ENCR
European Network
of Cancer Registries

Position paper of the Survival in ECIS Working Group

Version 1 October 2025

Contents

Authors of the Position paper of the Working Group "Survival in ECIS"

Volker Arndt, Baden-Württemberg Cancer Registry, Germany, Working Group Coordinator
Otto Visser, Netherlands Cancer Registry, IKNL, Coordinator of sub-Working Groups 1 and 2
Roberta De Angelis, EUROCARE project, Istituto Superiore di Sanità, Rome, Italy, Coordinator of sub-Working Group 3
David Pettersson, European Commission Joint Research Centre (JRC), Coordinator of sub-Working Group 4
Joanna Bartnicka, European Commission Joint Research Centre (JRC)
Manola Bettio, European Commission Joint Research Centre (JRC)
Cristina Bosetti, European Commission Joint Research Centre (JRC)
Laura Botta, EUROCARE project, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
Maria Dolores Chirlaque, Murcia Cancer Registry, Department of Epidemiology, Regional Health Council of Murcia, Spain
Elena Demuru, EUROCARE project, Istituto Superiore di Sanità, Rome, Italy
Tadeusz Dyba, European Commission Joint Research Centre (JRC)
Manuela Flego, European Commission Joint Research Centre (JRC)
Gemma Gatta, EUROCARE project, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
Lina Jansen, Baden-Württemberg Cancer Registry, Germany
Paul Christopher Lambert, Cancer Registry of Norway
Roberto Lillini, EUROCARE project, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
Carmen Martos, Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO), Spain
Irmina Michalek, Department of Cancer Pathology Maria Sklodowska-Curie National Research Institute of Oncology (MSCI), Poland
Tor Åge Myklebust, Cancer Registry of Norway
Raquel Negrão Carvalho, European Commission Joint Research Centre (JRC)
Giorgia Randi, European Commission Joint Research Centre (JRC)
Silvia Rossi, EUROCARE project, Istituto Superiore di Sanità, Rome, Italy
Mark Rutherford, University of Leicester, England, UK
Geert Silversmit, Belgian Cancer Registry, Belgium
Grit Sommer, Childhood Cancer Registry of Switzerland
Annalisa Trama, EUROCARE project, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

Acknowledgements

The authours of the position paper gratefully acknowledge Claudia Allemani, Michel Coleman and Veronica Di Carlo from the CONCORD project, London School of Hygiene & Tropical Medicine, England, UK, for their expert advice.

Members of the Working Group "Survival in ECIS"

Volker Arndt, Baden-Württemberg Cancer Registry, Germany, Working Group Coordinator
Otto Visser, Netherlands Cancer Registry, IKNL, Coordinator of sub-Working Groups 1 and 2
Roberta De Angelis, EUROCARE project, Istituto Superiore di Sanità, Rome, Italy, Coordinator of sub-Working Group 3
David Pettersson, European Commission Joint Research Centre (JRC), Coordinator of sub-Working Group 4
Claudia Allemani, CONCORD project, London School of Hygiene & Tropical Medicine, England, UK
Joanna Bartnicka, European Commission Joint Research Centre (JRC)
Manola Bettio, European Commission Joint Research Centre (JRC)
Cristina Bosetti, European Commission Joint Research Centre (JRC)
Laura Botta, EUROCARE project, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
Maria Dolores Chirlaque, Murcia Cancer Registry, Department of Epidemiology, Regional Health Council of Murcia, Spain
Michel Coleman, CONCORD project, London School of Hygiene & Tropical Medicine, England, UK
Elena Demuru, EUROCARE project, Istituto Superiore di Sanità, Rome, Italy
Veronica Di Carlo, CONCORD project, London School of Hygiene & Tropical Medicine, England, UK
Tadeusz Dyba, European Commission Joint Research Centre (JRC)
Manuela Flego, European Commission Joint Research Centre (JRC)
Gemma Gatta, EUROCARE project, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
Lina Jansen, Baden-Württemberg Cancer Registry, Germany
Paul Christopher Lambert, Cancer Registry of Norway
Roberto Lillini, EUROCARE project, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
Carmen Martos, Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO), Spain
Irmina Michalek, Polish National Cancer Registry, Poland
Tor Åge Myklebust, Cancer Registry of Norway
Raquel Negrão Carvalho, European Commission Joint Research Centre (JRC)
Giorgia Randi, European Commission Joint Research Centre (JRC)
Silvia Rossi, EUROCARE project, Istituto Superiore di Sanità, Rome, Italy
Mark Rutherford, University of Leicester, England, UK
Geert Silversmit, Belgian Cancer Registry, Belgium
Grit Sommer, Childhood Cancer Registry of Switzerland
Annalisa Trama, EUROCARE project, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

1. Introduction

The European Cancer Information System (ECIS), a key platform for disseminating cancer burden indicators across Europe-namely incidence, mortality, survival, and prevalence has been developed and maintained by the European Commission's Joint Research Centre (JRC), in collaboration with the European Network of Cancer Registries (ENCR). As part of its commitment to delivering timely and relevant cancer information to stakeholders, the European Commission has recognized the need to update and enhance the survival estimates available through ECIS. Reporting cancer survival estimates offers crucial insights into the effectiveness of early detection and treatment strategies. As of now, the survival included in ECIS is derived from the EUROCARE-5 project, which considers cancers diagnosed 1999 to 2007, thus necessitating update to include recent years.

To address this gap, the JRC, in collaboration with DG SANTE, has embarked on an initiative to compute and integrate up-to-date cancer survival data into ECIS. By aligning cancer survival estimates with the current ECIS frameworks for incidence and mortality—considering geographical specifics, cancer site definitions, age ranges, and timeliness—the initiative aims to provide more granular and accurate survival information. This task involves a collaborative effort with European cancer registries and European projects (e.g. EUROCARE, CONCORD).

In addition to the timeliness requirements, the existing data is insufficiently detailed to provide a comprehensive regional overview. Thus, there is a pressing need to refine these indicators to offer insights at the regional level and to better capture survival trends over time, which could further inform policy and healthcare decisions.

A survival strategy document has been drafted by the JRC-ENCR Management Committee (Annex 1). In line with the objectives, the Committee also decided to establish a dedicated working group (WG) of experts to agree on the data standards, quality checks and methods, for computation and publication of survival in ECIS. This WG represents a first concrete JRC-ENCR initiative towards combining expertise and reaching agreement between survival experts and stakeholders for the routinely provision of survival in ECIS. The WG was charged with devising the methodological framework needed to validate data and conduct survival analyses for ECIS dissemination. Its work aimed to ensure alignment with ECIS's historical incidence and mortality, establishing the necessary expertise and technological infrastructure for the sustainable and timely delivery and sharing of enhanced survival and prevalence data.

Meanwhile, a Direct Grant to Member States' authorities (Joint Action CR-g-24-40) was launched in 2024, aimed at supporting quality improvement of cancer registry data feeding ECIS. Its second main objective focuses on the improvement of survival and prevalence indicators in ECIS, by:

"*…building on the already available mechanisms in Member States to feed ECIS with record-based data held by cancer registries.*

This part of the Joint Action should consist in validating and processing data collected via data calls to European registries and deriving up-to-date survival and prevalence indicators in alignment with the current ECIS settings for incidence and mortality figures, in terms of geographical detail, cancer sites definition, age range availability and timeliness".

Based on the Joint Action schedule, cancer registries are tasked with submitting data that enable the computation of updated survival and prevalence indicators, ensuring their availability in ECIS by the end of 2026. This work includes developing and piloting processes for the ongoing provision of

current data to both ECIS and the European Cancer Inequalities Registry, which requires a level of data timeliness not currently met by previous projects.

The mandate of the survival WG was therefore enlarged for also laying the groundwork for the activities of the CancerWatch Joint Action.

2. Methodology

A Call for expression of interest to experts in the field was published on the ENCR website and in the ENCR newsletter. In addition, other experts were invited directly because of their participation in international research projects on survival, or renowned expertise in the field.
The working group “Survival in ECIS”, chaired by Volker Arndt, had its first online meeting in June 2024 and met regularly until July 2025, with a total of nine meetings. The experts identified four main topics that the working group should address and decided to work in dedicated sub-working groups(sub-WG), as follows:

1. Methods and output (Chaired by Otto Visser): on definition of parameters (observed/period, relative, net, conditional survival) and addressing granularity/subgroups (length of survival, tumour sites, histology) proposed to be published in ECIS.
2. Data quality criteria (Chaired by Otto Visser): addressing data quality criteria (for reporting data quality of CR in ECIS and assess inclusion of data/registries, data validation and plausibility checks that are specifically needed for survival analysis.
3. Data release to third parties (Chaired by Roberta de Angelis): to elaborate on the appropriate data format for releasing ECIS survival data by the JRC-ENCR to third-party research projects.
4. Aggregated data delivery to ECIS (Chaired by David Pettersson): on how to deal with registries not sharing record-level data to estimate survival indicators in ECIS

Each sub-group had a number of online meetings and regularly briefed the whole working group during plenary meetings. Each sub-group produced a document reporting on the consensus or describing issues that remained unresolved. This document consolidates all outputs, incorporating comments and revision from the full ‘Survival in ECIS’ WG.

3. Survival methods and output

A set of parameters was proposed by sub-WG1 for the survival estimation and the granularity/sub-groups that should be presented on ECIS, including the length of survival, the tumour definitions and histology.

3.1. Methods for survival estimation

The proposed survival indicator for dissemination in ECIS is the age standardized net survival for each subsequent follow-up year up to five years after diagnosis (1-year survival, 2-year survival, 3-year survival, 4-year survival, 5-year survival).
For complete periods, the cohort method shall be used. For incomplete periods of follow-up, the period method shall be used.
The issue of net survival beyond five years remains under discussion, and will be resolved later.
Similarly, it was agreed that additional outcome measures (e.g. crude survival, relative survival, 10-year survival, etc.) or parameters (e.g. morphological subgroups) would be possible, but their inclusion is not required at this time.

3.1.1. Minimum number of cases for publication of survival estimates in ECIS

In sub-WG 1, there was no consensus on the minimum number of cases needed for considering the survival estimate in ECIS. Applying a threshold consisting of at least 50, 100 or even 200 cases in the analysed population stratum would allow more stable estimates. Although the majority of the WG members agreed with the introduction of a threshold, others were opposed to any threshold.

Arguments against a threshold

- data from areas with small populations would not be shown
- data for rare cancers would not be shown (e.g. children)

Arguments in favor of the threshold

- more stable estimates
- comparisons between registries/countries would be more reliable
- given confidence intervals are difficult to understand for the public, in most situations, only the estimate number would be presented (which may result in wrong conclusions)
- to avoid exclusion of small populations, a user can combine strata (e.g. years/periods to increase the number of cases even though the problem persists for very small countries and/or rare entities
- for children a lower threshold can be used, e.g. 10

Regardless of the application of a threshold, confidence intervals providing the statistical uncertainty of estimations, can be added to allow proper interpretation of the estimate, if not by default, at least as a selectable option to professionals and other stakeholders able to interpret them.

3.2. Parameters for survival dissemination

3.2.1. Period of diagnosis

- Each individual incidence year as of 1981 (as far as available)
- 5-year incidence periods (1981-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005, 2006-2010, 2011-2015, 2016-2020, etc..)
- 10-year incidence periods (1981-1990, 1991-2000, 2001-2010, 2011-2020, etc.)

3.2.2. Geographical area

- Europe and EU-27
- Country
- Optional for the large countries (e.g. France, Germany, Italy, Poland, Spain, UK): NUTS1 level
  To be decided by country

3.2.3. Sex

- Males
- Females
- Males+females

3.2.4. Age groups

- Children (0-14) and adolescents (15-19)
  - 0-14
    - 0
    - 1-4
    - 5-9
    - 10-14
  - 15-19
- Adults (15 or older)
  - 15-39
  - 40-54
  - 55-64
  - 65-74
  - 75+

  Note: these age groups differ from the age groups for age standardization.

3.2.5. Cancer entities

3.2.5.1. Cancer in children and adolescents (0-18)

ICCC main groups; see the latest version for the definitions

- All sites
- I Leukaemia
- II Lymphoma

- III CNS
- IV Neuroblastoma
- V Retinoblastoma
- VI Renal tumours
- VII Hepatic tumours
- VIII Bone
- IX Soft tissue
- X Germ cell tumours
- XI Other
- XII Unspecified

3.2.5.2. For adults (15 or older)

Topography and morphology based on ICD-O-codes; exact codes available in ECIS

- All sites excluding keratinocytic skin cancers
- Carcinoma of head & neck
- Carcinoma of lip
- Carcinoma of oral cavity
- Carcinoma of salivary glands
- Carcinoma of oropharynx
- Carcinoma of nasopharynx
- Carcinoma of hypopharynx
- Carcinoma of nasal cavity and middle ear
- Carcinoma of accessory sinuses
- Carcinoma of larynx
- Carcinoma of oesophagus
- Carcinoma of stomach
- Carcinoma of small Intestine
- Carcinoma of colon and rectum
- Carcinoma of colon
- Carcinoma of rectum
- Carcinoma of anus and anal canal
- Carcinoma of liver and intrahepatic bile ducts
- Carcinoma of gallbladder
- Carcinoma of other and unspecified parts of biliary tract
- Carcinoma of pancreas
- Gastro-entero-pancreatic neuroendocrine tumour (GEP-NET)
- Carcinoma of trachea
- Carcinoma of bronchus and lung
- Carcinoma of thymus
- Mesothelioma
- Sarcoma of bones, joints and articular cartilage
- Sarcoma of soft tissues and viscera
- Melanoma of skin
- Merkel cell carcinoma of skin
- Adnexal and skin appendage neoplasm
- Kaposi Sarcoma
- Carcinoma of breast
- Carcinoma of vulva

- Carcinoma of vagina
- Carcinoma of cervix uteri
- Carcinoma of corpus uteri
- Carcinoma of ovary
- Carcinoma of the penis
- Carcinoma of prostate
- Germ cell tumours of testis
- Carcinoma of kidney
- Carcinoma of renal pelvis and ureter
- Carcinoma of bladder
- Malignant melanoma of uvea
- Malignant melanoma of conjunctiva
- Malignant brain and other CNS tumours
- Carcinoma of thyroid gland
- Mucosal melanoma
- Precursor cell lymphoblastic leukaemia/lymphoma
- Non Hodgkin lymphoma, incl. CLL
- Hodgkin lymphoma
- Plasma cell neoplasms
- Myeloproliferative neoplasms
- Myelodysplastic neoplasms
- Myeloid leukaemia / biphenotypic leukaemia

Stratifications of survival by histological cancer type/subtype and stage are expected among the available visualisation of ECIS cancer statistics once these two dimensions are included in the visualisation of incidence indicators.

4. Data quality

4.1. Quality checks for survival estimation

The sub-working group 2 discussed the data quality criteria for inclusion of data/registries in survival computation and specified data validation and plausibility checks. Results of quality checks can be classified differently depending on the impact this has on specific analyses (e.g. incidence, survival etc.). The same check that could be irrelevant for incidence analysis could qualify as important for the survival analysis, requiring thus the exclusion of the case (e.g missing/unknown vital status).

This chapter aims to highlight situations in which results of the quality control have no impact on incidence but have an important impact on survival. In other words, we aim to highlight situations in which starting from the same dataset, the cases included in the incidence and survival analyses could, rightly, differ. The validation rules described below will guide the preparation of the study dataset for survival analysis.

In addition, this chapter proposes data imputation procedures to be performed for some variables.

4.1.1. The main steps in the preparation of the study dataset

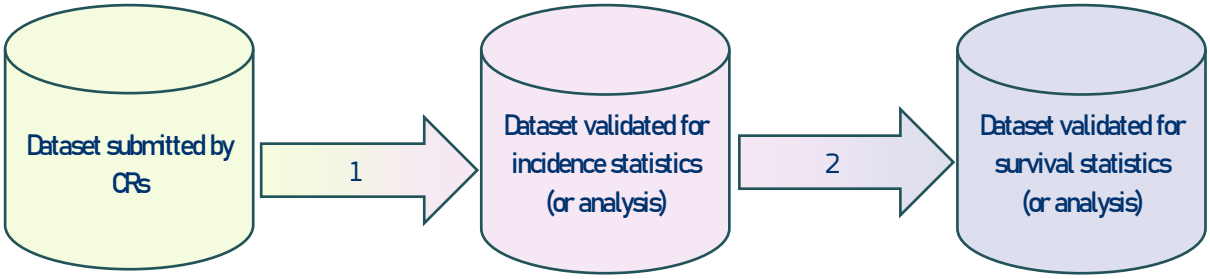

Figure 1. ***Diagram showing the pathway from the original dataset to the study dataset for survival analysis.*** *Arrow 1 represents the preparation process of the validated dataset, according to the JRC-ENCR data quality checks for incidence analysis. Starting from the validated dataset, arrow 2 represents the preparation process of the study dataset, which includes additional quality checks specific for survival analysis.*

The preparation of the study dataset for survival analysis is done in two main steps. The quality checks and validation rules applied to the original datasets, as submitted by the Cancer Registries (CRs) to the JRC (see arrow 1 in Figure 1 above), are described in the Annex 3 of this document. The result of this process is a validated dataset for incidence analysis. This validated dataset is the starting point for further preparation of the study dataset (see arrow 2 in Figure 1 above), in this case a dataset for survival analysis. The additional quality checks specific for survival analysis are described in the main body of this chapter and detailed in Annex 4.

To be noted that exclusions are study specific and do not regard the validated dataset. The data cleaning procedures do not lead to the deletion of the record from the dataset, except in the case of mis-recordings. All inconsistencies or errors not solved but approved by the registry do remain in the dataset, and they are essential for reporting analysis specific quality indicators (see the final document of Sub-working group 3 “European Cancer Information System: mechanisms and formats for sharing cancer survival data”).

In addition to the validation of the incidence file, the life tables submitted by the CRs are routinely checked with respect to the concordance with the other files in the geographical and temporal coverage, and whether the values provided are in range.

4.1.2. Definitions

In this document, the following definitions of inconsistencies in cancer registry data coding are applied:

- **Compulsory revisions**: invalid/missing values in core survival variables (variables regarding and the date of last ascertainment of vital status, survival time and vital status) or invalid combinations of two or more core variables. Such inconsistencies impair the computation of survival estimates, thus implying the exclusion of the record from the survival analysis, if the value cannot be imputed or is not corrected.

- **Additional revisions**: invalid/missing values in variables that are not core or unlikely/rare combinations of two or more variables. Such records are also reported to the CR, however, if the CR cannot correct them, they are included in the study dataset for survival analysis with an appropriate error label.

Note: in this chapter, these labels and the related definitions are used in the context of preparation of the study dataset for survival analysis. Their meaning is not the same as the label "error" and "warning" used in the JRC-ENCR Quality Checks Software (QCS), since the QCS is primarily a tool for flagging incorrect/unlikely codes in the datasets submitted by the cancer registries. Further to the QCS identification of data inconsistencies, the JRC is then applying additional scripts that process the QCS output to build the validated dataset.

The preparation of the study dataset for survival analysis will be based on the results of the QCS output files and the aim of the specific survival analyses, which will determine the inclusion of an additional label for compulsory or optional corrections sent to the registries. Only after consultation with the registry, the result of each univariate or multivariate check (see below) will define an appropriate label that may lead to exclusion or inclusion of records.

4.1.3. Types of Quality checks

Quality checks in the QCS are of three types:

- **Univariate checks**: to validate the consistency of the single variables, check for systematic anomalies in variables, and detect values not compliant with the data collection protocol;

- **Multivariate checks (intra-record)**: to validate the consistency between variables in each individual record and detect invalid or unlikely combinations of variables;

- **Multivariate checks (inter-record):** to validate the consistency of multiple records for the same individual and detect duplicated records (duplicate patient ID and tumour ID) or false multiple primaries (not compliant with the IARC-ENCR international rules for multiple primary tumours).

The first two types, vertical and horizontal checks for the preparation of the study dataset for survival analysis, are addressed in this document. They are further documented with examples in Annex 4, along with the relevant QCS labels. Univariate and multivariate checks addressed in the previous step of preparation of the validated dataset for incidence analysis (see arrow 1 in Figure 1) are described in Annex 3.

4.1.3.1. QCS Univariate checks

1. **The last known vital status (Vit_stat)**

This variable is crucial for survival analysis. CRs that do not provide valid information on the last known vital status must be excluded from survival studies. If unknown/missing/invalid values are present, they must be corrected or excluded. High proportions of unknown/missing/invalid values show a systematic bias of the CR, which cannot be ignored and must be addressed.

In the JRC-ENCR preparation of the study dataset for survival analysis, missing, unknown or invalid vital status will be reported as a compulsory revision to CRs, which will determine exclusion of non-corrected records from survival analysis.

2. **Date of follow up: year of the last known vital status (YoF) and month of the last known vital status (MoF)**

The follow-up date is a crucial variable in survival analysis and missing/unknown/invalid values are not allowed. If cases with these YoF/MoF values are present, they must be corrected by the CR, imputed from survival time, or if this is not possible, excluded from analysis. High proportions of unknown/missing/invalid values show systematic bias of the CR, which cannot be ignored and should be addressed.

In the JRC-ENCR preparation of the study dataset for survival analysis, all records with missing, unknown or invalid YoF will be reported as an additional revision to CRs, with an appropriate label. Records with unknown/missing date of follow-up, as well as survival time, will be identified by the QCS check W-SUMU (described later) which will determine exclusion of non-corrected records from the survival analysis.

3. **Survival time in days (Surv_time)**

The survival time is a crucial variable in survival analysis and missing/unknown/invalid values are not allowed. If cases with these survival time values are present, they must be corrected by the CR, imputed from the dates of incidence and follow-up, or if this is not possible, excluded from analysis. High proportions of unknown/missing/invalid values show systematic bias of the CR, which cannot be ignored and should be addressed.

In the JRC-ENCR preparation of the study dataset for survival analysis, all records with missing, unknown or invalid survival time will be reported as an additional revision to CRs, with an appropriate label. Records with unknown survival time, as well as the date of follow-up, will be identified by the QCS check W-SUMU (described later) which will determine exclusion of non-corrected records from survival analysis.

4. **Incidental finding of cancer at autopsy (Autopsy variable)**

Autopsy cases must be excluded from survival analysis. To be noted that only a minimum part of records can be incidentally identified at autopsy. If unknown/missing/invalid values are present, they must be reported to the CR. High proportions of cases with these values show systematic bias of the CR, which cannot be ignored and should be addressed.

In the JRC-ENCR preparation of the study dataset for survival analysis, missing, unknown or invalid values for cases incidentally discovered at autopsy will be dealt with in two different ways, depending on the survival time variable:

- Cases with survival time > 0 are likely non-autopsy cases so they will be reported as additional revision to CRs, which will anyway result in inclusion of non-corrected records for the preparation of the study dataset for survival, with an appropriate label.
- For cases with survival time = 0, and which are not DCO cases, the status of autopsy is ambiguous. These cases will be reported as a compulsory revision to CRs, which will determine exclusion of non-corrected records from the survival analysis thanks to an appropriate error label.

4.1.3.2. QCS Multivariate checks

1. **Consistency between "date of follow-up" and "date of incidence" or "date of birth"**

QCS code: E-CoDV.

In the JRC-ENCR preparation of the study dataset for survival analysis, cases with the "date of follow-up" falling before the "date of incidence" or before the "date of birth" will be reported as a compulsory revision to CRs, which will determine exclusion of non-corrected records from the survival analysis.

2. **Consistency between vital status =1 (alive), autopsy and basis of diagnosis**

QCS code: E-VSBD.

In the JRC-ENCR preparation of the study dataset for survival analysis, cases with the following combination of variables:

- Vital status = 1 (alive) AND

- either autopsy = 1 (incidental finding at autopsy), or basis of diagnosis = 0 (DCO)

will be reported as compulsory revisions to CRs, which will determine exclusion of non-corrected records from survival analysis.

3. **Consistency between autopsy =1 (incidental finding at autopsy), vital status, survival time, dates of incidence and follow-up**

QCS code: E-AUVS.

In the JRC-ENCR preparation of the study dataset for survival analysis, cases with the following combination of variables:

- Autopsy=1 (incidental finding at autopsy) AND

- either Vital status not 2 (dead), or survival time not 0, or date of incidence not equal to date of follow up,

will be reported as a compulsory revision to CRs, which will determine exclusion of non-corrected records from survival analysis.

4. **Consistency between basis of diagnosis (BoD) = Death Certificate Only (DCO) and "last known vital status", "survival duration in days", "dates of incidence/follow-up"**

QCS code: E-BDVS.

In the JRC-ENCR preparation of the study dataset for survival analysis, cases with the following combination of variables:

- BoD=0 (DCO) AND

- either Vital status not 2 (dead), or survival time not 0, or date of incidence not equal to date of follow up,

will be reported as compulsory revisions to CRs, which will determine exclusion of non-corrected records from survival analysis.

5. **Missing/unknown "survival duration in days" and "date of follow-up" or "date of incidence"**

QCS code: W-SUMU.

Survival duration in days is essential in the survival analysis. If not already calculated and provided by the cancer registry, it can be calculated and imputed by the JRC provided that both dates of incidence and follow-up are present. If one of the variables "date of incidence" or "date of follow-up" are not present, survival duration in days is considered as missing/unknown.

In the JRC-ENCR preparation of the study dataset for survival analysis, missing or unknown “survival duration in days” AND (“date of follow-up” OR “date of incidence”), will be reported as a compulsory revision to CRs, which will determine exclusion of non-corrected records from survival analysis.

6. **Inconsistency between “survival duration in days” and “date of follow-up” or “date of incidence”**

QCS code: E-SUDA.

Accurate survival duration in days is essential in survival analysis. This check labels records, where the survival time (in months) differs by more than one month from the distance (in months) between the incidence date and the follow-up date.

In the JRC-ENCR preparation of the study dataset for survival analysis, cases with the following combination of variables:

- Survival time in months = x (defined by Survival time in days / 30.43675) AND

- Date of follow-up – date of incidence > x+1 OR Date of follow-up – date of incidence < x-1

will be reported as compulsory revisions to CRs, which will determine exclusion of non-corrected records from survival analysis.

* Average number of days in a calendar month

4.1.3.3. Data standardization procedures - imputation of missing follow-up dates or survival time

Some missing variables are reconstructed to complete the information when possible.

1. **Imputation of dates of follow-up**

This is done for the month and year of follow-up if missing/unknown/invalid, as complete dates of follow-up are required by statistical software to estimate net or relative survival. To this purpose, the availability of survival duration in days and the date of incidence is crucial. Completeness of the date of incidence is already ensured following validation of the dataset for incidence computation.

When performing imputation of the date of follow-up, it is assumed that the day of incidence falls on the 15th of the month.

2. **Imputation of survival time**

When survival duration in days is missing/unknown/invalid, but the dates of incidence and follow-up are complete in month and year, the imputation of survival time in days is done assuming 15th of the month for the incidence day and the 16th for the follow-up day.

4.1.4. Additional steps during survival analysis

In addition to the described checks and imputations, other steps are required for survival analysis:

- exclusion of Death Certificate Only cases;
- exclusion of cases discovered by incidental finding at autopsy;
- records with survival time coded as 0 days: these records can either be excluded or included in the analysis; the action to be taken will depend on the vital status and proportion of such cases in the dataset/analysis.
    - cases with survival time 0 and vital status dead (not DCO, not autopsy) are possible as they can potentially be patients who die shortly after or during diagnosis. These cases should be included in the survival analysis
    - cases with survival time 0 and vital status alive could be censored.

The number of excluded cases is tracked and used to compute the quality indicators described in the section 4.2.

Note: in case the statistical software requires that the survival time is > 0 for the record to be included, records with survival time = 0 (which are not DCO and not autopsy cases) could be modified to have survival time = 1 day.

4.1.5. Conclusions and next steps

The validation rules described above constitute the ENCR recommendation on how the ECIS validated dataset for incidence should be further worked to prepare the study dataset for survival.

On top of inconsistencies already detected and addressed at the stage of preparation of the validated dataset for incidence computation, the QCS will be used to detect the additional inconsistencies relevant for survival. Most of the checks described above are already implemented in the QCS, and additional ones proposed and agreed in the framework of the WG on survival will be added by the JRC.

Dedicated additional tools for the implementation of the validation rules described in this document (e.g. selection of records for compulsory revision by CRs, imputation of variables, etc) will be addressed in a second stage, possibly in the context of the upcoming JA CancerWatch.

4.2. Quality indicators

The sub-working group on data quality also discussed the data quality indicators for the data/registries that should be computed to accompany the survival estimates in ECIS.

Quality indicators for survival aim to detect disturbing factors that may influence the survival estimates, i.e. factors that result in over- or underestimation of survival rates.

4.2.1. Selective (in)completeness of cancer incidence

Incompleteness in its own will have limited influence on survival estimates. However, if incompleteness is not evenly distributed among the different prognostic groups, this will result in over- or underestimation of survival rates, depending on the prognostic group that is (partially) missing. This may concern:

- Stage or grade (higher stage or grade has lower survival)
- Age groups (older age groups mostly have lower survival)
- Morphology groups (e.g. NET or SCLC/NSCLC)

Quality indicators should give an impression of the presence/absence of selective incompleteness.

4.2.1.1. Death certificate only (DCO)

A high proportion of DCO-cases may result in over- or underestimates of survival, depending on the (im)possibility to trace back and the stage/age distribution of the 'missed' cases (i.e. cases that were not a DCO and not notified to the registry).

4.2.1.2. Microscopic (pathological) verification (MV)

A low proportion of MV-cases indicates a poor quality of morphology information and may lead to comparability issues. Excluding cases without MV may result in overestimation of survival, as cases without MV generally have lower survival, as per the following examples:

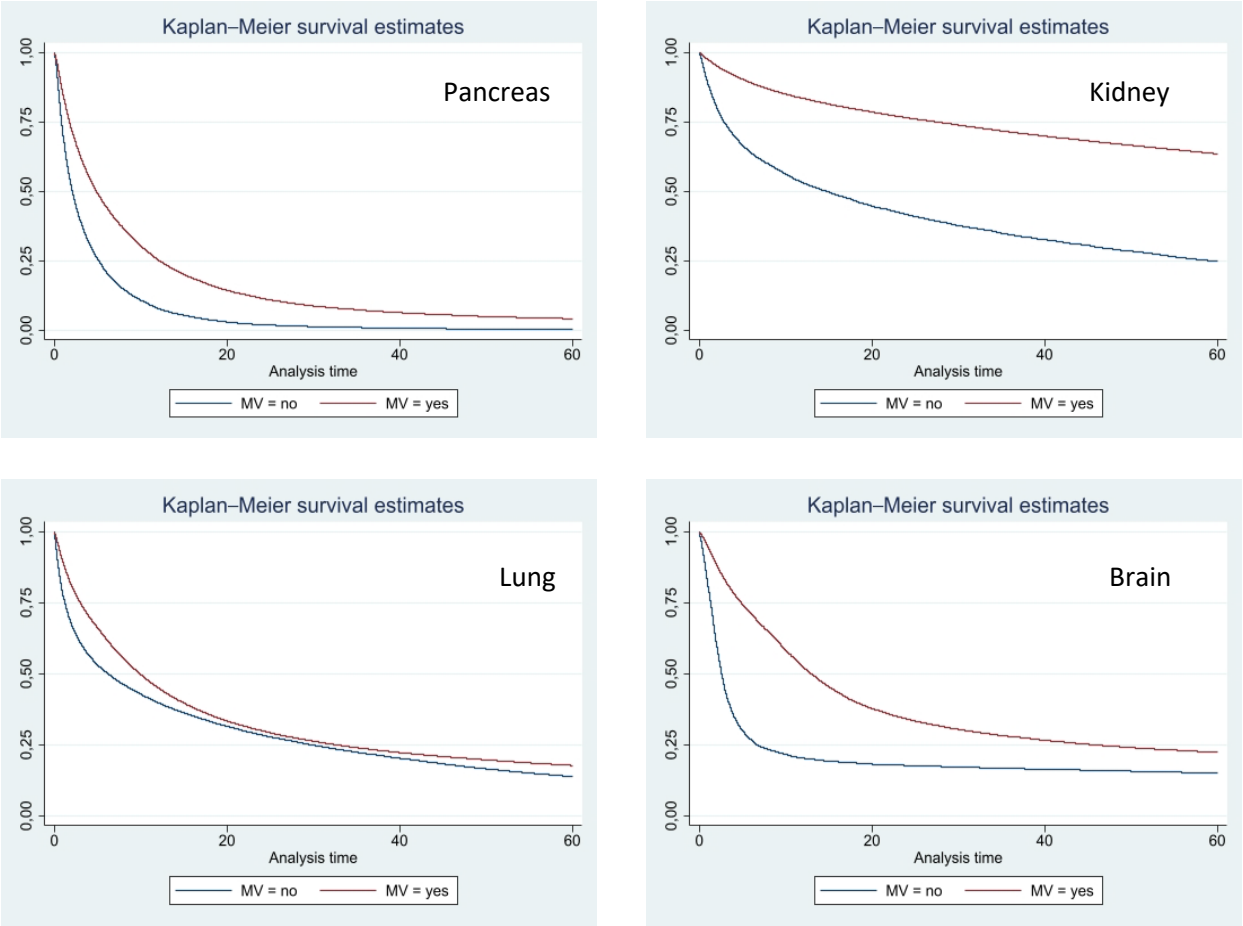

Figure 2: *Survival estimates with or without morphological verification for pancreas, kidney, lung and brain.*

4.2.2. Poor follow-up procedures

Poor follow-up procedures (= death events 'missed' by the registry) result in overestimation of survival. Methods to detect poor follow-up are:

4.2.2.1. Immortals

Count the number of 'immortals' (e.g. proportion surviving patients born > 100 years before date of last follow-up)

4.2.2.2. Groups with expected poor survival

If survival rates for groups for which (very) poor survival can be expected (untreated patients, stage IV patients, pancreas, SCLC, glioblastoma, 75+, etc.), is relatively high this may be an indication for incompleteness of follow-up. Higher survival than expected on its own is not a problem, as long as there is a plausible explanation.

4.2.2.3. Censored cases before the end of follow-up

Cases that were censored before the end of follow-up period may be in different categories, depending on the follow-up procedures in a cancer registry.

- For cancer registries that are able to do a record linkage to the population register (passive follow-up) there are no patients ‘lost to follow-up’. However, there may be cases that were censored before the end of follow-up because of emigration from the registration area: % emigration
- For cancer registries that do active follow-up cases that could not be followed until the end of the follow-up period: % lost-to-follow-up

If a registry does not reach a tentative threshold this should be a reason to contact the registry for an explanation.

5. Aggregated data delivery to ECIS

A dedicated sub-working group discussed the challenges of survival estimation for registries contributing to ECIS survival indicators without sharing record-level data

In order to publish up-to-date survival estimates for the European area, the European Cancer Information System (ECIS) needs a software solution (Cancer Registries Aggregation Tool for ECIS on Survival CRATE 2.0) for federated calculation of survival estimates for those registries and countries that for different reasons are unable to provide record-level data to the JRC. A subset of ENCR registries have already been involved in an aggregated-data submission of cancer incidence data to ECIS in 2024 and 2025 using the CRATE software.

Important requirements for such a solution are 1) the ability to calculate estimates identical to those calculated by JRCs cancer information team on record-level data, 2) a reasonable amount of resources from the cancer registries to run the software, and 3) provision of quality indicators and logs to ensure that the process runs properly and that comparability can be assessed.

The Nordic countries have, for several years, used a distributed approach to collect indicators (including survival) for the NORDCAN web application using the nordcan.R software. The overall experience of this approach has been good according to those involved in collecting of data and updating the NORDCAN web application. A critical assessment of the NORDCAN solution, addressing pros and cons of its implementation in the ECIS context, is the scope of the current document.

One objective of the CancerWatch JA starting in September 2025 is to develop a software that can be used for a distributed/federated collection of ECIS survival estimates. The Joint Action also aims at streamlining the different software (quality check tools) currently in use by different stakeholders using cancer registry data (ENCR, IARC, NORDCAN and several individual cancer registries). In this context, a useful outcome of the survival WG subgroup 4 consists of drafting specifications for the desired functionalities of a CRATE 2.0 software for federated calculation of cancer survival estimates. The current document aims to outline such a specification while only discussing quality checks specifically related to survival calculations.

Furthermore, it is advisable and desirable that software tools developed for the collection of data for ECIS is flexible enough to allow use for broader purposes by the cancer registry community. For this reason, there is some discussion about software flexibility in the specification.

5.1. The NORDCAN cancer statistics database and the transition to distributed data processing

NORDCAN is a database of cancer statistics for the Nordic countries and includes a web tool for presenting cancer incidence, mortality, prevalence and survival in the Nordic countries. It was created in the mid 1990's and has been published on the web since early 2000's. NORDCAN is partly funded by the association of the Nordic Cancer societies (NCU) and partly through in-kind contributions from the cancer registries within the association of the Nordic cancer registries (ANCR). The ANCR board of directors also functions as the steering committee for NORDCAN. In addition, a working group consisting of representatives from the Nordic registries meets two times per year with focus on operations and development of the NORDCAN project. The NORDCAN Secretariat has been

hosted by the Norwegian cancer registry since 2019 when it was moved there from the Danish Cancer Society.

Around 2016, a discussion about moving the Secretariat was started, and, in relation to this and to the emerging GDPR regulations, to change the way data were collected to a distributed/federated model where the statistics were to be calculated at the individual cancer registries using a common software that was to be jointly developed by the Nordic countries. Until then, record-level data had been sent to the NORDCAN Secretariat in Copenhagen where all quality checks and calculation of indicators were done. In connection with this transition, an agreement was also made with IARC that they would develop a new web application for presenting the statistics.

The reasoning behind the decision to change to the distributed model of data processing was that it had become more and more problematic to send record-level data between countries due to stringent data protection regulations and the sensitivity of personal health data, and that this was a possibility to secure a sustainable model for future collection of indicators for NORDCAN. A software (nordcan.R) for the distributed data processing was then jointly developed by several R-programmers from the Nordic cancer registries.

The transition to a distributed processing of the cancer indicators was successful. The last record-level data sent to the Secretariat in Copenhagen was for the incidence year 2016 and cancer indicators for 2017 and onwards have been calculated at the individual cancer registries using the nordcan.R software.

5.2. Description of the NORDCAN technical solution

A summary of the NORDCAN technical solution can be found in Larønningen et al (2023)[1] with details on the Cancer Registry of Norway's a Git Repository. Note that there are details about the data call and some of the programs and methods used by nordcan.R on the Wiki page. Below is a summary of key details.

After the data have been prepared according to the instructions in the data call (which was specified to be quite closely related to the ECIS call for data existing in 2020), nordcan.R can be run. This will run other R scripts and call other software (IARCcrgTools and Stata).

The nordcan.R scripts will do the following,

- Run the preprocessing package to check if the input files adhere to data call. For example, check if all mandatory variables are present, variable names are correct, variable types are correct, check no missing data for certain variables, check variable ranges, check date formats etc.
    - Some additional checks using the IARCcrgTools program.

[1] Larønningen et al (2023) Nordcan.R: a new tool for federated analysis and quality assurance of cancer registry data. Vol 13 | https://doi.org/10.3389/fonc.2023.1098342

    - The input files for IARCcrgTools are created by nordcan.R and the output files are read back into R.
- Enrich the original data. For example, calculate month/year of birth/incidence/death/follow-up, survival time in days, ICD10 code, NORDCAN cancer entity groupings and exclusion details (missing information, benign, DCO, autopsy etc.)
- Incidence / Mortality calculations are performed in R
- Survival estimates are calculated by calling Stata from nordcan.R
    - Uses the stnet program
    - Age standardization (ICSS) uses the individual weighting approach.
    - For temporal trends, period analysis is used for last period(s).
    - For overall estimates there must be a minimum of 30 patients alive at start of follow-up with a minimum 3 in any one age group.
    - For age group specific estimates there must be a minimum of 30 patients alive at start of follow-up.
    - There is a threshold on the number of remaining patients at risk at any time (to be agreed)
- Create aggregated output files to be sent to IARC.
- Create summary files/graphs (technical information, comparison summary, comparisons between two runs).
    - These are for each registry to check consistency of data with previous years- For example, comparing the newly created counts to those created for the existing data in NORDCAN.

5.3. Critical assessment of the NORDCAN solution in the ECIS/ENCR environment

There are some important conditions present in the NORDCAN example. The Nordic cancer registries have a long history of collaboration, the ANCR being established already in the 1960's, and the NORDCAN project already running for many years before the transition to a distributed processing of data. The Nordic cancer registries are running from the 1940's and 50's, and have acquired high data-holder maturity. The registries are similar and the differences that do exist are well known and, importantly, the cancer indicators had been calculated centrally on record-level data for many years which offered the possibility to compare with historical time series. Furthermore, the individual registries were, since many years, already familiar with the initial preparation of data before feeding it into nordcan.R as these were essentially the same data preparations that were earlier required when sending record-level data to Denmark.

These preconditions are largely absent for data collection in the broader ECIS/ENCR context where the data holder maturity and resources such as money, staffing, experience, computer hardware and software may look very different between European registries.

The steps that must be performed by the registries using a software for distributed processing of data are 1) preparation of data to be fed into the software, 2) installation and running of the software 3) handling of possible errors that occur when running the software, and 4) assessing the correctness of the calculated indicators before publication.

Given the heterogeneity of ENCR countries/registries described above, it is therefore predictable that several registries might need support concerning one or more of these steps if the solution would be extended to the entire ENCR community. For the same reasons, transition to a distributed or federated collection of data for all registries might not be feasible for all registries. Nevertheless, Nordic registries and those having already participated in the ECIS aggregated-data submission may have the adequate expertise and resources to contribute to ECIS survival indicators using a nordcan.R inspired tool.

Besides the need for good written instructions, support in the form of online meetings and seminars may be needed as already done for the ECIS aggregated-data submission. Technical support for installing and running the software will also be important. The need for support will be greatest the first year the registry uses the federated method for submitting survival cancer indicators but will gradually decrease as the registry acquires experience with the process.

5.4. Requirements for a software for distributed/federated survival contribution to ECIS (CRATE 2.0)

The nordcan.R software is written in R but also calls routines in Stata, and the IARC/IACR quality check tool is used in the preparation of the data. While this has worked well in the Nordic setting it should be streamlined for the use in a broader European setting. Furthermore, cancer registries should not have to buy new statistical software to be able to run the CRATE 2.0 and the software used to develop CRATE 2.0 should ideally be generally familiar among cancer registry staff, such as statisticians.

Although there are several alternatives, a viable solution is to develop the software in R as it is free and that there is a high probability to find statisticians with sufficient experience in R to adequately run the software in most cancer registries. This was also the solution already adopted to develop CRATE software used for the ECIS data call. Alternatively, the tool can be developed in a stand-alone software solution in Java. Nevertheless, even if Java is widely used and a free software, the experience with the JRC-ENCR quality check tool, written in Java, is that registries sometimes experience problems installing and running the software. An outstanding problem that has to be resolved in collaboration with the Joint Action CancerWatch is that thePohar-Perme estimator as period analysis as well as individual weighting are not yet implemented in R. Period analysis is suggested by sub-working group 1 for incomplete 5-year periods and Individual weighting could be useful as it makes the survival analysis less sensitive to small numbers in individual age strata for the age adjusted measures. Regardless of the chosen technology, it is important to safeguard future sustainability of the tool in terms of continuous maintenance and development, particularly after the Joint Action (JA) CancerWatch has finished. The choice of technology must be compatible with the IT-environment where it is going to be hosted later.

The CRATE 2.0 software should, at a minimum, be able to calculate net survival indicators according to the outcome of the work of the survival in ECIS sub-working group 1. It is, however, reasonable to construct from the beginning a flexible software, allowing for a set of survival indicators that could be wider than the subset strictly needed for the publication in the ECIS web application.

In order for the CRATE 2.0 software to also be a go-to tool for the cancer registries in their own projects, the software should be flexible regarding the possible setting of, for example, age inclusion, minimum number of persons alive at the start of the follow up and minimum number of cases in each age group required to produce the survival estimates. Furthermore, the software should allow for inclusion of additional stratification variables besides the stratification used in ECIS to make it possible for cancer registries to stratify on, for example, stage, level of education, deprivation indexes or country of birth.

The default setting should be the ones used in the calculation of survival estimates for ECIS web application. The session log should indicate the settings used, and the session log number should be included in the output to enable unambiguous identification of the correct session log.

Data preparation, correction and cleaning of inconsistencies should be done before running CRATE 2.0, using the JRC-ENCR Quality Checks Software. CRATE 2.0 should only perform final checks on input data such as checks on variable names, variable type, missing data, variable range, date format, and date order. The CRATE 2.0 should also calculate quality indicators in accordance with the outcome in JRC-ENCR survival in ECIS sub-WG 2.

6. Release of the ECIS data to third parties for research

A dedicated sub-working group discussed the formats of survival data that ECIS could share with Third Parties for research purposes, and the mechanisms by which such sharing could be arranged. This document reflects the perspectives of the WG and has been developed to inform decisions on data access procedures. The chapter also reports some elements to frame the context in which data sharing by ECIS is allowed.

6.1. Rationale for ECIS data sharing with Third Parties for research

- To enhance the value of the work done by European cancer registries in the collection and curation of primary data by allowing the secondary use of their data in cancer research and surveillance by other entities (Third Parties), such as research groups.
- To improve the quality and comparability of epidemiological estimates of the cancer burden in Europe by using a central dataset that is validated, harmonised and maintained by the Joint Research Center (JRC) according to common rules agreed with the European Network of Cancer Registries (ENCR) and together with the cancer registries that provide data.
- To be coherent with the Europe's Beating Cancer Plan strategy, pointing to the need to more effectively integrate data provided by population-based cancer registries in Europe into the cancer research ecosystem, and in line with the spirit of the European Health Data Space (EHDS).

6.2. Legal basis for the record-level data transfer

The ECIS Data Protection Record (DPR-EC-00417) envisages the possibility to disclose pseudonymised record-level data to Third Parties for research purposes. In Art. 5, the ECIS DPR identifies eligible Third Parties and the legal and administrative conditions for data sharing. In particular, the JRC may disclose cancer data to researchers for the purpose of scientific studies, and to national and international bodies in the field of cancer epidemiology, prevention and treatment, subject to the provisions of Art. 9 of Regulation (EU) 2018/1725.

Conditions

- Approval of data owners (the registries)
- JRC concludes a separate agreement with the recipient of the pseudonymised data.

6.3. Data sharing formats

Aggregated record-level data, e.g. mini-sets of a small number of records aggregated according to certain criteria (age, cancer entity, etc.), so-called Summary Anonymised Records (SAR), allows computation of incidence and mortality rates because every individual fits the same criteria in the aggregated set. However, it is not possible to determine the duration of survival of members in a given SAR, therefore such data format is unsuitable for estimating survival.

Data can be shared in two different formats, depending on received request:

6.3.1. Pseud*onymised record-level data*

These data are needed for analytical and descriptive studies on cancer survival in Europe. Typical examples are:

- Estimation of trends and inequalities in survival within and between countries (in addition to what will be already published in ECIS)
- Analytical studies that use regression models to assess the role of the available prognostic factors (sex, age, tumour biology, geographical area, stage at diagnosis, treatment) on trends and inequalities in cancer survival
- Studies aimed at estimating additional indicators related to cancer survival, such as cure indicators

In addition, we need to consider that in a research activity the study design details (e.g. cohort to be analysed, ICD-O-3 morphology grouping, data stratification choices granting adequate numbers for survival comparisons, …) cannot always be fully defined 'a priori' (except for the basic elements) but often need to be defined through a process, requiring the possibility to compare and validate different alternatives. The final selections, stratifications and estimators depend on the results of preliminary analyses and sensitivity analyses made on record-level data; thus, researchers may need a broader selection of data than what will be included in the final models. This request is to be specified in the analysis protocol, will be evaluated and possibly processed according to the data minimisation principle.

6.3.2. *Stratified survival estimates other than those published in the ECIS web application*

Stratified survival estimates on strata other than those in the ECIS web application (for instance, for different cancer entities, calculated over a specific year interval, on different age groups, etc.) can be the subject of specific requests.

These requests would be subject to the submission of a protocol for research, to be assessed and approved in first instance by the JRC-ENCR Management Committee. The approved protocol would then be shared with the contributing CRs, informing them about the proposed research and asking agreement for inclusion of their own data in the add-on analysis performed by JRC on request of the third party (see Paragraphs 5,6,7 for details). Such type of requests are also subject to resource availability from the JRC for the analysis.

This document describes the sharing of individual data from the ECIS database and does not consider federated analysis of cancer registry data.

6.4. Development of the record-level dataset to be shared

The quality of the ECIS data is checked using the JRC-ENCR Quality Check (QCS) Software and additional procedures implementing ECIS project decisions on core variables included in the statistical analysis, to be further confirmed by each registry. Inconsistencies detected by the procedures are iteratively discussed by the JRC team with the ENCR Steering Committee first and then with the registries, who are ultimately responsible for possibly modifying the dataset and correcting errors or inconsistencies.

At this stage, records that have been mis-recorded (in whole or in part) in the registry dataset (e.g. records including metastases, multiple primaries, duplicate cases) are identified and flagged to the registry for correction, to be implemented both by the JRC in the central database as well as locally by the registry in its database. Registries send corrected records to the JRC which are further verified with the QCS.

Once this process is complete and data are not to be corrected further, the dataset should be considered final and ready for analysis (validated registry dataset). An extract from this dataset, based on the specific study protocol and according to the data minimization principle (study dataset), may be shared with Third Parties.

The study dataset, as an extraction from the validated registry data, is free from mis-recordings and systematic errors, but it is not fully free of errors and inconsistencies that need to be adequately reported in a research study. Therefore, additional checks might be applied by Third Parties to the study dataset to derive study-specific quality indicators to report the data quality of the study-specific cohort and to support the interpretation of survival comparison (e.g. Survival across countries or across age groups) (**Figure 2**). When assessing researchers' need for data, their request to calculate study-specific quality indicators will be taken into consideration.

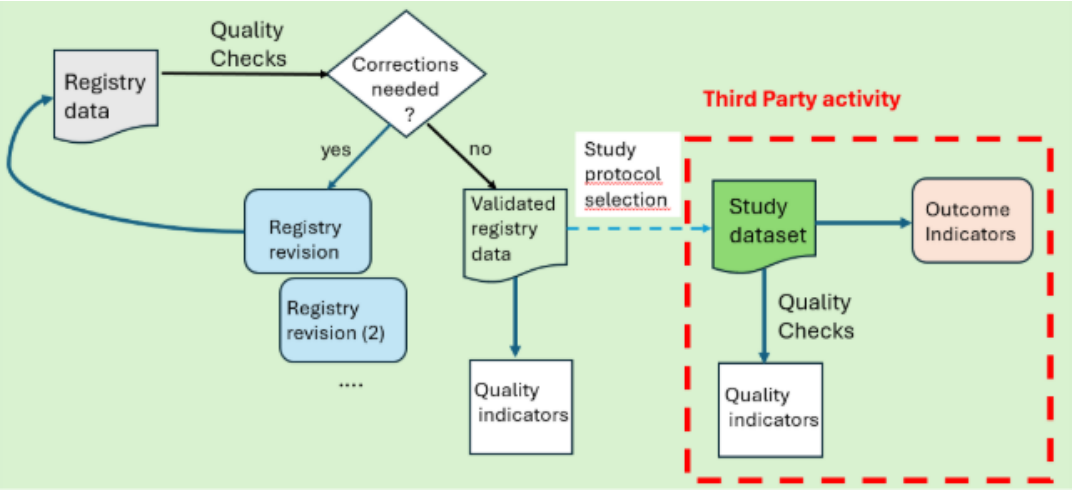

Figure 3. *Overall data processing: from first registry data submission to outcome indicators*

In brief, the study dataset data should include:

- both correct records and records with errors/inconsistencies (validated registry data, namely the dataset approved by the registry as its final data submission) albeit subject to data minimization principles.
- a summary of the revisions/corrections made on the first original submitted dataset.

6.5. Data request procedure

Data requests to ECIS should be submitted with a study protocol that includes the rationale, aims, methods, specifications of the needed data and expected results of the proposed study. The

conditions for authorship should also be specified, i.e. whether contributing registries and JRC are invited for co-authorship or be listed in an acknowledgement quoting the source of the data. Furthermore, if ethical approval is required for legal processing of the personal data in research in the jurisdiction where the research will be conducted, all such relevant documentation shall be submitted together with the request. The protocol should clarify how the analyses will differ from what is already available in ECIS, and if record-level data is requested, the documentation must specify and justify the need for pseudonymised record-level data.

The JRC-ENCR management committee should define the elements that are required for the assessment of the study protocol. A data request template should be made available to guide potential applicants. Metadata should also be made available to understand the registry data available and the ECIS variables in order to allow the definition of the study protocols.

In case of stratified survival estimates, further to approval the request is subject to resources available for the analyses at the JRC.

In case of record-level data, further to approval the request is subject to the signature of the Agreement between JRC and data recipients.

6.5.1. Evaluation of the request for data

The scientific appropriateness of the request should be evaluated by *a board of experts specifically appointed* by the JRC-ENCR management committee. It is advised that the board encompasses a range of expertise (biostatistics, epidemiology, cancer registration and other domain experts). The board evaluates the technical feasibility, scientific relevance and soundness of the proposal, based on previously agreed criteria.

6.5.2. Approval mechanism for cancer registries

A mechanism should be envisaged to inform the registries about the submission of requests and the outcome of the evaluation process (positive or negative). Participating registries would only be required to consider requests for data that have been positively evaluated by the board.

A procedure should be designed and implemented to allow the registries to express their decision about the use of their data in the proposed study, allowing the possibility for the registry to specify reasons in case of refusal. No reply from the registry side shall be considered as refusal. The same procedure should allow the registries to declare their interest in co-authoring the scientific articles produced in the study, if this is foreseen in the protocol.

6.5.3. Transmission of the record-level dataset

The dataset to be shared should be prepared based on the criteria defined in the study protocol and the number of registries that have adhered to it. The dataset should be prepared starting from the final validated registry dataset described above. The data should be prepared and released as soon as the registries have expressed their adherence or not to the protocol.

Delays in the data collection and revision tend to have a high impact on population-based registry data, which are often considered to be 'out of date'. These delays, when added to the time required for centralised validation and analysis, seriously reduce the value of registry data for cancer control. Therefore, the release of data to Third Parties for research purposes (other than those already foreseen in ECIS) shall proceed without undue delay as soon the data request has been approved, provided the availability of the validated dataset. All relevant conditions to the release shall be specified in the agreement between the JRC and the Third Party, which will precede the data transfer.

7. Conclusions

The WG on Survival in ECIS has developed the foundation for the inclusion of survival estimates in ECIS. The working group reached consensus regarding proposed survival indicators to be presented in the ECIS web application, methods for calculations, and related quality indicators.

The WG also developed a basis for understanding the requirements for sharing data to research consortia, as well as a foundation for the specifications of the CRATE 2.0 software to compute survival indicators.

A few questions were not completely resolved, as either the WG did not reach consensus or there was still not enough information to make final decisions:

- the minimum number of persons living at start of follow up required to calculate and present survival estimates in ECIS.
- on how to use the thresholds of the quality indicators considered by sub-WG 2.
- the final choice of technology to be used in CRATE 2.0, due to the outstanding issues regarding the implementation of the Pohar-Perme net survival estimator in R, maybe requiring a solution like that in nordcan.R where the R-program calls a Stata function for the survival calculation.
- inclusion/exclusion from the analysis records with survival time 0, which are not DCO and autopsy cases

These issues are left for further discussion and decision.

Future development and management of survival indicators in ECIS

The WG agreed that addition of other outcome measures such as crude survival, relative survival, 10-year survival and conditional survival, or additional parameters, for example morphological subgroups, should be possible in the future. It was also noted that CRATE 2.0 should ideally be designed with sufficient flexibility to allow for inclusion of new parameters (e.g. stage, morphology etc.) and measures.

The CRATE 2.0 software will be developed in the context of the JA CancerWatch in alignment with the ECIS requirements. An important question concerns future development and sustainability of the software, which must be secured also after the JA CancerWatch has come to conclusion. Sustainability must be a key consideration in all design choices for CRATE 2.0.

Annex 1. JRC-ENCR Strategy document on Survival in ECIS

JRC-ENCR Strategy document for provision of reliable and up-to-date survival in the European Cancer Information System (ECIS)

Motivated by the recent initiatives by the European Commission (EC) under the EU4Health programme[1], there is an increasing need to strengthen the role of the European Cancer Information System (ECIS) in supporting the information requests coming from the policy side. This strategy paper focuses on cancer survival statistics, identified as the most critical area in ECIS where further work is needed, and reports on the strategy to support and address information needs.

1 Information needs on survival

Identified needs for reliable and up-to-date survival in ECIS are the following:

- to address requests from the policy side;
- to ensure regular (preferably annual) updates based on most recent data;

[1] https://health.ec.europa.eu/non-communicable-diseases/cancer_en

ENCR is hosted by the European Commission's Joint Research Centre
Via E. Fermi 2749 | 21027 Ispra | Italy | jrc-encr@ec.europa.eu | www.encr.eu

- to align with ECIS data on cancer incidence and mortality, in terms of timeliness, cancer entities' definitions, age groups, geographical units;
- to feed the European Cancer Inequalities Registry.

The survival estimates currently published in ECIS are outdated and a procedure for regular updates, not yet in place, is specified in this document.

2 The JRC-ENCR strategy for up-to-date survival in ECIS

As per discussion and agreement at the 86th Joint Research Centre-European Network of Cancer Registries (JRC-ENCR) Management meeting[2], survival will be computed by the JRC based on the data routinely submitted by the cancer registries for ECIS. This will allow the regular provision of timely indicators in ECIS, once up-to-speed validation processes are in place.

3 JRC-ENCR initiative to support survival in ECIS

An ENCR working group (WG) will start its activity as of June 2024 to agree on the data standards, quality checks and methods, for computation and publication of survival in ECIS. This WG represents a first concrete JRC-ENCR initiative towards combining expertise and reaching agreement between survival experts and stakeholders for the routinely provision of survival in ECIS. The WG will address data validation with a focus on survival variables, define the methodology for estimates computation and address survival data dissemination for ECIS.

4 Survival in ECIS supported by the DG SANTE's direct grant to Member States

A Direct Grant to Member States' authorities (Joint Action CR-g-24-40) was launched in 2024, aimed at supporting quality improvement of cancer registry data feeding ECIS. Its second main objective focuses on the improvement of survival and prevalence indicators in ECIS, by:

"*...building on the already available mechanisms in Member States to feed ECIS with record-based data held by cancer registries.*

This part of the Joint Action should consist in validating and processing data collected via data calls to European registries and deriving up-to-date survival and prevalence indicators in alignment with the current ECIS settings for incidence and mortality figures, in terms of geographical detail, cancer sites definition, age range availability and timeliness".

The WG on survival should closely interlink with the Joint Action activities, and can provide a strong basis for their work.

[2] https://www.encr.eu/sites/default/files/meetings/ENCR-SC-meetings/Summary_Minutes%2086th%20SC_final.pdf

5 Sharing data with research consortia

Collaborative research projects on population-based cancer survival in Europe and worldwide have a long tradition and consolidated expertise in the field. These projects, as well as others, are invited to share their knowledge by taking part in the ENCR WG on survival.

Once data validation and methodological issues are specified by the survival WG, the JRC will implement and incorporate them in its ECIS data management tools, so that ECIS data validation and workflow will also include specific survival checks. After publication of the survival in ECIS, related data in an appropriate format may be made available by the JRC-ENCR for research by third parties. This further data sharing from the JRC-ENCR for research purposes is subject to approval of the cancer registries and may involve a collaboration agreement with the JRC.

6 Roadmap

The ENCR WG on survival will start its activity as of June 2024. Updated survival (and prevalence) indicators should be available in ECIS according to the ECIS timetable and at the latest by the end of 2026. Methodology established by the WG will allow the exercise to be repeated on a rolling basis.

Annex 2. Quality indicators for completeness

Table A2.1. Quality indicators for completeness of incidence

Quality indicator	Tentative threshold	Netherlands		Remarks
% DCO	<5%		-	Not available in the Netherlands
% MV for solid cancers in children (<15 years)	<97%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	94% 95% 95% 93% 91% **93%**	Haematological malignancies are excluded (100% MV)
% MV for cancer of the liver, gallbladder, bile ducts and pancreas (C22-C25) in adults (15 years or older)	<80%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	67% 66% 65% 68% 70% **68%**	
% MV for lung cancer (C34) in adults (15 years or older)	<95%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	93% 93% 92% 91% 88% **91%**	
% MV for kidney cancer (C64) in adults (15 years or older)	<90%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	87% 84% 84% 87% 86% **86%**	
% MV for malignant brain tumours (C71) in adults (15 years or older)	<90%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	82% 83% 79% 82% 82% **81%**	

There was no consensus on the application of specific thresholds nor how the information would be displayed in case the data does not meet these quality indicators thresholds.

Table A2.2. Quality indicators for completeness of follow-up

Quality indicator	Tentative threshold	Netherlands		Remarks
Immortals: % surviving males born >100 year before last date of follow-up	<0.5%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	0.04% 0.03% 0.04% 0.08% 0.00% 0.04%	• Includes all males born before 1923 with follow-up until 1-1-2023 • Exclude cases censored because of emigration/lost to follow-up • Threshold may be higher in countries with a very high life expectancy (e.g. 1%) • See appendix 3 for a sample table
Immortals: % surviving females born >100 year before last date of follow-up	<0.5%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	0.06% 0.07% 0.13% 0.19% 0.30% 0.10%	• Includes all females born before 1923 with follow-up until 1-1-2023 • Exclude cases censored because of emigration/lost to follow-up • Threshold may be higher in countries with a very high life expectancy, (e.g. 2%)
% of clinically diagnosed gastrointestinal cancers surviving > 5 years	<5%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	2% 1% 2% 1% 3% **3%**	• Exclude cases censored within 5 years after diagnosis because of emigration • Basis of diagnosis 1-4 • Topography C15-C20
% of pancreatic cancers (C25) surviving > 5 years	<5%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	2% 1% 1% 2% 3% **2%**	• Exclude cases censored within 5 years after diagnosis because of emigration • Exclude island cell tumours (topography C25.4 and/or morphology 8240-8249, 8150-8159)
% of clinically diagnosed pancreatic cancers (C25) surviving > 5 years	<5%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	1% 1% 0% 0% 1% **1%**	• Exclude cases censored within 5 years after diagnosis because of emigration • Exclude islet cell tumours (topography C25.4) • Basis of diagnosis 1-4
% of small cell lung cancer in patients aged 75 or older surviving > 5 years	<5%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	1% 1% 1% 2% 2% **2%**	• Exclude cases censored within 5 years after diagnosis because of emigration • Topography C34 • Morphology 8041-8045
% of mesothelioma surviving > 5 years	<5%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	3% 2% 3% 3% 4% **3%**	• Exclude cases censored within 5 years after diagnosis because of emigration • Morphology 9050-9053 • All sites •
% of malignant brain tumours in patients aged 75 or older surviving > 5 years	<5%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	1% 1% 2% 1% 2% **2%**	• Exclude cases censored within 5 years after diagnosis because of emigration • Exclude cases without microscopic confirmation • Behaviour code 3

Quality indicator	Tentative threshold	Netherlands		Remarks
% of acute leukaemia in patients aged 75 or older surviving > 5 years	<5%	1991-1995 1996-2000 2001-2005 2006-2010 2011-2017 **1991-2017**	0% 1% 1% 1% 1% **1%**	• Exclude cases censored within 5 years after diagnosis because of emigration • Include all acute leukaemia (ALL, AML, as well as biphenotypic, undifferentiated & unspecified acute leukaemia) • Exclude APL (M9866)
% emigration	>0%			• For cancer registries with passive follow-up
% lost-to-follow-up	<5%			• For cancer registries with active follow-up

Table A2.3. Sample table for calculating 'immortals'

Sample table of males (diagnosed with cancer in 1991-2017) born >100 year before the last date of follow-up (=born before 1923) according to vital status at date of last follow-up (=1-1-2023)

Year of birth	Year of death													Alive*	Total
	<=2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022		
1888	2														2
1890	1														1
1891	5														5
1892	5														5
1893	12														12
1894	15														15
1895	19														19
1896	44														44
1897	68														68
1898	113														113
1899	184														184
1900	220														220
1901	337														337
1902	508														508
1903	674														674
1904	922														922
1905	1.155	1	1												1.157
1906	1.543				1										1.544
1907	1.885	2	2			1									1.890
1908	2.381	3		3											2.387
1909	2.926	5	2	1											2.934

Year of birth	Year of death													Alive*	Total
	<=2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022		
1910	3.502	7	3	1	2		1								3.516
1911	4.110	9	8	2	3	2	1								4.135
1912	5.134	14	14	3	3	5	1	1						1	5.176
1913	5.860	20	19	9	12	5	3	3		1					5.932
1914	6.819	37	33	21	16	12	2	3	4	1	1				6.949
1915	7.130	63	39	28	23	14	7	7		2	1			2	7.316
1916	7.914	67	68	49	34	22	12	8	9	2		1		3	8.189
1917	8.803	128	112	56	43	38	35	12	10	6	6	2		1	9.252
1918	9.241	154	128	101	57	43	41	28	10	14	10	3	1	2	9.833
1919	10.386	231	194	143	106	81	76	46	29	22	10	7	4	4	11.339
1920	12.828	320	302	231	180	151	117	82	51	42	25	20	9	8	14.366
1921	13.153	411	384	302	250	212	157	118	73	56	36	20	21	14	15.207
1922	13.595	478	441	362	296	273	198	147	106	68	56	38	28	18	16.104
Total	121.494	1.950	1.750	1.312	1.026	859	651	455	292	214	145	91	63	53	130.355
														0.04%**	

*alive at 1-1-2023

**53 males out of a total of 130.355 were still alive at 1-1-2023, which equals 0.04%

Annex 3. JRC-ENCR Quality Checks Software for incidence variables

Univariate checks

Table A3.1 Univariate checks for Incidence variables

Variable name	Variable description	Coding	Missing/ unknown	QCS check	QCS flag	JRC action for incidence analysis	Comment
Age	Age at diagnosis in years	0-120	blank/999	unknown, missing, wrong format, out of range	W-UNKN, W-MISS, E-FORM, E-OUTR	Exclude if not corrected by CR (compulsory revision)	Age 999-> Unknown values are allowed in ECIS Data Call Protocol but not included in the incidence reporting
Sex	Sex at birth	1 →Male; 2 →Female	9	unknown, missing, wrong format, out of range	W-UNKN, E-MISS, E-FORM, E-OUTR	Exclude if not corrected by CR (compulsory revision)	Sex 3 →Other and Sex 9 -> Unknown are allowed in ECIS Data Call Protocol but not included in the incidence reporting
MoB	Month of birth	Range of allowed values: From 1 to 12	99	unknown, missing, wrong format, out of range	W-UNKN, W-MISS, E-FORM, E-OUTR	Include	
YoB	Year of birth	Range of allowed values: >1842 and ≤ the current year	9999	unknown, missing, wrong format, out of range	W-UNKN, W-MISS, E-FORM, E-OUTR	Include (if age is available)	Missing YoB is an issue if age cannot be calculated, see E-AGEC horizontal check)
MoI	Month of incidence	Range of allowed values: 1 - 12	99	unknown, missing, wrong format, out of range	W-UNKN, W-MISS, E-FORM, E-OUTR	Include	
YoI	Year of incidence	Range of allowed values: From 1941 to present	Not allowed	missing, wrong format, out of range	E-MISS, E-FORM, E-OUTR	Exclude if not corrected by CR (compulsory revision)	
Topography	ICD-O-3 topography code	Valid code in ICD-O-3	Not allowed	missing, wrong format, out of range	E-MISS, E-FORM, E-OUTR	Exclude if not corrected by CR or suggest corrections (compulsory revision)	Example of correction: imputation of "9" as the 4th digit of topography if

Variable name	Variable description	Coding	Missing/ unknown	QCS check	QCS flag	JRC action for incidence analysis	Comment
							only 3 digits code is provided (if the code exists).
Morphology	ICD-O-3 morphology code	Valid code in any ICD-O-3 version	Not allowed	missing, wrong format, out of range	E-MISS, E-FORM, E-OUTR	Exclude if not corrected by CR or suggest corrections (compulsory revision)	
Behaviour	ICD-O-3 behaviour	0→ Benign neoplasm; 1→ Neoplasm of uncertain and unknown behaviour; 2→ In situ neoplasm; 3→ Malignant neoplasm;	Not allowed	missing, wrong format, out of range	E-MISS, E-FORM, E-OUTR	Exclude if not corrected by CR or suggest corrections (compulsory revision)	Behaviour 6 and 9 can be recoded to behaviour 3 upon CR's compulsory revision and approval. Records with behaviour 5 excluded if not corrected.
BoD	Basis of diagnosis	0→Death certificate only (DCO); 1→Clinical; 2→Clinical investigation; 4→Specific tumour markers; 5→Cytology; 7→Histology; 8→Cytogenetic and/or molecular testing	9	unknown, missing, wrong format, out of range	W-UNKN, W-MISS, E-FORM, E-OUTR	Include	

Multivariate checks

Table A3.2 JRC-ENCR QCS multivariate checks for Incidence variables

Variables involved	Description	QCS flag	JRC action for incidence analysis	Comment
Age, DoI, DoB	Age is invalid and it is impossible to calculate age from date of incidence - date of birth	E-AGEC	Exclude if not corrected by CR (compulsory revision)	
Age, DoI, DoB	Date of incidence - date of birth different from age	E-AGED	Exclude if not corrected by CR (compulsory revision)	Records excluded if YoI-YoB different than age +/- 1 year
DoI, DoB	Date of incidence and date of birth not coherent (see p. 16 JRC Technical report)	E-CoDA	Exclude if not corrected by CR (compulsory revision)	
Topo, Sex	Topography + Sex not valid (see Table 3 JRC Technical Report)	E-SETO	Exclude if not corrected by CR (compulsory revision)	
Topo, Morpho	morphology and topography combinations are unlikely (see Table 8 of the JRC Technical Report)[2]	W-MOTO	Include but ask CR revision (additional revision). In some cases, Exclude if not corrected by CR or suggest corrections (compulsory revision)	Compulsory revision requested for a subset of incorrect combinations or possible metastases.
Morpho, Beh	Morphology and behaviour combinations are not included in the ICD-O-3 (except for those in Table 9 of the JRC Technical Report which are allowed)	W-MOBE	Include but ask CR revision (additional revision). In some cases, Exclude if not corrected or suggest correction (compulsory revision)	Compulsory revision requested for a subset of incorrect combinations.
Age, Topo, Morpho	Unlikely Age + tumour type (see Table 2 of the JRC Technical Report)	W-AGMT	Include but ask CR revision (additional revision)	
Topo, Morpho, Beh, BoD	Morphology too specific according to the basis of diagnosis (see Table 4 of the JRC Technical Report)	W-BDMO	Include but ask CR revision (additional revision)	

Variables involved	Description	QCS flag	JRC action for incidence analysis	Comment
Topo, Morpho, Beh, BoD	Morphology not specific enough according to the basis of diagnosis	W-BDMS	Include but ask CR revision (additional revision)	
Beh, pT, cT, Stage	Behaviour and TNM combination not valid (pg. 33 of the JRC Technical Report)	W-BTNM	Include but ask CR revision (additional revision)	Only a subset of cases with behaviour =3 was reported to Cancer Registry. This was to ensure that potential incosistency between behaviour and stage was not due to incorrect behaviour, which would affect incidence reporting.
Morpho, Beh, Grade	Morphology, behaviour and grade combinations are unlikely (see tables 6 and 7 of the JRC Technical Report)	W-MOGR	Include but ask CR revision (additional revision)	Only a subset of cases reported to Cancer Registry. Grade is used in only some of the ECIS definitions of cancer entities. Revisions were requested for these specific combinations.

[1] JRC Technical Report (https://www.encr.eu/sites/default/files/Recommendations/Cancer%20Data%20Quality%20Checks%20Procedure%20Report%202.0_v20240823.pdf)

Table A3.3 JRC-ENCR QCS multivariate inter-records checks for Incidence variables

Variables involved	Description	QCS flag	JRC action for incidence analysis
PAT, Tum	Duplicate PatientID-TumourID	E-DUPL	Compulsory revision to be done prior to data validation. Propose recoding (if possible) or ask CR to recode and resubmit the dataset.
Pat, Tum, Topo, Morpho	Not all recorded diagnoses should be considered for incidence analysis according to 2004 International Rules for Multiple Primary cancer.	W-MPMT, W-MPCR	Selection of appropriate records for incidence reporting, according to International Rules. Done after the data cleaning (all corrections applied). Selection shared with CRs but record-by-record revision not requested.

Annex 4. JRC-ENCR Quality Checks Software for survival variables

Univariate compulsory checks

Table A4.1. JRC-ENCR QCS univariate checks specific for Survival variables: compulsory revisions

Variable name	Variable description	Coding	Missing/unknown	QCS check	QCS flag	JRC decision for Incidence analysis	Action if not corrected by CR
Vit_stat	The last known vital status	1→ Alive; 2→ Dead	9	unknown, missing, wrong format, out of range	W-UNKN, W-MISS, E-FORM, E-OUTR	Include (not validated for Incidence analysis)	Exclude

Table A4.2. JRC-ENCR QCS univariate checks specific for Survival variables: additional revisions.

Variable name	Variable description	Coding	Missing/ unknown	QCS check	QCS flag	JRC decision for Incidence analysis	Action if not corrected by CR
Surv_time*	Duration of survival in days	Range of allowed values: ≥ 0 and <= (current year -1942) in days	99999	unknown, missing, wrong format, out of range	W-UNKN, W-MISS, E-FORM, E-OUTR	Include (not validated for Incidence analysis)	Include (if possible to calculate it); otherwise exclude
YoF*	Year of last known vital status	Range of allowed values: > 1941 and ≤ the current year	9999	unknown, missing, wrong format, out of range	W-UNKN, W-MISS, E-FORM, E-OUTR	Include (not validated for Incidence analysis)	Include (if possible to calculate it); otherwise exclude
MoF*	Month of last known vital status	Range of allowed values: From 1 to 12	99	unknown, missing, wrong format, out of range	W-UNKN, W-MISS, E-FORM, E-OUTR	Include (not validated for Incidence analysis)	Include (if possible to calculate it); otherwise exclude
Autopsy	Incidental finding of cancer at autopsy	0→No; 1→Yes	9	unknown, missing, wrong format, out of range	W-UNKN, W-MISS, E-FORM, E-OUTR	Include (not validated for Incidence analysis)	If survival time > 0 : onclude (if other errors not present) If survival time = 0: exclude if not corrected by the CR

* If this variable cannot be imputed, records are already excluded due to the W-SUMU label

Multivariate checks

Multivariate checks (intra record) are also performed between crucial variables for Incidence analysis (PAT ID, TUM ID, Age, Sex, DoI (MoI,YoI), DoB (MoB,YoB), BoD, Topo, Morpho, Beh). Example: consistency check between morphology and topography; consistency check between morphology and behaviour, etc. All these checks are considered compulsory for revision.

Table A4.3. E-CoDV rule: Consistency check for dates of follow-up, incidence and birth

E-CoDV rule					
Consistency check for dates of follow-up, incidence and birth		Valid combination of variables (no errors)			QCS error code if invalid combination
All dates (DoF, DoI, DoB) are complete		**Date of the last known vital status (MoF, YoF) > date of incidence (MoI, YoI)**			**E-CoDV**
		Date of the last known vital status (MoF, YoF) > date of birth (MoB, YoB)			
QCS output examples: E-CoDV					
Error_Code	**Var1_Name**	**Var1_Value**	**Var2_Name**	**Var2_Value**	**Action if not corrected by CR**
E-CoDV	DoF	Sep-06	DoB	Oct-06	Exclude
E-CoDV	DoF	Sep-06	DoI	May-12	Exclude

Table A4.4. VSBD rule: From Table 10 of JRC technical report (rule applied if all variables are: not missing/unknown, in range, correct format)

E-VSBD rule		
Consistency check if **Vital Status (Vit_stat)=1**	Valid combination of variables (no errors)	QCS error code if invalid combination
Vital status = 1 (alive)	**Autopsy** (incidental finding of cancer at autopsy) **≠ 1** (incidentally diagnosed at autopsy) **Basis of diagnosis ≠ 0** (DCO-Death Certificate Only)	**E-VSBD**

QCS output examples: E-VSBD (these cases can potentially produce also E-AUVS)							
Error_Code	Var1_Name	Var1_Value	Var2_Name	Var2_Value	Var3_Name	Var3_Value	**Action if not corrected by CR**
E-VSBD	Vit_stat	1	BoD	0	Autopsy	1	Exclude
E-VSBD	Vit_stat	1	BoD	0	Autopsy	0	Exclude
E-VSBD	Vit_stat	1	BoD	7	Autopsy	1	Exclude

Table A4.5. E-AUVS rule: From Table 12 of JRC technical report (rule applied if all variables are: not missing/unknown, in range, correct format)

E-AUVS rule		
Consistency check if **Autopsy=1**	Valid combination of variables (no errors)	QCS error code if invalid combination
Autopsy = 1 (yes)	**Vital status** = **2** (dead)	**E-AUVS**
	Survival (in days) = **0**	
	Date of incidence (MoI, YoI) **= Date of the last known vital status** (MoF, YoF)	

QCS output examples: E-AUVS (these cases can potentially produce also E-VSBD)

Error_Code	Var1_Name	Var1_Value	Var2_Name	Var2_Value	Var3_Name	Var3_Value	Var4_Name	Var4_Value	Var6_Name	Var6_Value	**Action if not corrected by CR**
E-AUVS	Autopsy	1	Vit_stat	2	Surv_time	57	DoI	Sep-06	DoF	Nov-06	Exclude
E-AUVS	Autopsy	1	Vit_stat	2	Surv_time	2254	DoI	Jul-10			
E-AUVS	Autopsy	1	Vit_stat	1	Surv_time	3134	DoI	May-12			

Multivariate checks to be implemented

Table A4.6. Update of the current E-BDVS rule: From Table 11 of JRC technical report (rule applied if all variables are: not missing/unknown, in range, correct format)

E-BDVS rule (to be implemented)		
Consistency check if **Basis of Diagnosis** (**BoD**)=0	Valid combination of variables (no errors)	QCS error code if invalid combination
Basis of diagnosis = 0 (DCO)	**Vital status = 2** (dead)	**E-BDVS**
	Survival (in days) **= 0**	
	Date of incidence (MoI, YoI) **= Date of the last known vital status** (MoF, YoF)	

QCS output examples: E-BDVS (these cases can potentially produce also E-VSBD)

Er-ror_Code	Var1_Name	Var1_Va-lue	Var2_Name	Var2_Va-lue	Var3_Name	Var3_Va-lue	Var4_Name	Var4_Va-lue	Var6_Name	Var6_Va-lue	**Action if not corrected by CR**
E-BDVS	BoD	0	Vit_stat	2	Surv_time	1	DoI	Jun-91	DoF	Jun-91	Exclude
E-BDVS	BoD	0	Vit_stat	2	Surv_time	143	DoI	May-10	DoF	Oct-10	Exclude
E-BDVS	BoD	0	Vit_stat	2	Surv_time	550	DoI	Jul-03	DoF	Jan-05	Exclude
E-BDVS	BoD	0	Vit_stat	1	Surv_time	3134	DoI	May-12	DoF	Dec-20	Exclude

Table A4.7. New W-SUMU rule

NEW W-SUMU rule to be implemented		
New check between Surv_time and DoI, DoF if Surv_time is missing/unknown	Valid combination of variables (no warnings)	QCS error code if invalid combination
Survival time = 99999 OR missing	**Date of the last known vital status (MoF, YoF) ≠ missing/unknown**	**W-SUMU**

QCS output examples: W-SUMU

Error_Code	Var1_Name	Var1_Value	Var2_Name	Var2_Value	Action if not corrected by CR
W-SUMU	Surv_time	99999	DOF	**99-15**	Exclude
W-SUMU	Surv_time		DOF		Exclude

Table A4.8. New W-SUDA rule

NEW W-SUDA rule to be implemented							
New consistency check between Surv_time and DoF, DoI if Surv_time is provided in days	Valid combination of variables (no errors)				QCS error code if invalid combination		
Survival time in days	**DoF - DoI = Survival time in days / 30.43675 ± 1**				**E-SUDA**		
QCS output examples: W-SUDA							
Error_Code	Var1_Name	Var1_Value	Var2_Name	Var2_Value	Var3_Name	Var3_Value	**Action if not corrected by CR**
E-SUDA	Surv_time	45	DoI	Sep-06	DoF	Sep-06	Exclude
E-SUDA	Surv_time	45	DoI	Jun-06	DoF	Sep-06	Exclude