen the growing body of evidence on their long-term health risks.

6 Lessons learned

Major problems encountered in the European pilot applications regard national or regional regulations to access health administrative data sources, to link registries patients data and to share individual patients' data with collaborating partners. These problems are due to different interpretation of GDPR across European countries, lack of formalised authorization procedures, complexity of authorization procedures, timing for obtaining permissions or linked data.

Other issues were related to the completeness, standardisation and quality of administrative data sources. Difficulties in the study protocols application related to the use of national standards were also encountered. Finally, interoperability of the different datasets within and across countries resulted complex and time consuming.

From all pilots implementations we learned that the linkage with administrative data sources is challenging but is a way to promote the adoption of common standards, data re-use for research purposes (secondary use of data) and better inter-operability.

Collaboration with several experts, including clinicians, computer scientists, statisticians, cancer registration experts and owners of administrative databases should be promoted.

Quality checks of integrated data sources should be also standardised.
It is vital to include all stakeholders to reach consensus regarding the legal framework, obligations and benefits for all actors and to ensure data integration

Another important theme emerged regards IT infrastructure that should be updated to support the secondary use of data through electronic linkage. Also a common understanding of the legal framework to ensure data integration should be promoted

The role of CRs in clinical and public health research should be strengthen and widened.

7 Recommendations for implementation and adoption

General recommendations:

We support efforts to increase the visibility of the role of CRs at EU and national level, to broaden their scope and the range of data they collect and their involvement in a growing range of clinical and translational research fields, as well as cancer planning and monitoring.

Efforts at EU level need to be continued to emphasize to national policy makers the need and urgency of CR development and to provide political and financial support. CRs should increasingly become key

players in the EU health data space to reach more comprehensive cancer information systems, joining both clinical and public health components.

A standard definition of AYA and of AYA's cancer would greatly improve the comparability of data across EU countries..

IPAAC pilots and international experience (Medicare, medic-aid, Canadian) support the integration of all available health data sources, socioeconomic and demographic data sources (such as census data), cost data sources also when intended for different purposes as the best way forward to enrich CRs data.

Open issues	Recommendations
Access: many different authorization procedures, time- and resource- consuming, GDPR local interpretations	Light and fast procedures to access the data Data governance open to data sharing within country and across countries Discussion with national/European competent authority to build trust on data coming from dataset linkages
Quality: completeness and population covered by the external multiple datasets	Infrastructure updating: AI technique, text recognition (for timeliness and feasibility of data collection) Ensure comparability The more data are used the better in terms of quality they become!
Interoperability: problems of standardisation, lack of common standards	Innovative methods to ensure interoperability of the different dataset within and across countries should be developed starting from available experience
Costs: data linkage and data analyses require dedicated personnel	enrichment of current registries datasets should be adequately financially supported

8 References and documentation

- Results of the survey to European Cancer Registries to census data sources available for linkage. Task 7.1 of WP 7: https://www.ipaac.eu/res/file/outputs/wp7/advancing-registries-data-integration-administrative-data-sources.pdf
- Protocol of the pilot study 7.2 of WP7: https://www.ipaac.eu/res/file/outputs/wp7/integrating-cancer-registry-data-quality-of-care.pdf
- Study protocol Task 3 WP 7: https://www.ipaac.eu/res/file/outputs/wp7/piloting-integration-data-cancer-costs.pdf
- Protocol of the pilot study 7.4 of WP7: https://www.ipaac.eu/res/file/outputs/wp7/piloting-registries-data-integration-cancer-survivorship-adolescents-young-adults.pdf
- Bernasconi A, et al. Adolescent and Young Adult Cancer Survivors: Design and Characteristics of the First Nationwide Population-Based Cohort in Italy. J Adolesc Young Adult Oncol. 2020 Oct;9(5):586-593.
- Labianca R et al. Early colon cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2013;24:vi64-72.
- Van Cutsem E et al. Metastatic colorectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2014;25:iii1-9.
- Glynne-Jones R et al. Rectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2017;28:iv22-40
- Dummer R et al. Cutaneous melanoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2015;26: v136-132

- Ducreux M et al. Cancer of the pancreas: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2015;26: v56-68.
- Barbera et al. Quality of end-of-life cancer care in Canada: a retrospective four-province study using administrative health care data. Curr Oncol. 2015;22:341-55.
- Gigli A, Francisci S, Capodaglio G, et al. The Economic Impact of Rectal Cancer: A Population-Based Study in Italy. Int J of Environ Res Public Health 2021, 18, 474.
- Gigli A, Francisci S, Guzzinati S, Hall A, Hackey M, Scoppa S, Mariotto A. Cancer prevalence by phase of care: an indicator for assessing health service needs. Tumori. 2021 Aug;107(4):311-317.
- Busco S, Tavilla A, Gigli A, Lopez T, Pierannunzio D, Mallone S, Guzzinati S, Capodaglio G, Giusti F, Angiolini C, Francisci S. A direct method for the identification of patterns of care using administrative databases: the case of breast cancer. Eur J Health Econ. 2021 Dec;22(9):1477-1485.
- Francisci S, Guzzinati S, Capodaglio G et al. Patterns of care and cost profiles of women with breast cancer in Italy: EPICOST study based on real world data. Eur J Health Econ 2020 Sep; 21(7):1003-1013.
- ENCR-JRC Call for Data 2015: https://encr.eu/sites/default/files/ENCR-JRC2015_Protocol.pdf
- EUROCARE-6 protocol for updating population-based cancer survival in Europe, June 2015 – available at: http://www.eurocare.it/Eurocare6/ProtocolsEU6/tabid/93/Default.aspx
- Lillini R, Rugge M, Guzzinati S, Bonfarnuzzo S, Magri MC, Minicozzi P, Sant M. Indicators of standard care for skin melanoma, computed by combining Cancer Registry information and administrative health data in Veneto. First results from the iPAAC 7.2 Pilot Study. Oral presentation. 45th Annual Meeting of the Group of Cancer Epidemiology and Registration in Latin Language Countries, 12-14th May 2021 Luxembourg: https://234d9f1c-3134-4e15-8ed4-2ea0a4757907.filesusr.com/ugd/6510a0_219994d3f07f43a4b08ad9f8d8906116.pdf
- Trama A, et al. Excess risk of subsequent malignant neoplasms in adolescent and young adult cancer survivors: Results from the first Italian population-based cohort. Cancer. 2021 Sep 28. doi: 10.1002/cncr.33931
- Trama A, et al. Late mortality, subsequent malignant neoplasms and hospitalisations in long-term survivors of adolescent and young adult haematological cancers. Under review in Frontiers Oncology