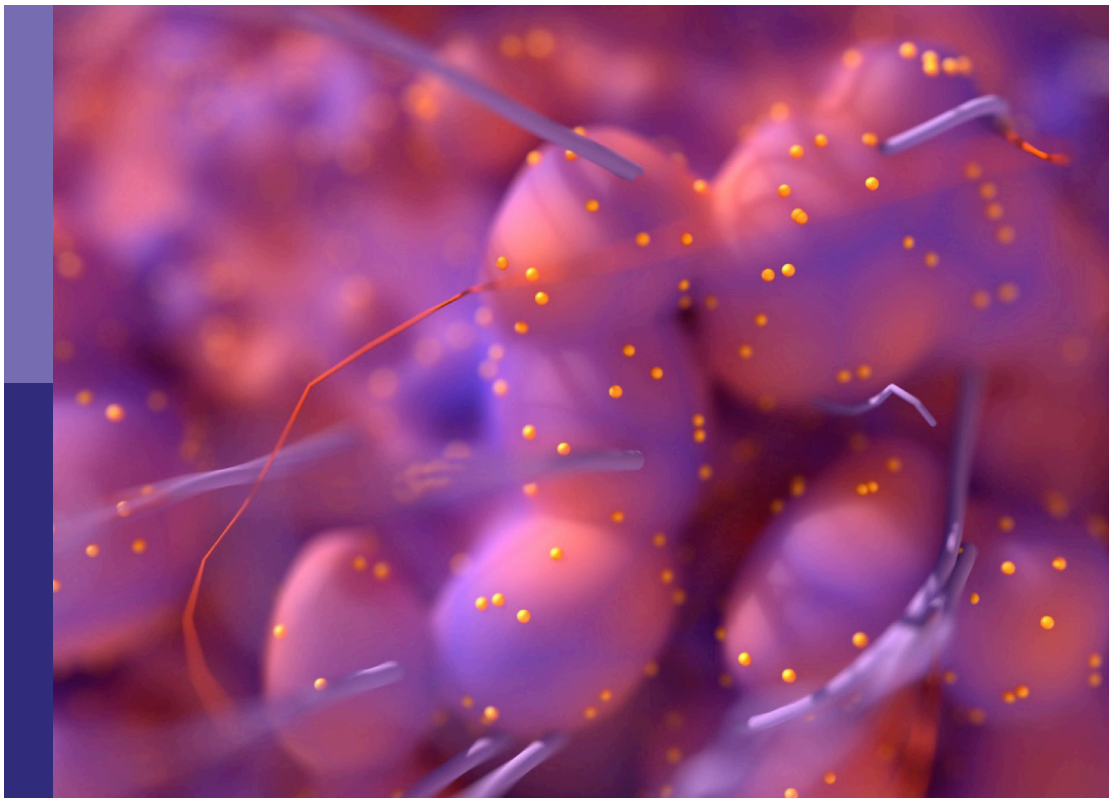
frontiers | Research Topics

Joining efforts to improve data quality and harmonization among European population-based cancer registries

Edited by
Francesco Giusti, Liesbet Van Eycken, Otto Visser and Carmen Martos

Published in
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-8325-5582-8
DOI 10.3389/978-2-8325-5582-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Joining efforts to improve data quality and harmonization among European population-based cancer registries

Topic editors

Francesco Giusti — European Commission, Joint Research Centre (JRC), Italy
Liesbet Van Eycken — Belgian Cancer Registry, Belgium
Otto Visser — Netherlands Comprehensive Cancer Organisation (IKNL), Netherlands
Carmen Martos — Public Health Research Center, Biomedical Research of Health and the Valencian Community (FISABIO), Spain

Citation

Giusti, F., Van Eycken, L., Visser, O., Martos, C., eds. (2024). *Joining efforts to improve data quality and harmonization among European population-based cancer registries*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5582-8

Table of contents

frontiers | Frontiers in Oncology

TYPE Editorial
PUBLISHED 09 October 2024
DOI 10.3389/fonc.2024.1496574

OPEN ACCESS

EDITED AND REVIEWED BY
Sharon R. Pine,
University of Colorado Anschutz Medical Campus, United States

*CORRESPONDENCE
Carmen Martos
JRC-ENCR@ec.europa.eu;
carmen.martos@fisabio.es

RECEIVED 14 September 2024
ACCEPTED 24 September 2024
PUBLISHED 09 October 2024

CITATION
Martos C, Giusti F, Van Eycken L and Visser O (2024) Editorial: Joining efforts to improve data quality and harmonization among European population-based cancer registries. *Front. Oncol.* 14:1496574. doi: 10.3389/fonc.2024.1496574

Editorial: Joining efforts to improve data quality and harmonization among European population-based cancer registries

Carmen Martos[1,2]*, Francesco Giusti[1], Liesbet Van Eycken[3] and Otto Visser[4]

[1]European Commission, Joint Research Centre (JRC), Ispra, Italy, [2]Rare Diseases Research Unit, Foundation for the Promotion of Health and Biomedical Research in the Valencian Region (FISABIO), Valencia, Spain, [3]Belgian Cancer Registry, Brussels, Belgium, [4]Department of Registration, Netherlands Comprehensive Cancer Organisation (IKNL), Utrecht, Netherlands

KEYWORDS

population-based cancer registries, data quality, harmonization, data comparability, Europe

Editorial on the Research Topic
Joining efforts to improve data quality and harmonization among European population-based cancer registries

The aim of population-based cancer registries (PBCRs) is to collect information from all new cases of cancer that occur in a defined population (1). They play an essential role in cancer surveillance, quantifying the burden of cancer in terms of incidence, prevalence and survival at population level, describing geographical variation and time trends. In addition, PBCRs are an important information source for planning and evaluating cancer control policies and healthcare systems (2, 3).

The reliability, use and comparability of the data provided by PBCRs depend on their quality as well as the harmonization of data collection and processing, coding and case definition.

The aim of this Research Topic was to share experiences on cancer data quality and harmonization in Europe, focusing on: 1) challenges in data comparability among PBCRs; 2) description of tools and activities for improving cancer data quality and harmonization; 3) Assessment of data quality in PBCRs; 4) challenges in data quality and harmonization related to national data protection regulations; 5) impact of data quality and harmonization on cancer indicators; and 6) epidemiological and statistical methods for improving data comparability.

Three of the fifteen articles included in the Research Topic focus on tools for checking internal consistency of cancer registry data. Giusti et al. give an overview of the Joint Research Centre-European Network of Cancer Registries Quality Check Software (JRC-ENCR QCS), describing its role in processing data files submitted by PBCRs contributing to the European Cancer Information System (ECIS) and its functionalities. The JRC-ENCR QCS is a Java standalone desktop software developed and updated by the JRC to support

the validation of cancer registry data. It can be freely downloaded from the ENCR website (4).

Tagliabue et al. compared the functional features and the output differences between the JRC-ENCR QCS and the IARC/IACR CHECK program developed by the International Agency for Research on Cancer (IARC).

Nicholson et al. presented the design of an ontology approach to model the ENCR rules (5) for validating childhood tumors, including some examples of how the ontology handles the ENCR data-validation requirements.

Indicators related to the four dimensions of data quality have been used to evaluate PBCR data: completeness, validity, comparability and timeliness (6, 7). The article "*Quality indicators: completeness, validity and timeliness of cancer registry data contributing to the European Cancer Information System*" Giusti et al. reported the quality indicators from 130 European PBCRs and their time trends using the data collected in the 2015 ENCR-JRC data call. The results provided by this paper could be used as the baseline for monitoring PBCRs data quality indicators in Europe over time.

Two articles by Galceran et al. and by Visser et al. included the current ENCR Recommendations for recording/reporting urothelial tumors and the ENCR Recommendations for coding the basis of diagnosis, respectively. The ENCR Recommendations (8) provide common definitions and rules to improve the data comparability among European PBCRs.

The role of the PBCR in cancer surveillance in term of incidence is shown in two papers by (Giusti et al.) and (Trallero et al.). The article by Giusti et al. highlights geographical and time trend differences in esophageal and gastric cancer in Europe by sub-sites and morphology subgroups. A wide variability in oesophago-gastric cancers was observed, with a corresponding improvement in accuracy of registration in the analyzed period. Trallero et al. described the incidence of hematological malignancies among children in Spain during the period 1983-2018 and compared their results with other Southern European countries. Main diagnostic sub-groups of the International Classification of Childhood Cancer (2017 update) were used for reporting their results.

Three papers focused on prevalence methodology. Demuru et al. explored the validity of alternative versus standard completeness indexes for estimating complete cancer prevalence in Europe. Toffolutti et al. described the procedures to derive complete prevalence and some indicators of cancer cure using data provided by Italian PBCRs. Francisci et al. proposed a new method for estimating short term projections on cancer prevalence by phase of care (initial, continuing and final) that applies to geographical areas covered by cancer registration.

Technological advances and record linkage have contributed to the improvement of the data provided by the PBCR (9, 10). Stage and treatment variables are recommended by the ENCR to be recorded in the European PBCRs (11).

The article by Giusti et al. gives an overview of reporting and using cancer treatment data provided by the European PBCRs. A literature review, conference proceedings and data from 125 European cancer registries contributing to the 2015 ENCR-JRC data call were used to explore the current situation of cancer treatment registration in Europe.

Lopez-Cortes et al. reported the experiences of the International Benchmarking of Childhood Cancer Survival by Stage (BENCHISTA) project to ensure data quality, harmonization and comparability among the CRs participating in the project.

The application of the European General Data Protection Regulation (GDPR) (12) since 2018 has complicated the sharing of health data among European countries, in particular in the Nordic countries due to a stricter interpretation of the GDPR. Larønningen et al. described a new GDPR-compliant federated analysis programme (nordcan.R) and how to use it for computing statistics for the Nordic cancer statistics web platform NORDCAN. The programming languages used for nordcan.R were R and Stata.

Finally, Giusti et al. highlight the recent and ongoing activities of the ENCR, the JRC and the European PBCRs in data quality and harmonization.

In summary, the fifteen articles (9 original research, 3 technology and code, 2 method and 1 perspective) published on this Research Topic provide an overview of the efforts and collaborations among European PBCRs, stakeholders, the ENCR and the JRC to improve data quality and harmonization of European cancer registries. This will contribute to the knowledge of cancer epidemiology in Europe and improve insights in cancer inequalities among European countries and regions. In addition, the Research Topic "Joining Efforts to Improve Data Quality and Harmonization Among European Population-Based Cancer Registries" could provide some important elements for the current Joint Action EU4H-2024-JA-IBA-03, direct grants to Member States' authorities: to support quality improvement of cancer registry data feeding the European Cancer Information System (13).

Author contributions

CM: Writing – original draft, Writing – review & editing. FG: Writing – original draft, Writing – review & editing. LE: Writing – original draft, Writing – review & editing. OV: Writing – original draft, Writing – review & editing.

Acknowledgments

The editors highly appreciated all submissions and would like to thank all contributing authors and reviewers for this Research Topic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

1. Jensen OM, Whelan S. Planning a cancer registry. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, editors. *Cancer registration: principles and methods*. International Agency for Research on Cancer (IARC Scientific Publications, No. 95, Lyon (1991). p. 22–8.

2. Parkin DM. The evolution of the population-based cancer registry. *Nat Rev Cancer*. (2006) 6:603–12. doi: 10.1038/nrc1948

3. Parkin DM. The role of cancer registries in cancer control. *Int J Clin Oncol*. (2008) 13:102–11. doi: 10.1007/s10147-008-0762-6

4. European Network of Cancer Registries. Tools for registries . Available online at: https://encr.eu/tools-for-registries (Accessed 18 July 2024).

5. Martos C, Giusti F, Van Eycken E, Visser O. A common data quality check procedure for European cancer registries (2023). Ispra, Italy: European Commission. Available online at: https://encr.eu/sites/default/files/Recommendations/Cancer%20Data%20Quality%20Checks%20Procedure%20Report%202.0_v20240506.pdf (Accessed 18 July 2024).

6. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *Eur J Cancer*. (2009) 45:747–55. doi: 10.1016/j.ejca.2008.11.032

7. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods. Part II: Completeness. *Eur J Cancer*. (2009) 45:756–64. doi: 10.1016/j.ejca.2008.11.033

8. ENCR recommendations. Available online at: https://encr.eu/ENCR-Recommendations (Accessed 18 July 2024).

9. Tucker TC, Durbin EB, McDowell JK, Huang B. Unlocking the potential of population-based cancer registries. *Cancer*. (2019) 125:3729–37. doi: 10.1002/cncr.32355

10. Arndt V, Holleczek B, Kajüter H, Luttmann S, Nennecke A, Zeissig SR, et al. Data from population-based cancer registration for secondary data analysis: methodological challenges and perspectives. *Gesundheitswesen*. (2020) 82:S62–71. doi: 10.1055/a-1009-6466

11. ENCR. Recommendations for a standard dataset for the european network of cancer registries. Available online at: https://encr.eu/sites/default/files/Recommendations/ENCR-Recommendation-standard-dataset_Mar2023.pdf (Accessed 18 July 2024).

12. European Union. General data protection regulation - GDPR (2018). Available online at: https://gdpr-info.eu/ (Accessed 12 March 2024).

13. EU4Health Programme (EU4H). EU4H-2024-JA-IBA-03 — Direct grants to Member States' authorities: to support quality improvement of cancer registry data feeding the European Cancer Information System (CR-g-24-40) . Available online at: https://www.euro-access.eu/_media/file/547_call-fiche_eu4h-2024-pj-02_en.pdf (Accessed 30 July 2024).

frontiers | Frontiers in Oncology

TYPE Methods
PUBLISHED 23 November 2022
DOI 10.3389/fonc.2022.1046239

OPEN ACCESS

EDITED BY
Paulo S. Pinheiro,
University of Miami, United States

REVIEWED BY
Lois Dickie,
National Cancer Institute (NIH, United States
Ann Chao,
National Cancer Institute (NIH), United States

*CORRESPONDENCE
Jaume Galceran
jaume.galceran@salutsantjoan.cat
Carmen Martos
JRC-ENCR@ec.europa.eu

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

RECEIVED 16 September 2022
ACCEPTED 28 October 2022
PUBLISHED 23 November 2022

CITATION
Galceran J, Parada D, Eden M, Tumino R, Warren AY, Martos C, Neamtiu L, Visser O and Daubisse-Marliac L (2022) The 2022 ENCR Recommendations on recording and reporting of urothelial tumours of the urinary tract.
Front. Oncol. 12:1046239.
doi: 10.3389/fonc.2022.1046239

The 2022 ENCR Recommendations on recording and reporting of urothelial tumours of the urinary tract

Jaume Galceran[1,2,3*†], David Parada[4,2,3†], Michael Eden[5], Rosario Tumino[6], Anne Yvonne Warren[7], Carmen Martos[8*], Luciana Neamtiu[8], Otto Visser[9] and Laetitia Daubisse-Marliac[10,11,12,13†]

[1]Cancer Epidemiology and Prevention Service, Hospital Universitari Sant Joan de Reus, Reus, Spain, [2]Institut d'Investigacions Sanitàries Pere Virgili (IISPV), Reus, Spain, [3]Faculty of Medicine and Health Sciences, Universitat Rovira i Virgili, Reus-Tarragona, Spain, [4]Pathology Service, Hospital Universitari Sant Joan de Reus, Reus, Spain, [5]NHS Digital, Cambridge, United Kingdom, [6]Cancer Registry and Histopathology Department, Provincial Health Authority (ASP), Ragusa, Italy, [7]Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom, [8]European Commission, Joint Research Centre (JRC), Ispra, Italy, [9]Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, Netherlands, [10]Claudius Regaud Institute, IUCT-O, Tarn Cancer Registry, Toulouse, France, [11]CERPOP, Toulouse University, Inserm UMR 1295, UPS, Toulouse, France, [12]FRANCIM, Network of French Cancer Registries, Toulouse, France, [13]University Hospital Center, IUCT-O, Cancer Coordination Center, Toulouse, France

An updated European Network of Cancer registries (ENCR) Recommendations on Recording and Reporting of Urothelial Tumours of the Urinary Tract had been published in 2022. After the publication by the ENCR of the "Recommendations for coding bladder cancers" in 1995, knowledge about the biology and pathology of urinary tract tumors and their classification has varied and increased substantially. On the other hand, several studies have shown that cancer registries use different definitions, criteria for inclusion and coding of urothelial tumors. This great variability among registries affects not only the criteria for recording (registration, coding and classification) but also the criteria of reporting (counting in the statistics of incidence and survival) urinary tract tumors. This causes difficulties in the data comparability from different registries. Recording and reporting of urothelial tumors requires the application of standard criteria that must take into account the combination of the multiple aspects as the primary topography, the histological type, the grade, the extent of invasion, the multi-centricity, the progressions and the time interval between tumors. This led to the creation of a Working Group of the ENCR that developed these recommendations on the recording and reporting of urothelial tumors of the urinary tract. This article reports these recommendations and the rationale for each.

KEYWORDS

urothelial tumors, recommendations, bladder cancer, recording, reporting, registration practices, cancer registry, Europe

Introduction

In 1995 the European Network of Cancer Registries (ENCR) distributed the "Recommendations for coding bladder cancers" (1). These recommendations were elaborated because of the special characteristics of urothelial tumors and, especially, the difficulties of clinicians and pathologists to correctly determine their morphology, level of invasion and grade and, which makes it impossible to correctly and precisely classify them.

Among the characteristics that make it difficult to record, code and report urothelial tumors are their multicentricity, their great capacity for recurrence and progression, difficulties in correctly determining their grade and level of invasion, and the existence of variants and types that can be confused with other tumors.

After the publication of these Recommendations, knowledge of the biology and pathology of urinary tract tumors has increased substantially and, therefore, their classification has been modified (2, 3). On the other hand, several studies have shown that cancer registries use different definitions, criteria for inclusion and coding of urothelial tumors (4). A recently published study confirms that this variability is still relevant today (5). This wide variability among registries affects not only the criteria for recording (registration, coding and classification) but also the criteria of reporting (counting in the statistics of incidence and survival) urinary tract tumors. This makes it difficult to compare urothelial tumor burden between cancer registries.

The recording and reporting of urothelial tumors requires the application of standard criteria. The combination of multiple aspects must be taken into account: the primary topography, the histological type, the grade, the extent of invasion, the multi-centricity, the recurrences and progressions and the time interval between tumors, the difficulties in the obtaining of the result of biopsies, the recording stage, the existence of tumors diagnosed before the registry's period of recording, the residence of patients at the time of diagnosis of each tumor and the standard criteria for multiplicity. All this led to the creation of a new ENCR Working Group that has reviewed and updated the ENCR Recommendations published in 1995. These new recommendations were published/distributed in June 2022 under the title "ENCR Recommendations on Recording and Reporting of Urothelial Tumours of the Urinary Tract" and European population-based cancer registries must apply them to all urothelial tumors with an incidence date of 1st January 2022 or later (6).

These recommendations are based on current knowledge about the biology, anatomical pathology and epidemiology of urinary tract tumors reflected in the fourth edition of the WHO Classification of Tumors of the Urinary System and Male Genital Organs of 2016 (2) and also in new knowledge on urothelial tumors published more recently (7–9). Although WHO 2016 classification has been used, these recommendations include all the aspects listed in the previous paragraph for recording and reporting these tumors in a harmonized way in the European cancer registries.

These recommendations will enable population-based cancer registries to improve the quality of their data and the comparability of incidence and survival data, while providing useful information to clinicians and policymakers. This document reports these recommendations and the rationale for each one of them.

Methods

In 2017, the cancer registries of Tarn (France) and Tarragona and Girona (Spain) launched a survey to European cancer registries on the practices of registration, coding and reporting of urothelial tumors. For example, in cases in which the tumor presented various levels of progression. The survey was answered by 42 registries. The conclusions of the survey were that there was an urgent need to define clear rules for the registration these tumors. As an example, in cases where the tumor had various levels of progression from a low-grade non-invasive tumor to an invasive tumor, 8 recorded only the first tumor, one recorded only the last (invasive), 13 recorded the first (low-grade non-invasive) and the last; 11 recorded combinations that included the first and the last, and 9 recorded all tumors. In relation to reporting, there was also great variability: 18 reported only the first, 13 reported only the last, 10 reported combinations of one or several tumors, and one did not report any tumor. In addition to questions on inclusion (recording) and reporting criteria, the survey also included questions on coding composite tumors such as urothelial carcinomas with squamous, adenocarcinomatous or neuroendocrine component, and neuroendocrine carcinoma with urothelial carcinoma. In this aspect, the degree of discordance was lower.

In June 2018, the ENCR launched an offer of expressions of interest from member registries to join an Urothelial Cancers Task Force that would include both cancer registry and clinical representation. The aims of the Working Group (WG) were to address the difficulties in the registration of urothelial cancers and to update the ENCR recommendations published in 1995. The new ENCR Recommendations would improve incidence and survival data comparability across different European registries and countries.

Once the WG was established, a first meeting was held by teleconference on December 5, 2018. On July 12, 2019, a first face-to-face meeting of the Working Group was held at the Join Research Center (JRC) in Ispra, Italy, at which it was decided to update and to draft the ENCR Recommendations. A second face-to-face meeting was also held in Ispra on November 8, 2019. After the second meeting, the WG continued its work virtually, introducing modifications to previous versions of the document.

During the meetings and during the virtual work of the Working Group, each of the decisions on recording and coding

issues were made by consensus of at least 7 of the 8 members of the group.

Once the WG finalized the draft Recommendations, it was sent to the ENCR Steering Committee, which reviewed it and proposed some modifications. Once the Working Group and the Steering Committee agreed on the document, it was sent to all ENCR members for revision and feedback. Some registries sent their comments and asked for clarifications. All the questions asked were answered and some of the registries' proposed modifications were introduced. Finally, on June 8, 2022, the Steering Committee approved the final version of the Recommendations that were published on the ENCR website a few days later (https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation_UT_Jun2022_EN.pdf). Figure 1 shows the scheme of this process from the offer of expression of interest to the publication of the Recommendations.

Results and discussion

In cancer registration and especially in the registration of some types of tumors with high rates of recurrence and progression such as urothelial tumors of the urinary tract, it is important to differentiate between **recording** (registration) and **reporting** (counting) tumors. A cancer registry can record several tumors of the urothelium (of different site, grade or invasion) of the same patient but according to international criteria and for the purposes of comparability, only one or a part of them is actually reported.

Recommendations for recording urothelial tumors

The recommendations for recording of urothelial tumors are based on three general principles

First, these Recommendations apply to all urothelial tumors (transitional cell tumors) and their variants regardless of tumor topography (renal pelvis, ureter, urinary bladder, or urethra – International Classification of Diseases for Oncology, Third Edition (ICD-O-3) codes C65 to C68–). Therefore, they apply to the pure urothelial carcinomas, to urothelial carcinomas with divergent (squamous, glandular, trophoblastic and other) differentiation, and to all other variants (nested, microcystic, lymphoepithelioma-like, plasmacytoid/signet ring cell, sarcomatoid, giant cell, lipid-rich, clear cell and poorly differentiated) of urothelial carcinomas. Sarcomas and other histologic types of cancer (e.g., adenocarcinomas, squamous cell carcinomas, or neuroendocrine tumors) of the urinary tract are not included in these recommendations, although they do occur in the urinary tract and should also be recorded by registries.

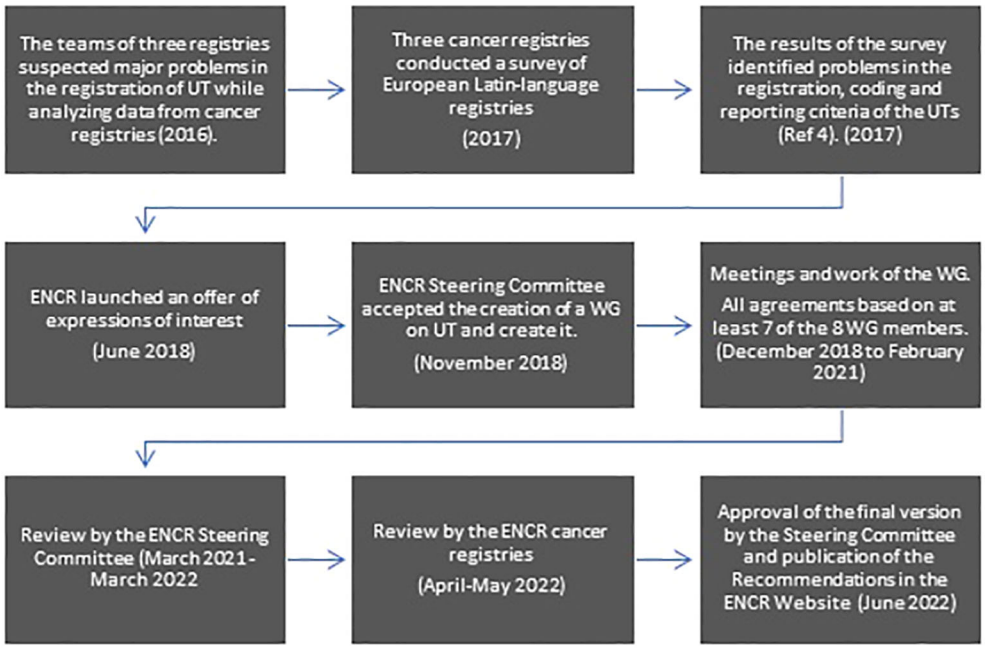


FIGURE 1
Scheme of the process of preparing the ENCR Recommendations on Urothelial Cancers.

Secondly, in order to correctly record and code urothelial tumors, it is essential to have access to pathological examinations (reports) since knowledge of the topography, morphological type, behavior and grade of the tumor is required. The non-existence or non-availability of anatomopathological reports prevents, in many cases, knowing whether the tumor should be registered and, in all cases, accurate coding of some variables (morphology, behavior, grade…).

Third, although in cancer registration, the usual definition of synchronous tumors includes all tumors of the same location that appear in a period of less than or equal to 3 months, these specific recommendations for urothelial tumors define synchronous tumors of the same location and laterality as those that present in a period of less than or equal to 4 months. This criterion also applies to urothelial tumors whose resection is performed in two phases since, in many of these cases, the initial resections are not complete or the second revision is sometimes delayed, particularly in elderly patients.

Criteria for the inclusion (registration) of urothelial tumors

In the following paragraphs, the 11 rules or criteria for the inclusion (registration) of urothelial tumors in the cancer registry are described. Each of the inclusion criteria is indicated in italics and indented as they are in the European Network of Cancer Registries Recommendations document (6).

Types of tumors to be included

Cancer registries must record all invasive and non-invasive urothelial carcinomas including those without histological confirmation. Obviously, the urothelial concept includes any type of urothelial carcinoma and any of its variants.

The 4th and 5th editions of the WHO Classification of Tumours of the Urinary System (2, 3) allow a clear differentiation between malignant tumors (invasive or not) and non-malignant tumors. According this new WHO Classification, papillary urothelial neoplasms of low malignant potential (PUNLMP), urothelial papillomas, inverted urothelial papillomas, urothelial proliferation of uncertain malignant potential and urothelial dysplasia are not considered malignant, and are therefore not recommended for registration in a cancer registry. However, cancer registries that for whatever reason are interested in any of these entities may register them if they wish, but they should never be included in the *incidence computation.*

"*The following types of tumors arising in the urinary tract must be recorded:*

1. Non-invasive papillary urothelial carcinoma, low-grade

2. Non-invasive papillary urothelial carcinoma, high-grade

3. Urothelial carcinoma in situ *(carcinoma* in situ*)*

4. All invasive carcinomas

5. Tumour with histologic examination but invasion cannot be assessed

6. Tumour with cytological examination only (see rule 2.b, page 6)

7. Tumour with no microscopic confirmation (see rule 2.c, page 6)"

Multiples sites

The International Rules for Multiple Primary Cancer edited jointly by the International Agency for Research on Cancer, the World Health Organization, the International Association of Cancer Registries and the European Network of Cancer Registries in 2004 (10) are for "reporting" data on cancer incidence and survival, so that cancer risk and outcome are comparable between different populations. The same Rules indicate that for collection, it is recommended that registries collect and register more detailed data and, in fact, cancer registries use different rules for defining multiple primaries when registering cancer cases. Such cases should be collapsed to conform to the international rules for analysis.

The WG that prepared these updated ENCR Recommendations considered that, in order to be able to analyze many aspects of these tumors, it is necessary to have information on all tumors with different three-digit ICD-O-3 topography. Therefore, the recommendation is

"if a patient presents with several (synchronous or metachronous) urothelial tumors in different sites, record all tumours of different three-digit sites (C65-C68) and laterality (if renal pelvis or ureter). If a metachronous tumor is diagnosed in the ureter or urethra after cystectomy, it should not be recorded if it has arisen at the surgical margin because it should be considered as a local recurrence of the removed tumor in the urinary bladder except if it is a progression."

Progressions

A characteristic of urothelial carcinomas is their high capacity for recurrence and progression. Reported 5-year rates of non-muscle-invasive bladder carcinoma recurrence range from 50% to 70% and reported 5-year rates of progression range from 10% to 30%. Factors associated with recurrence and progression include, among others, high grade, high stage, large tumor size, multifocality, high number of previous recurrences and presence of concomitant carcinoma *in situ* (11), and histological variants. Tumor grade, stage, and carcinoma *in situ* are the most important variables for progression (12) Taking into account this ability of urothelial

tumors to progress, and their different prognosis depending on their grade, level of invasion and morphology, it has been considered necessary that cancer registries record progressions in order to be able to correctly compare survival among different populations.

Studies have suggested that invasive urothelial tumors develop along at least two molecular pathways, *via* either high-grade papillary tumors or carcinoma *in situ* (7, 13). For this reason, when a new urothelial tumor is diagnosed in a patient who already has previous tumors, it can be difficult to define whether or not the new tumor represents progression. Therefore, in these recommendations the process of progression was determined not on the basis of the molecular pathway but on the basis of the severity of the tumor and its ability to progress further. Thus, carcinoma *in situ* was considered as progression of high-grade non-invasive carcinoma and the recommendation was defined as follows:

"If a patient presents with several urothelial tumors in the same three-digit topographical site that includes some progression of the disease, register the first tumor and then subsequently only those tumours that represent a chronological progression. The following series shows the order that represent a progression:

Non-invasive, low grade (TaG1) → Non-invasive, high grade (TaG3) → In situ *(Tis) → Invasive, superficial (T1) → Muscle-invasive (T2+).*

Due to the special characteristics of urothelial tumours, the recording of the different stages should be done for these tumours in order to know their progression. Remember that all known steps of this progression should be recorded. Therefore, for example, the recording of a T2+ invasive tumor does not replace the recording of a T1 invasive tumor if the latter is known."

Recurrences

It has already been mentioned that urothelial tumors have a great tendency to present with recurrences and progressions. Multiplicity, tumor size, and prior recurrence rate are the most important variables for recurrence (12). Recurrences do not significantly change the patient's prognosis. Therefore the fourth rule specifies that:

"Tumours that represent recurrences (not progressions) with the same or lower level of invasion and degree do not have to be recorded".

Synchronous urothelial tumors of the same site and laterality

Tumor multifocality, that is, the existence of two or more non-contiguous tumor formations separated by a macroscopically non-tumorous tissue area, is common in urothelial carcinomas. In carcinoma *in situ*, involvement of the surface urothelium is usually multifocal (14). In non-invasive urothelial tumors, multifocality is one of the factors determining clinical risk of recurrence and disease progression (2). Due to this characteristic of urothelial tumors, the WG agreed that the presence of multifocality at the same topography has to be registered as a single tumor and if the different tumors have a different level of aggressiveness (grade, level of invasion -T-), the one to be registered is the most aggressive one, to ensure a correct survival analysis. This standard should apply to synchronous tumors and, as discussed in the general principles of urothelial tumor registration, this means all tumors within a maximum period of 4 months between them.

"If a patient presents with more than one urothelial tumour in the same three-digit topographical site and laterality (if renal pelvis or ureter) in a short period of time (≤4 months – i.e. synchronous–), record only the most aggressive of them (based on the progression scheme in point 3 above) but with the date of diagnosis taken from the first tumour.

This criterion also applies to tumours whose resection is performed in two phases. In these cases, the temporal course of clinical investigation should also be considered because sometimes initial resections are not complete or the second look is sometimes delayed, particularly in old patients."

Codes of site in synchronous tumors of bladder:

Two or more tumors may arise synchronously in the bladder, with similar or different aggressiveness. In this case, if the two (or more) tumors are in the same subsite of the bladder, this subsite should be coded, but if the tumors are in different subsites, code C67.8 should be recorded to follow ICD-O criteria on "Tumors involving more than one topographic category or subcategory".

"Record synchronous tumors of the bladder using the synchronous tumor rule (rule 5). If the highest level of progression is present in more than one tumour and in more than one subsite (four-digit topography), code the site as C67.8 even if the tumours are not contiguous. If they appear in the same subsite, codify the corresponding subsite."

Synchronous urothelial tumors of different site

Although the International Rules for Multiple Primary Cancers (ICD-O Third Edition) for reporting tumors consider tumors of the renal pelvis (C65), ureter (C66), urinary bladder (C67) and other and unspecified urinary organs (C68.9) as belonging to a single topographic site, it is highly recommended that tumors from different three-digit ICD-O-3 sites be recorded as different tumors. First, the separate

registration of the multiple tumors allows registries to better describe the incidence. Second, for survival studies, the knowledge of the existence of multiple tumors and their site is fundamental since the prognosis depends, among other factors, on the primary site where the tumor has developed.

"If a patient presents with more than one urothelial tumor in different three-digit topographical sites in a short period of time (≤4 months –synchronous–), record each tumor separately, each one with its corresponding topography, morphology, behavior codes and incidence date (do not use grouping code C68.9 for registration purpose)".

Bilateral tumors

Unlike synchronous urothelial tumors of different site, bilateral tumors share the same site code. However, having a single tumor at a paired site (pelvis or ureter) does not carry the same prognosis as having a tumor at each of the paired sites. Furthermore, their aggressiveness (grade, level of invasion, morphological type) may be different. For these reasons, and although only one tumor should be counted for the calculation of incidence, it is recommended that bilateral tumors of the same site be recorded according to the following criteria:

"If a patient presents with several (synchronous or metachronous) urothelial tumours in both sides of the same paired organ (e.g. right and left pelvis or right and left ureter), record all the tumors of each side of each three digit site following rules 3 to 6 (e.g. 1st urothelial carcinoma in right ureter and its progressions, and 1st urothelial carcinoma in left ureter and its progressions)."

Mixed situations of multiplicity, progressions and synchronicity/ metachronicity

Due to the multifocality and progressive characteristics of urothelial tumors, there are many possible combinations of multiplicity, progression and temporality. Consider the example of a patient who presents with a mixed combination of multiple synchronous and metachronous urothelial tumors in the same and different three-digit topographies. This patient presented in this chronological order with a "Non-invasive low-grade carcinoma" of bladder (1) followed by a synchronous "Invasive carcinoma" of bladder (2) followed by an "*In situ* carcinoma" of right renal pelvis (3) followed by a "Non-Invasive high-grade carcinoma" of right renal pelvis (4) followed by an "Invasive carcinoma" of bladder (5).

Tumor 1 and tumor 2 are synchronous at the same site, so only the more aggressive, in this case the invasive one should be recorded (with the date of diagnosis of the first tumor). Tumor 3 should be recorded because it appeared in a different site. Tumor 4, on the other hand, should not be recorded because it must be considered a recurrence of tumor 3. Finally, tumor 5 should not be recorded either because it is a recurrence of tumor 2.

"If a patient presents with a combination of synchronous and metachronous multiple urothelial tumors in the same and/or different three-digit sites, record them according to rules 2 to 8."

First tumor occurring outside the area of registration

In cancer registries it is possible to find cancers of the same topography separated in time. The higher the incidence of a type of cancer and its survival, the more likely it is to find this phenomenon. Colorectal and breast cancers are a good example of this. But, once again, due to their high capacity for recurrence and progression, urothelial carcinomas are the ones that present this phenomenon most frequently, except for non-melanoma skin cancers.

People can change their residence throughout their lives and each tumor is registered in association with the patient's residence at the time of diagnosis. This may result in a first tumor being diagnosed when the patient resides outside the registry area and the next one(s) being diagnosed when the patient resides in the registry area. In this situation, if we do not record the first tumor, we will not be aware that the second is not an incident case but rather a prevalent one and, therefore, we will mistakenly count it as incident. This will cause an overestimation of the incidence. So, the recording of a first tumor diagnosed outside the area of registration allows the registry to know if a subsequent tumor is a recurrence or progression (recorded but not reported as incident) thus avoiding over-reporting.

"A patient can move from one residence to another, so place of residence should be related to the tumours and not to the patient. If information is available showing a patient resident in the coverage area of the registry has been previously diagnosed with a urothelial tumor(s) when resident outside the registration area, record all of them (the ones occurring outside the area of registration and the ones diagnosed being resident in the area of the registry) according to rules 2 to 8 (that enables the tumours to be flagged as 'Extra-regional' for reporting purposes)."

First tumor occurring before the operation period of the registry

A similar situation occurs when the first tumor is diagnosed before the registry operation period. If we do not record the first tumor diagnosed before the registration period, it is impossible

to know that successive cancers from the same site are recurrences or progressions and not cancers that should be reported as incident. So, the recording of tumors diagnosed before the period of operation of the registry allows the registry to know whether subsequent tumors should be recorded as progression or recurrence (recorded but not reported as incident) to prevent over-reporting.

"If information is available showing a patient resident in the coverage area of the registry has been diagnosed with one or more urothelial tumors before the operation period of the registry, record all their tumors (the ones diagnosed before and the one diagnosed after first date of operation of the registry) according to rules 2 to 8."

Recommendations for classification and coding

In the following paragraphs, the recommendations for the coding and classification of urothelial tumors are described. As has been done with the inclusion criteria section, recommendations are indicated in italics and indented as they are in the ENCR Recommendations document (6).

Classification used in the cancer registries

Cancer registries usually code topography, morphology, behavior, and grade of the tumor according to the International Classification of Diseases for Oncology (ICD-O). This classification has evolved over the years with several editions and revisions that allow coding of newly defined cancer entities. Thus, as far as possible, it is recommended that registries adapt their coding criteria to the new editions and revisions of the ICD-O.

The recommended version of the ICD-O until the end of 2019 was the International Classification of Diseases for Oncology, 3rd edition, 1st revision (ICD-O-3.1) (15) and the second revision (ICD-O-3.2) which is not yet published is recommended to be used for tumors diagnosed on or after January 1, 2020 as reported on the IACR website (16).

The new versions of the ICD-O try to adapt as much as possible to the most recent versions of the WHO Classification of Tumours, which include morphological codes. In June 2022, a new version of the "WHO Classification of Tumours: Urinary and Male Genital Tumours" has been published (3). In some cases, the latest version of the ICD-O does not include a morphological category. For example, nested urothelial carcinoma is not covered by OCD-O-3.2. However, the latest edition of the WHO Classification of June 2022 indicates that it should be coded with code 8120/3.

"All urothelial tumors must be coded according to the most recent version of the International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3) (these classifications are almost equivalent to the WHO classification)."

Morphology, behavior and grade

a) Codes of the most frequent morphological categories when histology is available

Any of the morphological types can be found in any of the topographies of the urinary tract (renal pelvis, ureter, bladder, and urethra). Likewise, apart from the existing difficulties in determining the exact subtype of urothelial carcinoma, often the greatest difficulty is in determining the level of tumor invasion. This occurs because the evaluation of biopsies and transurethral resections of the bladder (TURB) can be extremely difficult for several reasons (9). First, proper pathology reporting is extremely dependent on the quality of the submitted material. Cautery artifact may hinder accurate staging at initial TURB for large tumors by understaging up to 6% of patients (17). Second, pathologists can have difficulty recognizing superficial invasion of the lamina propria and differentiating invasion of the muscularis propria from invasion of the muscularis mucosae (18).

When a tumor has been examined histologically but it has not been possible to determine the level of invasion, a dilemma arises in the cancer registry between coding behavior/2 (non-invasive carcinoma) or/3 (invasive carcinoma). In these situations, behavior/3 is never assigned by default, but only if there is a clinical impression of an invasive tumor (then use the code 8120/3). Otherwise (i.e. no obvious invasion on clinical/paraclinical examination), the code 8130/2 (non-invasive papillary tumor) must be used when the term papillary is mentioned in the pathology report or when the TURB report mentions papillary appearance. However, if the term papillary is not mentioned or there is no information about the appearance of the tumor, then it cannot be coded as papillary and code 8120/

Tumor type	*Morphology/ Behavior*	*Grade*
Non-invasive (papillary) urothelial carcinoma, low-grade		*1*
Non-invasive (papillary) urothelial carcinoma, high-grade	*8130/2**	*3*
Non-invasive (papillary) urothelial carcinoma, grade unknown		*9*
Urothelial carcinoma (with histologic examination), but invasion cannot be assessed		
• *Papillary term mentioned or papillary appearance (exophytic lesion)*	*8130/2***	

(Continued)

Continued

Tumor type	*Morphology/ Behavior*	*Grade*
• *Papillary term not mentioned or no information about appearance*	*8120/2 ****	*1/3/9*
• *The clinical impression is of invasive disease*	*8120/3*****	*3*
Urothelial carcinoma in situ (carcinoma *in situ*)	*8120/2*	*3******
Invasive carcinoma, not otherwise specified (NOS) (1)	*8010/3*	*3******
Invasive urothelial carcinoma	*8120/3*	*3******

(1) Although most carcinomas of urinary tract are urothelial, there are also other carcinomas such as squamous or adenocarcinoma. Therefore, if urothelial or transitional cell is not specified on the pathological report, code "8010/3". But if non-invasive urothelial carcinoma was previously diagnosed, record (code) as urothelial carcinoma (8120/3), provided that prostate carcinoma invading the urinary bladder is ruled out. Also, if the concept urothelial is in the tumor description, code 8120/3 even if not specified in the final diagnosis.
(*) When the term "papillary" is not specified in the pathological report but the pathology report indicates an urothelial carcinoma with pTa stage, code 8130/2 (plus grade, if specified)
(**) In this case, code pTa.
(***) In this case the code pT is pTX (and not pTis), so as not to be confused with Carcinoma in situ.
(****) If the clinical impression is of invasive disease, then code with/3 behavior code and grade 3.
(*****) All in situ and invasive carcinomas must be recorded as high grade. Although the pathology report may indicate "low grade" or not indicate a grade, if it is an in situ or invasive tumor, it must be considered high grade."

2 should be used. In this case the code pT is pTX (and not pTis), to avoid confusion with carcinoma *in situ*.

All *in situ* and invasive carcinomas should be coded as grade 3 even if the pathology report indicates "low grade" or does not indicate the grade. The reason is explained in section 2.e of these grading and coding criteria.

b) Codes when only cytological examination is available*

Although all cases should have histological examination, in a few cases only cytological examination can be found. This may be because the patient has not had a histologic examination or because the cancer registry does not have it available. In these cases, it is recommended to use the "Paris System reporting for urine cytology (19–22).

This System have the following seven diagnostic categories: 1. Non-diagnostic/Unsatisfactory; 2. Negative for high-grade urothelial carcinoma (NHGUC); 3. Atypical urothelial cells (AUC); 4. Suspicious for high-grade urothelial carcinoma (SHGUC); 5. High-grade urothelial carcinoma (HGUC); 6. Low-grade urothelial neoplasm (LGUN), and 7. Other: primary and secondary malignancies and miscellaneous lesion. Of these, only categories 4 and 5 should be considered as high-grade urothelial carcinomas.

In these high-grade tumors diagnosed by cytological examination only, a consensus has been agreed upon for high-grade urothelial carcinoma to be coded as behavior/2 although it was acknowledged there is a limited evidence base to support either this or coding to behavior code/3. In any case, in cancer registration if there is the clinical impression, e.g. with imaging, that the tumor is invasive then it should be coded with the behavioral code/3. In these cases, an effort should also be made to ascertain whether the tumor has a papillary appearance (8130) or not (8120) by reviewing the imaging.

Non-urothelial malignant cells may also be found on cytology. Evidently, in these cases, non-urothelial malignant cells seen on cytology should be coded according to the pathology report and clinical information. This would be the case, for example, for non-urothelial urinary tract tumors (squamous, glandular, Müllerian type, neuroendocrine, melanocytic, mesenchymal…) and metastases from tumors outside the urinary tract.

Cytology results	*Morphology*/ Behavior ***	*Grade*
High grade urothelial carcinoma or "suspicious for high-grade urothelial carcinoma" (SHGUC of the Paris classification). *(See ANNEX 2, section "Paris System reporting for urine cytology", paragraph "Behavior of high grade tumors diagnosed by cytology only").*	*8130/2 (papillary appearance) or 8120/2*	*3*

(*) If you only have cytological examination, try to find out if the tumor has a papillary appearance (8130) or not (8120) by reviewing the imaging.
(**) If the clinical impression (e.g. scans) is of invasive disease, then code with/3 behavior code.
Non-urothelial malignant cells seen on cytology should be coded according to the pathology report and clinical information.
If the topography of the tumor is highlighted on radiology/imaging, code the specific site. Otherwise, code the topography C68.9 (urinary tract, NOS)."

Of course, an effort should also be made to identify and code the exact topography of the tumor by radiology/imaging. If this is not known, topography 68.9 (Urinary tract, not otherwise specified (NOS) should be coded as the tumor can be located at any point between the renal pelvis and the urethra.

c) Codes when only non-microscopic confirmation is available (histo/ cytopathological evidence unavailable)

In other rare situations, neither histological nor cytopathological evidence is available. In these cases, only tumors with a clinically malignant appearance can be recorded, which can be coded as 8000/3 because the morphologic result is not available,

Tumor type	*Morphology/ Behavior*	*Grade*
No microscopic confirmation: Tumor clinically malignant	*8000/3*	*9*
No microscopic confirmation: Tumor NOS	*Do not record**	

(*) If recorded, code: 8000/1 Grade 9."

and grade 9 because it is also unknown. If the tumor has no malignant appearance or its appearance of malignancy is doubtful, it is not necessary to register it and, if it is decided to register it, code it as 8000/1 grade 9.

"When histo/cytopathological evidence is unavailable but clinical appearance is confirmed by the clinician, use the following codes.

d) Codes of behavior for unknown level of invasion

When there is a histologic examination but the exact level of invasion is unknown, it is usually because either subepithelial connective tissue or muscularis propria is not present in the specimens received by the pathologist. In either of these cases the first thing to do, if possible, is to consult the pathologist for advice/assessment.

In case of urothelial papilloma, papillary urothelial neoplasms of low malignant potential (PUNLMP) or urothelial proliferation of uncertain malignant potential, the recommendation is not to register these entities as already mentioned in the point "1.1. Types of tumors to be included" of the Criteria for the inclusion (registration) of urothelial tumors.

What should be done when subepithelial connective tissue is not present? As the lack of subepithelial connective tissue does not preclude the diagnosis of non-invasive carcinomas, if it diagnosed as a "Non-invasive papillary urothelial carcinoma" or a "Carcinoma *in situ*" code behavior/2. However, if morphological characteristics are not specified, code behavior/2 because it is the maximum aggressiveness that can be assumed (behavior/3 should never be assigned by default). In relation to the morphology, the code to use depends on the appearance at endoscopy: 8120 (no papillary appearance) or 8130 (papillary appearance).

And what to do when muscularis propria is not present? If sub-epithelial connective tissue is invaded, code behavior/3. But, otherwise, code behavior/2 according to the morphological characteristics (papillary or not).

"d1) "Subepithelial connective tissue" is not present in resection.

First of all, ask for pathologist assessment. If it is not possible or the pathologist can't give an answer:

- If "Urothelial papilloma":/0 (there is no recommendation to record this tumor).
- If "Papillary urothelial neoplasm of low malignant potential (PUNLMP)":/1 (there is no recommendation to record this tumor but if it is recorded, code 8130/1 without grade and pT) (some pathologists can erroneously code pTa in PUNLMP. pTa should be used only in carcinomas).
- If "Urothelial proliferation of uncertain malignant potential":/1 (there is no recommendation to record this entity).
- *If "Non-invasive papillary urothelial carcinoma" or "Carcinoma* in situ*":/2*
- *If morphological characteristics are not specified:/2 (Codify morphology 8120 (no papillary appearance) or 8130 (papillary appearance) depending on the appearance at endoscopy).*

d2) *"Muscularis propria" is not present in resection.*

First of all, ask for pathologist assessment. If it is not possible or the pathologist can't give an answer:

- *If sub-epithelial connective tissue is invaded:/3.*
- *Otherwise, code behavior/2 (according to the morphological characteristics)."*

e) Grade

Grade registration is especially important for the non-invasive papillary urothelial carcinomas where it is necessary to distinguish between the high-grade (code 3) and the low-grade (code 1) tumors. As in 2004, the 2016 WHO Classification recommends the use of the grading classification first put forth by ISUP in 1997 (2). This 2-tiered grading system—high versus low grade—is intended to simplify clinical decision making in daily practice over the 3-tiered 1973 system. It also provides congruence between histology and cytology reports, and highlights the prompt therapeutic requirement for all high-grade lesions (flat or papillary) (23). Moreover this system does not outperform the 1973 system in prognostic value, but shows higher reproducibility (24). If the pathology report does provide tumor grades according to both 2016 and 1973 systems or does not indicate whether the tumor is low grade or high grade, but rather indicates the grade based on the three categories of level 1, 2 and 3, the following table of these Recommendations shows the correspondence between the two classifications. As a result, code 2 will no longer be used to code the grade.

In relation to the invasive urothelial carcinomas, the overwhelming majority of invasive urothelial carcinomas are high grade (25). However, some variants (e.g. large nested variant of urothelial carcinoma) may present a "pseudo-benign" (deceptively bland) appearance, but this appearance is misleading,

Description in the pathology report	*Code*
Grade 1	*Low grade (1)*
Grade 1/2 (low grade or no grade mentioned)	*Low grade (1)*
Grade 2 low grade	*Low grade (1)*
Grade 2 high grade	*High grade (3)*
Grade 2/3 (high grade or no grade mentioned)	*High grade (3)*
Grade 3	*High grade (3)*

since these variants have a poor outcome (26–28). On this basis, all invasive urothelial tumors should be recorded as 'Grade 3'.

"*Codes according to the description in the pathological report:*

Codes for urothelial carcinomas with other morphological terms

Urothelial carcinoma has long been known to have a remarkable propensity for divergent differentiation (29), which is seen most commonly in association with high-grade and locally advanced disease. Common morphologic manifestations of divergent differentiation are along squamous and secondly glandular lines, but also along trophoblastic lines. Around 10% of cases have multiple mixed histologic types (30).

This remarkable propensity for morphological diversity is due both to divergent differentiation and to the existence of histological subtypes. Much literature has been devoted to the characterization and definition of histological entities, but only few prospective data exist (31). Recently, molecular classification (i.e. on basis of expression and genetic alterations) has enriched our understanding of bladder cancer and provided us with a new framework for stratification and assessing response to different therapy regimens (32). It is important to understand that when talking about divergent differentiation or subtypes, a therapeutic implication exists. Therefore, the pathologist must be aware of the diagnostic criteria and accurately report them (8).

Urothelial carcinoma with squamous cell divergent differentiation (with an squamous component): We must differentiate pure squamous cell carcinoma from urothelial carcinoma with squamous cell divergent differentiation (with an epidermoid component) because they are a different tumor type and are treated differently (7). Therefore, we must code the morphology of the first one as 8070 and the latter as 8120."

a) Urothelial cell carcinoma with epidermoid component (squamous divergent differentiation): 8120

Code squamous carcinoma only if it is a pure squamous carcinoma: 8070 "Pure squamous carcinomas" should be registered separately from urothelial carcinomas because they are a different tumor type from urothelial carcinomas and are treated differently (1, 2), even if the 2004 International Rules for Multiple Primary Cancers include this two tumors in the same morphology group.

Urothelial carcinoma with an adenocarcinomatous component (glandular divergent differentiation): The same applies to the urothelial carcinoma with an adenocarcinomatous component (glandular divergent differentiation). So, we must code as adenocarcinoma only if it is a pure adenocarcinoma (8140) and for urothelial cell carcinoma with adenocarcinomatous component, the code must be 8120.

b) Urothelial cell carcinoma with adenocarcinomatous component (glandular divergent differentiation): 8120

Code adenocarcinoma only if it is a pure adenocarcinoma: 8140

"Pure adenocarcinomas" should be registered separately from urothelial carcinomas because they are a different tumor type from urothelial carcinomas.

Urothelial cell carcinoma subtypes and ICD-O-3 specific code: ICD-O-3 and the 2016 WHO Classification, contains some subtype codes for urothelial tumors. These codes should be used whenever they are reported in pathology reports. These codes are: micropapillary: 8131, lymphoepithelioma-like: 8082, sarcomatoid: 8122, giant cell: 8031 and undifferentiated: 8020.

c) Urothelial cell carcinoma subtypes and ICD-O-3 specific code *(new specific codes may appear in subsequent versions of ICD-O/WHO Classification):*

- *Micropapillary: 8131*
- *Lymphoepithelioma-like: 8082*
- *Sarcomatoid: 8122*
- *Giant cell: 8031*
- *Undifferentiated: 8020*

Urothelial cell carcinoma without specific subtype in ICD-O-3 classification: The other subtypes of urothelial carcinomas without a specific morphological code in the ICD-O-3 classification (e.g. nested, microcystic, plasmacytoid, signet ring cell, diffuse, lipid-rich, clear-cell) must be coded as 8120. However, it is possible that some of these subtypes may have specific codes in subsequent versions of ICD-O/WHO Classification. If this occurs, it is recommended to use the new codes that appear. This is in fact already the case since the fifth edition of the WHO classification was recently published (2022), shortly after the release of these recommendations. This new classification assigns the morphological code 8122/3 not only to sarcomatoid urothelial tumors but also to plasmocytoid, signet ring cell and diffuse urothelial tumors (3).

d) Urothelial cell carcinoma without specific subtype in ICD-O-3 classification (e.g. nested, microcystic, plasmacytoid, signet ring cell, diffuse, lipid-rich, clear-cell) *(some of these may have specific codes in subsequent versions of ICD-O/WHO Classification): 8120*

Urothelial cell carcinoma with neuroendocrine component (neuroendocrine differentiation): A very different case is that of neuroendocrine tumors. Neuroendocrine tumors are classified into well-differentiated neuroendocrine tumor and neuroendocrine carcinoma which includes both large- and small-cell neuroendocrine carcinoma. Whereas well-differentiated neuroendocrine tumors occur in pure form, neuroendocrine carcinomas are often admixed with some form of non-neuroendocrine carcinoma that is most frequently urothelial carcinoma (33). Both large- and small-cell neuroendocrine carcinomas can arise within the bladder. Large-cell neuroendocrine carcinoma is extremely uncommon whereas the incidence of small-cell neuroendocrine carcinoma is

only 0.5-1.0% (34). The cell of origin of neuroendocrine carcinoma is unclear (35). Microscopically, large-cell neuroendocrine carcinomas are usually high-grade and poorly differentiated. Approximately 50% of cases of small-cell neuroendocrine carcinomas show an admixture of small-cell neuroendocrine carcinoma with non-small-cell carcinoma components (33, 36) and the ratio of neuroendocrine and non-neuroendocrine components may vary and the amount of the neuroendocrine carcinoma component may be important to outcomes.

The term "Neuroendocrine carcinoma" should be used in all tumors with small or large cell neuroendocrine histology in any proportion of the tumor (37). Recording the histological tumor type using the 2016 WHO classification is a required element as this parameter often has prognostic and therapeutic significance. Therefore, the code assigned to the tumor morphology should most accurately reflect the pathological diagnosis from among the following: 8041 (small cell neuroendocrine carcinoma), 8013 (large cell neuroendocrine carcinoma), 8045 (small and large cell carcinoma), 8240 (neuroendocrine carcinoma well-differentiated or low-grade), 8249 (neuroendocrine carcinoma moderately-differentiated or high grade) and 8246 (neuroendocrine carcinoma, NOS).

A tumor is classified as urothelial carcinoma if there is any urothelial differentiation [including associated urothelial carcinoma *in situ* (CIS)], with any other types present reported with an estimated percentage. Thus, a carcinoma showing 20% urothelial differentiation and 80% glandular differentiation should be reported under the histological tumor type "Urothelial carcinoma". An exception to this rule is for cases with any amount of neuroendocrine component (small cell neuroendocrine carcinoma or large cell neuroendocrine carcinoma) where classification is now in the neuroendocrine tumor category. Thus, a mixed tumor with 30% small cell neuroendocrine carcinoma and 70% urothelial carcinoma should be reported under the histological tumor type as neuroendocrine tumor (small cell neuroendocrine carcinoma). This is a controversial issue, as reflected by the different approaches recommended by WHO 2016 in chapters on the neuroendocrine tumors and urothelial carcinoma variants. The International Collaboration on Cancer Reporting (ICCR) recommends the latter approach but recognizes that the percentage of the neuroendocrine component could inform patient management, particularly with newer treatment modalities such as immunotherapy.

e) Urothelial cell carcinoma with neuroendocrine component (neuroendocrine differentiation):

Always encode neuroendocrine carcinoma independently of the amount of the neuroendocrine component (See Annex 2: Comments. Neuroendocrine tumors).

– Small cell neuroendocrine carcinoma: 8041

– *Large cell neuroendocrine carcinoma: 8013*

– *Composite small and large cell neuroendocrine carcinoma: 8045*

– *Neuroendocrine carcinoma well-differentiated or low-grade NET: 8240*

– *Neuroendocrine carcinoma moderately-differentiated or high-grade NET: 8249*

– *Neuroendocrine carcinoma, NOS: 8246"*

Non-urothelial specific carcinomas

Unlike urothelial tumors with squamous, glandular or other types of differentiation, there are non-urothelial tumors of the urinary tract such as (pure) adenocarcinomas, (pure) squamous carcinomas, and neuroendocrine, melanocytic, mesenchymal or lymphoid tumors (38) which must be recorded separately from urothelial tumors following the general criteria for other tumors.

Table 1 summarizes the main criteria of inclusion (according to invasion, grade and existence of progression). For each site (right and left pelvis, right and left ureter, bladder and urethra), this table summarizes, which tumors should be registered by application of rules 2 to 8. In summary: after recording the first tumor (/2 or/3) of each site, only record subsequent tumors that represent progression, according to the grouping of categories (columns 1, 2, 3, 4 and 5).

Coding the Basis of Diagnosis

Considering the current methods for diagnosing urothelial tumors, the possible usable codes are as follows. Evidently, as with all other cancers, if the cancer registry only has a record of the case by death certificate, the code to be used is "0" (Death certificate only).

- Histology (Biopsy or surgical resection or autopsy specimen) 7
- Cytology only (urine) 5
- Only imaging or cystoscopy without biopsy or autopsy without a tissue diagnosis 2
- Death certificate only 0

In case of doubt, see the ENCR Recommendations on Basis of Diagnosis. It should be noted that the current recommendations on the basis of diagnosis were distributed in 1999, and it is likely that new ENCR recommendations on this subject will be published soon. Finally, if in the future the diagnostic methods for urothelial tumors are expanded and,

TABLE 1 Summary table of main criteria of inclusion (according to invasion, grade and existence of progression).

STEPS of PROGRESSION				
1. Non-invasive low grade/grade unknown	**2. Non-invasive high-grade or invasion cannot be assessed**	**3. *In situ***	**4. Invasive (T1)**	**5. Invasive (T2+)**
8130/2 G1 or 8130/2 G9	**8130/2 G3 or 8120 or 8130/2 G3 or 8120/2 G3**	**8120/2 G3**	**8010/3 G3 or 8120/2 G3 or 8000/3 G9**	**8010/3 G3 or 8120/3 G3 or 8000/3 G9**
Non-invasive Papillary Carcinoma, Low Grade or Non-invasive Papillary Carcinoma, Grade unknown	Non-invasive Papillary Carcinoma, High Grade or High grade urothelial carcinoma on cytology or Suspicious for high grade urothelial carcinoma on cytology	Urothelial Carcinoma *In situ* or Urothelial carcinoma with histologic examination but invasion cannot be assessed	Invasive carcinoma NOS or Invasive urothelial carcinoma or No microscopic confirmation: Tumour clinically malignant	Invasive carcinoma NOS or Invasive urothelial carcinoma or No microscopic confirmation: Tumour clinically malignant

consequently, these codes are modified or expanded, it is recommended to follow the modifications that may be defined by the ENCR.

Coding stage

Stage at diagnosis is one of the most important prognostic factors for the vast majority of tumors and this is also true for urothelial tumors (39). For this reason, survival analyses should be performed not only by sex and age but also by stage in order to distinguish whether differences in survival over time or between populations are due to a different distribution of cases by stage or to differences in cancer care.

In urothelial carcinomas, it is important to distinguish tumor invasion of the smaller, discontinuous, slender smooth muscle fibers of the muscularis mucosae (T1) from invasion of the larger, compact bundles of the muscularis propria (40, 41) and, as already commented, pathologists can have difficulty recognizing focal, superficial invasion of the lamina propria and differentiating invasion of the muscularis propria from invasion of the muscularis mucosae -ie, stage T1 from T2, which has immense implications for patient care (18).

On the other hand, although the combination of morphological, behavioral and grade codes are sufficient to distinguish between carcinomas *in situ* (CIS) and noninvasive papillary carcinomas, recording the T category of noninvasive tumors (pTis or pTa) in cancer registries validates the correctness of the data.

"Record "TNM-stage" (1, 4) whenever possible and, at least the "T-category".

This is important to allow Tis tumors to be easily distinguished from other tumors with behavior/2."

Recommendations for reporting urothelial tumors

Due to the complexity of urothelial tumors, these Recommendations are mostly devoted to recording criteria (registration, coding and classification). However, recommendations for the reporting of these tumors are also important for data comparability and are discussed below.

In order to follow the IARC/IACR/ENCR "International Rules for Multiple Primary Cancers" for computing incidence, only the first urothelial tumor regardless of the behavioral code (/2 or/3) should be counted. This will ensure incidence comparability between registries

The most important fact to note is that following the recommendations for recording provides the raw data that can be analyzed later. By doing so, data from cancer registry databases can be used to perform multiple analyses as part of local cancer surveillance and service assessment or can be transmitted for National, European or International projects.

The objectives of international projects can be very varied, so the "data call protocol" from international projects should define very accurately the criteria for inclusion of the data to be submitted and should also explain in detail how the data will be analyzed for incidence and survival estimations.

The following two examples show how the objectives and, consequently, the use of data for analysis can vary:

1. Counting the incidence of urinary bladder cancer: will a patient's first urothelial tumor be counted regardless of whether it is invasive or non-invasive, or will only invasive urothelial tumors be counted? Will non-urothelial bladder tumors also be included in the calculation?

2. Urinary bladder cancer survival computation: will the first tumor from any patient regardless of her behavior be included in the analysis or will only invasive tumors be considered?

Table 2 summarizes the general principles and criteria for inclusion (registration), and list of coding issues in the 2022 ENCR recommendations on urothelial tumours.

In conclusion, due to the great variability in the criteria for registration, coding and reporting of urothelial tumors among the different cancer registries, it is very difficult to determine the quantitative impact on incidence and survival rates of these new "ENCR Recommendations on the registration and reporting of urothelial tumors of the urinary tract". This would only be possible with the performance of a prospective study performed by all cancer registries applying both the old and the new criteria and evaluating the differences. These ENCR Recommendations imply a higher workload for the registry teams but will provide the framework to ensure comparability of outcomes for this tumor type across cancer registries in Europe and to enable a broader spectrum of analysis of incidence, survival and prevalence data for urothelial tumors. In the medium-term, an evaluation to review if the updated recommendations had any impact on the incidence and the quality of registered of urothelial tumors by registries would be desirable.

TABLE 2 Summary of the general principles and criteria of inclusion (registration), and list of coding issues in the 2022 ENCR recommendations on urothelial tumours.

General principles
- These recommendations should be applied to the pure urothelial carcinomas, to urothelial carcinomas with divergent differentiation, and to all other variants.
- Do everything possible to have access to pathological examinations (reports)
- Synchronous urothelial tumors are considered to be all those in the same site that appear in a period of less than or equal to 4 months.
Recommendations for recording urothelial tumors
Criteria for inclusion
- Types of tumors to be included: all invasive and non-invasive urothelial carcinomas including those without histological confirmation.
- Record all synchronous or metachronous urothelial tumors in different sites and laterality.
- Record progressions of the same three-digit sites.
- Do not record recurrences.
- In case of synchronous tumors of the same three-digit site and laterality, record only the most aggressive one
- In case of some synchronous bladder tumors of different subsite with the same level of progression, code as C67.8
- In case of some synchronous urothelial tumors of different three-digit site, code as C67.8.
- Record all the tumors of each side of each three digit site following the previous rules.
- If a patient has been previously diagnosed with an urothelial tumor(s) when resident outside the registration area, record all of them (the ones occurring outside the area of registration and the ones diagnosed being resident in the area of the registry) according to previous rules.
- If a patient has been diagnosed with one or more urothelial tumor(s) before the operation period of the registry, record all their tumors (the ones diagnosed before and the one diagnosed after first date of operation of the registry) according to previous rules.
Coding
- Code according to the most recent version of the International Classification of Diseases for Oncology
- Codes of the most frequent morphological categories when histology is available
- Codes when only cytological examination is available
- Codes when only non-microscopic confirmation is available
- Codes of behavior for unknown level of invasion
- Coding of Grade
- Codes for urothelial carcinomas with other morphological terms
• Urothelial carcinoma with squamous cell divergent differentiation
• Urothelial carcinoma with an adenocarcinomatous component
• Urothelial cell carcinoma subtypes and ICD-O-3 specific code
• Urothelial cell carcinoma without specific subtype in ICD-O-3 classification
- Non-urothelial specific carcinomas
- Coding the Basis of Diagnosis
- Coding stage
Recommendations for reporting urothelial tumors
- The Recommendations for recording provide the raw data which can be subsequently analysed.
- Follow IARC/IACR rules to calculate incidence (according to the "International Rules for Multiple Primary Cancers")
- At the local level, analyze the data recorded and coded with the new Recommendations according to the defined objectives.
- In international projects, define very precisely the inclusion criteria for the data to be submitted and explain in detail how the data will be analyzed for incidence and survival estimates.

Finally, the rules we propose for the registration of urothelial tumors, which constitute the most complex example of multiple, recurrent or progressive tumors, could be extended to tumors of other locations that may present these characteristics, in particular tumors for which screening programs exist, such as breast or colon-rectal tumors.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

JG, LM, DP, ME, RT, AW, CM and LN contributed to the development of the ENCR Recommendations. OV critically reviewed the ENCR Recommendations. The first draft of the manuscript was written by JG, LM and DP. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Pheby D, Martínez C, Roumagnac M, Schouten LENCR Working Group. *ENCR recommendations for coding bladder cancers* (1995). ENCR. Available at: https://www.encr.eu/sites/default/files/pdf/bladeng.pdf (Accessed July 15, 2022).

2. Moch H, Humphrey PA, Ulbright TM, Reuter VE. *WHO classification of tumors of the urinary system and Male genital organs. 4th ed. world health organization classification of tumors*. Lyon, France: IARC Press (2016).

3. Tsuzuki T, Compérat EM, Netto GJ, Raspollini PM, Moch H, Srigley JR, et al. Tumours of the urinary tract, in: *WHO classification of tumours editorial board. urinary and male genital tumours* (2022). Lyon (France: International Agency for Research on Cancer. Available at: https://tumourclassification.iarc.who.int/ (Accessed July 15, 2022).

4. Ferlay J, Kraywinkel K, Rous B, Znaor A. Chapter 3: Classification and coding, in: *Cancer incidence in five continents, vol. XI (electronic version)* (2017). Lyon: International Agency for Research on Cancer. Available at: https://ci5.iarc.fr (Accessed July 15, 2022).

5. Daubisse-Marliac L, Grosclaude P, Carulla M, Parada D, Vilardell L, Ameijide A, et al. Registration of urothelial tumors in cancer registries: How to improve and make it more useful? *Int J Environ Res Public Health* (2022) 19:2714. doi: 10.3390/ijerph19052714

6. Eden M, Daubisse-Marliac L, Galceran J, Martos C, Neamtiu L, Parada D, et al. *ENCR recommendations. recording and reporting of urothelial tumors of the urinary tract. European network of cancer registries* (2022). Available at: https://www.encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation_UT_Jun2022_EN.pdf (Accessed July 15, 2022).

7. Audenet F, Attalla K, Sfakianos JP. The evolution of bladder cancer genomics: What have we learned and how can we use it? *Urol Oncol* (2018) 36:313–20. doi: 10.1016/j.urolonc.2018.02.017

8. Park S, Reuter VE, Hansel DE. Non-urothelial carcinomas of the bladder. *Histopathology* (2019) 74:97–111. doi: 10.1111/his.13719

9. Compérat E, Oszwald A, Wasinger G, Hansel DE, Montironi R, van der Kwast T, et al. Updated pathology reporting standards for bladder cancer: biopsies, transurethral resections and radical cystectomies. *World J Urol* (2022) 40:915–27. doi: 10.1007/s00345-021-03831-1

10. International Agency for Research on Cancer, World Health Organization, International Association of Cancer Registries and European Network of Cancer Registries. *International rules for multiple primary cancers (ICD-O third edition). internal report no. 2004/02* (2004). Lyon: International Agency for Research on Cancer. Available at: http://www.iacr.com.fr/images/doc/MPrules_july2004.pdf (Accessed July 15, 2022).

11. Cambier S, Sylvester RJ, Collette L, Gontero P, Brausi MA, van Andel G, et al. EORTC nomograms and risk groups for predicting recurrence, progression, and disease-specific and overall survival in non-muscle-invasive stage Ta-T1 urothelial bladder cancer patients treated with 1-3 years of maintenance bacillus calmette-guérin. *Eur Urol* (2016) 69:60–9. 10.1016/j.eururo.2015.06.045

12. van Rhijn BWG, Burger M, Lotan Y, Solsona E, Stief CG, Sylvester RJ, et al. Recurrence and progression of disease in non-muscle-invasive bladder cancer: from epidemiology to treatment strategy. *Eur Urol* (2009) 56:430–42. doi: 10.1016/j.eururo.2009.06.028

13. Guo C, Czerniak B. Bladder cancer in the genomic era. *Arch Pathol Lab Med* (2019) 143:695–704. doi: 10.5858/arpa.2018-0329-RA

14. Farrow GM, Utz DC, Rife CC. Morphological and clinical observations of patients with early bladder cancer treated with total cystectomy. *Cancer Res* (1976) 36:2495–501.

15. World Health Organization. *International classification of diseases for oncology, third edition, first revision*. Geneva: World Health Organization (2013).

16. *ICD-O-3.2* (2021). Available at: http://www.iacr.com.fr/index.php?option=com_content&view=category&layout=blog&id=100&Itemid=577 (Accessed July 15, 2022).

17. Truong M, Liang L, Kukrej J, O'Brien J, Jean-Gilles J, Messing E. Cautery artifact understages urothelial cancer at initial transurethral resection of large bladder tumours. *Can Urol Assoc J* (2017) 11:E203–6. doi: 10.5489/cuaj.4172

18. Paner GP, Brown JG, Lapetino S, Nese N, Gupta R, Shen SS, et al. Diagnostic use of antibody to smoothelin in the recognition of muscularis propria in transurethral resection of urinary bladder tumor (TURBT) specimens. *Am J Surg Pathol* (2010) 34:792–9. doi: 10.1097/PAS.0b013e3181da7650

19. Barkan GA, Wojcik EM, Nayar R, Savic-Prince S, Quek ML, Kurtycz DF, et al. The Paris system for reporting urinary cytology: The quest to develop a standardized terminology. *Adv Anat Pathol* (2016) 23:193–201. doi: 10.1097/PAP.0000000000000118

20. VandenBussche CJ. A review of the Paris system for reporting urinary cytology. *Cytopathology* (2016) 27:153–6. doi: 10.1111/cyt.12345

21. Rosenthal DL, Wojcik EM, Kurtycz DFI eds. *The Paris system for reporting urinary cytology*. Switzerland: Springer International Publishing (2016).

22. Owens CL, VandenBussche CJ, Burroughs FH, Rosenthal DL. A review of reporting systems and terminology for urine cytology. *Cancer Cytopathol* (2013) 121:9–14. doi: 10.1002/cncy.21253

23. Compérat E, Amin M, Reuter V. Reply re: Murali varma, Brett delahunt, theodorus van der kwast. grading noninvasive bladder cancer: World health organisation 1973 or 2004 may be the wrong question. two decades of world health organisation/international society of urological pathology bladder cancer grading: time to reflect on accomplishments and plan refinement in the molecular era, not regress to readoption of a 45-year-old classification. *Eur Urol* (2019) 76:416–7. doi: 10.1016/j.eururo.2019.07.017

24. Soukup V, Čapoun O, Cohen D, Hernández V, Babjuk M, Burger M, et al. Prognostic performance and reproducibility of the 1973 and 2004/2016 world health organization grading classification systems in non-muscle-invasive bladder cancer: a European association of urology non-muscle invasive bladder cancer guidelines panel systematic review. *Eur Urol* (2017) 72:801–13. doi: 10.1016/j.eururo.2017.04.015

25. Nishiyama N, Kitamura H, Maeda T, Takahashi S, Masumori N, Hasegawa T, et al. Clinicopathological analysis of patients with non-muscle-invasive bladder cancer: prognostic value and clinical reliability of the 2004 WHO classification system. *Jpn J Clin Oncol* (2013) 43:1124–31. doi: 10.1093/jjco/hyt120

26. Amin MB, Smith SC, Reuter VE, Epstein JI, Grignon DJ, Hansel DE, et al. Update for the practicing pathologist: The international consultation on urologic disease-European association of urology consultation on bladder cancer. *Mod Pathol* (2015) 28:612–30. doi: 10.1038/modpathol.2014.158

27. Lopez-Beltran A, Henriques V, Montironi R, Cimadamore A, Raspollini MR, Cheng L. Variants and new entities of bladder cancer. *Histopathology* (2019) 74:77–96. doi: 10.1111/his.13752

28. McKenney JK. Precursor lesions of the urinary bladder. *Histopathology* (2019) 74:68–76. doi: 10.1111/his.13762

29. Grace DA, Winterm CC. Mixed differentiation of primary carcinoma of the urinary bladder. *Cancer* (1968) 21:1239–43. doi: 10.1002/1097-0142(196806)21:6<1239::AID-CNCR2820210627>3.0.CO;2-E

30. Wasco MJ, Daignault S, Zhang Y, Kunju LP, Kinnaman M, Braun T, et al. Urothelial carcinoma with divergent histologic differentiation (mixed histologic features) predicts the presence of locally advanced bladder cancer when detected at transurethral resection. *Urology* (2007) 70:69–74. doi: 10.1016/j.urology.2007.03.033

31. Veskimäe E, Espinos EL, Bruins HM, Yuan Y, Sylvester R, Kamat AM, et al. What is the prognostic and clinical importance of urothelial and nonurothelial histological variants of bladder cancer in predicting oncological outcomes in patients with muscle-invasive and metastatic bladder cancer? a European association of urology muscle invasive and metastatic bladder cancer guidelines panel systematic review. *Eur Urol Oncol* (2019) 2:625–42. doi: 10.1016/j.euo.2019.09.003

32. Kamoun A, de Reyniès A, Allory Y, Sjödahl G, Robertson AG, Seiler R, et al. A consensus molecular classification of muscle-invasive bladder cancer. *Eur Urol* (2020) 77:420–33. doi: 10.1016/j.eururo.2019.09.006

33. Abrahams NA, Moran C, Reyes AO, Siefker-Radtke A, Ayala AG. Small cell carcinoma of the bladder: A contemporary clinicopathological study of 51 cases. *Histopathology* (2005) 46:57–63. doi: 10.1111/j.1365-2559.2004.01980.x

34. Trias I, Algaba F, Condom E, Español I, Seguí J, Orsola I, et al. Small cell carcinoma of the urinary bladder. presentation of 23 cases and review of 134 published cases. *Eur Urol* (2001) 39:85–90. doi: 10.1159/000052417

35. Wang X, MacLennan GT, Lopez-Beltran A, Cheng L. Small cell carcinoma of the urinary bladder-histogenesis, genetics, diagnosis, biomarkers, treatment and prognosis. *Appl Immunohistochem Mol Morphol* (2007) 15:8–18. doi: 10.1097/01.pai.0000213106.12731.d7

36. Grignon DJ, Ro JY, Ayala AG, Shum DT, Ordoñez NG, Logothetis CJ, et al. Small cell carcinoma of the urinary bladder. a clinicopathologic analysis of 22 cases. *Cancer* (1992) 69:527–36. doi: 10.1002/1097-0142(19920115)69:2<527::AID-CNCR2820690241>3.0.CO;2-7

37. Varma M, Srigley JR, Brimo F, Compérat E, Delahunt B, Koch M, et al. Dataset for the reporting of urinary tract carcinoma-biopsy and transurethral resection specimen: recommendations from the international collaboration on cancer reporting (ICCR). *Mod Pathol* (2019) 33:700–12. doi: 10.1038/s41379-019-0403-9

38. Humphrey PA, Moch H, Cubilla AL, Ulbrigth TM, Reuter VE. The 2016 WHO classification of tumors of the urinary system and Male genital organs–part b: Prostate and bladder tumors. *Eur Urol* (2016) 70:106–19. doi: 10.1016/j.eururo.2016.02.028

39. Magers MJ, Lopez-Beltran A, Montironi R, Williamson SR, Kaimakliotis HZ, Cheng, et al. Staging of bladder cancer. *Histopathology* (2019) 74:112–34. doi: 10.1111/his.13734

40. Paner GP, Ro JY, Wojcik EM, Venkataraman G, Datta MW, Amin MB. Further characterization of the muscle layers and lamina propria of the urinary bladder by systematic histologic mapping: implications for pathologic staging of invasive urothelial carcinoma. *Am J Surg Pathol* (2007) 31:1420–9. doi: 10.1097/PAS.0b013e3180588283

41. Ro JY, Ayala AG, el-Naggar A. Muscularis mucosa of urinary bladder. importance for staging and treatment. *Am J Surg Pathol* (1987) 11:668–73. doi: 10.1097/00000478-198709000-00002

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 08 February 2023
DOI 10.3389/fonc.2023.1109978

OPEN ACCESS

EDITED BY
Jerry Polesel,
Aviano Oncology Reference Center (IRCCS), Italy

REVIEWED BY
Stefano Guzzinati,
Azienda Zero, Italy
Maria Paula Curado,
A. C. Camargo Cancer Center, Brazil

*CORRESPONDENCE
Francesco Giusti
Francesco.Giusti@registreducancer.org;
francescogiusti@hotmail.com

†PRESENT ADDRESS
Francesco Giusti,
Belgian Cancer Registry, Brussels, Belgium
Carmen Martos,
Foundation for the Promotion of Health and Biomedical Research in the Valencian Region (FISABIO), Valencia, Spain

SPECIALTY SECTION
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

RECEIVED 28 November 2022
ACCEPTED 20 January 2023
PUBLISHED 08 February 2023

CITATION
Giusti F, Martos C, Trama A, Bettio M, Sanvisens A, Audisio R, Arndt V, Francisci S, Dochez C, Ribes J, Fernández LP, Gavin A, Gatta G, Marcos-Gragera R, Lievens Y, Allemani C, De Angelis R, Visser O, Van Eycken L and the ENCR Working Group on Treatment Data Harmonisation (2023) Cancer treatment data available in European cancer registries: Where are we and where are we going?
Front. Oncol. 13:1109978.
doi: 10.3389/fonc.2023.1109978

Cancer treatment data available in European cancer registries: Where are we and where are we going?

Francesco Giusti[1,2*†], Carmen Martos[1,3†], Annalisa Trama[4], Manola Bettio[1], Arantza Sanvisens[5], Riccardo Audisio[6], Volker Arndt[7], Silvia Francisci[8], Carine Dochez[2], Josepa Ribes[9], Laura Pareja Fernández[9], Anna Gavin[10], Gemma Gatta[4], Rafael Marcos-Gragera[5], Yolande Lievens[11], Claudia Allemani[12], Roberta De Angelis[13], Otto Visser[14], Liesbet Van Eycken[2] and the ENCR Working Group on Treatment Data Harmonisation

[1]European Commission, Joint Research Centre (JRC), Ispra, Italy, [2]Belgian Cancer Registry, Brussels, Belgium, [3]Foundation for the Promotion of Health and Biomedical Research in the Valencian Region (FISABIO), Valencia, Spain, [4]Evaluative Epidemiology Unit, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, [5]Epidemiology Unit and Girona Cancer Registry, Catalan Institute of Oncology, Oncology Coordination Plan, Department of Health, Autonomous Government of Catalonia; Girona Biomedical Research Institute (IDIBGI), Girona, Spain, [6]Department of Surgery, Institute of Clinical Sciences, Sahlgrenska University Hospital, Göteborg, Sweden, [7]Epidemiological Cancer Registry Baden-Württemberg (M110) & Unit of Cancer Survivorship (C071), Division of Clinical Epidemiology and Aging Research (C070), German Cancer Research Center (DKFZ), Heidelberg, Germany, [8]National Centre for Disease Prevention and Health Promotion, Istituto Superiore di Sanità, Rome, Italy, [9]Catalan Cancer Plan, Department of Health of Catalonia, Hospitalet del Llobregat, Barcelona, Spain, [10]Northern Ireland Cancer Registry, Centre for Public Health, Queen's University Belfast, Belfast, Ireland, [11]Department of Radiation Oncology, Ghent University Hospital and Ghent University, Ghent, Belgium, [12]Cancer Survival Group, London School of Hygiene and Tropical Medicine, London, United Kingdom, [13]Department of Oncology and Molecular Medicine, Istituto Superiore di Sanità, Rome, Italy, [14]Department of Registration, Netherlands Comprehensive Cancer Organisation (IKNL), Utrecht, Netherlands

Population-based cancer registries are responsible for collecting incidence and survival data on all reportable neoplasms within a defined geographical area. During the last decades, the role of cancer registries has evolved beyond monitoring epidemiological indicators, as they are expanding their activities to studies on cancer aetiology, prevention, and quality of care. This expansion relies also on the collection of additional clinical data, such as stage at diagnosis and cancer treatment. While the collection of data on stage, according to international reference classification, is consolidated almost everywhere, data collection on treatment is still very heterogeneous in Europe. This article combines data from a literature review and conference proceedings together with data from 125 European cancer registries contributing to the 2015 ENCR-JRC data call to provide an overview of the status of using and reporting treatment data in population-based cancer registries. The literature review shows that there is an increase in published data on cancer treatment by population-based cancer registries over the years. In addition, the review indicates that treatment data are most often collected for breast cancer, the most frequent cancer in women in Europe, followed by colorectal, prostate and lung cancers, which are also more

common. Treatment data are increasingly being reported by cancer registries, though further improvements are required to ensure their complete and harmonised collection. Sufficient financial and human resources are needed to collect and analyse treatment data. Clear registration guidelines are to be made available to increase the availability of real-world treatment data in a harmonised way across Europe.

KEYWORDS

cancer registry, data harmonisation, questionnaire, big data, Europe, cancer registry data, cancer treament

1 Introduction

Among non-communicable diseases, cancer remains one of the most important causes of death worldwide. In 2020, 4 million new cases were estimated to be reported in Europe, with around 1.9 million deaths (1). Although improvements in cancer survival over time are being observed, wide variations between European countries still persist (2–4).

Population-based cancer registries (CRs) are responsible for collecting high-quality population-based incidence and survival data on all reportable neoplasms within a defined catchment-area. Starting from the 1940s, population-based CRs have been operational in an increasing number of European countries, adhering to international standards set by the International Association of Cancer Registries (IACR), in collaboration with the International Agency for Research on Cancer (IARC) (5–8).

Following the European Commission's 1985 "Europe Against Cancer" Programme, the European Network of Cancer Registries (ENCR) has been operating since 1990 to strengthen the collaboration among CRs, aiming to improve the quality, comparability and availability of cancer incidence data; to provide information on and to monitor cancer incidence and mortality in Europe; and to encourage the use of CRs data in cancer control, health-care planning and research. Since 2012, the ENCR Secretariat has been hosted in Ispra, Italy, by the Directorate-General Joint Research Centre (JRC), the science and knowledge centre of the European Commission. The JRC supports the ENCR with the dissemination and harmonisation of cancer data, with the overall aim of accurately comparing data between European countries. CRs can be members of the ENCR if they are based in countries within the United Nations geographical definition of Europe, plus Cyprus. Currently, nearly 200 population-based CRs are active in Europe, of which 189 are full members and 4 are associate members of ENCR (9, 10). Finally, the JRC has been developing, maintaining and expanding the European Cancer Information System (ECIS) as the infrastructure hosting, processing and disseminating European CR data (1). Harmonised cancer burden indicators across European areas computed from CR data are released in the ECIS web application (11).

During the last decades, the role of CRs has evolved beyond providing cancer incidence and survival data. Depending on available resources, CRs are now becoming more involved in different areas of cancer control, including aetiology of cancer, evaluation of screening programmes, and monitoring quality and outcomes of cancer care and trends in cancer survival (12). In Europe, data collection on cancer treatment modalities (surgery, radiotherapy, chemotherapy, etc…) is very heterogeneous. Several CRs are collecting cancer treatment related data on a continuous or regular basis, while other CRs collect them on an *ad hoc* basis or only for specific projects. Some CRs collect treatment data for all tumours, others only for specific tumours. Data can be collected from medical records and administrative medical claims (such as hospital discharge records and drug prescriptions) (13).

Treatment data collected by CRs allows for the: (1) Monitoring of treatment patterns; (2) Assessment of the compliance with clinical practice guidelines; (3) Evaluation of the impact of new treatments at population level; and (4) Evaluation of access to treatment. Recommended treatment for a specific cancer strongly depends on its stage at diagnosis, as specific treatment modalities and strategies are indicated for selected stages only. The availability of data on stage is therefore a prerequisite for the use and proper interpretation of treatment data collected by CRs (13–15).

Only two previous projects (EUROCHIP-3 and EUROCOURSE) provided an overview on the availability of three main indicators in European population-based CRs: stage at diagnosis, cancer treatment delay and compliance with cancer guidelines. While overall treatment data collection was rather low (30% of CRs), an increase in data collection has been observed (43% of CRs) over time between the two projects (4, 5, 16).

In addition to stage, biomarkers have been playing an important role in guiding treatment options and in the prognosis of several tumour types such as breast, oropharyngeal and lung cancer. Although a constant increase in the number of publications on biomarkers from CRs has been observed in recent years, there is still the need of an harmonisation of such data, and possibly an increased interaction with clinicians and hospital-based registries (17, 18)

Since information about availability and comparability of treatment data is lacking, this article aims to give an overview of the current registration status for cancer treatment data among population-based CRs in Europe. The outcome of the study represents a basis for drafting recommendations to CRs either to initiate data treatment collection or to continue and improve

treatment data collection, coding and reporting to assure data comparability among European CRs.

2 Methods

To explore the current situation of cancer treatment registration in Europe, a literature search was conducted, including both peer-reviewed articles and mainly cancer registration-related conference proceedings. In addition, treatment data collected in the framework of the 2015 ENCR-JRC data call were explored, and are here summarised.

2.1 Literature review

2.1.1 Peer-reviewed literature

A literature review was performed on Pubmed to identify peer-reviewed publications mentioning treatment data from CRs in the title and/or abstract. The first selection was done with keywords "*cancer registry*", "*cancer registries*", "*tumor registry*", "*tumor registries*", "*tumour registry*", "*tumour registries*", "*oncological registry*", "*oncology registry*", together with "*treatment*", "*surgery*", "*radiotherapy*", "*chemotherapy*", "*therapy*". Keywords with the English language names of all European countries were applied. There was no specific starting period selected, while the end period was set at October 2, 2022.

The results of the Pubmed search were imported in the Rayyan literature review web-tool for further screening (19). As a final step, the results were imported in the statistical software SAS Version 9.3 (SAS Institute Inc., Cary, NC, USA) in order to perform string searches through the *PRXMATCH* Function. A specific search string was also used to look at age groups reported in the publications.

Articles were excluded when the registry was located outside Europe, was not population-based, did not include treatment data, and when the study was not about CR data (e.g. clinical trials).

A consistency check of the selection criteria was independently performed on a sample of 100 articles on which agreement was reached on 97 out of 100 articles. After discussion on the remaining 3 articles, the resulting criteria were applied to the search algorithm.

Articles from population-based CRs in European countries (plus Cyprus) reporting on treatment data were included in the analysis. Articles from CRs operating in more than one country were also included. Figure 1 describes the flowchart of the included articles.

2.1.2 Conference proceedings

The title and the abstract of presentations given during the period 2016-2019 in scientific international CRs conferences [European Network of Cancer Registries (ENCR), Group for Cancer Epidemiology and Registration in Latin Language Countries (GRELL), the International Association of Cancer Registries (IACR)] plus the European Society for Medical Oncology (ESMO) were also screened.

2.2 The 2015 ENCR-JRC data call

In 2015 a first ENCR-JRC data call was launched by the ENCR Steering Committee and JRC, as the source for data feeding the ECIS (1). Outputs at registry level reported in the ECIS web application include incidence and mortality by cancer entity, sex, age group and geographical area. Besides this information, the protocol for the 2015 ENCR-JRC data call investigated also on cancer treatment data (20). Additional information could be retrieved by answers to dedicated questions of the questionnaire accompanying the data submission.

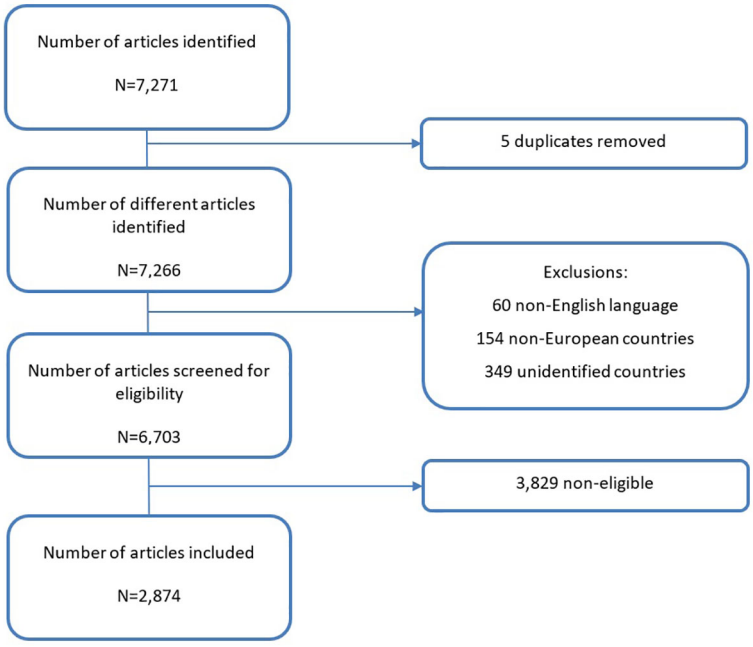


FIGURE 1
Flowchart of the identification, screening and eligibility of articles included in the literature review.

General (all-sites) and childhood CRs contributing to the ECIS and having answered to the 2015 ENCR-JRC data call questionnaire were included in the current analysis. Site-specific registries, and regional registries overlapping with a national CR were excluded.

2.2.1 The 2015 ENCR-JRC data call questionnaire

Filling the accompanying questionnaire was an essential requirement to complete the data submission. The questionnaire comprised 4 sections (Cancer case file; Population data; Mortality data and Life tables), and included in section 1 on the Cancer case file the following questions related to the registration of treatment data:

- 1.21 Do you record information about treatment in the registry?
- 1.21.1 Please, provide a description of the variables, if they are different than those in the protocol:
- 1.21.2 Please specify the sources of data on treatment:

The questionnaire was sent through the EUSurvey, the online survey management tool of the European Commission (21). Data from the submitted questionnaires were stored and analysed with Microsoft Excel.

2.2.2 The 2015 ENCR-JRC data call: Treatment data reported by the CRs

The protocol of the 2015 ENCR-JRC data call included 4 variables investigating the first course of cancer therapy after diagnosis by using the following variables:

- Surgery (including any surgery to remove all or part of the cancer. Biopsy which is followed by definitive surgery was not to be included; other biopsies, where the cancer was completely excised, could be included);
- Systemic cancer therapy, including chemotherapy, targeted therapy, immunotherapy and hormone therapy;
- Radiotherapy;
- Bone marrow transplantation.

All variables were recorded with yes, no or unknown.

Data were submitted by CRs to the JRC through the ENCR-JRC Portal, checked for consistency and harmonised by JRC, using the JRC-ENCR Quality Check Software (QCS), Stata and SAS statistical software (22, 23).

A quality evaluation was performed on the four most common cancer entities (1): breast, colorectal, prostate and lung. The percentage of cancer cases with surgery was calculated by CR for each site, and compared with data previously observed in studies from CRs reporting treatment patterns.

3 Results

3.1 Literature review

3.1.1 Peer-reviewed literature

A total of 2,874 articles out of 7,271 returned by the search (from year 1975 to October 2022) were included in the analysis.

The majority of papers with treatment data information came from five countries: Netherlands (632 articles - 22% of the total), Sweden (290 - 10%), United Kingdom (225 - 8%), Germany (197 - 7%) and Norway (188 - 7%) for a total of 1532 articles (53% of the total). In addition, registries operating in another 23 countries authored 912 publications (32%).

A total of 430 publications were international (15%), with data from at least two European countries (Figure 2). The latest ranged from large international studies such as the European Cancer Registry based study on survival and care of cancer patients (EUROCARE), the Surveillance of Rare Cancers in Europe (RARECARE/RARECAREnet), the CONCORD programme for the global surveillance of cancer survival and the European Registration of Cancer Care (EURECCA), to collaborations between CRs from as little as two different European countries.

Since many CRs started operating later than others in the period of interest (1975-2022), the analysis was also performed for the most recent period (2013-2022), for which the percentage contribution remained unchanged (Table 1).

The highest number of articles reporting information on treatment was related to breast cancer (442 articles - 15% of the total), followed by colorectal (413 - 14%), prostate (159 - 6%) and lung cancer (155 - 5%). Additional single cancer entities were addressed in 804 articles (28%), 603 articles (21%) reported on more than one cancer entity, whereas for 298 articles (10%) the search string could not find any specific cancer entity (Figure 3).

Out of the total number of articles reporting cancer treatment data, 385 (13% of the total) were published between 1975 and 2002. A steep increase in the number of articles was observed in subsequent five-years periods: 269 (9%) in 2003-2007, 408 (14%) in 2008-2012, 774 (27%) in 2013-2017 and 1,038 (36%) in the latest period (January 2018-October 2, 2022) (Figure 4).

The selected articles were published in 599 different Journals. The 10 Journals with the highest number of papers published altogether 796 articles (28%). Epidemiology, Oncology, Surgery, specific cancer

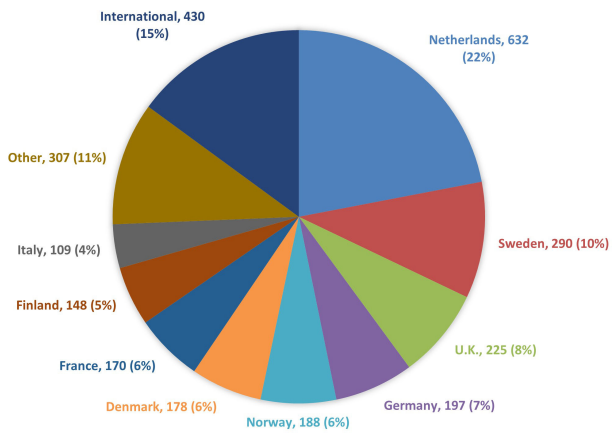


FIGURE 2
Number of articles on cancer treatment by CR country of operation, 1975-2022.

10.3389/fonc.2023.1109978

TABLE 1 Number of articles on cancer treatment by CR country of operation, January 2013-October 2022.

Country	Number of articles	Percentage
Netherlands	496	27.4
Sweden	182	10.0
Germany	129	7.1
Norway	117	6.5
U.K.	109	6.0
Denmark	91	5.0
France	83	4.6
Finland	82	4.5
Italy	54	3.0
Switzerland	39	2.2
Ireland	36	2.0
Spain	27	1.5
Belgium	18	1.0
Lithuania	15	0.8
Poland	15	0.8
Czech Republic	9	0.5
Iceland	8	0.4
Portugal	8	0.4
Slovenia	7	0.4
Hungary	6	0.3
Estonia	5	0.3
Austria	4	0.2
Croatia	4	0.2
Russia	3	0.2
Bulgaria	1	0.1
Ukraine	1	0.1
International	263	14.5
Total	1812	100.0

entities were the most common focus of the 599 Journals. In order to compare Journals with different periods of publication, number of articles was checked for period 2013-2022, which was already covered by the majority of Journals (Table 2).

A specific search string was used to look at age groups reported in the treatment publications. The majority, 2,043 (71%), was not focused on a specific age group (in Figure 5, this group is shown as "All ages"). This group was in fact mainly composed of studies reporting only on adults, although this was not specifically investigated by the search string. Out of the remaining publications, 595 (21%) reported data on elderly populations with 70 years as a common threshold. Childhood populations were addressed in 201 articles (7%), and a further 35 articles (1%) reported data from both elderly and children.

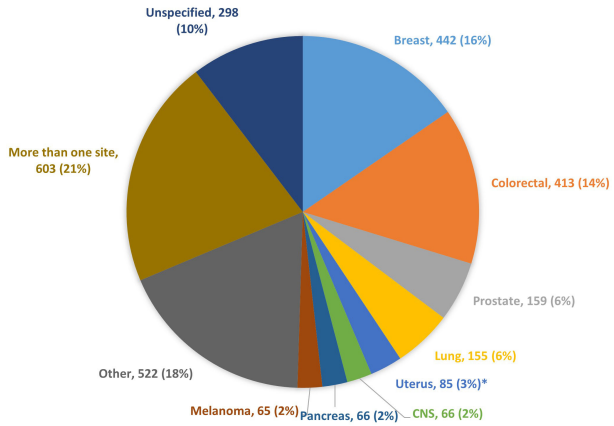


FIGURE 3
Number of articles reporting CRs information on cancer treatment, by cancer entity. * Endometrial and cervical cancers.

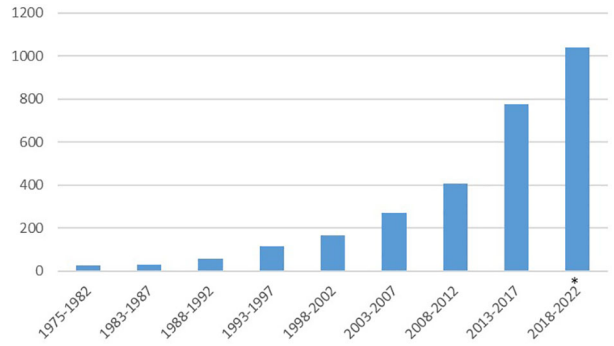


FIGURE 4
Number of articles on cancer treatment by year of publication. * Up to October 2, 2022.

3.1.2 Conference proceedings

The results of the overview from the following CRs scientific meetings, having taken place between 2016 and 2019 are presented: ENCR 2016 and 2018, GRELL 2016-2019, IACR (restricted to European contributions) 2016-2019. Presentations on treatment data given by European CRs at the European Society for Medical Oncology (ESMO) Congresses 2016-2018 were also included, for a total of 213 studies (7, 24–26).

Out of 495 oral and poster presentations given at the (mainly European) GRELL and ENCR conferences in 2016-19, 132 (27%) were related to treatment.

Out of 135 CRs presentations on treatment at ENCR, IACR and ESMO (GRELL not being considered for this specific evaluation as it is only related to Latin language countries), 26 (19%) were from the Netherlands, 19 (14%) from the U.K., 11 (8%) from Belgium, 9 (7%) each from Spanish and Italian CRs. Sixteen (12%) presentations were from international studies, for the majority high-resolution ones.

Thirty-nine (18%) out of the 213 ENCR, GRELL, IACR and ESMO considered presentations were on breast cancer, 24 (11%) on colorectal cancer, 16 (8%) on lung cancer, 11 (5%) on pancreatic

10.3389/fonc.2023.1109978

TABLE 2 Number of articles on cancer treatment published in the 10 most frequent Journals, 2013-2022.

Journal (starting year publication)	Number of articles	Percentage
Acta Oncologica (1963)	89	4.9
European Journal of Cancer (1965)	77	4.3
European Journal of Surgical Oncology (1975)	74	4.1
International Journal of Cancer (1966)	57	3.2
BMC cancer (2001)	51	2.8
Breast Cancer Research and Treatment (1981)	43	2.4
Cancer Epidemiology (1976)	39	2.2
British Journal of Surgery (1913)	35	1.9
Annals of Surgical Oncology (1994)	33	1.8
Colorectal disease (1999)	33	1.8

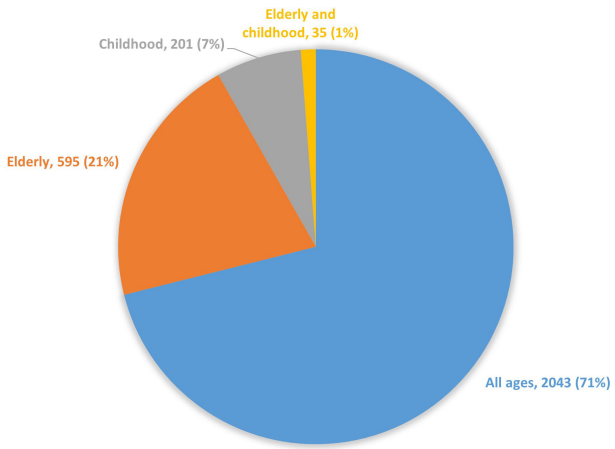


FIGURE 5
Number of articles on cancer treatment by age group, 1975-2022.

cancer, 10 (5%) on prostate cancer, whereas 13 (6%) took into account more than one cancer entity.

Seventy-five (35%) presentations focused on reporting treatment practice, without specific reference to guidelines, 30 (14%) on quality of care and adherence to guidelines, 26 (12%) on survival by type of treatment. Other topics addressed were the evaluation of recurrences, late effects of treatment, evaluation of new treatments at population level, new methodologies for gathering treatment data, quality of life, end-of-life care.

3.2 The 2015 ENCR-JRC data call

3.2.1 The 2015 ENCR-JRC data questionnaire

Overall, a total of 119 general (all ages and all cancer sites) and 6 specialised childhood CRs submitted data to feed the ECIS, and responded to the 2015 data call questionnaire. Eleven additional registries submitted data to the ECIS but did not fill in the questionnaire, thus not contributing to its evaluation. Out of the 125 population-based CRs included in the analysis, 21 were national CRs while 104 were regional ones, representing a total of 30 countries.

Out of the 125 CRs, 61 (49%) replied "*Yes*" to question "*1.21 Do you record information about treatment in the registry?*", while 64 CRs (51%) replied "*No*".

Specifically, 76% (15 out of 21) of the national general CRs reported recording treatment data, as compared to 41% (40 out of 98) of the regional CRs (Figure 6). In addition, all six childhood CRs reported dealing with treatment data.

Of the 61 CRs declaring to record treatment data, 59 specified the source of data on treatment. The most referenced sources were hospital discharge records (N = 23), clinical records (N = 23), both sources (N = 4) and notifications from physicians and hospitals (N = 6).

Forty-two out of the 61 CRs reporting treatment data provided additional information (question "*1.21.1 Please, provide a description of the variables, if they are different than those in the protocol:*"). Twelve registries reported that data are available but were not submitted or will only be made available for specific studies or upon request. Eleven registries commented that they had more data available than those requested in the data call, such as starting date of therapy, or additional clinical data for selected cancers and subgroups. Regarding the question about systemic treatment, CRs reported to be able to provide detailed data on the specific type of therapy, e.g. chemotherapy, hormone therapy, immunotherapy and targeted therapy. Five registries reported that they only record treatment data for specific cancers (on colorectal cancer (5 CRs), breast cancer (4 CRs), lung cancer (2 CRs), skin melanoma (1 CR) and lymphoma (1 CR)).

In addition, out of the 125 CRs included in the analysis, 98 (78%) registries reported to collect data on cancer stage. Ninety-one collected pathological or clinical TNM (Tumour/Nodes/Metastasis), 2 childhood CRs reported using specific childhood staging only, 3 CRs were only collecting summary extent of disease and 1 'condensed TNM', while 1 did not provide further information on staging (27, 28).

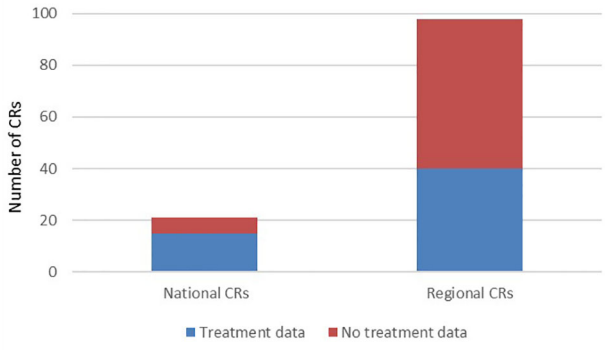


FIGURE 6
Number of general CRs reporting the recording of treatment data.

3.2.2 The 2015 ENCR-JRC data call: Information on cancer treatment

Overall 130 registries (124 general and 6 specialised on childhood cancers) registries contributed to the ECIS database, with a total of 34,610,818 individual cancer cases as of 16/10/2020. Out of them, 30 registries (22%) - 28 general and 2 childhood CRs- provided treatment data for all or part of the period of incidence (Figure 7).

The number of cases provided by the 28 general CRs submitting treatment data was 12,872,032 (37% of the total). From what reported in the data call questionnaire, additional 29 general registries (22% of the total) declared to record treatment information although they did not submit it, whereas two CRs submitted treatment data but did not reply to the questionnaire.

Seven registries out of 28 provided information on all 4 treatment types as defined in the protocol for data collection (surgery, systemic therapy, radiotherapy and bone marrow transplantation), 20 registries provided data on surgery, systemic therapy and radiotherapy, and one registry provided data on surgery only. As for the cancer entities with reported data on treatment, twenty-five out of 28 CRs submitted information on surgery for lung cancer, prostate, bladder, corpus uteri, melanoma, pancreas and other cancer entities. An additional registry submitted data also on surgery for colorectal cancer; all 28 registries submitted information on surgery for breast cancer.

Out of 1,491,881 breast cancer cases submitted by the 28 general registries in the period of data availability, 82% were treated with surgery (interquartile range 79%-88%). Out of 1,464,389 colorectal cancer cases provided by the 26 registries, 71% underwent surgery (interquartile range 70%-79%). For prostate, out of 1,033,071 cases from 25 CRs, 36% received surgical treatment (interquartile range 33%-51%). As for lung cancer, out of 1,388,712 cases from 25 registries, only 19% received surgery (interquartile range 15%-22%).

4 Discussion

To the best of our knowledge, this is the first analysis combining data from a literature review and conference proceedings, together with data reported by the European CRs in the 2015 ENCR-JRC dataset to get the evolution over time of the status of collecting and reporting cancer treatment data in population-based CRs. Our study highlighted that population based CRs collecting treatment data either (1) do not report the data; (2) report the data but do not publish in peer-reviewed Journals; and (3) report and publish the data.

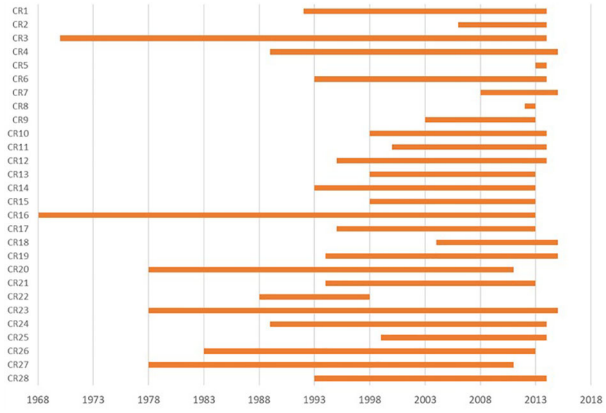


FIGURE 7
Availability of treatment information by incidence year and contributing CR in the ECIS database as of 16/10/2020.

The literature review shows that there is an increase in published data on cancer treatment by population-based CRs over the years. Most articles are from CRs operating in Western and Northern European countries, notably countries with either a national CR and/or with a long history of cancer registration. In particular, in Nordic countries an extensive record linkage between different national data sources is routinely performed by CRs, which allows detailed treatment data collection and reporting (29).

In addition to the increase of publications in peer-reviewed Journals, we also notice that the vast majority of publications were in specialised Journals for clinicians, surgeons, radiation oncologists, making this data more and more relevant, at least in some European countries. This growing interaction and collaboration of clinicians and CRs could signal an increasing benefit for both the epidemiological and clinical environment.

One reason of the scarcity of articles from some European areas might be partly due to the fact that treatment data are only reported by CRs in national or regional reports, often in their respective local languages (10). In addition, many CRs gather treatment data only for *ad hoc* projects, such as the EUROCARE, RARECAREnet and CONCORD high resolution studies or the EURECCA studies (30–32).

Moreover, limited resources in some European countries and regions could play a role in the difference in reporting and using treatment data among European CRs as well as less developed or absent national linkage/database structures (33).

As for the specialised childhood registries, given the much lower number of incident cases (e.g. 16.000 estimated in 2020, compared to 4 million for adults), registration and use of treatment data is more widespread than for the general CRs. The literature review showed that 8% of articles report data on treatment for the paediatric age groups, whereas childhood cancers represent only 0.4% of total incidence.

The literature review also revealed that a consistent (21%) proportion of publications from CRs is reporting data on elderly patients. Cancer cases in people aged 70 years or above represented more than 47% of the total EU-27 estimated incidence for 2020 (1), making this group underrepresented in the literature. This underreporting might be also related to the fact that in some countries treatment of elderly cancer patients is administered in settings such as hospices or to homecare services that are not reported in health care records and are not regularly accessible to the CRs. It is anyhow important to have identified such publications, since elderly people are usually even more underrepresented in clinical trials. CRs can indeed offer an added value and complementarity with clinical studies, helping exploring treatment strategies in the elderly (32).

The literature review and conference proceedings revealed that treatment data are most often collected for breast cancer, the most frequent cancer in women in Europe. Treatment data are often

collected for colorectal, prostate and lung cancers, which are also among the most common. Data in the literature review are also consistent with the results derived from the dataset of the 2015 ENCR-JRC data call, where all registries reported treatment data for breast cancer, while most registries reported treatment data for colorectal, prostate and lung cancer.

Prior to the 2015 ENCR-JRC data collection, two other projects (EUROCHIP-3 and EUROCOURSE) provided information on the availability and use of treatment data in European CRs. The EUROCHIP-3 survey was carried out during 2010 and presented an overview on the availability of three main indicators in European population-based CRs: stage at diagnosis, cancer treatment delay and compliance with cancer guidelines. Information on treatment data was available in 30% of the 86 responding registries (4). The second project, EUROCOURSE (2010-2012), reported that 43% of the 106 responding registries gathered information on first treatment (5). The 2015 ENCR-JRC questionnaire reported that 49% of the 125 responding registries collect cancer treatment data. This proportion is higher (52%) if the site-specific and regional CRs overlapping with national ones is considered. A steady increase in the percentage of CRs collecting treatment data is therefore observed over the three data collection periods. A possible reason for this is the rising number of European countries and regions using electronic health records, which can be used for research purposes (33).

Evidence from the results of the ENCR-JRC 2015 data collection suggests that national registries are collecting cancer treatment data more frequently as compared to regional registries. This is consistent with the fact that usually national CRs have more resources available, either technical, financial and/or human. It was also observed that while 61 CRs reported in the questionnaire to collect treatment data, only 28 CRs actually submitted such data in the 2015 ENCR-JRC data call. Twelve registries indeed mentioned in the questionnaire that they collect treatment data but did not submit them, mainly motivating this with data incompleteness. This underreporting behaviour calls for increased awareness among the CRs on the importance in reporting treatment information.

While half of the general CRs responding to the 2015 call reported to collect treatment data, four out of five reported to collect data on cancer stage. According to the data call questionnaire, the six childhood CRs included in our analysis report to collect data on both cancer stage and treatment. Overall, more CRs are reporting data on cancer stage as compared to treatment. This finding is consistent with earlier research reporting that 61% of responding CRs collected data on cancer stage, while 43% reported cancer treatment data (5). The higher number of CRs recording stage compared to those reporting stage and treatment, is likely related to the fact that stage information has been standardised with the introduction of the TNM classification system already in the 1940s. There are also extensive training materials and activities on TNM coding, which explain its diffusion among European CRs. Such standardisation has not yet been performed thoroughly for the coding and registration of treatment data. Consensus guidelines for staging childhood cancers (the Toronto Paediatric Cancer Stage Guidelines) have been developed and endorsed for use by CRs. The international project 'BENCHISTA' involves most of the European CRs and is a good example of how to standardise the collected information for clinical variables like treatment and stage (34, 35).

The analysis of treatment data provided by the 28 general CRs contributing to the ECIS database revealed that for the main solid tumours the proportion of cancer patients treated with surgery was: 82% for breast cancer, 71% for colorectal, 36% for prostate and 19% for lung cancer. These results are consistent with previously published evidence on the impact of surgical treatment in Europe and in the USA (36–40).

In the recent years exploratory analyses on treatment in Europe by stage, age group, sex, period of incidence and geographical area from the ECIS database were carried out, addressing specifically breast, colorectal, prostate, endometrium and glioblastoma. Such analyses investigated to what extent some selected clinical and treatment patterns by age group, stage and period could be monitored using the 2015 ENCR-JRC dataset (40–45).

A limitation in the literature review could be the focus on the proceedings from scientific conferences of only four international societies, and the lack of other grey literature such as reports on CRs websites, or in languages other than English.

A further limitation was given by the use of search strings: although checks were performed on the results, the search method reduced the level of detail of the results from the review. Lastly, only titles and abstracts were reviewed, thus losing potential information from the full articles' text.

Regarding the 2015 ENCR-JRC dataset, one limitation consisted in the impossibility to distinguish between chemotherapy and other types of systemic therapy. This issue has been addressed in the new 2022 ECIS call for data protocol, where information on timing (neoadjuvant or adjuvant therapy), crucial for monitoring clinical care, has been added; systemic therapy information has been split in different variables (chemotherapy, targeted therapy, immunotherapy, hormone therapy, other/unspecified), and surgery has been detailed between local surgery and operative surgery (46).

In order to use treatment information and to ensure its quality and comparability at European level, a more harmonised collection of these variables among European population-based CRs is required. In fact, the availability of comparable information on treatment (and stage at diagnosis) is crucial to improve the interpretation of cancer outcome disparities between populations, therefore bringing valuable real life information for patients, clinicians, policymakers and other stakeholders.

5 Conclusion and way forward

Treatment data are increasingly being reported by CRs, though further improvements are needed to ensure complete and harmonised coverage of such important information. Sufficient technical, financial and human resources are needed to collect treatment data in a harmonised way, while clear guidelines for treatment data collection need to be developed.

To address these challenges, the ENCR Working Group on Treatment Data Harmonisation was set up in June 2021, with the aim of bringing together European experts in cancer registration, epidemiology and from the clinical field to discuss and draft guidelines for improved data collection and harmonisation of treatment data among European population-based CRs. This ongoing activity will be a key step to provide cross-comparisons

between European regions and countries, contributing to design actions to ensure better integrated and comprehensive cancer care and addressing unequal access to optimal care, namely the ultimate goal of the European Commission's Europe's Beating Cancer Plan (47).

6 The ENCR working group on treatment data harmonisation

Francesco Giusti (European Commission/Belgian Cancer Registry), Carmen Martos (European Commission/FISABIO, Valencia, Spain), Otto Visser (Netherlands Comprehensive Cancer Organisation - IKNL), Lisbeth Van Eycken (Belgian Cancer Registry), Claudia Allemani (London School of Hygiene and Tropical Medicine), Volker Arndt (German Cancer Research Centre - DKFZ), Riccardo Audisio (Sahlgrenska University Hospital, Göteborg, Sweden), Manola Bettio (European Commission), Roberta De Angelis (Istituto Superiore di Sanità, Rome, Italy), Henna Degerlund (Finnish Cancer Registry), Silvia Francisci (National Health Institute, Rome, Italy), Gemma Gatta (Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy), Anna Gavin (Northern Ireland Cancer Registry), Tom Børge Johannesen (Cancer Registry of Norway), Yolande Lievens (Ghent University Hospital and Ghent University, Ghent, Belgium), Margit Mägi (Estonian Cancer Registry), Rafael Marcos-Gragera (Epidemiology Unit and Girona Cancer Registry, Catalan Institute of Oncology, Girona, Spain), Eva Morris (Nuffield Department of Population Health, Big Data Institute, University of Oxford, UK), Regina Nanieva (National Institute for Cancer Epidemiology and Registration - NICER, Switzerland), Raquel Negrão Carvalho (European Commission), Laura Pareja Fernandez (Department of Health of Catalonia, Hospitalet del Llobregat, Barcelona, Spain), Francesco Pignatti (European Medicines Agency), Josepa Ribes (Department of Health of Catalonia, Hospitalet del Llobregat, Barcelona, Spain), Silvia Rossi (Istituto Superiore di Sanità, Rome, Italy), Arantza Sanvisens (Epidemiology Unit and Girona Cancer Registry, Catalan Institute of Oncology, Girona, Spain), Annalisa Trama (Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy), Maciej Trojanowski (Greater Poland Cancer Registry), Ulrich Wagner (National Institute for Cancer Epidemiology and Registration – NICER, Switzerland), Paul Walsh (National Cancer Registry Ireland), Vesna Zadnik (Epidemiology and Cancer Registry, Institute of Oncology Ljubljana, Slovenia).

Data availability statement

The datasets presented in this article are not readily available because aggregated data has been used; individual patient data is not available. Requests to access the datasets should be directed to francescogiusti@hotmail.com.

Author contributions

The first draft of the manuscript was written by FG, CM and LV. All authors contributed to the article and approved the submitted version.

Acknowledgments

Grateful acknowledgement is given to the European cancer registries for contributing to the initiatives and studies included in ECIS.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. ECIS - European Cancer Information System. (2022). European Union. Available at: https://ecis.jrc.ec.europa.eu (Last accessed on 30 September 2022).

2. Sant M, Minicozzi P, Primic-Žakelj M, Otter R, Francisci S, Gatta G, et al. Cancer survival in Europe, 1999–2007: Doing better, feeling worse? *Eur J Cancer* (2015) 51:2101–3. doi: 10.1016/j.ejca.2015.08.019

3. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): Analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* (2018) 391:1023–75. doi: 10.1016/S0140-6736(17)33326-3

4. Siesling S, Kwast A, Gavin A, Baili P, Otter REUROCHIP-3 Workpackage 5. Availability of stage at diagnosis, cancer treatment delay and compliance with cancer guidelines as cancer registry indicators for cancer care in Europe: Results of EUROCHIP-3 survey. *Int J Cancer* (2013) 132:2910–7. doi: 10.1002/ijc.27957

5. Siesling S, Louwman WJ, Kwast A, van de Hurk C, O'Callaghan M, Rosso S, et al. Uses of cancer registries for public health and clinical research in Europe: Results of the European network of cancer registries survey among 161 population-based cancer registries during 2010–2012. *Eur J Cancer* (2015) 51:1039–49. doi: 10.1016/j.ejca.2014.07.016

6. International Association of Cancer Registries. (2022). Available at: https://www.iacr.com.fr/ (Last accessed on 20 October 2022).

7. International Agency for Research on Cancer. (2022). Available at: https://www.iarc.who.int/events/ (Last accessed on 20 October 2022).

8. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: Principles and methods. part I: Comparability, validity and timeliness. *Eur J Cancer* (2009) 45:747–55. doi: 10.1016/j.ejca.2008.11.032

9. European Network of Cancer Registries. (2022). Available at: https://encr.eu/ (Last accessed on 22 September 2022).

10. Forsea A-M. Cancer registries in Europe–going forward is the only option. *Ecancer* (2016) 10:641. doi: 10.3332/ecancer.2016.641

11. Bettio M, Negrao Carvalho R, Dimitrova N, Dyba T, Flego M, Giusti F, et al. Measuring the cancer burden in Europe: The European cancer information system (ECIS). *EMJ Oncol* (2019) 7(1):48–9. doi: 10.1093/annonc/mdz263.014

12. Zanetti R, Sacchetto L, Coebergh JW, Rosso S. To accelerate cancer prevention in Europe: Challenges for cancer registries. *Eur J Cancer* (2018) 104:151–9. doi: 10.1016/j.ejca.2018.09.001

13. Sankila R, Black R, Coebergh JW, Démaret E, Forman D, Gatta G, et al. *Evaluation of clinical care by cancer registries. IARC technical publication 37*. Lyon:World Health Organization (2003).

14. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, et al. Cancer survival in Europe 1999–2007 by country and age: Results of EUROCARE-5–a population-based study. *Lancet Oncol* (2014) 15(1):34–4. doi: 10.1016/S1470-2045(13)70546-1

15. O'Sullivan B, Brierly J, Byrd D, Bosman F, Kehoe S, Kossary C, et al. The TNM classification of malignant tumours – towards common understanding and reasonable expectations. *Lancet Oncol* (2017) 18(7):849–51. doi: 10.1016/S1470-2045(17)30438-2

16. Coebergh JW, van den Hurk C, Rosso S, Comber H, Storm H, Zanetti R, et al. EUROCOURSE lessons learned from and for population-based cancer registries in Europe and their programme owners: Improving performance by research programming for public health and clinical evaluation. *Eur J Cancer* (2015) 51(9):997–1017. doi: 10.1016/j.ejca.2015.02.018

17. Duffy MJ. Tumor markers in clinical practice: A review focusing on common solid cancers. *Med Princ Pract* (2013) 22(1):4–11. doi: 10.1159/000338393

18. Voith von Voithenberg L, Crocetti E, Martos C, Dimitrova N, Giusti F, Randi G, et al. Cancer registries - guardians of breast cancer biomarker information: A systematic review. *Int J Biol Markers* (2019) 34(2):194–9. doi: 10.1177/1724600819836097

19. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan - a web and mobile app for systematic reviews. *Syst Rev* (2016) 5:210. doi: 10.1186/s13643-016-0384-4

20. ENCR-JRC Project 2015. *Updating population-based cancer incidence and mortality in Europe* (2015). Available at: https://www.encr.eu/sites/default/files/ENCR-JRC2015_Protocol.pdf (Last accessed on 20 October 2022).

21. EU Survey – European Commission. (2022). Available at: https://ec.europa.eu/eusurvey/home/welcome (Last accessed on 20 October 2022).

22. European Network of Cancer Registries. *Tools for registries* (2022). Available at: https://encr.eu/tools-for-registries (Last accessed on 20 October 2022).

23. Giusti F, Martos C, Adriani S, Flego M, Brunetto A, Dyba T, et al. *The JRC-ENCR Quality Check Software (QCS) for the validation of cancer registry data: User compendium – version 2.0, Ispra: European Commission 2022, JRC127031*. (2022).

24. Group for cancer epidemiology and registration in Latin language countries, in: *Annual meetings* (2021). Available at: https://www.grell-network.org/past-meetings (Last accessed on 20 October 2022).

25. European Network of cancer registries, in: *Conferences* (2022). Available at: https://www.encr.eu/conferences (Last accessed on 20 October 2022).

26. European Society for medical oncology, in: *ESMO congresses* (2022). Available at: https://www.esmo.org/meetings/esmo-congresses (Last accessed on 20 October 2022).

27. Brierley J, Gospodarowicz MK, Wittekind C. *The TNM classification of malignant tumours. 8*. Oxford: Wiley Blackwell (2017).

28. European Network of cancer registries, in: *ENCR RECOMMENDATIONS condensed TNM for coding the extent of disease*. Available at: https://www.encr.eu/sites/default/files/pdf/extentofdisease.pdf (Last accessed on 20 October 2022).

29. Pukkala E, Engholm G, Højsgaard Schmidt LK, Storm H, Khan S, Lambe M, et al. Nordic Cancer registries – an overview of their procedures and data comparability. *Acta Oncol* (2018) 57(4):440–55. doi: 10.1080/0284186X.2017.1407039

30. Gatta G, Capocaccia R, Botta L, Mallone S, De Angelis R, Ardanaz E, et al. Burden and centralised treatment in Europe of rare tumours: results of RARECAREnet-a population-based study. *Lancet Oncol* (2017) 18(8):1022–39. doi: 10.1016/S1470-2045(17)30445-X

31. Allemani C, Sant M, Weir HK, Richardson LC, Baili P, Storm H, et al. Breast cancer survival in the US and Europe: A CONCORD high-resolution study. *Int J Cancer* (2013) 132:1170–81. doi: 10.1002/ijc.27725

32. Derks MGM, Bastiaannet E, Kiderlen M, Hilling DE, Boelens PG, Walsh PM, et al. Variation in treatment and survival of older patients with non-metastatic breast cancer in five European countries: A population-based cohort study from the EURECCA breast cancer group. *Br J Cancer* (2018) 119(1):121-9. doi: 10.1038/s41416-018-0090-1

33. Oderkirk J. *Readiness of electronic health record systems to contribute to national health information and research. In: OECD health working papers, no. 99.* (2013) (Paris: OECD Publishing). doi: 10.1787/9e296bf3-en

34. Gupta S, Aitken JF, Bartels U, Brierley J, Dolendo M, Friedrichet P, et al. Paediatric cancer stage in population-based cancer registries: The Toronto consensus principles and guidelines. *Lancet Oncol* (2016) 17(4):e163–72. doi: 10.1016/S1470-2045(15)00539-2

35. Botta L, Gatta G, Didone F, Lopez Cortes A, Pritchard-Jones K. International benchmarking of childhood cancer survival by stage at diagnosis: The BENCHISTA project protocol. *PloS One* (2022) 17(11):e0276997. doi: 10.1371/journal.pone.0276997

36. American Cancer Society. *Breast cancer facts & figures 2019-2020*. Atlanta: American Cancer Society, Inc (2019).

37. van Walle L, Punie K, Van Eycken E, de Azambuja E, Wildiers H, Duhoux FP, et al. Assessment of potential process quality indicators for systemic treatment of breast cancer in Belgium: A population-based study. *ESMO Open* (2021) 6(4):100207. doi: 10.1016/j.esmoop.2021.100207

38. Gatta G, Sant M, Coebergh JW, Hakulinen Tthe EUROCARE Working Group. Substantial variation in therapy for colorectal cancer across Europe: EUROCARE analysis of cancer registry data for 1987. *Eur J Cancer* (1996) 32:831–5. doi: 10.1016/0959-8049(95)00642-7

39. Benitez Majano S, Di Girolamo C, Rachet B, Maringe C, Grønlie Guren M, Glimelius B, et al. Surgical treatment and survival from colorectal cancer in Denmark, England, Norway, and Sweden: a population-based study. *Lancet Oncol* (2019) 20(1):74–87. doi: 10.1016/S1470-2045(18)30646-6

40. Jakobsen E, Palshof T, Osterlind K, Pilegaard H. Data from a national lung cancer registry contributes to improve outcome and quality of surgery: Danish results. *Eur J Cardiothorac Surg* (2009) 35(2):348–52. doi: 10.1016/j.ejcts.2008.09.026

41. Giusti F, Martos C, Crocetti E, Dimitrova N, Randi G, Neamtiu L, et al. Patient data to monitor clinical patterns in early and advanced breast cancer in Europe. *Ann Oncol* (2019) 30:v75. doi: 10.1093/annonc/mdz240.048

42. Giusti F, Martos C, Neamtiu L, Randi G, Dyba T, Flego M, et al. Evaluating colorectal cancer treatment patterns by stage in European cancer registries. *Ann Oncol* (2020) 31:S455. doi: 10.1016/j.annonc.2020.08.618

43. Giusti F, Martos C, Neamtiu L, Randi G, Flego M, Dyba T, et al. Risk of second malignancies and survival differences by treatment pattern in prostate cancer: A population-based study. *Eur Urol Open Sci* (2020) 21(Suppl 3):S126. doi: 10.1016/S2666-1683(20)36189-9

44. Giusti F, Martos C, Neamtiu L, Randi G, Flego M, Dyba T, et al. *Exploring endometrial cancer treatment patterns and outcomes in Europe: an analysis with population-based cancer registry data. cancer registration saves lives, ENCR scientific meeting 2021, 16-18 November 2021, virtual event*. Available at: https://www.encr.eu/sites/default/files/inline-files/ENCR%20Scientific%20Meeting%202021%20Conference%20Book_updated%2017%20Nov.pdf (Last accessed on 20 October 2022).

45. Giusti F, Martos C, Scoccianti S, Neamtiu L, Randi G, Flego M, et al. Exploring changes in glioblastoma treatment patterns in Europe and the USA with population-based cancer registry data. *Ann Oncol* (2020) 31:S398. doi: 10.1016/j.annonc.2020.08.474

46. European Network of Cancer Registries. *Call for data protocol for European population-based cancer registries* (2022). Available at: https://encr.eu/sites/default/files/inline-files/ECIS%20call%20for%20data%20protocol_20220317_0.pdf (Last accessed on 20 October 2022).

47. European Commission. *Europe's beating cancer plan*. Available at: https://health.ec.europa.eu/system/files/2022-02/eu_cancer-plan_en_0.pdf (Last accessed on 20 October 2022).

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 24 April 2023
DOI 10.3389/fonc.2023.1114701

OPEN ACCESS

EDITED BY
Liesbet Van Eycken,
Belgian Cancer Registry, Belgium

REVIEWED BY
Jean-Michel Billette,
Statistics Canada, Canada
Angela Mariotto,
National Cancer Institute (NIH),
United States

*CORRESPONDENCE
Roberta De Angelis
roberta.deangelis@iss.it

SPECIALTY SECTION
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

RECEIVED 02 December 2022
ACCEPTED 24 March 2023
PUBLISHED 24 April 2023

CITATION
Demuru E, Rossi S, Ventura L, Dal Maso L, Guzzinati S, Katalinic A, Lamy S, Jooste V, Di Benedetto C, De Angelis R and the EUROCARE-6 Working Group (2023) Estimating complete cancer prevalence in Europe: validity of alternative vs standard completeness indexes.
Front. Oncol. 13:1114701.
doi: 10.3389/fonc.2023.1114701

Estimating complete cancer prevalence in Europe: validity of alternative vs standard completeness indexes

Elena Demuru[1], Silvia Rossi[1], Leonardo Ventura[2], Luigino Dal Maso[3], Stefano Guzzinati[4], Alexander Katalinic[5], Sebastien Lamy[6], Valerie Jooste[7], Corrado Di Benedetto[8], Roberta De Angelis[1]* and the EUROCARE-6 Working Group

[1]Department of Oncology and Molecular Medicine, Istituto Superiore di Sanità, Rome, Italy, [2]Clinical and Descriptive Epidemiology Unit, Istituto per lo Studio, la Prevenzione e la Rete Oncologica (ISPRO), Firenze, Italy, [3]Cancer Epidemiology Unit, Centro di Riferimento Oncologico (CRO), Istituto di Ricerca e Cura a Carattere Scientifico (IRCCS), Aviano, Italy, [4]Veneto Cancer Registry, Azienda Zero, Padova, Italy, [5]Cancer Registry of Schleswig-Holstein, Institute for Social Medicine and Epidemiology, University of Lübeck, Lübeck, Germany, [6]Tarn Cancer Registry, Claudius Regaud Institute - Center for Epidemiology and Research in Population Health (CERPOP U1295), University of Toulouse - Inserm, Toulouse, France, [7]Digestive Cancer Registry of Burgundy, Dijon University Hospital, INSERM UMR1231, Dijon, France, [8]IT Service, Istituto Superiore di Sanità, Rome, Italy

Introduction: Comparable indicators on complete cancer prevalence are increasingly needed in Europe to support survivorship care planning. Direct measures can be biased by limited registration time and estimates are needed to recover long term survivors. The completeness index method, based on incidence and survival modelling, is the standard most validated approach.

Methods: Within this framework, we consider two alternative approaches that do not require any direct modelling activity: i) empirical indices derived from long established European registries; ii) pre-calculated indices derived from US-SEER cancer registries. Relying on the EUROCARE-6 study dataset we compare standard vs alternative complete prevalence estimates using data from 62 registries in 27 countries by sex, cancer type and registration time.

Results: For tumours mostly diagnosed in the elderly the empirical estimates differ little from standard estimates (on average less than 5% after 10-15 years of registration), especially for low prognosis cancers. For early-onset cancers (bone, brain, cervix uteri, testis, Hodgkin disease, soft tissues) the empirical method may produce substantial underestimations of complete prevalence (up to 20%) even when based on 35-year observations. SEER estimates are comparable to the standard ones for most cancers, including many early-onset tumours, even when derived from short time series (10-15 years). Longer observations are however needed when cancer-specific incidence and prognosis differ remarkably between US and European populations (endometrium, thyroid or stomach).

Discussion: These results may facilitate the dissemination of complete prevalence estimates across Europe and help bridge the current information gaps.

KEYWORDS

cancer prevalence, cancer registries, cancer survivors, cancer survivorship, EUROCARE, Europe, SEER program

10.3389/fonc.2023.1114701

1 Introduction

Cancer prevalence statistics enumerate the number, or the proportion, of people in a population living after a cancer diagnosis at a specific date. Unlike other surveillance metrics based on cancer registries' observations, such as incidence or survival, direct measures of prevalence are intrinsically incomplete, as they cannot include the cancer survivors diagnosed before the start of registration. Complete prevalence must be necessarily estimated to recover long term survivors, especially when the period of registration is limited.

The completeness index method is one of the most accurate and used methods to estimate complete prevalence starting from limited-duration prevalence measured by cancer registries (1). Based on incidence and relative survival modelling and on their relationship with prevalence, this method provides a correction factor, the so-called completeness index, or R-index, to complete cancer-specific registries observations.

The completeness index method has been systematically validated and applied since many years in the USA (2), where complete prevalence statistics are published annually as an integral part of the SEER Cancer Statistics (3). A software to implement the method is distributed by the National Cancer Institute, along with completeness indexes derived from the SEER registries datasets (4).

Conversely, in Europe complete prevalence estimates are not systematically available in all countries with active population-based cancer registries. European cancer prevalence estimates by country are made available by GLOBOCAN (5), however they are limited to 5-years since diagnosis (6). Occasionally, on a project basis, the completeness index method has been applied to European CRs data to derive complete prevalence of rare cancers (7–9) or frequent cancers by European country and area (10, 11). Complete prevalence is periodically estimated through the completeness index approach only in Italy (12, 13), where the method was first proposed. Experiences in other countries refer to limited-duration prevalence (14) or to different methods (15–19). Only some European registries operating since the 50s, such as those in Nordic countries or Slovenia, are able to measure a virtually complete prevalence without any estimation (20, 21).

Integrating traditional surveillance metrics with accurate complete prevalence estimates is of increasing importance, given the remarkable growth of cancer survivors in all ageing societies. They represent a heterogeneous population, in terms of healthcare needs and quality of life, that should be better quantified and qualified (22–27). Given this background, closing the existing gaps in Europe is one of the priorities in cancer surveillance.

Promoting the use and dissemination of complete cancer prevalence indicators by country in Europe was one of the goals of the European Joint Action on Cancer iPAAC (Innovative Partnership for Action Against Cancer) (28). Exploring the feasibility of viable solutions to facilitate the use of completeness indexes was part of the project's activities.

With this purpose, in the present study we compared the standard method of deriving prevalence completeness index in Europe (by modelling incidence and survival data from European populations) with alternative approaches that do not require any statistical modelling, namely: i) empirical indexes derived from the longest prevalence data available from European registries; ii) publicly available model-based indexes estimated from SEER-US data (4). The study aims to assess under which conditions of application (registration time length and cancer type) these "non-standard" approaches may adequately surrogate the reference method, which remains the "gold standard".

Nowadays, indeed, cancer prevalence observations are available for time series and populations to a much greater extent than when R-indexes were first proposed (1). Assessing application conditions of empirical R-indexes may facilitate the use and dissemination of complete prevalence estimates across Europe and contribute to bridge the present information gaps. For the same reasons it is worth exploring the application limits to European data of SEER-US indexes that are publicly available and ready to be used.

2 Materials and methods

The study relies on the dataset of the EUROCARE-6 project, a wide collaborative study on cancer survival and prevalence in Europe (29) based on cancer registries data. The dataset includes pseudonymised individual data on cancer patients' incidence and life status, as well as life tables and resident population in each registry.

For the purpose of the study we selected 62 general cancer registries from 27 European countries (21 with national population coverage) providing prevalence data up to 1/1/2013, the most recent common prevalence index date available in the dataset. At this date the maximum duration of registration ranged from 5 to 35 years, with median at 20 years.

The following four different types of analyses were conducted each using a specific dataset depending on the scope. Cancer registries included 5% to 50% coverage of the 27 countries' population (Table 1).

a) *Empirical completeness indexes*. Pooled prevalence data from 8 registries with an observation period of 35 years (maximum available duration of registration) were used to estimate European empirical completeness indexes.

b) *Model-based completeness indexes*. Pooled incidence and relative survival data from 11 registries with at least 30 years of observation were used to derive standard European model-based completeness indexes.

c) *Validation of completeness indexes*. Registry-specific prevalence from the registries with at least 20 years of observation were the reference to validate European model-based completeness indexes (gold-standard method) estimated in step b. Registries in dataset b) were excluded from the validation dataset.

d) *Complete prevalence estimation*. Registry-specific observed prevalence from all eligible 62 registries, up to their maximum registration duration (from 5 to 35 years), were used to estimate complete prevalence in each registry according to standard and alternative methods.

To compare complete prevalence values estimated from the different completeness indexes we performed distinct analyses for a

10.3389/fonc.2023.1114701

TABLE 1 Description of the registries included in each analysis-specific dataset.

Dataset	Type of analysis	Registration length (years)	Number of registries	Registries	Population (% study coverage)
a)	Empirical index	35	8	**Denmark, Estonia, Finland, Iceland, Norway, Scotland,** Geneva *(Switzerland)*, Parma *(Italy)*	23,592,911 (5%)
b)	Model-based index	>=30	11	*Registries in dataset a) plus:* ***Austria, Slovenia,*** *Tarragona (Spain)**	34,806,065 (8%)
c)	Validation of completeness indexes	>=20	20	**Bulgaria, Lithuania, Malta, The Netherlands, Northern Ireland, Wales,** Balearic Islands, Basque Country, Granada *(Spain)*, Graubünden and Glarus, Eastern Switzerland *(Switzerland)*, Bas Rhin, Doubs, Haut-Rhin, Isere, Somme, Tarn *(France)*, Modena, Ragusa, Romagna *(Italy)*	44,230,482 (10%)
d)	Complete prevalence estimation	>=5	62	*Registries in dataset c) and b) plus:* **Belgium, Cyprus, Czech Republic, England, Ireland, Latvia, Poland,** Herault, Lille, Poitou Charentes *(France)*, Bremen, Federal States (BR,MW-PSA,THU), Hamburg *(Germany)*, Bergamo, Puglia Barletta Andria-Trani, Catania-Messina-Enna, Latina, Monza-Brianza, Napoli, Nuoro, Palermo, Piacenza, Reggio Emilia, Siracusa, Sondrio, Taranto, Umbria *(Italy)*, Southern Portugal *(Portugal)*, Castellon, Girona *(Spain)*, Friburg, Ticino *(Switzerland)*	231,214,391 (51%)

*The registry of Tarragona is included in dataset b) and not in d) because limited-duration prevalence is available at 1/1/2012.
Population covered by the registries in each dataset and percent coverage of the 27 European countries that participated in EUROCARE-6 included in the study are shown. National registries are in bold.

selection of 30 common index cancers. Cancer entities were defined according to the Third Revision of the International Classification of Diseases for Oncology (ICDO-3). Only malignant primary cancers were included, except for brain and urinary bladder (Supplementary Materials, Table A1). Non-malignant tumours proportion by registry ranges from 0 to 28% for brain cancer and from 0 to 54% for urinary bladder, thus reflecting varying registration criteria across Europe. The first primary tumour for each cancer entity was considered, meaning that each person was counted only once and that people with multiple primary cancers affecting different sites contribute to prevalence counts of different entities. Consequently, cancer-specific counts do not sum up to counts of all cancers combined.

2.1 Observed limited-duration prevalence

Limited-duration prevalence observed in each registry population was computed at the index date with the counting method, available in the SEER*Stat software (30) by enumerating the number of patients known to be alive at the index date. Life-table survival probabilities stratified by registry, sex, grouped age at diagnosis (0-59, 60-74, 75+), cancer site and 5-year period of diagnosis, were attributed to patients lost to follow-up to count those estimated alive at the prevalence index date. Age at the prevalence date was detailed in 5-year groups and 85+. The proportion of lost to follow-up is generally very low, below 2% in most countries.

2.2 Completeness index estimation (R-index)

R-index at duration d (R_d) is defined as the ratio of prevalence at duration d to estimated complete prevalence. It expresses an estimation of percent completeness of a given limited-duration prevalence. Complete prevalence is therefore estimated dividing the number of observed prevalent cases at a given duration d (N_d) by the corresponding R-index at the same duration (1).

For each cancer we derived R-index by sex, age at prevalence date (i) in 5-year age groups and annual registration duration (d). Model-based and empirical approaches were both considered.

i) European empirical R-index (EU emp)

Empirical R-indexes were obtained from the pool of registries in dataset a) (Table 1) as the ratio of the observed prevalent cases at duration d to the observed prevalent cases at the maximum duration (35 years), namely $R_{i,d} = N_{i,d}/N_{i,35}$. Age at prevalence date was grouped in 5-year classes except for extreme ages (0-29 and 80+) for which wider groupings were used to avoid random fluctuations due to the scarce number of cases. Using these empirical indexes is to assume that observed 35-year limited duration prevalence equals (i.e. is sufficiently close to) complete prevalence.

ii) Standard European model-based R-index (EU mod)

For the pool of registries in dataset b) (Table 1) we computed incidence rates and relative survival (RS) with the SEER*Stat software (30). RS, the ratio of observed survival in a group of cancer patients to the expected survival in a comparable group from the general population, was determined using the Ederer 2 cohort method. Incidence and survival data were stratified by cancer type, sex, 5-year period of diagnosis (1980-1984, 1985-1989, 1990-1994, 1995-1999, 2000-2004, 2005-2009, 2010-2014) and age at diagnosis (5-year and 85+ for incidence; cancer-specific strata for relative survival are given in Table A1 Supplementary Materials). We modelled pooled incidence and relative survival data following the standard methodology (2). We fitted a mixture "cure-model" of Weibull type to RS data. These models assume that only a fraction of patients will die of the disease, with time to death following a Weibull

distribution, while the others are considered as cured. The non-linear regression procedure (NLIN) available in the SAS Software (SAS System for Windows, version 9.4; SAS Institute, Cary, NC) was used to estimate model parameters.

We fitted two alternative logistic age-cohort models to incidence rates stratified by age and period of diagnosis. Non-parametric cohort-effect was modelled through 10-year groups and parametric dependency on age at diagnosis was assumed by using respectively an exponential or a six-degree polynomial. Both models were estimated with the SAS LOGISTIC procedure.

Parameters of survival and incidence models were then imported in the software implementing the completeness index standard method (COMPREV) (4) to produce European model-based R-indexes.

iii) SEER model-based R-index (SEER mod)

Model-based R-indexes, estimated by the US National Cancer Institute (NCI) from the SEER-Program cancer registries data, were extracted from the COMPREV software (4).

2.3 Validation of the completeness indexes

The completeness index method allows to estimate any limited-duration prevalence beyond the longest observed period. Prevalence at any duration d_2 can be estimated dividing observed prevalence at maximum available duration d_1 by the ratio of the two corresponding R-indexes: R_{d1}/R_{d2}.

We used this property to validate R-indexes estimated by modelling European data, i.e. by using the gold standard method. For each eligible registry observed, 20-year prevalence was compared with estimated 20-year prevalence. To simulate a registration activity shorter than 20 years, observed prevalence was artificially truncated at durations d=5,10,15 years. The goodness of fit was measured separately for each cancer type as the weighted average percent relative difference in absolute value between estimated (N') and observed (N) 20-year number of prevalent cases (*APRD*):

$$APRD = \sum_r \left(\frac{\left|N'_{20,r} - N_{20,r}\right|}{N_{20,r}}\right) w_r \times 100$$

Registry-specific proportions of cancer cases (w_r) were used as weights. The absolute value of the relative difference avoids compensations between under- and over-estimations and provides a maximum average discrepancy compared to observations. The registries used for this validation (dataset c in Table 1) did not coincide with those used for estimating European model-based R-indexes (dataset b in Table 1).

2.4 Comparison of complete prevalence estimates

Cancer-, sex-, age- and duration-specific prevalence completeness indexes were applied to observed prevalence at maximum available duration in each of the 62 registries in dataset d) to obtain estimates of complete prevalence at 1/1/2013. Standard model-based complete prevalence estimates were compared to those obtained with alternative R-indexes (EU emp or SEER mod).

Weighted average percent relative difference between alternative and standard estimates of complete prevalence (PRD) was analysed by cancer site, sex and grouped registration duration (10-14 years, 15-19 years, 20-24 years, 25-35 years). The resident population covered by each registry was used as weight in the average.

3 Results

3.1 Incidence and relative survival models

In general, mixture cure models fitted data well and observed relative survival generally lied within the confidence limits estimated for predicted survival (examples are reported in the Supplementary Materials, Figure A1). Moreover, in most cases the survival curves reached a plateau within 20 years of follow-up, meaning that the cure assumption is satisfied in this time interval.

Diagnostic plots and values of the Akaike Information Criterion (AIC) showed that polynomial models fitted incidence data much better than exponential models for all the considered cancer types (Supplementary Materials, Figure A2). This is particularly evident for cancers at early onset or with bimodal age at diagnosis. Age polynomials provide indeed higher flexibility in modelling age trends compared to the exponential model.

3.2 Trends of the completeness indexes

Some examples of cancer-specific completeness indexes trends by age at prevalence date and duration of registration are shown in Figures 1–3. The comparison of the three different methods (SEER mod, EU mod and EU Emp) is restricted to the age range 30-79 years for which R-index can be estimated for all methods by 5-year age classes. Wider groups (0-29 and 80+) are in fact needed to compute empirical indexes for extreme age ranges with few cases.

Completeness index increases with the length of registration period and is higher for cancers at low prognosis (Figure 1) than for those at high to medium prognosis (Figure 2). A reduced survival implies indeed a more complete observed prevalence. Generally, R-index is close to 100% at young age and decreases with advancing age at prevalence date. For early onset tumours (Figure 3), however, young survivors can be partly not observable depending on the length of registration activity. Prevalence completeness is highest for low prognosis cancers diagnosed mainly in the elderly (Figure 1). At 15 years of registration, R-index is above 80-90% with minimum values for the eldest survivors. The empirical index trend is less smooth compared to model-based R-indexes because, being based on observations, it is more subject to random fluctuations, as also proven by confidence intervals (not shown in the graphs). At 35-years of registration all methods provide R-index values around 100%, meaning that such duration is sufficiently long to detect practically all survivors.

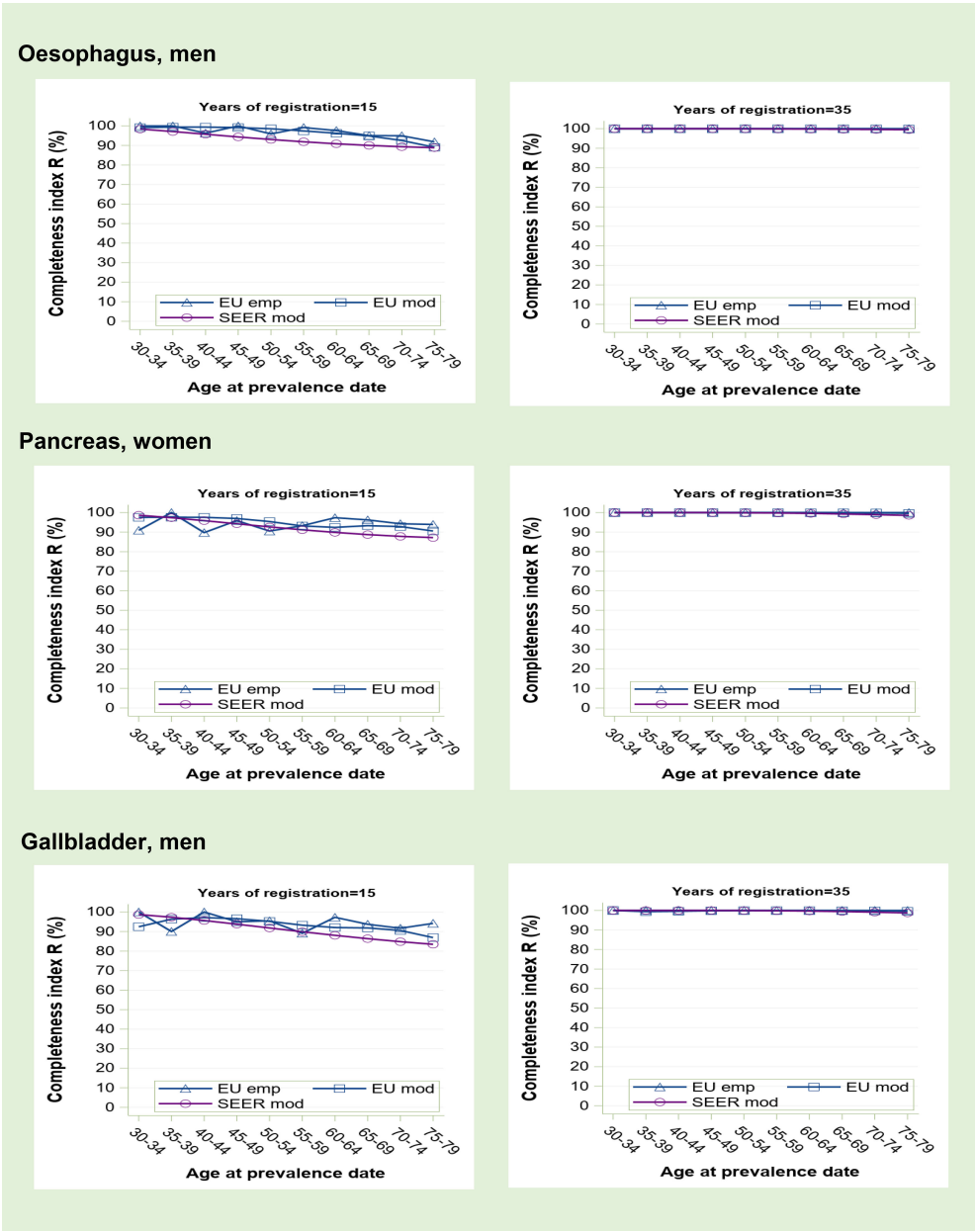


FIGURE 1
Prevalence completeness index (R index) at 1st January 2013 estimated for some tumours at low prognosis (oesophagus, pancreas, gallbladder) according to alternative methods: SEER model-based, EU model-based, EU empirical by age at prevalence date and registration time length (15 and 35 years).

Prevalence completeness is intermediate for higher prognosis cancers diagnosed in middle to old age (Figure 2). In the examples shown (breast, colorectal and corpus uteri cancers), at 15 years of registration, R-index varies from 95-100% to 50-70% as a function of age at prevalence date. SEER R-index values are slightly lower compared to those based on European data, reflecting a more favourable prognosis for US patients. At 35 years, model-based R-indexes tend to converge to 100% (95-98% for the eldest age group).

Cancers at early onset show the lowest R-index values and the most marked variations (Figure 3). At 15 years, observed prevalence is far from being complete for most age groups, particularly for bone cancers that are almost equally diagnosed at all ages. A registration period of 35 years appears insufficient to observe all long-term survivors, as shown by the residual gap (up to 50%) between empirical and model-based R-index estimates. By contrast, SEER and standard R-index, which are both model-based, show a quite similar age profile.

3.3 Validation of the completeness indexes

Tables 2A, B report observed 20-year prevalence proportion per 100,000 for the pool of registries in the validation dataset, for male and female populations, respectively. The weighted average percent relative differences, in absolute value, between registry-specific 20-year observed and standard estimated prevalence (APRD) is also

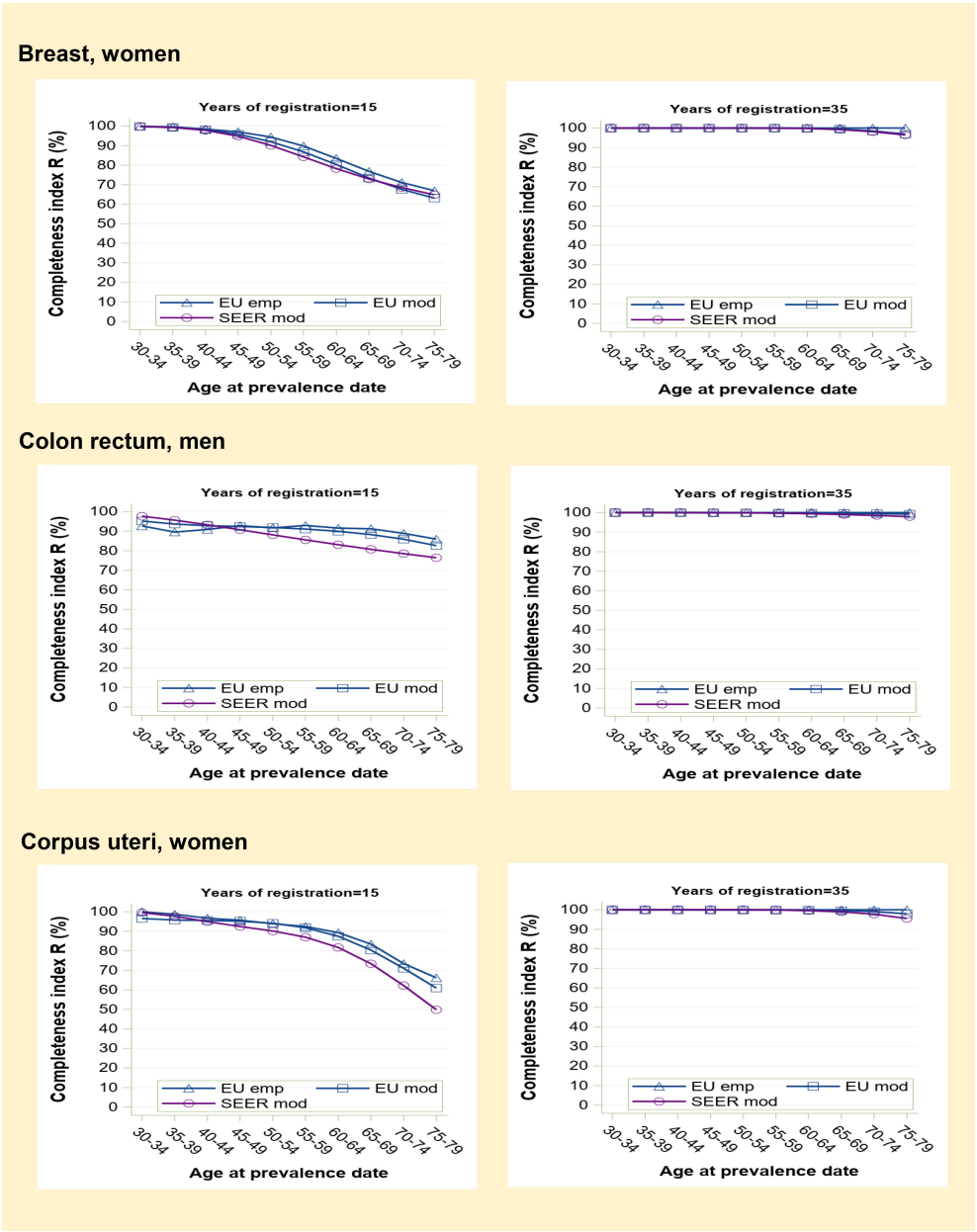


FIGURE 2
Prevalence completeness index (R-index) at 1st January 2013 estimated for some frequent medium-high prognosis tumours (breast, colon-rectum, corpus uteri) according to alternative methods: SEER model-based, EU model-based, EU empirical by age at prevalence date and registration time length (15 and 35 years).

reported and is obtained by artificially truncating observed prevalence at 5,10 and 15 years.

Average discrepancies between estimates and observations decrease as registration length increases. Particularly with registration times of 15 years the fit to observations is always good (APRD are well below 5%, maximum 6.3% for cervical cancer). At 10 years the validation is equally satisfying for all cancers examined (APRD values do not exceed 5%) except for young-onset cancers (cervix uteri, thyroid, brain and, to lesser extent, skin melanoma, bones, testis and Hodgkin lymphoma), suggesting that 15-year observed prevalence provides a more robust basis for this class of tumours.

Conversely prevalence observations limited to 5-years lead to less precise estimates in most of the cases (APRD exceed 5%) especially, but not only, for young-onset cancers (21% for cervical cancer, 12.5% for prostatic cancer).

3.4 Comparative assessment of complete prevalence estimates

Empirical (EU Emp) and SEER (SEER mod) complete prevalence estimates were compared to the standard model-based estimates (EU mod) for all 62 eligible cancer registries (dataset d).

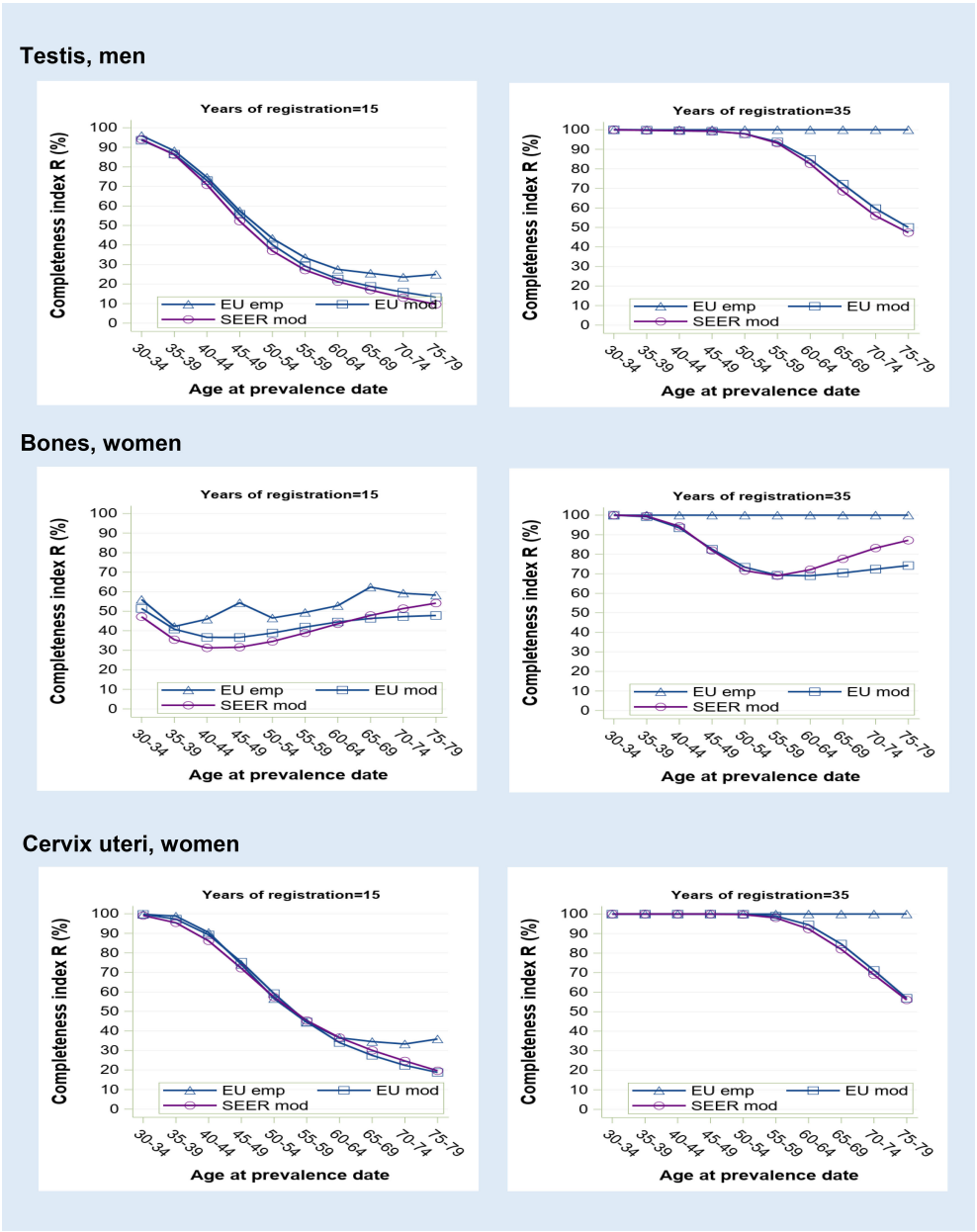


FIGURE 3
Prevalence completeness index (R-index) at 1st January 2013 estimated for some tumours diagnosed at young age (testis, bones, cervix uteri) according to alternative methods (SEER model-based, EU model-based, EU empirical by age at prevalence date and registration time length (15 and 35 years).

PRD between alternative and standard complete prevalence estimates of some index tumours is plotted in Figure 4 by registration time length (from 5 to 35 years).

Consistently with Figures 1, 2, when considering cancers at late age at onset with low (pancreas, lung) or good prognosis (colon-rectum and breast), the empirical estimates (Figure 4, blue crosses) approach model-based estimates as registration length increases. PRD values between -5% and 0 are indeed reached already after 10 years of registration. Conversely, for testicular and cervical cancers empirical indexes provide complete prevalence estimates that are systematically lower than model-based estimates (PRD at about -10% or -20% respectively) regardless of the registration time length, consistently with R-index patterns for early-onset tumours (Figure 3).

Differences between SEER and standard European complete prevalence estimates (Figure 4, purple circles) are almost null at all durations for pancreatic and breast cancers, and after 20 years of observation, for colorectal and lung cancers. Being model-based, SEER R-indexes reproduce standard estimates better than the empirical indexes for cervical and testicular cancers (PRD approaching zero with growing registration time).

A complete picture of percent relative differences between alternative and standard complete prevalence estimates is given in Tables 3A, B (EU Emp vs EU mod) and Tables 4A, B (SEER mod Vs

10.3389/fonc.2023.1114701

TABLE 2A Validation of European model-based R-index, *men*.

Cancer site	Observed 20-y Prevalence	APDR, 20-y prevalence estimated by truncating registries observations at		
		15 years	10 years	5 years
	Proportion x 100,00	*%*	*%*	*%*
All sites	**2,999**	*2.0*	*5.0*	*6.4*
Prostate	**1,082**	*1.1*	*3.3*	*12.5*
Colon Rectum	**488**	*1.2*	*2.8*	*6.0*
Bladder	**354**	*1.1*	*2.9*	*5.5*
Skin melanoma	**161**	*2.1*	*5.9*	*13.8*
Lung	**157**	*1.3*	*3.1*	*3.3*
Kidney	**138**	*1.4*	*4.5*	*8.4*
Non-Hodgkin lymphoma	**122**	*0.9*	*3.1*	*5.7*
Testis	**103**	*2.1*	*5.0*	*10.9*
Larynx	**79**	*1.6*	*2.9*	*6.2*
Head and neck	**76**	*1.1*	*2.9*	*5.0*
Stomach	**61**	*2.0*	*4.4*	*9.7*
CLL/SLL	**49**	*1.6*	*4.2*	*7.5*
Hodgkin lymphoma	**38**	*3.1*	*5.4*	*9.3*
Thyroid	**33**	*4.2*	*8.3*	*14.3*
Brain	**32**	*4.6*	*9.5*	*13.8*
Multiple myeloma	**31**	*1.0*	*3.5*	*6.1*
Soft tissues	**30**	*2.2*	*3.1*	*4.3*
Oesophagus	**27**	*1.5*	*3.4*	*5.7*
Liver	**19**	*0.9*	*1.7*	*4.2*
Pancreas	**16**	*1.3*	*5.1*	*6.8*
Penis	**15**	*1.8*	*2.6*	*5.4*
AML	**12**	*2.3*	*3.0*	*4.8*
CML	**10**	*2.4*	*4.1*	*5.8*
Gallbladder	**9**	*1.3*	*4.7*	*9.6*
Bones	**9**	*3.0*	*6.3*	*10.4*

CLL/SLL, Chronic lymphocytic leukaemia/small lymphocytic lymphoma; AML, Acute myeloid leukaemia; CML, Chronic myeloid leukaemia.
The validation is limited to cancer registries with at least 20 years of observations at 1/1/2013 that were not used to estimate R-index. Pooled observed 20-y prevalence proportion and weighted average percent relative differences in absolute value (APRD, %) between observed and estimated 20-years prevalence by cancer site are shown. Estimates are derived by applying completeness indexes to observed prevalence truncated at 5, 10 and 15 years and the average is weighted using the registry-specific number of prevalent cases.

EU mod), as a function of the duration of registration, starting from the group of 10 registries in operation for 10-14 years to the group of 17 registries active for 25-35 years. Mean standard complete prevalence proportion and PRD values in each pool of registries are reported by sex and cancer site. Negative values of PRD indicate an average underestimation of complete prevalence compared to the standard method.

The empirical R-index underestimates compared to the gold standard (Tables 3A, B) but the difference declines as registration time increases. The two methods lead to similar complete prevalence (PRD not exceeding 5% in absolute value) already after 10 or 15 years of registration for most cancers of the elderly, including those at highest prevalence (breast, prostate, colon and rectum, bladder) and those at poorest prognosis (e.g. oesophagus, larynx, gallbladder, pancreas, multiple myeloma) that show the lowest discrepancies. Most tumours at early onset represent an exception to this general pattern. PRD values reach 10-20% (testis, brain, bones, soft tissues and cervical cancers, Hodgkin lymphoma) and are scarcely sensitive to the duration of registration. On the contrary, more comparable estimates were observed for skin melanoma and thyroid cancers, both at early onset and with remarkably rising incidence across Europe.

10.3389/fonc.2023.1114701

TABLE 2B Validation of European model-based R-index, *women*.

Cancer site	Observed 20-y Prevalence	APDR, 20-y prevalence estimated by truncating registries observations at		
		15 years	10 years	5 years
	Proportion x 100,00	*%*	*%*	*%*
All sites	**3,471**	*1.8*	*3.9*	*5.9*
Breast	**1,537**	*0.9*	*1.9*	*3.5*
Colon Rectum	**414**	*1.4*	*3.0*	*4.8*
Corpus Uteri	**274**	*2.3*	*4.7*	*7.0*
Skin melanoma	**218**	*2.3*	*6.1*	*11.8*
Cervix uteri	**144**	*6.3*	*14.3*	*21.0*
Thyroid	**119**	*5.4*	*12.0*	*16.3*
Ovary	**110**	*0.7*	*1.1*	*5.7*
Non-Hodgkin lymphoma	**105**	*1.4*	*2.8*	*4.6*
Bladder	**94**	*2.0*	*4.8*	*7.7*
Lung	**88**	*1.3*	*3.5*	*4.8*
Kidney	**84**	*1.5*	*3.7*	*6.9*
Stomach	**41**	*2.1*	*5.5*	*9.3*
Head and Neck	**37**	*1.2*	*2.7*	*4.9*
CLL/SLL	**36**	*1.6*	*4.1*	*8.2*
Hodgkin lymphoma	**31**	*1.9*	*5.1*	*8.7*
Brain	**26**	*4.9*	*10.8*	*18.1*
Multiple myeloma	**26**	*1.1*	*4.1*	*7.3*
Soft tissues	**22**	*1.6*	*3.8*	*7.5*
Pancreas	**15**	*1.1*	*4.1*	*6.6*
AML	**12**	*2.3*	*4.0*	*8.1*
Larynx	**11**	*1.6*	*3.6*	*4.8*
Oesophagus	**10**	*2.1*	*4.0*	*5.6*
Gallbladder	**10**	*2.5*	*6.4*	*9.7*
Bones	**8**	*2.0*	*3.5*	*6.2*
CML	**8**	*3.1*	*5.5*	*10.7*
Liver	**7**	*1.5*	*3.8*	*5.2*

CLL/SLL: Chronic lymphocytic leukaemia/small lymphocytic lymphoma; AML, Acute myeloid leukaemia; CML, Chronic myeloid leukaemia.
The validation is limited to cancer registries with at least 20 years of observations at 1/1/2013 that were not used to estimate R-index. Pooled observed 20-y prevalence proportion and weighted average percent relative differences in absolute value (APRD, %) between observed and estimated 20-years prevalence by cancer site are shown. Estimates are derived by applying completeness indexes to observed prevalence truncated at 5, 10 and 15 years and the average is weighted using the registry-specific number of prevalent cases.

SEER R-indexes may provide either under- or over-estimations of standard complete prevalence (Tables 4A, B) that diminish as registration time grows. They provide similar estimates to the standard method after 10 or 15 years of registration for most tumours and, being based on models as well, even for most of early onset tumours (Hodgkin lymphoma, soft tissues, bones, cervix uteri, skin melanoma). Wider discrepancies were instead found when incidence and survival patterns in US and European populations determine differences between standard and SEER R-index values (non-Hodgkin lymphomas, thyroid, corpus uteri, testis, brain, larynx and stomach cancers). Notably PRD values (within 5%) for male brain cancer do not properly reflect the actual differences between SEER and standard R-index by age (under- and over- estimations are compensated in the weighted average) regardless of the duration of registration.

This comparative assessment of the alternative methods to derive complete cancer prevalence is summarised in Table 5 to facilitate readability and use of the results.

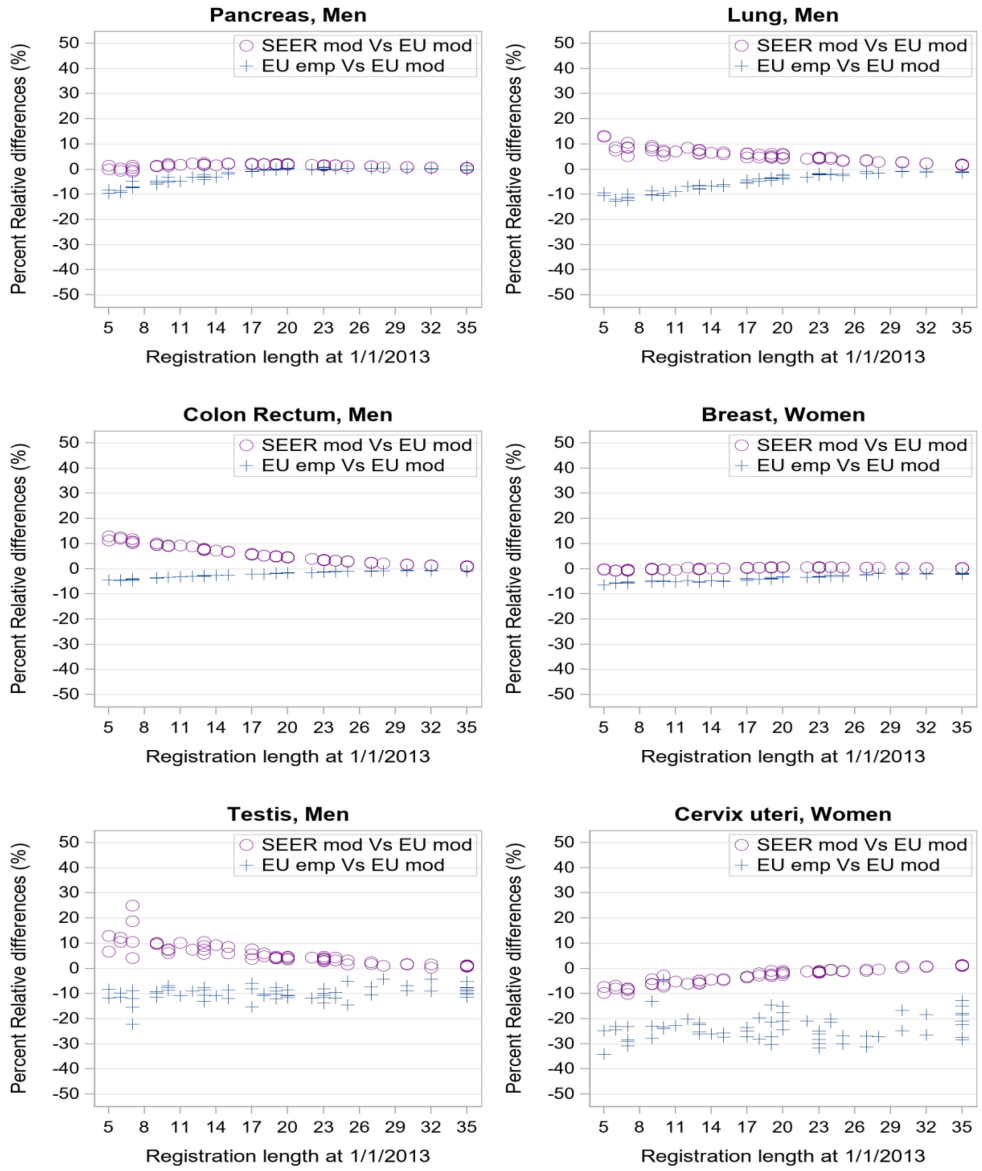


FIGURE 4
Percent relative difference (%) by registration length at 1/1/2013 of complete prevalence estimates obtained with SEER model-based or EU empirical R-index against EU model-based estimate as reference value. Each point corresponds to one of the 62 registries in dataset d of Table 1.

4 Discussion

To our knowledge, this is the first study exploring the validity of alternative approaches to derive prevalence completeness indexes. The study relies on an exceptionally wide European population-based dataset covering 50% of the population of the 27 countries involved.

Model-based R-indexes were introduced more than 20 years ago (1). Nowadays observations of cancer prevalence are available for time series and populations of much greater extension, thus testing the validity of empirical indexes that have now become available is relevant for a wider application of the method. The completeness index method is indeed particularly suited for local registry-based applications that rely on the available observed limited-duration prevalence.

Other methods to estimate complete prevalence include those modelling prevalence as a function of cancer-specific incidence and survival, both derived from cancer registries' data. Unlike the completeness index method, these methods do not rely on observed limited-duration prevalence and are more suited to derive time projections of cancer prevalence or national estimates in countries with partial registration coverage (15–18, 25).

From the validation study, a registration time period of at least 10 years turned out to be necessary to safely apply the prevalence completeness index method, confirming this cut-off as a general recommendation.

In many situations empirical R-index was found to provide complete prevalence estimates comparable to the "gold-standard". Registries' observation time window, cancer specific incidence age

TABLE 3A Comparison between empirical and standard model-based complete prevalence at 1/1/2013 by cancer site for the 62 European registries included in the study grouped by registration time length (from 10-14 years to 25-35 years), Men.

Registration time interval in years (number of registries)	10-14 y (10)		15-19 y (12)		20-24 y (13)		25-35 y (17)		Registration time length with PRD ≤ 5%
Cancer site	Standard complete prev	PRD	Standard complete prev	PRD	Standard complete prev	PRD	Standard complete prev	PRD	
	Prop *100,000	%	Prop *100,000	%	Prop *100,000	%	Prop *100,000	%	*number of years*
All cancers	3,476	*-7.1*	3,757	*-4.5*	3,747	*-3.3*	3,692	*-2.2*	>15
Prostate	959	*-2.8*	1,190	*-1.5*	1,450	*-0.5*	1,225	*0.0*	>10
Colon Rectum	539	*-3.0*	565	*-2.0*	501	*-1.3*	550	*-0.8*	>10
Bladder	453	*-2.2*	456	*-2.1*	281	*-2.1*	462	*-1.9*	>10
Non-Hodgkin lymphoma	232	*-5.4*	251	*-4.8*	247	*-4.5*	245	*-4.0*	>15
Lung	195	*-7.3*	197	*-4.6*	180	*-2.3*	171	*-1.1*	>15
Kidney	182	*-5.6*	203	*-3.1*	173	*-2.2*	184	*-1.4*	>15
Testis	154	*-9.4*	160	*-10.5*	166	*-10.2*	167	*-8.4*	none
Skin melanoma	146	*-4.3*	171	*-2.7*	186	*-2.4*	213	*-2.4*	>10
Larynx	103	*-2.7*	97	*-3.6*	79	*-2.2*	95	*-1.5*	>10
Stomach	100	*-5.2*	106	*-4.2*	67	*-3.5*	96	*-2.8*	>10
Head and Neck	93	*-3.6*	85	*-2.4*	108	*-1.3*	89	*-0.6*	>10
Thyroid	89	*0.7*	79	*-0.9*	53	*-0.8*	73	*-3.9*	>10 *
Hodgkin lymphoma	70	*-10.6*	66	*-9.1*	59	*-8.1*	59	*-6.7*	none
CLL/SLL	55	*-4.2*	58	*-2.0*	64	*-1.4*	55	*-0.9*	>10
Liver	47	*-3.1*	49	*-1.1*	28	*-1.0*	36	*-0.7*	>10
Brain	45	*-14.4*	47	*-10.6*	42	*-9.7*	54	*-8.5*	none
Multiple myeloma	37	*-3.0*	39	*-1.3*	36	*-0.4*	38	*-0.2*	>10
Soft tissues	35	*-15.7*	37	*-13.8*	36	*-13.2*	37	*-10.7*	none
Pancreas	22	*-3.5*	24	*-0.4*	23	*0.2*	21	*-0.1*	>10
Oesophagus	18	*-2.0*	20	*-0.8*	28	*-0.4*	18	*0.0*	>10
Penis	19	*-4.5*	18	*-4.5*	17	*-3.6*	19	*-3.3*	>10
AML	16	*-1.6*	21	*-3.0*	16	*-3.2*	15	*-1.9*	>10
Bones	15	*-20.6*	18	*-20.7*	16	*-19.2*	18	*-18.3*	none
Gallbladder	13	*-4.1*	11	*-2.3*	10	*-0.7*	12	*-0.3*	>10
CML	11	*1.6*	13	*-0.7*	14	*-0.7*	12	*-0.2*	>10

CLL/SLL, Chronic lymphocytic leukaemia/small lymphocytic lymphoma; AML, Acute myeloid leukaemia; CML, Chronic myeloid leukaemia.
* for thyroid cancer values of PRD grow as registration length increases and for female thyroid cancer PRD slightly exceeds 5% after 25 years of registration (consistently with Empirical vs EU model-based R-index patterns).
Weighted average percent relative difference (PRD, %) between **empirical** and **standard model-based** complete prevalence estimates (the average is weighted using registries population). Registration time length with PRD values not exceeding 5%.

profile and prognosis act as modulating factors. For tumours mainly diagnosed in the elderly, EU empirical and EU model-based R-indexes led to similar results (within an average tolerance of 5%) when applied to prevalence data observed for at least 10 years.

By contrast, the empirical method underestimates very long-term survivorship for tumours with early age at onset, even when based on 35 years of observations. For this specific class of neoplasms, model-based methods are structurally more suited to

10.3389/fonc.2023.1114701

TABLE 3B Comparison between empirical and standard model-based complete prevalence at 1/1/2013 by cancer site for the 62 European registries included in the study grouped by registration time length (from 10-14 years to 25-35 years), Women.

Registration time interval in years (number of registries)	10-14 y (10)		15-19 y (12)		20-24 y (13)		25-35 y (17)		Registration time length with PRD ≤ 5%
Cancer Site	Standard complete prev	PRD	Standard complete prev	PRD	Standard complete prev	PRD	Standard complete prev	PRD	
	Prop *100,000	%	Prop *100,000	%	Prop *100,000	%	Prop *100,000	%	*number of years*
All cancers	4,380	*-8.4*	4,388	*-6.7*	4,305	*-5.4*	4,369	*-4.1*	>20
Breast	1,700	*-4.8*	1,831	*-4.1*	1,869	*-3.0*	1,777	*-2.0*	>10
Colon Rectum	477	*-4.0*	485	*-3.0*	443	*-2.4*	497	*-1.7*	>10
Thyroid	371	*-0.4*	278	*-2.2*	217	*-3.1*	257	*-6.1*	>10 *
Corpus Uteri	341	*-5.4*	305	*-4.0*	328	*-2.8*	343	*-2.6*	>10
Cervix Uteri	268	*-21.1*	225	*-20.2*	258	*-20.5*	230	*-21.3*	none
Non-Hodgkin lymphoma	203	*-6.2*	209	*-5.6*	207	*-5.4*	212	*-4.5*	>20
Skin melanoma	185	*-8.2*	232	*-5.8*	262	*-5.4*	289	*-5.4*	>20
Ovary	157	*-9.2*	145	*-9.8*	153	*-8.9*	151	*-8.4*	none
Bladder	114	*-4.4*	120	*-4.2*	77	*-3.6*	124	*-3.2*	>10
Kidney	115	*-7.1*	111	*-4.6*	109	*-4.0*	110	*-2.8*	>15
Lung	85	*-4.7*	90	*-2.8*	81	*-1.1*	100	*-0.5*	>10
Stomach	79	*-4.6*	77	*-5.0*	46	*-5.2*	75	*-5.1*	>10
Hodgkin lymphoma	61	*-12.3*	57	*-13.2*	49	*-11.0*	52	*-9.1*	none
CLL/SLL	41	*-2.4*	40	*-0.9*	48	*-0.4*	42	*-0.2*	>10
Brain	37	*-14.4*	38	*-9.8*	32	*-9.5*	47	*-8.0*	none
Head and Neck	39	*-5.8*	39	*-4.4*	42	*-2.9*	44	*-2.0*	>15
Multiple myeloma	32	*-2.2*	33	*-1.1*	31	*-0.5*	34	*-0.1*	>10
Soft tissues	31	*-18.4*	31	*-16.6*	31	*-16.5*	33	*-14.6*	none
Liver	20	*-5.5*	17	*-2.5*	9	*-2.2*	15	*-2.3*	>15
Pancreas	20	*-3.4*	22	*-0.3*	20	*0.2*	20	*-0.1*	>10
Gallbladder	17	*-1.5*	14	*-0.6*	11	*-0.4*	13	*-0.7*	>10
AML	15	*-6.7*	19	*-6.1*	18	*-4.9*	15	*-4.1*	>20
Bones	14	*-15.0*	15	*-18.3*	15	*-18.2*	14	*-17.3*	none
Larynx	11	*-2.6*	12	*-4.5*	11	*-2.9*	11	*-2.5*	>10
CML	9	*-1.8*	10	*-1.2*	10	*-0.2*	9	*-0.5*	>10
Oesophagus	6	*-1.9*	7	*-0.8*	9	*-0.3*	7	*-0.5*	>10

CLL/SLL, Chronic lymphocytic leukaemia/small lymphocytic lymphoma; AML, Acute myeloid leukaemia; CML, Chronic myeloid leukaemia.

* for thyroid cancer values of PRD grow as registration length increases and for female thyroid cancer PRD slightly exceeds 5% after 25 years of registration (consistently with Empirical vs EU model-based R-index patterns).

Weighted average percent relative difference (PRD, %) between **empirical** and **standard model-based** complete prevalence estimates (the average is weighted using registries population). Registration time length with PRD values not exceeding 5%.

TABLE 4A Comparison between SEER and standard model-based complete prevalence at 1/1/2013 by cancer site for the 62 European registries included in the study grouped by registration time length (from 10-14 years to 25-35 years), Men.

Registration time interval in years (number of registries)	10-14 y (10)		15-19 y (12)		20-24 y (13)		25-35 y (17)		Registration time length with PRD ≤ 5%
Cancer site	Standard complete prev	PRD	Standard complete prev	PRD	Standard complete prev	PRD	Standard complete prev	PRD	
	Prop *100,000	%	Prop *100,000	%	Prop *100,000	%	Prop *100,000	%	*number of years*
All cancers	3,476	*-1.2*	3,757	*-1.1*	3,747	*-1.3*	3,692	*-0.7*	>10
Prostate	959	*2.3*	1,190	*1.3*	1,450	*0.9*	1,225	*0.2*	>10
Colon Rectum	539	*8.7*	565	*5.3*	501	*3.8*	550	*1.3*	>15
Bladder	453	*7.1*	456	*4.1*	281	*2.6*	462	*0.6*	>15
Non-Hodgkin lymphoma	232	*-9.8*	251	*-5.0*	247	*-4.5*	245	*-2.4*	>15
Lung	195	*8.0*	197	*5.0*	180	*4.8*	171	*2.3*	>15
Kidney	182	*6.3*	203	*4.7*	173	*3.6*	184	*2.0*	>15
Testis	154	*7.8*	160	*5.7*	166	*3.8*	167	*1.4*	>20
Skin melanoma	146	*6.9*	171	*4.4*	186	*2.6*	213	*1.1*	>15
Larynx	103	*8.9*	97	*7.2*	79	*4.1*	95	*1.3*	>20
Stomach	100	*-6.8*	106	*-4.7*	67	*-2.5*	96	*-0.8*	>15
Thyroid	89	*20.2*	79	*14.4*	53	*10.6*	73	*5.3*	>20
Hodgkin lymphoma	70	*2.3*	66	*3.8*	59	*4.3*	59	*3.3*	>10
CLL/SLL	55	*-1.3*	58	*-0.5*	64	*-1.1*	55	*-0.8*	>10
Liver	47	*-0.4*	49	*-0.3*	28	*-0.6*	36	*-0.3*	>10
Brain	45	*3.2*	47	*4.4*	42	*4.6*	54	*3.1*	none *
Multiple myeloma	37	*-4.2*	39	*-1.9*	36	*-0.8*	38	*-0.1*	>10
Soft tissue	35	*2.6*	37	*2.4*	36	*0.8*	37	*-0.2*	>10
Pancreas	22	*2.1*	24	*1.9*	23	*1.7*	21	*0.7*	>10
Oesophagus	18	*4.9*	20	*2.1*	28	*1.6*	18	*0.4*	>10
AML	16	*-0.1*	21	*4.1*	16	*4.9*	15	*4.6*	>10
Bones	15	*1.0*	18	*-0.7*	16	*-1.6*	18	*-3.3*	>10
Gallbladder	13	*4.6*	11	*2.7*	10	*2.0*	12	*0.8*	>10
CML	11	*-8.5*	13	*-0.3*	14	*1.9*	12	*1.9*	>15

CLL/SLL, Chronic lymphocytic leukaemia/small lymphocytic lymphoma; AML, Acute myeloid leukaemia; CML, Chronic myeloid leukaemia.
* PRD for male Brain cancer reflect compensations of under/over-estimates between SEER and Standard R-index.
Weighted average percent relative difference (PRD, %) between **SEER** and **standard model-based** complete prevalence estimates (the average is weighted using registries population). Registration time length with PRD values not exceeding 5%.

capture unobserved survivors in the very long term. This limitation is also reflected in the estimation of all cancers that include a non-negligible proportion of juvenile cancers.

Using model-based completeness indexes derived from external rather than local patients' populations (SEER versus European) led to comparable prevalence estimates for the majority of cancers, even when applied to minimum registration periods (10 years). The list includes also most of the early onset tumours and, as a consequence, the complex of all cancer sites. Notable discrepancies were instead observed as a result of geographical differences in cancer incidence and survival patterns, regardless of the natural history of the disease (age at onset and prognosis). This, for instance, is the case of endometrial and thyroid cancers, or of brain tumours, as the inclusion criteria of non-malignant entities may vary between SEER and European registries, thus affecting the consistency of estimates.

TABLE 4B Comparison between SEER and standard model-based complete prevalence at 1/1/2013 by cancer site for the 62 European registries included in the study grouped by registration time length (from 10-14 years to 25-35 years), Women.

Registration time interval in years (number of registries)	10-14 y (10)		15-19 y (12)		20-24 y (13)		25-35 y (17)		Registration time length with PRD ≤ 5%
Cancer site	Standard complete prev	PRD	Standard complete prev	PRD	Standard complete prev	PRD	Standard complete prev	PRD	
	Prop *100,000	%	Prop *100,000	%	Prop *100,000	%	Prop *100,000	%	*number of years*
All cancers	4,380	*2.8*	4,388	*2.8*	4,305	*2.5*	4,369	*1.6*	>10
Breast	1,700	*0.1*	1,831	*0.5*	1,869	*0.6*	1,777	*0.3*	>10
Colon Rectum	477	*6.2*	485	*4.3*	443	*3.5*	497	*1.6*	>15
Thyroid	371	*7.1*	278	*6.2*	217	*5.3*	257	*3.0*	>20
Corpus Uteri	341	*22.3*	305	*14.9*	328	*9.6*	343	*3.8*	>25
Cervix Uteri	268	*-5.8*	225	*-2.1*	258	*-1.0*	230	*0.6*	>15
Non-Hodgkin lymphoma	203	*-9.1*	209	*-5.5*	207	*-4.9*	212	*-2.9*	>20
Skin melanoma	185	*6.6*	232	*4.9*	262	*3.7*	289	*1.9*	>15
Ovary	157	*4.1*	145	*3.9*	153	*3.7*	151	*2.6*	>10
Kidney	115	*0.9*	111	*1.4*	109	*1.4*	110	*0.7*	>10
Bladder	114	*5.2*	120	*3.2*	77	*2.2*	124	*0.5*	>10
Lung	85	*6.3*	90	*4.3*	81	*3.2*	100	*1.3*	>15
Stomach	79	*-13.7*	77	*-10.2*	46	*-7.4*	75	*-3.6*	>25
Hodgkin lymphoma	61	*2.0*	57	*2.9*	49	*3.0*	52	*2.2*	>10
CLL/SLL	41	0.0	40	0.9	48	0.6	42	0.3	>10
Brain	37	*13.3*	38	*14.8*	32	*13.1*	47	*7.7*	none
Multiple myeloma	32	*-1.6*	33	*0.8*	31	*1.3*	34	*0.7*	>10
Soft tissues	31	*-1.6*	31	*-2.2*	31	*-2.4*	33	*-1.9*	>10
Pancreas	20	*2.6*	22	*2.8*	20	*2.3*	20	*1.0*	>10
Liver	20	*7.1*	17	*4.5*	9	*3.7*	15	*1.2*	>15
Gallbladder	17	*-1.5*	14	*0.3*	11	*0.7*	13	*0.5*	>10
AML	15	*-2.5*	19	*1.9*	18	*3.3*	15	*2.8*	>10
Bones	14	*7.2*	15	*2.8*	15	*0.9*	14	*-3.0*	>15
Larynx	11	*1.4*	12	*2.0*	11	*0.8*	11	*0.1*	>10
CML	9	*-13.7*	10	*-3.9*	10	*-0.7*	9	*1.3*	>15
Oesophagus	6	*1.3*	7	*1.3*	9	*1.6*	7	*0.7*	>10

CLL/SLL, Chronic lymphocytic leukaemia/small lymphocytic lymphoma; AML, Acute myeloid leukaemia; CML, Chronic myeloid leukaemia.
Weighted average percent relative difference (PRD, %) between **SEER** and **standard model-based** complete prevalence estimates (the average is weighted using registries population). Registration time length with PRD values not exceeding 5%.

The results we obtained were coherent with the patterns of the relevant factors influencing cancer prevalence, e.g. age at prevalence date, low to high cancer prognosis, incidence age profile, length of the registration time period.

European model-based R-index values were slightly higher than those estimated from SEER data consistently with the prognostic differences between European and USA cancer patients, the latter generally reported to present more favourable survival levels (31). Differences are also partly due to incidence modelling choices. SEER R-indexes were indeed often derived by adopting exponential rather than polynomial incidence models (4). Finally, differences between IARC and SEER rules for identifying multiple primary tumours could also have an impact.

TABLE 5 Summary table reporting the registration time length (years) associated to comparable complete prevalence estimates (within a tolerance lower than 5%) between alternative (Empirical or SEER model-based) and standard completeness index method.

Registration time length (years)	Sex	European Empirical	SEER Model Based
>10	**M**	Bladder; Colon Rectum; Gallbladder; Head and Neck; Larynx; Liver; Multiple myeloma; Oesophagus; Pancreas; Penis; Prostate; Skin melanoma; Stomach; Thyroid; CLL/SLL; AML; CML	All cancers; Bones; Gallbladder; Hodgkin lymphoma; Liver; Multiple myeloma; Oesophagus; Pancreas; Prostate; Soft tissue; CLL/SLL; AML
	F	Bladder; Breast; Colon Rectum; Corpus Uteri; Gallbladder; Larynx; Lung; Multiple myeloma; Oesophagus; Pancreas; Stomach; Thyroid; CLL/SLL; CML	All cancers; Bladder; Breast; Gallbladder; Hodgkin's lymphoma; Kidney; Larynx; Multiple myeloma; Oesophagus; Ovary; Pancreas; Soft tissues; CLL/SLL; AML
>15	**M**	All cancers; Kidney; Lung; Non-Hodgkin lymphoma;	Bladder; Colon Rectum; Kidney; Lung; Non-Hodgkin lymphoma; Skin melanoma; Stomach; CML
	F	Head and Neck; Kidney	Bones; Cervix Uteri; Colon Rectum; Liver; Lung; Skin melanoma; CML
>20	**M**		Larynx; Testis; Thyroid
	F	All cancers; Non-Hodgkin lymphoma; Skin melanoma; AML	Non-Hodgkin lymphoma; Thyroid
>25	**M**		
	F		Corpus Uteri; Stomach
None	**M**	Bones; Brain; Hodgkin lymphoma; Soft tissues; Testis	Brain
	F	Bones; Brain; Cervix Uteri; Hodgkin lymphoma; Ovary; Soft tissues	Brain

CLL/SLL, Chronic lymphocytic leukaemia/small lymphocytic lymphoma; AML, Acute myeloid leukaemia; CML, Chronic myeloid leukaemia.
Results obtained from the 62 European registries included in the study.

Parametric mixed cure models of Weibull type were used for modelling survival (1, 2). More flexible cure fraction models could have been considered (32, 33) but the choice is limited to Weibull or exponential types in the COMPREV software.

The empirical indexes were derived by pooling data of 8 European registries with available 35-year observed prevalence at the index date. The limit at 35-years is arbitrary and just reflects the maximum available time span in the EUROCARE-6 dataset. However, it has been proven to provide a sufficient basis to estimate complete cancer prevalence for major cancers and for a variety of less frequent tumours with late age at onset. Lower values might be critical and extending this limit in applications to more recent prevalence index dates is advisable, considering the continuous progresses of cancer survival over time and the availability of longer registration time series.

Empirical indexes were subject to random fluctuations when based on sparse cases, for instance in correspondence of young age at prevalence date for tumours at late onset like pancreatic or prostatic cancer. However, such fluctuations are of scarce practical relevance because the index is applied to values of observed prevalence which are almost null in these circumstances.

R-indexes were generally positively validated on a fully independent dataset of 20 registries, therefore showing that the estimation datasets used to derive model-based completeness indexes were sufficiently representative of the prevalence patterns in other European populations. However, we cannot exclude that for some neoplasms the geographical heterogeneity of incidence or prognosis may have required area-specific R-indexes.

Notably the empirical completeness R-indexes are easy to compute but inevitably refer to a specific point in time (the index date of the maximum observable cancer prevalence). Thus they must be computed on a date which is reasonably close to the index date of the limited-duration prevalence we want to complete.

Conversely model-based R-indexes require higher computational effort to model incidence and relative survival trends, but they dynamically evolve over time (the period of diagnosis is parameterised in the models) and R-index values for varying prevalence index dates can be derived through the Comprev software (4).

In conclusion, the study tests the feasibility of using alternative formulations of the completeness index method to integrate limited-duration prevalence measured by population-based cancer registries. We focused on the European context where the lack of systematic data on the overall number of cancer survivors in many countries hinders the planning of health services and particularly survivorship care planning. This appears even more limiting in light of the future scenario in which the population of cancer survivors is indeed expected to increase significantly due to ongoing demographic changes and continued advances in therapies and diagnosis. Our results may facilitate the use and dissemination of complete cancer prevalence estimates across Europe and help to close the present information gaps.

Data availability statement

The datasets of empirical and standard model-based completeness indexes presented in this article are not readily available. Requests to access the datasets should be directed to RA, roberta.deangelis@iss.it.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

ED carried out the study and analysed the incidence and prevalence data. SR did quality checks, prepared the SEER*Stat study database and analysed the survival data. LV analysed the survival data. CB implemented the procedures to check the raw data and to generate the SEER*Stat study database. LM, SG, AK, SL, and VJ provided advice and revised the results. RA drafted the article, designed the study and the data quality checks. The EUROCARE-6 Working Group collected, prepared, and transmitted raw data for the study database, corrected data after quality controls, checked the results of the analyses and revised the final draft of the article. All authors interpreted results. All authors contributed to the article and approved the submitted version.

EUROCARE-6 working group

Austria: M. Hackl *(National CR)*; Belgium: E. Van Eycken; N. Van Damme *(National CR)*; Bulgaria: Z. Valerianova *(National CR)*; Croatia: M. Sekerija *(National CR)*; Cyprus: V. Scoutellas; A. Demetriou *(National CR)*; Czech Republic: L. Dušek; D. Krejici *(National CR)*; Denmark: H. Storm *(National CR)*; Estonia: M. Mägi; K. Innos* *(National CR)*; Finland: N. Malila; J. Pitkäniemi *(National CR)*; France: M. Velten *(Bas Rhin CR)*; X. Troussard *(Basse Normandie, Haematological Malignancies CR)*; A.M. Bouvier; V. Jooste* *(Burgundy, Digestive CR)*; A.V. Guizard *(Calvados, General CR)*; G. Launoy *(Calvados, Digestive CR)*; S. Dabakuyo Yonli *(Cote dOr, Gynaecologic (Breast) CR)*; M. Maynadié *(Cote dOr, Haematological Malignancies CR)*; A.S. Woronoff *(Doubs CR)*; J.B. Nousbaum *(Finistere, Digestive CR)*; G. Coureau *(Gironde, General CR)*; A. Monnereau* *(Gironde, Haematological Malignancies CR)*; I. Baldi *(Gironde, Central Nervous System CR)*; K. Hammas *(Haut-Rhin CR)*; B. Tretarre *(Herault CR)*; M. Colonna *(Isere CR)*; S. Plouvier *(Lille Area CR)*; T. D'Almeida *(Limousin CR)*; F. Molinié; A. Cowppli-Bony *(Loire-Atlantique/Vendée CR)*; S. Bara *(Manche CR)*; A. Debreuve *(Marne-Ardennes, Thyroid CR)*; G. Defossez *(Poitou-Charentes CR)*; B. Lapôtre-Ledoux *(Somme CR)*; P. Grosclaude; L. Daubisse-Marliac *(Tarn CR)*; Germany: S. Luttmann *(Bremen CR)*; R. Stabenow *(Common CR of 4 Federal States (Brandenburg, Mecklenburg-West Pomerania, Saxony-Anhalt, Thüringen))*; A. Nennecke *(Hamburg CR)*; J. Kieschke *(Lower Saxony CR)*; S. Zeissig *(Rhineland-Palatinate CR)*; B. Holleczek *(Saarland CR)*; A. Katalinic* *(Schleswig-Holstein CR)*; Iceland: H. Birgisson *(National CR)*; Ireland: D. Murray; P.M. Walsh *(National CR)*; Italy: G. Mazzoleni; F. Vittadello *(Alto Adige CR)*; F. Cuccaro *(Barletta-Andria-Trani CR)*; R. Galasso *(Basilicata CR)*; G. Sampietro *(Bergamo CR)*; S. Rosso *(Biella CR)*; C. Gasparotti; G. Maifredi *(Brescia CR)*; M. Ferrante; R. Ragusa *(Catania-Messina-Enna CR)*; A. Sutera Sardo *(Catanzaro CR)*; M.L. Gambino; M. Lanzoni *(Province of Varese and Como CR)*; P. Ballotari; E. Giacomazzi *(Cremona and Mantova CR)*; S. Ferretti *(Ferrara CR)*; A. Caldarella; G. Manneschi *(Firenze-Prato CR)*; G. Gatta*; M. Sant*; P. Baili*; F. Berrino*; L. Botta; A. Trama; R. Lillini; A. Bernasconi; S. Bonfarnuzzo; C. Vener; F. Didonè; P. Lasalvia; G. Del Monego; L. Buratti; G. Tagliabue *(Fondazione IRCCS Istituto Nazionale dei Tumori, Milan)*; D. Serraino; L. Dal Maso *(Centro di Riferimento Oncologico, IRCCS, Aviano for the Friuli Venezia Giulia CR)*; R. Capocaccia* *(Epidemiologia & Prevenzione Board)*; R. De Angelis*; E. Demuru; C. Di Benedetto; S. Rossi*; M. Santaquilani; S. Venanzi; M. Tallon *(Istituto Superiore di Sanità, Rome)*; L. Boni *(Genova CR)*; S. Iacovacci *(Latina CR)*; V. Gennaro *(Liguria, mesotheliomas CR)*; A.G. Russo; F. Gervasi *(Province of Milan and Lodi CR)*; G. Spagnoli *(Modena CR)*; L. Cavalieri d'Oro *(Monza and Brianza CR)*; M. Fusco; M.F. Vitale *(Napoli 3 South CR)*; M. Usala *(Nuoro CR)*; W. Mazzucco *(Palermo CR)*; M. Michiara *(Parma CR)*; G. Chiranda *(Piacenza CR)*; G. Cascone; C.P. Rollo *(Ragusa CR)*; L. Mangone *(Reggio Emilia CR)*; F. Falcini *(Romagna CR)*; R. Cavallo *(Salerno CR)*; D. Piras *(Sassari CR)*; A. Madeddu; F. Bella *(Siracusa CR)*; A.C. Fanetti *(Sondrio CR)*; S. Minerba *(Taranto CR)*; G. Candela; T. Scuderi *(Trapani CR)*; R.V. Rizzello *(Trento CR)*; F. Stracci *(Umbria CR)*; M. Rugge *(Veneto CR)*; A. Brustolin *(Viterbo CR)*; Latvia: S. Pildava *(National CR)*; Lithuania: G. Smailyte *(National CR)*; Malta: M. Azzopardi *(National CR)*; Norway: T.B. Johannesen* *(National CR)*; Poland: J. Didkowska; U. Wojciechowska *(National CR)*; M. Bielska-Lasota* *(National Institute of Public Health-National Institute of Hygiene-National Research Institute, Warsaw)*; Portugal: A. Pais *(Central Portugal CR)*; J. Rodrigues; M.J. Bento *(Northern Portugal CR)*; A. Miranda *(Southern Portugal CR)*; Slovakia: C. Safaei Diba *(National CR)*; Slovenia: V. Zadnik; T. Zagar (National CR); Spain: C. Sánchez-Contador Escudero; P. Franch Sureda *(Balearic Islands, Mallorca CR)*; A. Lopez de Munain; M. De-La-Cruz *(Basque Country CR)*; M.D. Rojas; A. Aleman *(Canary Islands CR)*; A. Vizcaino *(Castellon CR)*; R. Marcos-Gragera; A. Sanvisens *(Girona CR)*; M.J. Sanchez *(Granada CR)*; M.D. Chirlaque; A. Sanchez-Gil *(Murcia CR)*; M. Guevara*; E. Ardanaz *(Navarra CR, CIBERESP)*; A. Ameijide; M. Carulla *(Tarragona CR)*; Switzerland: Y. Bergeron *(Fribourg CR)*; C. Bouchardy *(Geneva CR)*; S. Mohsen Mousavi; P. Went *(Graubünden and Glarus CR)*; S. Mohsen Mousavi; M. Blum *(Eastern Switzerland CR)*; A. Bordoni *(Ticino CR)*; The Netherlands: O. Visser* *(National CR)*; UK-England: S. Stevens; J. Broggio *(National CR)*; UK-Northern Ireland: A. Gavin* *(National*

CR); UK-Scotland: D. Morrison *(National CR)*; UK-Wales: D. W. Huws* *(National CR)*. *EUROCARE Steering Committee.

Funding

This study was funded by the European Commission (Grant Agreement no. 801520 HP-JA-2017, *Innovative Partnership for Action Against Cancer*, iPAAC Joint Action) and by the Italian Association for Cancer Research (AIRC) (Grant no. 21879). The funding agency played no role in designing the study, collecting, analyzing or interpreting the data, writing the report, or deciding whether or not to submit the article for publication.

Acknowledgments

The authors would like to thank Tonino Sofia for the English-language revision.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1114701/full#supplementary-material

References

1. Capocaccia R, De Angelis R. Estimating the completeness of prevalence based on cancer registry data. *Stat Med* (1997) 16(4):425–40. doi: 10.1002/(SICI)1097-0258(19970228)16:4<425::AID-SIM414>3.0.CO;2-Z

2. Merrill RM, Capocaccia R, Feuer EJ, Mariotto A. Cancer prevalence estimates based on tumour registry data in the surveillance, epidemiology, and end results (SEER) program. *Int J Epidemiol* (2000) 29(2):197–207. doi: 10.1093/ije/29.2.197

3. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, et al. *SEER cancer statistics review, 1975-2017, National Cancer Institute*. Bethesda, MD (2020). Available at: https://seer.cancer.gov/csr/1975_2017/.

4. *Surveillance research program. national cancer institute COMPREV software*. Available at: https://surveillance.cancer.gov/comprev/.

5. *Global cancer observatory*. Available at: https://gco.iarc.fr.

6. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int J Cancer* (2013) 132(5):1133–45. doi: 10.1002/ijc.27711

7. Mallone S, De Angelis R, van der Zwan JM, Trama A, Siesling S, Gatta G, et al. Methodological aspects of estimating rare cancer prevalence in Europe: the experience of the RARECARE project. *Cancer Epidemiol* (2013) 37(6):850–6. doi: 10.1016/j.canep.2013.08.001

8. Gatta G, van der Zwan JM, Casali PG, Siesling S, Dei Tos AP, Kunkler I, et al. Rare cancers are not so rare: the rare cancer burden in Europe. *Eur J Cancer* (2011) 47(17):2493–511. doi: 10.1016/j.ejca.2011.08.008

9. *Rarecare-net project, work package on rare cancers epidemiology*. Available at: http://rarecarenet.istitutotumori.mi.it/rarecarenet/.

10. Micheli A, Mugno E, Krogh V, Quinn MJ, Coleman M, Hakulinen T, et al. *Cancer prevalence European registry areas. Ann Oncol* (2002) 13(6):840–65. doi: 10.1093/annonc/mdf127

11. Capocaccia R, Colonna M, Corazziari I, De Angelis R, Francisci S, Micheli A. Mugno e on behalf of the EUROPREVAL working group. Measuring cancer prevalence in Europe: the EUROPREVAL project. *Ann Oncol* (2002) 13(6):831–9. doi: 10.1093/annonc/mdf152

12. Guzzinati S, Virdone S, De Angelis R, Panato C, Buzzoni C, Capocaccia R, et al. Characteristics of people living in Italy after a cancer diagnosis in 2010 and projections to 2020. *BMC Cancer* (2018) 18(1):169. doi: 10.1186/s12885-018-4053-y

13. AIRTUM Working Group. Italian Cancer figures, report 2014: Prevalence and cure of cancer in Italy. *Epidemiol Prev* (2014) 38(6 Suppl 1):1–122. doi: 10.19191/EP14.6.S1.113

14. Haberland J, Bertz J, Wolf U, Ziese T, Kurth BM. German Cancer statistics 2004. *BMC Cancer* (2010) 10:52. doi: 10.1186/1471-2407-10-52

15. Colonna M, Boussari O, Cowppli-Bony A, Delafosse P, Romain G, Grosclaude P, et al. Time trends and short term projections of cancer prevalence in France. *Cancer Epidemiol* (2018) 56:97–105. doi: 10.1016/j.canep.2018.08.001

16. Colonna M, Mitton N, Bossard N, Belot A, Grosclaude PFrench Network of Cancer Registries (FRANCIM). Total and partial cancer prevalence in the adult French population in 2008. *BMC Cancer* (2015) 15:153. doi: 10.1186/s12885-015-1168-2

17. Maddams J, Utley M, Møller H. Projections of cancer prevalence in the united kingdom, 2010-2040. *Br J Cancer* (2012) 107(7):1195–202. doi: 10.1038/bjc.2012.366

18. Rossi S, Crocetti E, Capocaccia R, Gatta GAIRTUM Working Group. Estimates of cancer burden in Italy. *Tumori* (2013) 99(3):416–24. doi: 10.1700/1334.14807

19. Herrmann C, Cerny T, Savidan A, Vounatsou P, Konzelmann I, Bouchardy C, et al. Cancer survivors in Switzerland: a rapidly growing population to care for. *BMC Cancer* (2013) 13:287. doi: 10.1186/1471-2407-13-287

20. *NORDCAN online data tool, version 9.2*. Available at: https://nordcan.iarc.fr/en/dataviz.

21. *Slovenian cancer registry annual reports*. Available at: https://www.onko-i.si/eng/crs.

22. de Moor JS, Mariotto AB, Parry C, Alfano CM, Padgett L, Kent EE, et al. Cancer survivors in the united states: prevalence across the survivorship trajectory and implications for care. *Cancer Epidemiol Biomarkers Prev* (2013) 22(4):561–70. doi: 10.1158/1055-9965.EPI-12-1356

23. Hao S, Östensson E, Eklund M, Grönberg H, Nordström T, Heintz E, et al. The economic burden of prostate cancer - a Swedish prevalence-based register study. *BMC Health Serv Res* (2020) 20(1):448. doi: 10.1186/s12913-020-05265-8

24. Sharp L, Deady S, Gallagher P, Molcho M, Pearce A, Alforque Thomas A, et al. The magnitude and characteristics of the population of cancer survivors: Using population-based estimates of cancer prevalence to inform service planning for survivorship care. *BMC Cancer* (2014) 14:767. doi: 10.1186/1471-2407-14-767

25. Bluethmann SM, Mariotto AB, Rowland JH. Anticipating the "Silver tsunami": Prevalence trajectories and comorbidity burden among older cancer survivors in the united states. *Cancer Epidemiol Biomarkers Prev* (2016) 25(7):1029–36. doi: 10.1158/1055-9965.EPI-16-0133

26. Dal Maso L, Santoro A, Iannelli E, De Paoli P, Minoia C, Pinto M, et al. Cancer cure and consequences on survivorship care: Position paper from the Italian alliance against cancer (ACC) survivorship care working group. *Cancer Manag Res* (2022) 14:3105–18. doi: 10.2147/CMAR.S380390

27. Colonna M, Grosclaude P, Bouvier AM, Goungounga J, Jooste V. Health status of prevalent cancer cases as measured by mortality dynamics (cancer vs. noncancer): Application to five major cancers sites. *Cancer* (2022) 128(20):3663–73. doi: 10.1002/cncr.34413

28. *iPAAC joint action, work package 7 on cancer information and registries*. Available at: https://www.ipaac.eu/en/work-packages/wp7/.

29. *European Network of cancer registries (ENCR) call for data protocol 201*. Available at: https://encr.eu/call-for-data.

30. *Surveillance research program, national cancer institute SEER*Stat software version 8.3.9.2.* Available at https://seer.cancer.gov/seerstat/.v.

31. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz MCONCORD-3 Working Group. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): Analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countrie. *Lancet* (2018) 391(10125):1023–75. doi: 10.1016/S0140-6736(17)33326-3

32. Andersson TM, Dickman PW, Eloranta S, Lambert PC. Estimating and modelling cure in population-based cancer studies within the frame-work of flexible parametric survival models. *BMC Med Res Methodol* (2011) 11:96. doi: 10.1186/1471-2288-11-96

33. Yu XQ, De Angelis R, Andersson TML, Lambert PC, O'Connell DL, Dickman PW. Estimating the proportion cured of cancer: Some practical advice for users. *Cancer Epidemiol* (2013) 37(6):836–42. doi: 10.1016/J.CANEP.2013.08.014

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 25 May 2023
DOI 10.3389/fonc.2023.1197942

OPEN ACCESS

EDITED BY
Otto Visser,
Integral Cancer Center Netherlands (IKNL), Netherlands

REVIEWED BY
Bart Van Gool,
Belgian Cancer Registry, Belgium
Matthias Lorez,
National Institute for Cancer Epidemiology and Registration (NICER), Switzerland

*CORRESPONDENCE
Giovanna Tagliabue
giovanna.tagliabue@istitutotumori.mi.it

RECEIVED 31 March 2023
ACCEPTED 15 May 2023
PUBLISHED 25 May 2023

CITATION
Tagliabue G, Perotti V, Fabiano S, Tittarelli A, Barigelletti G, Contiero P, Mazzucco W, Fusco M, Bidoli E, Vicentini M, Pesce MT, Stracci F and The Collaborative Working Group (2023) Comparison between two cancer registry quality check systems: functional features and differences in an Italian network of cancer registries dataset.
Front. Oncol. 13:1197942.
doi: 10.3389/fonc.2023.1197942

Comparison between two cancer registry quality check systems: functional features and differences in an Italian network of cancer registries dataset

Giovanna Tagliabue[1]*, Viviana Perotti[1], Sabrina Fabiano[1], Andrea Tittarelli[1], Giulio Barigelletti[1], Paolo Contiero[2], Walter Mazzucco[3,4], Mario Fusco[5], Ettore Bidoli[6], Massimo Vicentini[7], Maria Teresa Pesce[8], Fabrizio Stracci[4,9] and The Collaborative Working Group

[1]Cancer Registry Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, [2]Environmental Epidemiology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, [3]Clinical Epidemiology Unit and Palermo Province Cancer Registry, University Hospital "P. Giaccone", Palermo, Italy, [4]Department of Oncology and Public Health, Executive Board of the Italian Network of Cancer Registries (AIRTUM), Milan, Italy, [5]Cancer Registry Unit, ASL Napoli 3 Sud, Naples, Italy, [6]Cancer Epidemiology Unit, Centro di Riferimento Oncologico (CRO), IRCCS, Aviano, Italy, [7]Epidemiology Unit, Azienda Unità Sanitaria Locale - IRCCS di Reggio Emilia, Reggio Emilia, Italy, [8]Cancer Registry Unit, ASL Caserta, Caserta, Italy, [9]Umbria Regional Cancer Registry, Department of Medicine and Surgery, University of Perugia, Perugia, Italy

Purpose: The aim of this study was to compare the functional characteristics of two computer-based systems for quality control of cancer registry data through analysis of their output differences.

Methods: The study used cancer incidence data from 22 of the 49 registries of the Italian Network of Cancer Registries registered between 1986 and 2017. Two different data checking systems developed by the WHO International Agency for Research on Cancer (IARC) and the Joint Research Center (JRC) with the European Network of Cancer Registries (ENCR) and routinely used by registrars were used to check the quality of the data. The outputs generated by the two systems on the same dataset of each registry were analyzed and compared.

Results: The study included a total of 1,305,689 cancer cases. The overall quality of the dataset was high, with 86% (81.7-94.1) microscopically verified cases and only 1.3% (0.03-3.06) cases with a diagnosis by death certificate only. The two check systems identified a low percentage of errors (JRC-ENCR 0.17% and IARC 0.003%) and about the same proportion of warnings (JRC-ENCR 2.79% and IARC 2.42%) in the dataset. Forty-two cases (2% of errors) and 7067 cases (11.5% of warnings) were identified by both systems in equivalent categories. 11.7% of warnings related to TNM staging were identified by the JRC-ENCR system only. The IARC system identified mainly incorrect combination of tumor grade and morphology (72.5% of warnings).

Conclusion: Both systems apply checks on a common set of variables, but some variables are checked by only one of the systems (for example, checks on patient follow-up and tumor stage at diagnosis are included by the JRC-ENCR system only). Most errors and warnings were categorized differently by the two systems, but usually described the same issues, with warnings related to "morphology" (JRC-ENCR) and "histology" (IARC) being the most frequent. It is important to find the right balance between the need to maintain high standards of data quality and the workability of such systems in the daily routine of the cancer registry.

KEYWORDS

data quality, population-based cancer registry, incidence, quality check systems, IARC, JRC-ENCR, cancer research

1 Introduction

One of the main objectives of population-based cancer registries is to collect complete and accurate data on cancers diagnosed in the population under registration. Data quality is an important issue in cancer registration because incomplete or poor-quality data generate flawed results.

Each cancer registry uses its own, internal rules for cancer coding and registration, as well as common rules developed and used by both the corresponding national registration network and the international registration networks, such as the European Network of Cancer Registries (ENCR) or the International Agency for Research on Cancer and the International Association of Cancer Registries (IARC/IACR). Registrars are encouraged to attend proposed training courses: for example, the North American Association of Central Cancer Registries (NAACCR) offers professional qualification and refresher courses, so that cancer registration is done in the most standardized way possible, with little variation due to personal interpretation or lack of up-to-date information.

In recent years, some registries have been using electronic health records for incidence calculation. Created for administrative purposes, electronic health records are timely and inexpensive but do not provide the same degree of clinical detail as medical records. They can be very useful, however, to improve the completeness and quality of cancer incidence data (e.g., pharmaceutical databases for drug treatment of cancer patients).

Data quality checks can be done at different points in time.

The NAACCR network in the US provides registries with a program that checks data quality at the time of data entry but also on already entered records (GenEDITS Plus) (1).

In Europe, IARC (2) and the European Commission Joint Research Center (JRC) in collaboration with the ENCR (3, 4) have made available to cancer registry operators two computer-based edit check systems: the IARC/IACR CHECK program and the JRC-ENCR quality check software. Both systems automatically check the quality of the data produced by the registries, leading to the definition of high-quality datasets standardized according to international criteria (3, 5, 6).

Each of these check systems has its own characteristics: both analyze common as well as system-specific variables and identify errors and deficiencies that, if corrected, will improve the quality of the generated data.

Every five years, IARC calls on cancer registries around the world to send in their data, so it can update the database it maintains and uses to monitor cancer. Based on the collective registry data IARC publishes the volume *Cancer Incidence in Five Continents*, an "invaluable source of information about the global burden and distribution of cancer" (2). In conjunction with this call, the Joint Research Center (JRC) of ENCR has also requested the submission of incidence databases from European registries, to build a large European database in the framework of the European Cancer Information System (7). To produce valid results, the submitted data must be comparable with each other, as complete as possible, and of good quality.

The aims of this study were two: to perform a quality evaluation of the data submitted to these international calls, and to compare the functional characteristics of the two most used systems to check the accuracy of cancer registry data.

The datasets of each registry participating in the study were checked with both systems. Outputs were compared to identify the characteristics and differences detected by each system in an effort to improve the quality of the recorded data and assess the functionality of each check system.

2 Materials and methods

2.1 Data sources

Twenty-two Italian population-based cancer registries affiliated with the Italian Network of Cancer Registries (AIRTUM) (8) participated in the study. The analyzed data spanned from 1986 to 2017, depending on the incidence periods recorded by each registry.

AIRTUM coordinates the national network of general and specialized (pediatric and pancreatic cancer) population-based cancer registries. It designs and conducts collaborative descriptive studies and research activities related to cancer epidemiology in Italy.

Italian cancer registries routinely collect data on incident cancer cases among all residents in the covered area through clinical records, regional mortality files, pathology files, pharmacology files, laboratory databases and hospital discharge databases (electronic health records). The data are collected by trained registrars according to established abstracting rules and standardized manuals such as the *International Classification of Diseases for Oncology*, third edition (ICD-O-3) and the *TNM Staging Manual* (9, 10). For the present study all registries sent in data on all primary tumors including data (if collected) of non-malignant tumors of the central nervous system and urinary bladder.

The registrars use all available pathologic and clinical information to document the date of diagnosis, ICD-O-3 cancer site (topography), histology (morphology), tumor behavior, stage, cancer-specific characteristics (e.g., human epidermal growth factor receptor-2, prostate-specific antigen, Gleason score), demographics and follow-up for vital status.

Data are structured as one record per person per cancer: persons with multiple cancers have multiple records.

2.2 Data quality

Measurement of the quality of registry data is based on four parameters: comparability, completeness, accuracy and timeliness (11). Our analysis was mainly focused on the accuracy of cancer registry data.

2.3 Quality checks

The data were processed using two computer-based data-checking systems developed to assess the quality of population-based cancer registry data.

The IARC/IACR CHECK program, produced by the World Health Organization, is freely available (5). It was created to assess the quality of data provided by registries from worldwide countries for the publication of *Cancer Incidence in Five Continents*. It validates code assignment (sex, incidence and birth date, ICD-O-3 topography, morphology and behavior) and checks the consistency between data items (age versus birth and incidence dates, chronology between birth and incidence dates, sex versus site, sex versus histology, age versus site, age versus histology, site versus histology, basis of diagnosis versus histology).

The JRC-ENCR quality check software (3) is produced by the JRC in collaboration with ENCR and is freely available for the quality control of cancer registry data. It checks for consistency within variables (patient record format, date of cancer incidence, basis of diagnosis, tumor characteristics and stage at diagnosis and patient follow-up) and consistency between variables (coherence between date of birth, date of incidence and date of last known vital status; consistency between age, tumor topography and morphology; consistency between basis of diagnosis, tumor morphology and behavior; consistency between tumor morphology and grade, between topography and laterality, between topography and morphology). Lastly, it offers the possibility of checking the consistency of vital status and autopsy, autopsy and basis of diagnosis, survival, date of incidence and follow-up.

The number of variables used in the checks is greater in the JRC-ENCR than in the IARC check system. For example, JRC-ENCR evaluates variables such as stage at cancer diagnosis, vital status and patient follow-up.

The two systems generate two types of indicators from the checked datasets: errors and warnings. These are specified with short labels that may differ depending on the system used (see Supplementary Materials).

Errors are defined as unacceptable values of variables or unacceptable combinations of variables (impossible code, impossible code combination, missing variable, wrong format or value of variable out of range), while warnings pertain to unusual codes or unlikely code combinations (possible but very rare code or possible but very rare code combination), which may, however, be accepted after specific verification.

Data analysis on quality checks performed by the JRC-ENCR and IARC systems for multiple primaries was not part of this study.

3 Results

3.1 AIRTUM cancer registries

Table 1 lists the contributing cancer registries and the number of cases provided by each registry. We analyzed 22 Italian population-based cancer registries for a total of 1,305,689 cases with different incidence periods (spanning from 1986 to 2017) depending on the registry.

3.2 Data quality checks

The median percentages of DCO cases (cancer with a diagnosis by death certificate only) and microscopically verified cases were 1.2% (range 0.03 to 3.06) for males and 1.4% (range 0.03 to 3.2) for females and 86.3% (range 81.7 to 93.9) for males and 87.3% (range 82.7 to 94.1) for females (data not shown), respectively. Standardized incidence and mortality rates, included temporal trends, where computed (data not shown); the integrated interpretation of these indicators add evidence of the good quality of cancer data of Italian registries.

In this analysis, only variables that presented problems are discussed. For the complete list of variables used by the two check systems, see the Supplementary Materials.

TABLE 1 The 22 Italian cancer registries participating in the study.

Cancer registry	Number of cases	Years of incidence
Avellino	16566	2010-2015
Benevento	13829	2010-2016
Bolzano	67045	1995-2017
Caserta	47509	2008-2016
Catania-Messina-Enna	174866	2003-2017
Friuli Venezia Giulia	60054	2013-2017
Genova	231368	1986-2016
Napoli 1	41128	2010-2015
Napoli 2	38862	2010-2016
Napoli 3	34231	2013-2017
Nuoro	4568	2013-2015
Palermo	106923	2003-2017
Pavia	63232	2003-2017
Ragusa-Caltanissetta	18980	2013-2017
Salerno	45796	2010-2016
Sassari	10695	2012-2015
Siracusa	44040	1999-2017
Sondrio	5702	2013-2015
Trapani-Agrigento	34813	2002-2013
Trento	84163	1995-2017
Umbria	156914	1994-2017
Valle d'Aosta	4405	2013-2017
Total	**1305689**	

3.3 General analysis

3.3.1 Errors

Both systems detected some errors in the checked cases. In the 1,305,689 cases checked, the JRC-ENCR system detected 2,248 errors (0.17%) and the IARC system 45 errors (0.003%). Table 2 lists the detected errors by type.

The proportion of true errors identified by the JRC-ENCR system was 98%, whereas this proportion was 2.2% with the IARC check system. Both system identified the same false errors (n=44).

3.3.2 Warnings

Among the 1,305,689 checked cases, the JRC-ENCR system reported 36,534 warnings (2.8%) and the IARC system 31,700 (2.4%) (Table 3).

The distribution of warnings by registry differed between the two check systems, from a maximum of 10.93% to a minimum of 0.38% with the JRC-ENCR system and from a maximum of 9.23% to a minimum of 0.12% with the IARC system (data not shown).

3.4 Comparison of JRC-ENCR and IARC check systems

This part of the analysis concerns comparisons between errors and warnings identified by the JRC-ENCR and IARC systems. A case may present one or more problems (errors and/or warnings) simultaneously, which may either be reported by both systems or by one of them only. When an error or warning detected by both check systems is identified, it means it has been categorized in the same way by both systems. The IARC check system detected 45 errors in the analyzed registry data; the errors categorized in the same way by the JRC-ENCR system were 42 (Table 4).

In the case series examined, 29,467 warnings (48.17% of total warnings) were detected only by the JRC-ENCR system and 24,633 (40.27% of total warnings) only by the IARC system, while 7,067 warnings (11.55% of total warnings) were detected and categorized in the same way by both systems (Table 5). The differences can be attributed to the different number of variables considered by the two check systems: the IARC system considered 10 variables and the JRC-ENCR system 39.

The types of warnings reported by the JRC-ENCR system only are presented in Table 6, while Table 7 lists the types of warnings reported by the IARC system only.

TABLE 2 Types of errors reported by the check systems (common types of errors between the two systems are aligned).

JRC-ENCR		n	%	IARC	n	%
E-CoDV	Date of last known vital status not valid	13	0.58			
E-FORM	Format error	433	19.26			
E-MISS	Value missing	60	2.67			
E-OUTR	Value out of range*	1741	77.45	ICD-O-3 (Topography)*	22	48.89
				ICD-O-3 (Morphology)*	22	48.89
E-SETO	Sex + Topography not valid	1	0.04	Sex/site combination	1	2.22
Total		2248	100		45	100

*Topography and Morphology IARC errors are included in E-OUTR category of JRC-ENCR system

TABLE 3 Types of warnings reported by the systems (common types of warnings between the two systems are aligned).

JRC-ENCR		n	%	IARC	n	%
W-AGMT	Unlikely age group + tumor type	632	1.73	Age/site/histology	490	1.55
W-BDMO	Morphology too specific	11711	32.06	Basis/histology	5895	18.60
				Behavior/histology	89	0.28
W-BDMS	Morphology not specific enough	10414	28.50			
W-BDMU	BoD + Morpho/Behavior	811	2.22			
W-BDpM	BoD + pM not valid	724	1.98			
W-BDpN	BoD + pN not valid	1119	3.06			
W-BDpT	BoD +pT not valid	1444	3.95			
W-BTNM	Behavior + TNM not valid	170	0.47			
W-MISS	A non-compulsory variables missing	23	0.06			
W-MOGR	Morphology + grade not valid	2906	7.95	Grade/morphology	22994	72.54
W-MOBE	Morphology + Behavior not valid	73	0.20			
W-MOTO	Morphology + Topo not valid	6483	17.75	Histology/site	2200	6.94
W-SEMO	Sex + Morphology not valid	12	0.03	Sex/histology	32	0.10
W-TNMS	Topo + TNM edition + T,N,M + Pathologic stage not valid	6	0.02			
W-UNKN	Unknown code found	6	0.02			
Total		36534	100		31700	100

BoD, basis of diagnosis.

Table 8 shows some of the most common combinations of topographies and morphologies flagged as warnings by the two check systems, listed by number and type. The JRC-ENCR system specifically flags the coding of morphologies of the hematopoietic system in tumors arising at sites other than bone marrow (429 warnings and 932 warnings depending on the morphology considered). The largest number of warnings with the IARC system (585) concerns certain morphologies of ovarian pertinence coded in tumors arising in the pancreas, peritoneum, and uterine cervix and body. The differences between the systems can be attributed to the different criteria defining the morphology-site combination.

4 Discussion

There is an obvious need to control the quality of data produced by cancer registries. Quality control takes place when data are used to carry out research, for example a survival study (9); to manage large databases of registry data (10); or to evaluate the performance of the registry itself (11). The present analysis addresses the quality control of population-based registry data by measuring the efficacy of two computer-based check systems. To our knowledge, this is the first published analysis of its kind.

4.1 Errors

The JRC-ENCR software tends to find a greater number of errors because, unlike the IARC system, it includes the evaluation of variables related to patient follow-up, vital status and TNM staging.

Both systems report errors such as the use of incorrect ICD-O-3 topography codes (for example, C22.9, C26.1, C45.0, liver unspecified, spleen, mesothelioma of pleura; all these are ICD-10 codes) (12). The IARC program also reports on morphologies it fails to recognize, for example those coded 8741, 8349, 8509 and 8348, which are new morphology codes included in the revised version of ICD-O-3 (13) and already in use by registries. This issue will be easily solved with updated checking algorithms.

4.2 Warnings

4.2.1 Morphology and topography

Both systems flag unusual combinations of morphology and topography, but use different criteria in the selection of such combination.

According to Berg, tumors with a primitive or mixed cell type may develop in any organ. They may arise from pluripotent stem cells remaining in the organ or by dedifferentiation, and this may

10.3389/fonc.2023.1197942

TABLE 4 Errors reported by both or one of the systems.

Errors	n	%
JRC-ENCR and IARC	42	1.87
Only JRC-ENCR	2206	98.00
Only IARC	3	0.13
Total	**2251**	100

TABLE 5 Warnings reported by both or one of the systems.

Warning	n	%
JRC-ENCR and IARC	7067	11.55
Only JRC-ENCR	29467	48.17
Only IARC	24633	40.27
Total	**61167**	100

explain why almost any type of cancer can be found in almost any site upon occasion (14).

The JRC-ENCR check system flags up certain combinations of morphology and topography that the IARC system does not identify as incompatible. For example, it rejects the combination of morphology 8000 (neoplasm) with topographies C42.0, C42.1 and C77 (blood, bone marrow and lymph node); it accepts morphology 8098 (adenoid basal carcinoma) only in the cervix uteri, while the IARC system accepts it also in C44 (skin); it accepts morphology 8124 (cloacogenic carcinoma) only for tumors in C21.2 (cloacogenic zone), whereas the IARC system accepts it at other sites of the gastrointestinal tract as well (C20.9 rectum, C21.1 anal canal, C21.8 overlapping lesion of rectum, anus and anal canal). Cloacogenic carcinoma, also called basaloid carcinoma, is an entity originating from the anal transitional epithelium. It is debated whether this neoplasm should be considered a separate entity from squamous cell carcinoma of the anal canal, given the differences in cells of origin, proteomic signatures and survival rates (15), or be classified as a carcinoma of squamous cell nature but manifesting a tendency toward glandular differentiation similar to that sometimes seen in tumors of the oral cavity, larynx or esophagus, currently designated as basaloid carcinomas (16). The specific expression of several types of cell keratins in the anal transitional zone is also found in epithelium of other squamocolumnar junctions such as the esophagogastric and endo-exocervical junctions (17, 18). The literature reports very rare cases of basaloid cell carcinoma arising in the colon and rectum (19).

The morphology code 8510 (medullary carcinoma) for tumors arising in the thyroid gland is accepted by the IARC but not the JRC-ENCR system. The JRC-ENCR system instead accepts code 8345 (medullary carcinoma with amyloid stroma) for thyroid cancer, reserving 8510 for cancers arising in breast, stomach and colon. This is justified by the fact that, despite some common morphologic features (lymphocytic infiltration, poorly differentiated cells), they are distinct entities. Medullary thyroid carcinoma is a neuroendocrine malignancy originating from parafollicular cells (C cells), whereas medullary carcinoma arising in other organs such as breast, stomach or colon is a very uncommon cancer (less than 5% of breast cancers and 0.05% of colon cancers) with neuroendocrine-like features, poorly differentiated aspects, microsatellite instability, lymphocytic infiltration and specific molecular characteristics (20). The American Network of SEER registries accepts both codes for

TABLE 6 Warnings reported by JRC-ENCR only.

JRC-ENCR	warnings	n	%
W-BDMS	Morphology not specific enough	10414	35.34
W-BDMO	Morphology too specific	6202	21.05
W-MOTO	Morphology + Topography not valid	6035	20.48
W-MOGR	Morphology + Grade not valid	2269	7.70
W-BDpT	BoD + pT not valid	1444	4.90
W-BDpN	BoD + pN not valid	1119	3.80
W-BDMU	BoD + Morpho/Behavior	811	2.75
W-BDpM	BoD + pM not valid	724	2.46
W-AGMT	Unlikely Age group + Tumor type	171	0.58
W-BTNM	Behavior + TNM not valid	170	0.58
W-MOBE	Morphology + Behavior not valid	73	0.25
W-MISS	A compulsory variable is missing	23	0.08
W-UNKN	Unknown code found	6	0.02
W-TNMS	Topo + TNM edition + T,N,M + Pathologic stage not valid	6	0.02
Total		**29467**	**100**

BoD, basis of diagnosis.

10.3389/fonc.2023.1197942

TABLE 7 Warnings reported by IARC only.

IARC warning	n	%
Grade/morphology	22357	90.76
Histology/site	1752	7.11
BoD/histology	386	1.57
Behavior/histology	89	0.36
Age/site/histology	29	0.12
Sex/histology	20	0.08
Total	**24633**	**100**

BoD, basis of diagnosis.

thyroid carcinoma with medullary histology (21). The IARC system classifies medullary carcinoma as "not site-specific carcinoma" and therefore accepts it for cancers arising at any site except bone, connective tissue and nervous system (C40-C42, C47, C48, C49, C70, C71, C72, C77).

The morphology codes 8370, 8700, 9490 and 9500 (adrenal cortical carcinoma, pheochromocytoma, ganglioneuroma, neuroblastoma) are not accepted by the JRC-ENCR system at the generic site C74.9 (adrenal gland NOS), but only at specific subsites of the adrenal gland such as C74.0 (cortex of adrenal gland) or C74.1 (medulla of adrenal gland), according to the specific morphology. Neuroblastoma is due to differentiation arrest of the neural-crest-derived sympathoadrenal lineage. The sympathoadrenal lineage is derived from neural crest cells that emigrate from the dorsal neural tube and migrate to distant sites during the early stages of embryogenesis (22). Clinically, neuroblastoma manifests as a primary tumor anywhere along the sympathetic nervous system, with >50% occurring in the adrenal medulla (C74.1) (23). The site of origin is therefore C74.1 (medulla of adrenal gland), as correctly indicated by the JRC-ENCR system. The IARC system, however, accepts coding of this morphology also at the generic site C74.9. There is a plausible reason for this: the IARC system has a global distribution, and there are geographic areas where it is difficult to obtain the information needed for

TABLE 8 Most common examples of morphology and site combinations (ICD-O-3 codes) reported as warnings.

Topography*	Morphology	No. of warnings	
		JRC-ENCR	IARC
C42/C77	8000	314	0
C38.1-C38.3/C41/C48/C71/C77	Epithelial morphology	85	31
≠ C34	8012 or 8041-8045	8	299
C44	8098	141	0
≠ C67	8120 or 8130	101	35
C20.9, C21.1, C21.8-C21.9	8124	91	0
C11, C21-C24, C30-C34, C56	8144	6	72
C22.0 or C24-C25	8160-8162	206	1
C50	8401	0	216
C25, C48, C53-C55	8441, 8460, 8470-8471	24	585
C73.9	8510	227	0
≠ C30.0, C44, C51, C60, C63.2, C69.0, C80	8770-8772	161	23
C74.9	8370, 8700, 9490, 9500	132	0
C71-C72	9530-9539	106	0
C80.9	9590-9591 or 9699	233	0
C42.1	9220, 9731, 9761, 9930	446	1
≠ C42.1	9732	429	57
C42.0 or C42.2 or C42.4	9820-9823, 9860-9863	932	0
C42.0 or C42.4	9590-9591, 9650-9653, 9670-9673, 9680, 9690	299	0
C64.9	8120, 8130, 8210-8211	70	0
C07-C09, C12-C17, C30-C33, C38, C42-C54, C61-C68, C71-C77	8240, 8243-8246, 8249	37	242

*See Table S5.

complete cancer incidence estimation, so all the collected information is used, even if it shows a lesser degree of accuracy (24).

Another difference is the use of code C80.9 (unknown primary site) in combination with hematopoietic morphologies (9590-9597 or 9699), which is accepted by the IARC system but not the JRC-ENCR system. The consistency check between topography and morphology brings to the fore two types of issues: the possible registration of an extranodal lymphoma whose precise organ of origin is unknown (the ICD-O-3 rule is to code lymphoma to C80.9, unknown primary site, if it is suspected to be extranodal and no site of origin is indicated) and the use of the code for unknown site rather than lymph node for a lymphoma of nodal origin. The JRC-ENCR system requires checking of all lymphoma cases coded with this topography, whereas the IARC system accepts any morphology of lymphoma at any site of origin because lymphomas are considered tumors with a non-specific site profile.

A further example involves morphology codes 8120 (transitional cell carcinoma) and 8130 (papillary transitional cell carcinoma), which according to the JRC-ENCR system are compatible with just a few sites of tumor origin (C56 ovary and C65-C68 renal pelvis, ureter, bladder, other urinary organs), while the IARC system accepts them for many other topographies (C11 nasopharynx; C14 other and ill-defined sites in lip, oral cavity and pharynx; C20 rectum, C21 anus and anal canal, C26 intestinal tract NOS; C30 nasal cavity and middle ear; C31 accessory sinuses; C53 cervix uteri; C61 prostate; C64 kidney). For tumors arising in the nasal cavity and accessory sinuses, the JRC-ENCR system accepts the morphology code 8121, Schneiderian (cylindrical [transitional] cell) carcinoma. Schneiderian carcinoma is a typical cancer of the nasal cavity and sinuses and is closely related to non-keratinizing squamous cell carcinoma. A typical feature is lack of maturation in the epithelial nests as in transitional cell carcinoma of the urinary tract, which this tumor subtype resembles.

Certain combinations of tumor morphology and topography are accepted by the JRC-ENCR system but trigger a warning from the IARC system. For example, morphology code 8401 (apocrine adenocarcinoma) is accepted for breast cancer by the JRC-ENCR system, while the IARC system accepts only C00 lip, C44 skin, C51 vulva, C60 penis, C63.2 scrotum NOS, C76 other and ill-defined sites as possible sites.

The IARC system reports tumors with a site-specific profile, but in most cases an unlikely combination between site and morphology according to IARC is accepted by JRC-ENCR. For example, adenocarcinoma, intestinal type (8144/3) is accepted for a larger number of sites by the JRC-ENCR than the IARC system.

Another difference in the selection of warnings concerns the use of morphology codes 8012 (large cell carcinoma NOS) or 8041-8045 (small cell carcinoma, oat cell carcinoma, small cell carcinoma fusiform cell, small cell carcinoma intermediate cell and combined small cell carcinoma) for tumors in sites other than the lung. Both systems limit the use of these morphologies to cancers arising in the respiratory system, but while the JRC-ENCR system considers coding these morphologies unusual only in tumors arising in C38, C40-C42, C47, C48.0, C49, C70-C72 and C77 (pleura, bone, joints and articular cartilage; peripheral nerves and autonomic nervous system; retroperitoneum and peritoneum; connective, subcutaneous and other soft tissues; meninges, brain and spinal cord, cranial nerves and other parts of the central nervous system; lymph nodes), the IARC system uses stricter limits and allows these morphologies only in cancers of the lung, ill-defined sites of the respiratory system, and intrathoracic organs (C34, C39.8, C39.9, C76.1 thorax, C76.7 other ill-defined sites, C76.8 overlapping lesion of ill-defined sites) in addition to unknown primary sites (C80.9). This results in the generation of a much greater number of warnings by the IARC system than the JRC-ENCR system.

The same applies to the use of morphology codes 8441 (serous cystadenocarcinoma), 8460 (papillary serous cystadenocarcinoma), 8470 and 8471 (mucinous cystadenocarcinoma and papillary mucinous cystadenocarcinoma) for cancers arising at sites C25 (pancreas), C48 (peritoneum and retroperitoneum) and C53, C54, C55 (cervix uteri, corpus uteri, uterus NOS). The IARC system accepts these morphologies only at the following sites: C56 (ovary), C57 (other and unspecified female genital organs) and C76 and C80 (abdomen, pelvis, other ill-defined sites, and unknown primary site). The JRC-ENCR system accepts morphology codes 8441, 8460 and 8471 also for cancers in C54 (corpus uteri), and code 8470 not only for cancers of the female genital tract (C56, C57) but also in C18 (colon) and C25 (pancreas). The result is a marked difference in warnings by the two systems: 585 by IARC and 24 by JRC-ENCR.

The IARC check system devised consistency checks between tumor site and morphology using the data collected in its large database, similar to what Berg did when he devised a system based on morphologic similarities and differences for the recognition of multiple tumors (14). In addition, the IARC checks refer to groups of morphologies that are accepted only for tumors arising in certain organs (tumors with a specific site profile) or that are not allowed in certain organs (tumors with an inverse site profile); there is also a group of morphologies that have no organ specificity and can be assigned to tumors arising in any organ (tumors with no specific site profile) (5). This leads to different choices in generating errors or warnings compared to the JRC-ENCR system. The IARC system only checks morphologies that are normally attributed on the basis of a cytologic/histologic diagnosis, whereas the JRC-ENCR system also performs the opposite check: generic morphology codes (8000, 9590, 9960) with a basis of cytologic/histologic diagnosis.

4.2.2 Staging and follow-up variables

The JRC-ENCR system gives out more warnings related to variables not considered in the IARC system (e.g., TNM stage, *TNM Staging Manual* edition, patient follow-up). Many of the reported warnings are due to incorrect coding of tumor stage or to the combination of a clinical basis of diagnosis and pathologic stage variables. Not all cancer registries can code tumor stage at diagnosis; moreover, the use of incorrect codes related to pathologic and clinical staging is frequent. Minicozzi's study of the presence and quality of staging at diagnosis in European population-based cancer registries showed that only half of the Italian registries participating in the study were able to provide staging information; particularly case records compiled in an automated manner or directly from pathology laboratory reports were lacking this variable (25).

These checks, along with demographic data, ensure appropriate staging of registered cases, making it possible to study cancer

survival and improving the accuracy of the registry's output. For example, it is unlikely that an advanced-stage neoplasm in the lung or pancreas will grant the patient who carries it a long survival time (26).

4.2.3 Behavior and stage

The unlikely combination of a tumor's behavior code and its registered stage (e.g., infiltrating carcinoma with behavior code/3 and *in situ* stage, pTis) (170 warnings) will lead the registry to review the case because of a suspected registration error. A possible scenario, on the other hand, is that of an *in situ* neoplasm developing aggressive behavior over time and ultimately generating metastases (27).

4.2.4 Histology and grade

Both the JRC-ENCR and IARC systems flag issues related to the incorrect combination of histology and tumor grade. Grade refers to differentiation in solid tumors (codes 1, 2, 3, 4, 9): it is a measurement of how closely the tumor cells resemble the parent tissue (organ of origin) (See Supplementary Materials). Well-differentiated tumor cells (grade 1) closely resemble the tissue from the organ of origin. Poorly differentiated (grade 3) and undifferentiated (grade 4) tumor cells are disorganized and abnormal looking. Codes 5, 6, 7 and 8 are cell indicators, because they describe the lineage or phenotype of the cell and are used only for hematopoietic and lymphoid neoplasms; code 9 indicates cell type not determined, not stated, or not applicable (13). Both systems follow a specific routine to identify incorrect or missing combinations to be flagged for revision (9).

The systems check the morphology codes of solid tumors requiring a specific grade (e.g., undifferentiated sarcoma 8805/3 with grade 4). The JRC-ENCR system will flag the combination of grades 5-8 and morphology codes outside the 9590-9992 range (hematopoietic system codes). The IARC system also performs the opposite check, flagging grade codes greater than or equal to 1 and less than or equal to 4 in combination with histology codes greater than or equal to 9590. Moreover, the IARC system flags more cases because it requires a specific grade for many hematopoietic neoplasms and does not accept the value 9 (not specified); for example, all B-cell lymphomas should have grade 6.

4.2.5 Histology and age

Another issue flagged by the systems concerns inappropriate combinations of tumor morphology and patient age at diagnosis, e.g., 9945 (chronic myelomonocytic leukemia) or 9876 (atypical chronic myeloid leukemia) for age less than 30 years; cancer site C51-C52 (vulva and vagina) for age less than 20 years, or cancer site C60 (penis) for age less than 30 years. Burkitt lymphoma (code 9687) is expected to be diagnosed in children aged less than 14 years, but the registries use the same morphology code for Burkitt-like lymphoma. The distinction between Burkitt and Burkitt-like lymphoma is morphologic: tumor cells in Burkitt-like lymphoma are slightly larger, with more nuclear variability and increased nucleolar prominence. This tumor may arise in patients with a median age of 47 years, but the use of this code in patients aged more than 14 years generated many warnings (28).

The JRC-ENCR system also takes into account patient age for some morphologies. For example, the system accepts basis of diagnosis 2 (clinical) for 8960 (Wilms tumor, nephroblastoma) at age 0-8 years, or basis of diagnosis 4 (specific tumor markers) for 9732 (multiple myeloma) at ages over 40 years; the IARC system does not consider age in these cases but only morphology.

4.2.6 Basis of diagnosis and morphology

The two systems differ in how they treat the variables "basis of diagnosis" and "morphology". The JRC-ENCR system marks a larger number of cases, because it flags some morphologies that are accepted by the IARC system with a clinical basis of diagnosis (1 or 2), e.g., 8170 (hepatocellular carcinoma), 9732 (multiple myeloma) and 9761 (Waldenstrom globulinemia). Moreover, for cases with death certificate only (DCO) as the basis of diagnosis, the JRC-ENCR system accepts morphologies that can be identified from the underlying cause of death code (ICD-10), while the IARC system flags all DCO cases with morphologies different from those accepted even without microscopic verification.

Another difference between the two systems is related to morphology code 8720 (melanoma): in the absence of histologic or cytologic examination the IARC system accepts only cases arising in C44 (skin) or C69 (eye and adnexa), whereas the JRC-ENCR system accepts melanoma arising at any site.

The large number of warnings detected by the systems is also due to the increased use of electronic health data. In hospital discharge records some cancer codes from the ICD-9-CM classification contain morphologic information (e.g., Hodgkin lymphoma, melanoma, myeloid leukemia, lymphoid leukemia, mycosis fungoides, non-Hodgkin lymphoma): to make incidence calculations, the registries use these codes associated with a clinical basis of diagnosis, but this is not accepted by the JRC-ENCR system in combination with such specific morphology.

The same applies to *in situ* neoplasms: these require a histologic basis of diagnosis, but the information may have come from hospital discharge records, where some tumor sites are labeled with a specific code when they exhibit *in situ* behavior (e.g., in ICD-9-CM, 233.0, carcinoma *in situ* of breast or 233.7, carcinoma *in situ* of bladder).

The JRC-ENCR system performs an additional check for basis of diagnosis 6 (histology of a metastasis): it considers it unlikely that a lymphoma or leukemia diagnosis is based on a metastasis (W-BDMU) (811 cases), whereas a bone marrow aspirate can be used as the basis of diagnosis for lymphomas.

4.2.7 Sex and histology

With regard to sex/histology consistency checks, IARC warnings are mostly due to unacceptable combinations, such as typical ovarian histology in cancer arising in C25 (pancreas) in a male patients while the JRC-ENCR system flags only cases in which ovarian morphology is not allowed in C25, e.g., 8471 (papillary mucinous cystadenocarcinoma).

5 Conclusion

The IARC/IACR CHECK program, intended for cancer registries worldwide, utilizes a less demanding checking system that is easy to use for all registries. At present its checking routine for histology requires updating with the new morphology codes included in the second revision of ICD-O-3. The JRC-JRC-ENCR quality check software carries out a number of additional checks compared to IARC. For this reason, it would be advisable to use both systems for data quality control, since they provide checks on different groups of variables (stage, follow-up) or on the same variables but with different modalities.

Finally, periodic checks are useful for identifying issues that inevitably arise when working with data. However, it is important to find the right balance between the need to maintain high standards of data quality – otherwise the data are useless – and the workability of such systems in the daily routine of the cancer registry.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The data on cancer cases used in this study were provided by cancer registries affiliated with AIRTUM and cannot be made freely available. Requests to access these datasets should be directed to not available.

Author contributions

(I) Conception and design: GT, PC. (II) Revising the work critically for important intellectual content: GT, PC, VP, SF, AT, GB, MV. (III) Collection and assembly of data: SF, VP, AT, GB. (IV) Data analysis and interpretation: GT, VP, SF, AT, GB, MV. (V) Manuscript writing: GT and VP. (VI) Final approval of manuscript: GT, PC, VP, SF, AT, GB, MV. (VII) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: GT, PC, VP, SF, AT, GB, MV. All authors contributed to the article and approved the submitted version.

Group members of Collaborative Working Group

Lorenza Boschetti, Stefano Marguati (Pavia CR); Giuseppa Candela, Tiziana Scuderi, Giovanna Fantaci (Trapani-Agrigento CR); Maurizio Zarcone, Sergio Mazzola, Barbara Ravazzolo (Palermo CR); Daniela Piras, Elisa Conca, Maria Rotella (Sassari CR); Pasquala Pinna, Roberta Bosu, Luisa Canu (Nuoro CR); Diego Serraino, Luigino Dal Maso, Federica Toffolutti (Friuli Venezia Giulia CR and CRO Aviano); Antonella Puppo, Claudia Casella, Enza Mariani (Genova CR); Serena Maria Curti, Salvatore Bongiorno, Claudio Trentin (Valle d'Aosta CR); Margherita Ferrante, Fiorella Paderni, Andrea Benedetto (Catania-Messina-Enna CR); Roberto Vito Rizzello, Maria Adalgisa Gentilini, Mariangela Pedron (Trento CR); Anna Clara Fanetti, Ivan Cometti, Paola Giumelli (Sondrio CR); Rosario Tumino, Francesca Bella, Antonino Colanino Ziino (Siracusa CR); Maria Francesca Vitale, Valerio Ciullo (Campania - Napoli 3 CR); Alessio Gili, Silvia Leite, Fortunato Bianconi (Umbria CR); Fabio Vittadello, Paolo Vian, Michael Mian (Bolzano CR); Giuseppe Cascone, Eugenia Spata, Antonella Ippolito (Ragusa-Caltanissetta CR); Giancarlo D'Orsi (Campania – Napoli 2 CR); Rosanna Ortolani, Valentina Cozza, Lucia De Coppi (Campania – Napoli 1 CR); Rossella Cavallo (Salerno CR); Maddalena Merola, Antonietta Minichino, Nadia Di Pascale (Caserta CR); Anna Rita Citarella (Benevento CR); Francesco D'Amelio (Avellino CR).

Funding

This research was funded by: Italian Ministry of Health "Ricerca Corrente" funds.

Acknowledgments

The authors would like to thank Marije de Jager for translation and revision of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1197942/full#supplementary-material

References

1. NAACCR. Available at: https://www.naaccr.org/edits-software/ (Accessed September 28, 2022).

2. IARC. Available at: https://ci5.iarc.fr (Accessed September 28, 2022).

3. ENCR. *JRC-ENCR data quality check software, version 2.0* . Available at: https://www.encr.eu/.

4. Giusti F, Martos C, Adriani S, Flego M, Brunetto A, Dyba T, et al. *The JRC-ENCR quality check software (QCS) for the validation of cancer registry data: user compendium - version 2.0*. Ispra, Italy: European Commission (2022).

5. Ferlay J. *IARCcrg tools, version 2.13* (2005). Available at: http://www.iacr.com.fr.

6. Martos C, Crocetti E, Visser O, Rous B, Giusti F. *Cancer data quality check working group. a proposal on cancer data quality checks: one common procedure for European cancer registries – version 1.1*. Luxembourg: EN Publications Office of the European Union (2018). doi: 10.2760/429053

7. ECIS. Available at: https://ecis.jrc.ec.europa.eu/.

8. AIRTUM. Available at: https://www.registri-tumori.it/cms/.

9. Fritz A, Percy C, Jack A, Shanmugarathan K, Sobin L, Parkin D, et al. *International classification of disease for oncology, ICD-O-3rd ed*. Geneva: WHO (2000).

10. Sobin LH, Gospodarowicz MK, Wittekind CInternational Union Against Cancer (UICC). *TNM classification of malignant tumours, 7th ed*. New York: Wiley-Blackwell (2009).

11. Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH WS. *Comparability and quality control in cancer registration IARC technical report no. 19*. Lyon: IARC (1994).

12. WHO. *ICD-10 : international statistical classification of diseases and related health problems: tenth revision, 2nd ed*. World Health Organization: Geneva (2004).

13. NAACCR. Available at: https://www.naaccr.org/icdo3/ (Accessed September 28, 2022).

14. Berg JW. Morphologic classification of human cancer. In: Shottenfeld D, Fraumeni J Jr., editors. *Cancer epidemiology and prevention, 2nd ed*. New York: Oxford University (1996).

15. Herfs M, Longuespée R, Quick CM, Roncarati P, Suarez-Carmona M, Hubert P, et al. Proteomic signatures reveal a dualistic and clinically relevant classification of anal canal carcinoma. *J Pathol* (2017) 241:522–33. doi: 10.1002/path.4858

16. Rosai J. *Rosai and ackerman's surgical pathology, vol. 1*. China, Mosby (2004). p. 258.

17. Wang X, Ouyang H, Yamamoto Y, Kumar PA, Wei TS, Dagher R, et al. Residual embryonic cells as precursors of a barrett's-like metaplasia. *Cell* (2011) 145:1023–35. doi: 10.1016/j.cell.2011.05.026

18. Herfs M, Yamamoto Y, Laury A, Wang X, Nucci MR, McLaughlin-Drubin ME, et al. A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer. *Proc Natl Acad Sci U.S.A.* (2012) 109:10516–21. doi: 10.1073/pnas.1202684109

19. Lee T, Yoon S, Kim M. Successful treatment of basaloid squamous cell carcinoma in the rectosigmoid colon: a case report and review of literature. *World J Clin cases* (2020) 8:3057–63. doi: 10.12998/wjcc.v8.i14.3057

20. Ottaiano A, Santorsola M, Perri F, Pace U, Marra B, Correra M, et al. Clinical and molecular characteristics of rare malignant tumors of colon and rectum. *Biol (Basel)* (2022) 11:1–15. doi: 10.3390/biology11020267

21. SEER. Available at: https://seer.cancer.gov/ (Accessed September 28, 2022).

22. Zeineldin M, Patel AG, Dyer MA. Neuroblastoma: when differentiation goes awry. *Neuron* (2022) 110:2916–28. doi: 10.1016/j.neuron.2022.07.012

23. Lundberg KI, Treis D, Johnsen JI. Neuroblastoma heterogeneity, plasticity, and emerging therapies. *Curr Oncol Rep* (2022) 24:1053–62. doi: 10.1007/s11912-022-01270-8

24. Severance TS, Njuguna F, Olbara G, Kugo M, Langat S, Mostert S, et al. An evaluation of the disparities affecting the underdiagnosis of pediatric cancer in Western Kenya. *Pediatr Blood Cancer* (2022) 69(10):1–6. doi: 10.1002/pbc.29768

25. Minicozzi P, Innos K, Sánchez MJ, Trama A, Walsh PM, Marcos-Gragera R, et al. Quality analysis of population-based information on cancer stage at diagnosis across Europe, with presentation of stage-specific cancer survival estimates: a EUROCARE-5 study. *Eur J Cancer* (2017) 84:335–53. doi: 10.1016/j.ejca.2017.07.015

26. Akushevich I, Yashkin A, Kovtun M, Yashin AI, Kravchenko J. Underlying mechanisms of change in cancer prevalence in older U.S. adults: contributions of incidence, survival, and ascertainment at early stages. *Cancer Causes Control* (2022) 33:1161–72. doi: 10.1007/s10552-022-01595-6

27. Tesch ME, Rosenberg SM, Collins LC, Wong JS, Dominici L, Ruddy KJ, et al. Clinicopathologic features, treatment patterns, and disease outcomes in a modern, prospective cohort of young women diagnosed with ductal carcinoma *in situ*. *Ann Surg Oncol* (2022) 29(13):8048-8057. doi: 10.1245/s10434-022-12361-y

28. Lacasce A, Howard O, Li S, Fisher D, Weng A, Neuberg D, et al. Modified magrath regimens for adults with burkitt and burkitt-like lymphomas: preserved efficacy with decreased toxicity. *Leuk Lymphoma* (2004) 45:761–7. doi: 10.1080/1042819031000141301

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 06 June 2023
DOI 10.3389/fonc.2023.1168325

OPEN ACCESS

EDITED BY
Francesco Giusti,
Belgian Cancer Registry, Belgium

REVIEWED BY
Saverio Virdone,
Thrombosis Research Institute,
United Kingdom
Vesna Zadnik,
Institute of Oncology Ljubljana, Slovenia

*CORRESPONDENCE
Luigino Dal Maso
epidemiology@cro.it
Stefano Guzzinati
stefano.guzzinati@azero.veneto.it

RECEIVED 17 February 2023
ACCEPTED 09 May 2023
PUBLISHED 06 June 2023

CITATION
Toffolutti F, Guzzinati S, De Paoli A, Francisci S, De Angelis R, Crocetti E, Botta L, Rossi S, Mallone S, Zorzi M, Manneschi G, Bidoli E, Ravaioli A, Cuccaro F, Migliore E, Puppo A, Ferrante M, Gasparotti C, Gambino M, Carrozzi G, Stracci F, Michiara M, Cavallo R, Mazzucco W, Fusco M, Ballotari P, Sampietro G, Ferretti S, Mangone L, Rizzello RV, Mian M, Cascone G, Boschetti L, Galasso R, Piras D, Pesce MT, Bella F, Seghini P, Fanetti AC, Pinna P, Serraino D, Dal Maso L and AIRTUM Working Group (2023) Complete prevalence and indicators of cancer cure: enhanced methods and validation in Italian population-based cancer registries. *Front. Oncol.* 13:1168325. doi: 10.3389/fonc.2023.1168325

Complete prevalence and indicators of cancer cure: enhanced methods and validation in Italian population-based cancer registries

Federica Toffolutti[1], Stefano Guzzinati[2*], Angela De Paoli[2], Silvia Francisci[3], Roberta De Angelis[4], Emanuele Crocetti[1], Laura Botta[5], Silvia Rossi[4], Sandra Mallone[3], Manuel Zorzi[2], Gianfranco Manneschi[6], Ettore Bidoli[1], Alessandra Ravaioli[7], Francesco Cuccaro[8], Enrica Migliore[9], Antonella Puppo[10], Margherita Ferrante[11], Cinzia Gasparotti[12], Maria Gambino[13], Giuliano Carrozzi[14], Fabrizio Stracci[15], Maria Michiara[16], Rossella Cavallo[17], Walter Mazzucco[18], Mario Fusco[19], Paola Ballotari[20], Giuseppe Sampietro[21], Stefano Ferretti[22], Lucia Mangone[23], Roberto Vito Rizzello[24], Michael Mian[25], Giuseppe Cascone[26], Lorenza Boschetti[27], Rocco Galasso[28], Daniela Piras[29], Maria Teresa Pesce[30], Francesca Bella[31], Pietro Seghini[32], Anna Clara Fanetti[33], Pasquala Pinna[34], Diego Serraino[1], Luigino Dal Maso[1*] and AIRTUM Working Group

[1]Cancer Epidemiology Unit, Centro di Riferimento Oncologico di Aviano (CRO) Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Aviano, Italy, [2]Epidemiological Department, Azienda Zero, Padua, Italy, [3]National Centre for Disease Prevention and Health Promotion, National Institute of Health, Rome, Italy, [4]Department of Oncology and Molecular Medicine, National Institute of Health, Rome, Italy, [5]Evaluative Epidemiology Unit, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy, [6]Tuscany Cancer Registry, Clinical Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network (ISPRO), Florence, Italy, [7]Emilia-Romagna Cancer Registry, Romagna Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Forlì, Italy, [8]Registro Tumori Puglia - Sezione Azienda Sanitaria Locale (ASL) Barletta-Andria-Trani, Epidemiologia e Statistica, Barletta, Italy, [9]Piedmont Cancer Registry, Centro di Riferimento per l'Epidemiologia e la Prevenzione Oncologica (CPO) Piemonte and University of Turin, Turin, Italy, [10]Liguria Cancer Registry, IRCCS Ospedale Policlinico San Martino, Genova, Italy, [11]Registro tumori integrato di Catania-Messina-Enna, Igiene Ospedaliera, Azienda Ospedaliero-Universitaria Policlinico G. Rodolico-San Marco, Catania, Italy, [12]Struttura Semplice Epidemiologia, Agenzia di Tutela della Salute (ATS) Brescia, Brescia, Italy, [13]Registro tumori ATS Insubria (Provincia di Como e Varese) Responsabile S.S. Epidemiologia Registri Specializzati e Reti di Patologia, Varese, Italy, [14]Emilia-Romagna Cancer Registry, Modena Unit, Public Health Department, Local Health Authority, Modena, Italy, [15]Umbria Cancer Registry, Public Health Section, Department of Medicine and Surgery, University of Perugia, Perugia, Italy, [16]Emilia-Romagna Cancer Registry, Parma Unit, Medical Oncology Unit, University Hospital of Parma, Parma, Italy, [17]Cancer Registry Azienda Sanitaria Locale (ASL) Salerno- Dipartimento di Prevenzione, Salerno, Italy, [18]Clinical Epidemiology and Cancer Registry Unit, Azienda Ospedaliera Universitaria Policlinico (AOUP) di Palermo, Palermo, Italy, [19]Registro Tumori ASL Napoli 3 Sud, Napoli, Italy, [20]Osservatorio Epidemiologico, ATS Val Padana, Mantova, Italy, [21]Servizio Epidemiologico ATS di Bergamo, Bergamo, Italy, [22]Emilia-Romagna Cancer Registry, Ferrara Unit, Local Health Authority, Ferrara, and University of Ferrara, Ferrara, Italy, [23]Emilia-Romagna Cancer Registry, Reggio Emilia Unit, Epidemiology Unit, Azienda Unità Sanitaria Locale - IRCCS di Reggio Emilia, Reggio Emilia, Italy, [24]Trento Province Cancer Registry, Unit of Clinical Epidemiology, Trento, Italy, [25]Innovation, Research and Teaching Service (SABES-ASDAA), Lehrkrankenhaus der Paracelsus Medizinischen Privatuniversität, Bolzano-Bozen, Italy,

[26]Azienda Sanitaria Provinciale (ASP) Ragusa - Dipartimento di Prevenzione -Registro Tumori, Ragusa, Italy, [27]Cancer Registry of the Province of Pavia, Pavia, Italy, [28]Unit of Regional Cancer Registry, Clinical Epidemiology and Biostatistics, IRCCS Centro di Riferimento Oncologico di Basilicata (CROB), Rionero in Vulture, Italy, [29]Nord Sardegna Cancer Registry, ASL, Sassari, Italy, [30]Monitoraggio rischio ambientale e Registro Tumori ASL Caserta, Caserta, Italy, [31]Siracusa Cancer Registry, Provincial Health Authority of Siracusa, Siracusa, Italy, [32]Emilia-Romagna Cancer Registry, Piacenza Unit, Public Health Department, AUSL Piacenza, Piacenza, Italy, [33]Sondrio Cancer Registry, Agenzia di Tutela della Salute della Montagna, Sondrio, Italy, [34]Nuoro Cancer Registry, RT Nuoro, Servizio Igiene e Sanità Pubblica, ASL Nuoro, Nuoro, Italy

Objectives: To describe the procedures to derive complete prevalence and several indicators of cancer cure from population-based cancer registries.

Materials and methods: Cancer registry data (47% of the Italian population) were used to calculate limited duration prevalence for 62 cancer types by sex and registry. The incidence and survival models, needed to calculate the completeness index (*R*) and complete prevalence, were evaluated by likelihood ratio tests and by visual comparison. A sensitivity analysis was conducted to explore the effect on the complete prevalence of using different *R* indexes. Mixture cure models were used to estimate net survival (NS); life expectancy of fatal (LEF) cases; cure fraction (CF); time to cure (TTC); cure prevalence, prevalent patients who were not at risk of dying as a result of cancer; and already cured patients, those living longer than TTC at a specific point in time. CF was also compared with long-term NS since, for patients diagnosed after a certain age, CF (representing asymptotical values of NS) is reached far beyond the patient's life expectancy.

Results: For the most frequent cancer types, the Weibull survival model stratified by sex and age showed a very good fit with observed survival. For men diagnosed with any cancer type at age 65–74 years, CF was 41%, while the NS was 49% until age 100 and 50% until age 90. In women, similar differences emerged for patients with any cancer type or with breast cancer. Among patients alive in 2018 with colorectal cancer at age 55–64 years, 48% were already cured (had reached their specific TTC), while the cure prevalence (lifelong probability to be cured from cancer) was 89%. Cure prevalence became 97.5% (2.5% will die because of their neoplasm) for patients alive >5 years after diagnosis.

Conclusions: This study represents an addition to the current knowledge on the topic providing a detailed description of available indicators of prevalence and cancer cure, highlighting the links among them, and illustrating their interpretation. Indicators may be relevant for patients and clinical practice; they are unambiguously defined, measurable, and reproducible in different countries where population-based cancer registries are active.

KEYWORDS

prevalence, cancer cure indicators, time to cure, Italy, survival, cure fraction, cure prevalence

1 Introduction

Unlike other indicators of cancer burden (i.e., incidence, survival, or mortality), complete prevalence cannot be directly observed by cancer registries (CRs) because cancer survivors diagnosed before the start of registration are not included in the CR databases. The more recently the CR started registration, the greater the number of unobserved survivors (1). Therefore, complete prevalence and indicators of cancer cure, almost always based on statistical models, are reported less frequently than other indicators of cancer burden.

In the last decade, some epidemiologic investigations have explored the issue of estimating cancer cure in high-income

countries (2–10), even if the usefulness to estimate indicators of cancer cure is held back by the lack of a shared definition of cure (11, 12). Nevertheless, several indicators of "cancer cure" have been proposed, particularly, the following: the *cure fraction* or the estimated probability of cure among incident cases (13, 14); the *time to cure*, the time necessary to make the excess risk of death due to cancer negligible (3, 4, 8, 10); *already cured* or the proportion of prevalent cases that have already reached the time to cure in a specific point in time (4); and *cure prevalence* or the proportion of all prevalent cases not expected to die due to their cancer (4, 15).

This article aimed to provide a complete and detailed description of the methodology and the procedures needed to derive complete prevalence and indicators of cancer cure from population-based CR data. The description has been accompanied by an application using the latest available Italian data. Improvement in the previously used algorithms (4, 5, 15, 16) to calculate cure indicators has been described, as well as validations of survival models and indicators. Finally, the epidemiological interpretation of indicators and the links among them are highlighted, with a discussion of assumptions made and their limitations.

2 Materials and methods

2.1 Study population

This study included 31 population-based Italian CRs with at least 9 years of registration and patient vital status ascertainment at least 1 year after the last incidence date. By the end of 2017, the maximum duration of registration ranged from 9 to 40 years, with a median of 22 years (Table 1). Twenty CRs are located in north-central Italy [i.e., homogeneous areas in terms of incidence and survival (16)] and 11 in the South-Islands. CRs coverage varied with regards to the population size (0.2 to 2.8 million inhabitants), and overall, they cover more than 28 million people of all ages (43% of the population in north-central Italy, 55% in the South-Islands, and 47% overall; Figure 1). Since a key methodological point for the estimation of cure indicators is the availability of reliable estimates of "long-term" incidence and survival in the population of interest, Italian CRs with at least 15 years of registration (Table 1) and complete follow-up at the end of 2018 were included for the estimation of model-based incidence and survival. The geographical representativeness of these CRs is similar (~30%) between the north-central area and the South-Islands. Up to 1 January 2018, nearly 3.3 million (3,276,906, Table 1) incidents of malignant cancer cases were diagnosed in nearly 3 million (2,957,828) men and women, of all ages, in areas covered by CRs. They were two times higher than the number of cases included in the previous Italian report (17), including 443,901 female breast cancer cases, and 420,726 colorectal and 370,034 lung cancers (Table 2). For breast and colorectal cancer patients, prevalence and indicators of cancer cure were also calculated by stage at diagnosis including information from CRs with<33% of missing stage information for at least 15 consecutive years (i.e., respectively from six CRs for breast cancer and five CRs for colorectal cancer, approximately 6% of the Italian population) (Table 1).

2.2 Cases and groupings

Prevalence and indicators of cancer cure were calculated for all malignant cancers and 62 types or their combinations (Table 2) using ICD-10 classification. In addition, ICD-O-3 topography and morphology codes were used to define specific subtypes (18). Urinary bladder cancers with benign or uncertain behavior and *in situ* tumors were also accounted for (ICD-10: D09.0, D30.3, D41.4), while non-melanoma skin cancers (ICD-10: C44) were excluded. To estimate cancer-specific prevalence for each patient, we considered only the first primary cancer occurring in that specific site. Multiple primary cancers in different organs diagnosed in the same person were included in each site-specific analysis. For the combinations of cancer types, only the first primary tumor was considered.

2.3 Quality checks

To ensure comparability and to verify the completeness of CR incidence and follow-up data and in agreement with well-established international guidelines and standards (16, 19), the following three quality indicators were calculated for each CR: the proportion of cases known by death certificate only (DCO), a common indicator for cancer registration accuracy and completeness; the proportion of microscopic verifications (MVs), an indicator of the quality of the documentation available to the registry; and the percentage of cases lost to follow-up before 5 years (<5% loss leads to little bias in survival analyses) (20).

2.4 Limited duration prevalence

Limited duration prevalence (LDP) on 1 January 2018 (i.e., index date) was computed from observed incidence and follow-up data for each CR. LDP includes only cases diagnosed after the start of the CR activity and was calculated up to the maximum registration period (between 9 and 40 years), stratified by cancer type, sex, 5-year age groups (from 0–4 to 80–84, and 85+), and years since diagnosis. The calculations were performed by counting the number of persons known to be alive at the index date and adjusting for those lost to follow-up, as implemented in the SEER*Stat software (21). For the eight CRs with the last year of incidence before 2017 (i.e., 2015 or 2016), LDP was calculated for the last 3 years available and projected to 1 January 2018 by CRs, cancer type, sex, age, and time since diagnosis, using a linear regression model with the calendar year as an independent variable (17, 22).

10.3389/fonc.2023.1168325

TABLE 1 Period of registration, population, and incident cases in Italian cancer registries, 1978–2017.

Cancer registry	Period of registration	Years of registration	Population on 1 January 2018 (×1,000)	Incident cases up to 2017[a]
Basilicata	2005–2017	13	563	38,934
Bergamo	2007–2017	11	1,111	73,172
Bolzano-Bozen[b]	1995–2017	23	528	59,084
Brescia[b]	1999–2017	19	1,162	128,909
Caserta	2008–2016	9	916	38,830
Catania-Messina-Enna[b]	2003–2017	15	1,870	140,024
Ferrara[b]	1991–2017	27	348	72,436
Firenze-Prato (Florence)[b]	1985–2016	32	1,269	237,326
Friuli Venezia Giulia[b]	1995–2017	23	1,211	200,985
Genova (Genoa)[b, c]	1993–2016	24	836	158,893
Mantova-Cremona[b]	1999–2016	18	763	75,897
Modena[b]	1988–2017	30	703	121,185
Napoli 3 Sud (Naples)[b, c]	1996–2017	22	1,179	79,628
Nord Sardegna[b]	1992–2015	24	329	38,879
Nuoro	2003–2015	13	209	14,678
Palermo[b]	2003–2017	15	1,205	90,021
Parma[b, c]	1978–2017	40	450	104,062
Pavia[b]	2003–2017	15	546	55,825
Piacenza	2006–2017	12	287	24,565
Puglia (Apulia)	2006–2017	12	2,760	179,070
Ragusa-Caltanissetta[b]	1981–2017	37	588	56,429
Reggio Emilia[b, c]	1996–2017	22	534	66,768
Romagna[b, c]	1993–2017	25	1,126	179,599
Salerno[b]	1996–2017	22	1,091	102,970
Siracusa (Syracuse)[b]	1999–2017	19	401	35,072
Sondrio[b, c]	1998–2017	20	181	21,943
Torino (Turin)[b]	1985–2015	31	861	171,960
Trento[b]	1995–2017	23	540	65,203
Umbria[b]	1994–2017	24	885	120,371
Varese-Como	1990–2015	26	885	124,192
Veneto[b]	1990–2017	28	2,122	355,631
All CRs			28,057	3,276,906
Italy			59,937	

[a]Malignant cancer except non-melanoma skin cancer (ICD-10: C00–C43, C45–C66, C68–C96) and bladder cancer (C67, D09.0, D30.3, D41.4).
[b]CRs with at least 15 years of incidence are included to estimate the completeness index (using model-based incidence and survival, 47% of all incident cases).
[c]CRs with information on the stage of colorectal and breast cancer (Reggio Emilia CR only for breast cancer).

Area	All CRs	CRs with ≥15 years of registration and follow-up at 2018
North-Centre	43%	33%
South-Islands	55%	28%
Italy	47%	31%

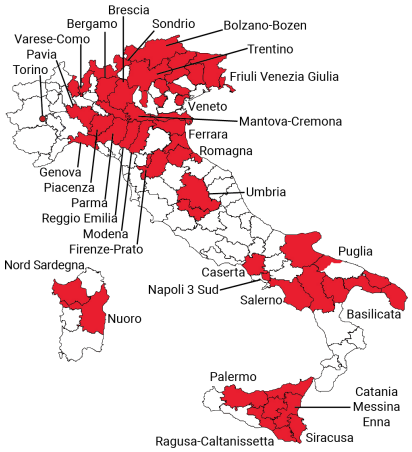


FIGURE 1
Areas and proportions of the Italian population included in the analyses. North-Centre includes Umbria and northern CRs.

TABLE 2 Cancer sites or types and number of cases included: Italian cancer registries, 1978–2017.

Site or type	ICD-10	ICD-O-3 (T and M) or TNM	Cases[a]
All malignant cancers but the skin	C00–43, C45–96, D09.0, D30.3, D41.4		3,276,906
Head and neck	C00–14, C30–32		115,794
Oral cavity	C01–14		60,917
Mouth (excluded Base of Tongue)	C02–06		26,870
Salivary glands	C07–08		7,012
Oropharynx	C01, C09–10		13,314
Nasopharynx	C11		4,933
Esophagus	C15		22,916
Stomach	C16		153,726
Small intestine	C17		10,203
Colorectal	C18–C21		420,726
Colorectal, Stage I[b]	C18–C21	Stages I	7,874
Colorectal, Stage II[b]	C18–C21	Stages II	12,229
Colorectal, Stages III–IV[b]	C18–C21	Stages III–IV	18,989
Colon	C18		291,678
Rectal	C19–20		119,832
Anus	C21		9,216
Liver	C22		110,888
Hepatocellular carcinoma		C22.0–C22.1, 8170–8175, 8970	48,964
Intrahepatic cholangiocarcinoma		C22.0–C22.1, 8013, 8020, 8041, 8154, 8160–8162, 8180, 8240, 8246, 8249, 8470	6,905

(Continued)

10.3389/fonc.2023.1168325

TABLE 2 Continued

Site or type	ICD-10	ICD-O-3 (T and M) or TNM	Cases[a]
Other hepatic cancer		C22.0–C22.1, any morphology except: 8013, 8020, 8041, 8154, 8160–8162, 8170–8175, 8180, 8240, 8246, 8249, 8470, 8970, 8800–8991, 9020, 9040–9044, 9050–9055, 9120–9133, 9140, 9150, 9170, 9180, 9220, 9231, 9240, 9251, 9260, 9364–9365, 9473, 9540, 9560–9571, 9580–9581, 9590–9989	55,137
Gallbladder	C23–24		43,842
Pancreas	C25		103,073
Larynx	C32		43,956
Lung bronchus trachea	C33–34		370,034
Bone	C40–41		6,453
Skin melanoma	C43		86,824
Mesothelioma	C45		13,659
Kaposi sarcoma	C46		8,007
Connective tissue	C47, C49		17,580
Soft tissue sarcoma[c]		All cancers sites except C40.0–C41.9, C32.3; C33.9; C34.0; C30.0; C30.1 (includes unknown primary sites): 8710, 8711, 8714, 8800, 8801, 8802, 8803, 8804, 8805, 8806, 8810, 8811, 8812, 8813, 8814, 8815, 8825, 8830, 8832, 8833, 8840, 8842, 8850, 8851, 8852, 8853, 8854, 8855, 8857, 8858, 8890, 8891, 8894, 8895, 8896, 8900, 8901, 8902, 8910, 8912, 8920, 8921, 8930, 8931, 8933, 8934, 8935, 8959, 8963, 8964, 8990, 8991, 9020, 9040, 9041, 9042, 9043, 9044, 9120, 9124, 9130, 9133, 9137, 9150, 9170, 9180, 9181, 9182, 9183, 9185, 9186, 9187, 9192, 9193, 9194, 9195, 9220, 9231, 9240, 9251, 9252, 9260, 9364, 9365, 9540, 9542, 9560, 9561, 9571, 9580, 9581 All cancer sites except C7–C8; C40.0–C41.9; C32.3; C33.9; C34.0; C30.0; C30.1; C60; C44; C63.2: 8940 C49 only: 8004 All cancer sites except C40.0–C41.9, C32.3; C33.9; C34.0; C30.0; C30.1, C56, C71, C72: 9473 All cancer sites except C40.0–C41.9, C32.3; C33.9; C34.0; C30.0; C30.1, C71, C72: 9503	33,054
Bone sarcoma[c]		C40.0–C41.9, C32.3; C33.9; C34.0; C30.0; C30.1: 8800, 8801, 8802, 8803, 8804, 8805, 8806, 8810, 8811, 8812, 8815, 8830, 8840, 8850, 8851, 8852, 8853, 8854, 8855, 8890, 8891, 8894, 8895, 8896, 8900,8901, 8902, 8910, 8912, 8920, 9040, 9041, 9042, 9043, 9044, 9120, 9124, 9130, 9133, 9150, 9170, 9180, 9181, 9182, 9183, 9184, 9185, 9186, 9187, 9192, 9193, 9194, 9195, 9220, 9221, 9230, 9231, 9240, 9242, 9243, 9250, 9260, 9364, 9473, 9540, 9560, 9561, 9571, 9580, 9581 Only in C40.0–41.9: 8004 All cancer sites: 9370, 9371, 9372	5,127
GIST[c]		8936	3,482
Breast (women only)	C50		443,901
Breast, Stage I[d]	C50	Stage I	25,050
Breast, Stage II[d]	C50	Stage II	18,493
Breast, Stages III–IV[d]	C50	Stages III–IV	8,568
Vagina and vulva	C51–52		12,789
Vulvar SCC		C51.0–C51.9, 8051–8084	8,073
Cervix uteri	C53		25,402
Corpus uteri	C54		72,447
Ovary	C56		48,830
Penis	C60		4,124
Prostate	C61		318,705
Testis	C62		17,646
Kidney	C64–66, C68		106,219
Bladder	C67, D09.0, D30.3, D41.4		236,967

(Continued)

TABLE 2 Continued

Site or type	ICD-10	ICD-O-3 (T and M) or TNM	Cases[a]
Brain and CNS	C70–72		51,609
Thyroid	C73		82,532
Thyroid, papillary	C73	8050, 8052, 8260, 8263, 8340–8344, 8350, 8450	65,989
Thyroid, follicular	C73	8290, 8330–8335	6,753
Thyroid, anaplastic	C73	8012, 8020–8035, 8190, 8337	1,456
Thyroid, medullary	C73	8246, 8345–8347, 8510	2,626
Hodgkin lymphoma		9650–9667	20,107
Non-Hodgkin lymphoma	C82–85, C96		112,808
CLL/SLL		9670, 9823	33,834
NHL, DLBC		9678–9684	35,040
NHL follicular		9675, 9690–9698	17,838
Myeloma (plasma cell)		9731–9734	47,029
Leukemia	C91–95		82,873
Precursor cell acute lymphoblastic leukemia		9727–9729, 9835–9837	8,121
Acute myeloid leukemia		9840, 9861, 9866–9867, 9870–9874, 9891–9931	25,516
Chronic myeloid leukemia		9863, 9875	8,960

[a]For the combinations of cancer types, only the first primary tumor was included.
[b]CRs of Genova, Sondrio, Parma, Romagna, and Napoli 3 Sud (5.3% of the Italian population), 1997–2017.
[c]RARECARE (18).
[d]CRs of Genova, Sondrio, Parma, Reggio Emilia, Romagna, and Napoli 3 Sud (6.1% of the Italian population), 1997–2017.

2.5 Survival

Reliable estimates of long-term (>15 years) survival are crucial for both the estimation of cure indicators and the complete prevalence through statistical modeling and completeness index estimation (see below). They should be representative of the population under study and sufficiently robust to allow modelization of survival in the distant past or near future.

Net survival (NS) is the probability that cancer patients survive their cancer up to a given time since diagnosis, after controlling for competing causes of death. NS allows comparison of populations as if the disease under study was the only possible cause of death. NS was calculated for cases of all ages diagnosed in 1991–2017 and follow-up until the end of 2018, using the cohort method and the Pohar Perme approach (23), as implemented by the SEER*Stat software (21).

DCO only and cases incidentally diagnosed at autopsy were excluded from the analysis.

Expected survival was computed from the regional life tables provided by the Italian National Institute of Statistics for each CR area, stratified by age (in years), sex, and calendar year (24).

For the pool of CRs with ≥15 years of incidence (Table 1) and follow-up until 2018, NS estimation was calculated by cancer type, sex, age at diagnosis (0–44, 45–54, 55–64, 65–74, 75+ years), and period of diagnosis (in 3-year periods from 1991–1993 to 2015–2017). For cancers with available stage information (i.e., breast and colorectal), NS estimation was calculated in the period 1997–2017 for a subset of CRs.

Conditional net survival (CNS) was calculated as the probability of surviving an additional number of years, given that patients already survived t years (16).

Model-based net survival was calculated using mixture cure models which consider a population as a mixture of two groups: the cured (i.e., patients who will have the same life expectancy as the general population) and not cured (i.e., the patients expected to die due to their cancer) (13). Consequently, the mixture cure model is a combination of two models which estimate both the proportion of cured patients (i.e., CF: the cure fraction) and the survival function of the remaining "not-cured" patients (i.e., fatal cases, $1 - CF$).

For any cancer type and sex, the model which best fit NS and CNS was explored starting from an age-stratified Weibull model. When this model did not converge, alternative models were

explored, i.e., Weibull without age stratification, age-stratified exponential, or exponential without age stratification. For rare cancer types, with few patients in some strata of sex or age, parameters were calculated by collapsing the relevant strata as specified in Supplementary Table 1. Parameters were estimated using the SAS NLIN procedure. The goodness of fit of "model-based" NS to "observed" NS was evaluated by likelihood ratio tests and by visual comparison (4, 25, 26), for each cancer type, period of diagnosis, sex, and age group.

2.6 Incidence

Incidence function is needed to describe the risk of being diagnosed with cancer, throughout the life span of each birth cohort in the population (i.e., to estimate the incidence before the start of registration by CRs and completeness index, see below). In the present study, a sixth-degree polynomial on age was the best-fitting model and was used to estimate incidence rates by cancer type and sex (27).

Age and cohort parameters of the incidence function were estimated using SAS logistic procedure by fitting crude incidence rates of patients diagnosed between 1990 and 2014 (in 5-year periods) in the same CRs used for survival modelization, between 1995 and 2014 for breast and colorectal cancers by stage. Incidence data were categorized according to cancer type, sex, 5-year age groups, and birth cohort (<1899, 1900–1904, …, 2000–2014). The goodness of fit of the incidence models was assessed by the Akaike information criterion (AIC) as well as by visual comparison between estimated and observed rates.

2.7 Completeness index

The completeness index (R_L) represents the proportion of prevalence observed from CRs with L years of registration, and it is necessary to calculate the complete prevalence as LDP/R_L (28, 29). R_L represents the percentage of completeness of LDP and varied between 0 and 1, depending on the prevalence observed by the registry. Values close to 1 indicate a high level of completeness and, therefore, a small correction to be applied to the observed prevalence. R_L was calculated by cancer type and sex, using the model-based net survival (NS) and incidence (I):

$$R_L(x) = \frac{\sum_{t=x-L}^{x} I(t)NS(t, x-t)}{\sum_{t=0}^{x} I(t)NS(t, x-t)}$$

where x is the age at prevalence and $x - t$ is the age at diagnosis. The completeness index was calculated using the ComPrev software (30).

To evaluate the effect of using different periods of incidence and survival on the completeness index estimates and complete prevalence, a validation was conducted using two registries with a long observation period: Veneto (28 years of duration, in the north, with high prevalence and relatively high incidence rate in comparison with all of Italy) and Ragusa (37 years, in the south, a low incidence and prevalence area). We compared the maximum observed LDP for the two CRs (LDP_{max} at 28 years for Veneto CR and at 37 years for Ragusa CR) with the LDP of the same duration ($\widehat{LDP_{max}}$) estimated by completing LDP at 15 years using three different completeness indexes $R_L(x)$: one based on the 1990–2017 incidence and survival, one on the 2003–2017, and using the $R_L(x)$ provided by the ComPrev software, estimated on SEER data. The calculation has been done as

$$\widehat{LDP_{max}} = \frac{LDP_{15}}{R_{15}} \cdot R_{max}$$

where R_{max} is the index at 28 years for Veneto CR and at 37 years for Ragusa CR.

2.8 Complete prevalence in 2018

Complete prevalence (Prev) was calculated on 1 January 2018. Estimation was based on observed LDP and, for the period before the start of registration, on the estimated fraction of prevalence not observed in the recorded data (28, 29). The estimated complete prevalence at age x ($Prev(x)$) includes all incident cases diagnosed at any age and can be split into two components, observed LDP (durations from $x - L$ to x years) and estimated unobserved ones (from 0 to $x - L - 1$):

$$Prev(x) = LDP_L(x) + Prev_L^{unobs}(x) = \frac{LDP_L(x)}{R_L(x)}$$

$Prev(x)$ was calculated as absolute numbers and proportions by CR, cancer type, sex, and age at prevalence.

For each registry with L<40 years, we also estimated the annual LDP up to 40 years after diagnosis:

$$LDP_d(x) = LDP_L(x) \cdot \frac{R_d(x)}{R_L(x)} \qquad \text{with d} = \text{L} + 1, \ \ldots 40$$

This estimation by years since diagnosis will be used for the calculation of *already cured* patients described in Section 2.11.

The absolute number of prevalent cases in Italy was obtained as the sum of proportions of prevalence estimates (age-, sex-, and cancer type-specific, obtained pooling CRs in the north-central area and in the South-Islands included in this study) multiplied by the corresponding Italian population in the same areas at the index date (24).

2.9 Complete prevalence projections

To obtain complete prevalence projections after 2018 for all CRs, and up to 2018 for CRs with missing incidence data in 2016 or 2017, the complete prevalence was estimated over the last three calendar years available by CR, cancer type, sex, and age. The number of prevalent cases was projected using a linear regression model with the calendar year as an independent variable, assuming that prevalence would follow a linear function. This simplified assumption (linear and constant trend) may not be valid for

long-term projections, but it is reasonable in the medium-term (e.g., 10 years) (17) for common cancer types. The proportions of prevalence estimates (age-, sex-, and cancer type-specific) from CRs in the north-central area and the South-Islands included in this study were multiplied by the corresponding Italian population in the same area at the index date by sex and age (24). It should be noted that the Italian population is observed until 2021 and forecasted in subsequent years when we used estimates based on the "median" forecast scenario.

2.10 Life expectancy of fatal cases, cure fraction, and time to cure

Life expectancy of fatal (LEF) cases is the survival experienced by the 50th percentile (i.e., median LEF) of fatal cases. In the example (Figure 2A) LEF was 1.8 years corresponding to NS = 75.7% half of those above the green dashed line. Not all cancer patients die because of their neoplasm and, for most cancer types, the NS curve reaches a plateau after a certain number of years (approximately 15 years). Notably, we can observe that a small or large proportion of patients will not die because of their neoplasm even if the plateau is not reached.

The CF represents the proportion of incident patients who experience, at diagnosis, the same life expectancy (mortality rates) as their peers in the general population (51%, Figure 2A). CFs have been calculated from mixture model-based NS and represent asymptotical values of NS when the time since diagnosis increases toward "infinity." Since the life expectancy of people with or without cancer is less than asymptotical, and to highlight connections and differences between CF and long-term NS, we also calculated NS at 50 years after diagnosis, at attained ages 90 and 100 years.

CF for all patients was calculated as a weighted average of age-specific CF, each weight being the proportion of incident cases in the corresponding age group. Changes in CF over time were estimated by using the *period* parameter of the survival function, which represents the effects of the "year of diagnosis" and can be modified assuming a linear effect of the period of diagnosis.

Figure 2B shows also the increase of 5-year CNS (blue curve) according to time since diagnosis. When 5-year CNS approaches 100%, patients reach the same life expectancy (mortality rates) as that observed in the general population who is free from cancer. The assumption is that time to cure (TTC) is reached when 5-year CNS becomes higher than 95% (3), thus assuming the residual 5% excess mortality to be clinically negligible. In the example (Figure 2B), the TTC is reached after 8.5 years.

2.11 Cure prevalence and already cured

Cure prevalence (CurePrev) is defined as the proportion of prevalent cancer patients who will not die as a result of cancer. This indicator was estimated by

$$CurePrev_t(x) = \frac{CF_{x-t} * Prev_t(x)}{[NS_{x-t}(t) + NS_{x-t}(t-1)]/2}$$

where CF_{x-t} and $NS_{x-t}(t)$ are, respectively, the cure fraction and the net survival of patients diagnosed at age $x-t$ and follow-up time t, to obtain $CurePrev_t(x)$, the cure prevalence at attained age x. In the present study, the mean NS at the beginning and the end of the year has been applied to each year since diagnosis. In other words, this indicator was computed as the number (or proportion) of prevalent cases having the same life expectancy (mortality rates) as the corresponding group (i.e., same sexes and age) in the general population, conditioned to be alive t years after diagnosis. For each cancer type and sex, the overall *CurePrev* was calculated as

$$CurePrev = \frac{\sum_x (\sum_t CurePrev_t(x))}{Prev_{TOT}}$$

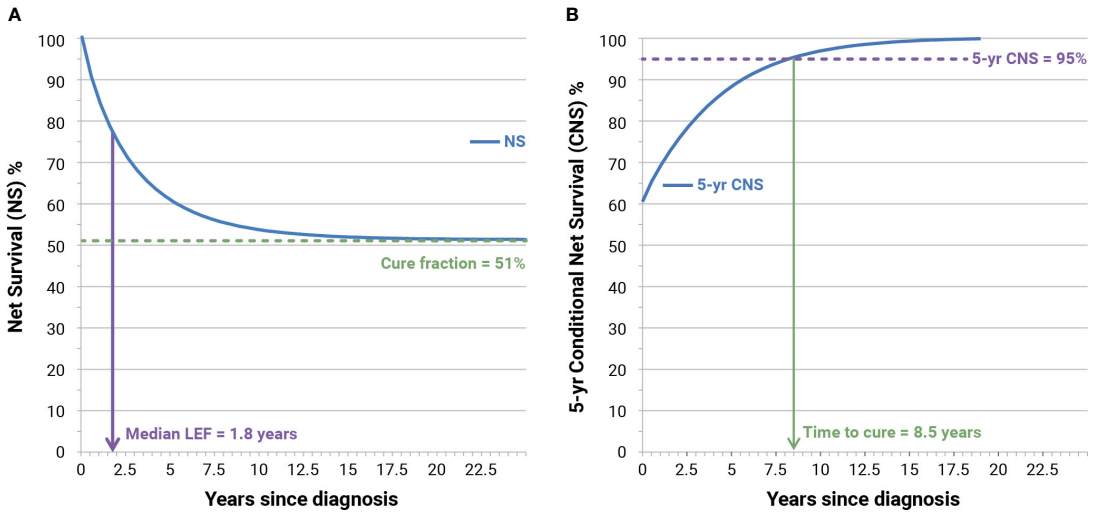


FIGURE 2
Examples of calculation of cure fraction, median life expectancy of fatal (LEF) cases **(A)**, and time to cure **(B)** for Italian patients (men and women) with colorectal cancer diagnosed in 1995 at age 55–64 years. NS, model-based net survival; CNS, conditioned NS.

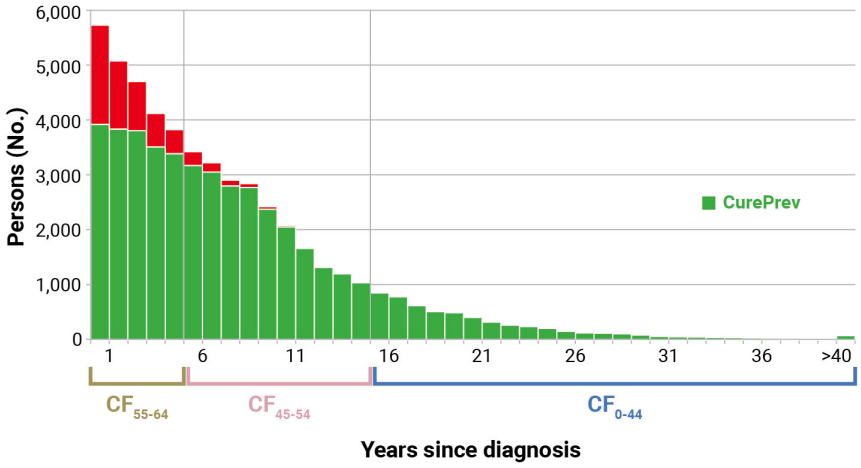


FIGURE 3
Calculation of cure prevalence (CurePrev) for Italian colorectal cancer patients (men and women), aged 55–64 years who were alive in 2018 (January 1st). Calculated applying to complete prevalence at attained age 55–64 the cure fraction (CF) calculated for age at diagnosis, according to years since diagnosis (Section 2.11). The red part of each bar includes cases who are expected to die because of their cancer.

summing up estimates over all ages at prevalence *(x)* where duration is up to the maximum 40 years after diagnosis and $Prev_{TOT}$ is the overall complete prevalence for all age groups considered.

Figure 3 shows an example of the calculation of *CurePrev* in which each annual vertical bar represents the number of patients alive *n* years after diagnosis. The green part of each bar includes cases having the same life expectancy as their peers in the general population (i.e., CF for those alive at that point) and markedly increases with time since diagnosis. Conversely, the red part of each bar includes cases who are expected to die because of their cancer and decreases with time since diagnosis.

To the same distribution of prevalent patients presented in Figure 3, TTC can be applied. Consequently, already cured (*Prev* (>*TTC*)) is defined as the proportion of patients who already reached TTC, defined here as 5-year CNS >95%. It was calculated as the sum of prevalent patients by more than TTC

$$Prev(>TTC) = \frac{\sum_x \sum_{t>TTC} Prev_t\ (x)}{Prev_{TOT}}$$

Estimates of *TTC* were calculated using age at diagnosis of patients, while $Prev_t$ was based on the age of prevalent cases. To overcome this discrepancy, we applied the TTC estimated at different ages at diagnosis to the distribution of prevalent cases at the attained age. In the example (Figure 4), prevalent patients at the attained age of 55–64 years (median 60 years) alive in 2017 had a TTC = 7 years (first 5 years) if diagnosed in the same age group, while they had TTC = 6 years if they were diagnosed at age 45–54 years (median 50 years). Consequently, patients prevalent at 60 years of age who were diagnosed at the same age can be considered cured after 7 years (not yet reached) and after 6 years if diagnosed

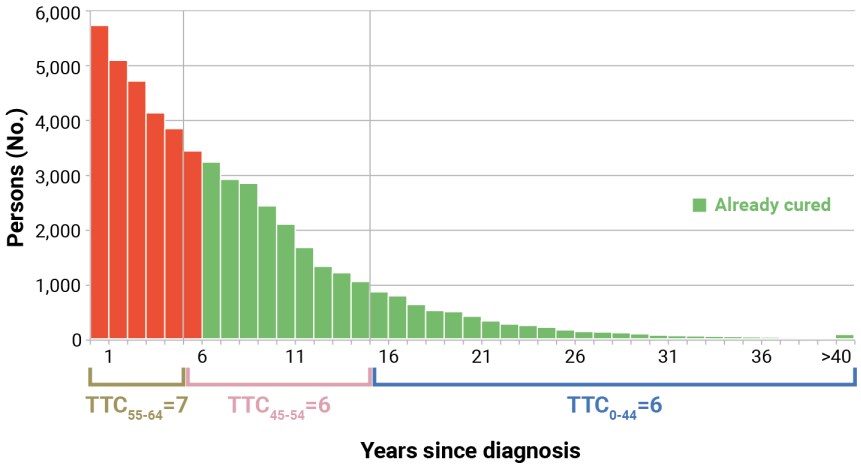


FIGURE 4
Calculation of already cured (Prev>TTC) for Italian colorectal cancer patients (men and women), aged 55–64 years who were alive in 2018 (January 1st). Calculated applying to complete prevalence at attained age 55–64 the time to cure (TTC) calculated for age at diagnosis, according to years since diagnosis (Section 2.11). The red part includes patients who have not yet reached TTC.

younger. Therefore, among these groups, those alive >6 years after diagnosis were considered already cured. The green part of Figure 4 includes already cured patients, while the red part includes those who have not yet reached TTC.

CurePrev included both patients surviving a shorter period than TTC (they will reach it in the future) and a small proportion (<5%, by definition) of already cured ($Prev(>TTC)$) with a small excess risk of death, in comparison with their peers in the general population. Notably, only $Prev(>TTC)$ patients can be individually identified.

In Supplementary Figure 1, the steps needed to calculate complete prevalence on 1 January 2018, projections for the following years, and indicators of cancer cure are summarized. The links among the indicators are also shown and which of them are preliminary to the estimation of the others. For instance, survival estimates are sufficient to calculate CF and TTC. Incidence estimates are also necessary for the calculation of the completeness index and, thus, the complete prevalence. Finally, both estimates of complete prevalence per year after diagnosis and estimates of TTC are needed to calculate the number of already cured patients.

2.12 Ethical approval

The Italian legislation identifies regional health authorities as collectors of personal data for surveillance purposes without explicit individual consent. The approval of a research ethics committee was not required, since this study is a descriptive analysis of pseudonymized cancer data collected by the registries, without any direct or indirect intervention on patients (31).

3 Results

3.1 Quality checks

Three major indicators of data completeness and quality of Italian CRs are shown in Table 3. In the last 10 years of registration (i.e., 2008–2017), the overall percentage of microscopically verified cases was 86.3% with only one CR<80%. The proportion of cases known by death certificate only or with an unknown base of diagnosis was 1.1% with only one CR with a proportion >2%. The percentage of cases lost to follow-up before 5 years was 0.6%, with only 7 out of 31 CRs >1%.

3.2 Validation of survival models

The comparisons of NS and 5-year CNS with corresponding model-based curves were made for all cancer types and sex. As an example, results for the cohort of breast cancer patients diagnosed in 1994–1996 and followed up until 24 years after diagnosis are shown by age groups in Figure 5. Overall, these comparisons and those for the 3-year period cohorts, from 1991–1993 to 2015–2017 (not shown), suggested a very good fit, not only for age-stratified Weibull models but also for exponential models, to estimate long-term model-based survival and cure indicators for breast cancer patients. In particular, for the 2,261 women with breast cancer at age 0–44, the 20-year NS was 64.4% and overlapping values emerged for the age-stratified Weibull models (NS WS = 64.7%) (Figure 5A, solid gold line). Some differences emerged for the age-stratified exponential models (NS ES = 63.0%) (solid blue line), broader for Weibull or exponential models without age stratification (dashed lines: 73.5% and 73.4%, respectively). The corresponding observed 5-year CNS 15 years after diagnosis was 93.9% (Figure 5B), slightly below the threshold for TTC (i.e., 95%), while they were 95.1% when calculated by the age-stratified Weibull or exponential models, 95.6% for Weibull, and 95.8% for the exponential models without age stratification. For patients with breast cancer diagnosed at ages 45–54 years (4,072 women) or 55–64 years (4,747 women), negligible differences emerged between observed and estimations of NS or 5-year CNS based on the age-stratified models (Weibull or exponential) (Figures 5C–F). The same applies at ages 65–74 years (5,355 women) at least until 15 years after diagnosis or attained at the age of 80–89 years (Figures 5G, H). The results of the observed and best-fitting model-based NS and 5-year CNS are also presented for patients with breast cancer by stage at diagnosis (Supplementary Figure 2) and for patients with colorectal (Supplementary Figure 3) or prostate cancers and soft tissue sarcomas (Supplementary Figure 4). A good fit emerged for all of them.

Supplementary Table 1 lists the survival model with the best fit by cancer type with appropriate adjustments for sex and age, if necessary.

3.3 Validation of incidence models

The comparisons between observed and model-based age-specific incidence rates are shown in Supplementary Figure 5. For all cancer types combined by sex, as well as for prostate and breast cancers diagnosed in the period 1990–2014, a very good fit emerged for incidence models to be included in the completeness index estimation. The same validations have been done for all cancer types, by sex and period.

3.4 Validation of the completeness index

In Table 4, frequent cancer types with relatively good prognoses (colorectal, breast, and thyroid cancers and skin melanoma) have been selected as examples in registries with relatively high (Veneto) or low (Ragusa) incidence rates. A less marked difference is expected for patients with poor prognosis or cancer types more frequently diagnosed at older ages when the proportion of patients living >15 years after diagnosis is low regardless. Differences<2% emerged for the four cancer types examined in the Veneto registry between the

10.3389/fonc.2023.1168325

TABLE 3 Quality indicators by cancer registry for cases[a] diagnosed in 2008–2017.

Cancer registry	Records (*n*)	Microscopic verifications (%)	DCO—unknown (%)	Lost to follow-up<5 years (%)
Basilicata	30,501	78.6	0.7	0.1
Bergamo	66,634	89.1	1.1	0.4
Bolzano-Bozen	27,819	91.0	0.9	0.0
Brescia	70,240	80.4	1.6	0.4
Caserta	38,830	86.5	1.6	0.3
Catania-Messina-Enna	97,131	87.4	1.6	1.4
Ferrara	28,756	84.9	0.4	0.9
Firenze-Prato	75,628	83.9	1.0	0.9
Friuli Venezia Giulia	89,495	89.9	0.4	0.5
Genova	58,776	84.6	1.0	0.2
Mantova-Cremona	46,150	86.3	0.7	0.2
Modena	46,445	87.8	0.6	0.0
Napoli 3 Sud	56,903	86.4	1.3	0.2
Nord Sardegna	15,003	84.9	1.7	0.1
Nuoro	9,476	82.8	1.7	0.1
Palermo	62,417	83.2	3.4	0.1
Parma	31,273	88.1	0.7	1.7
Pavia	37,100	82.5	0.9	0.5
Piacenza	20,186	81.2	1.2	1.6
Puglia	151,272	84.5	1.2	0.8
Ragusa-Caltanissetta	27,749	83.1	1.9	0.9
Reggio Emilia	32,221	90.1	0.2	1.1
Romagna	79,120	88.1	0.7	0.1
Salerno	51,616	83.2	1.4	2.5
Siracusa	19,877	82.9	1.8	1.3
Sondrio	11,050	83.7	2.0	0.0
Torino	48,933	88.8	1.6	1.4
Trentino	31,642	87.5	0.3	0.8
Umbria	54,518	92.9	0.4	0.6
Varese-Como	69,044	89.7	0.8	0.4
Veneto	138,489	87.9	0.8	0.4
All CRS	1,624,294	86.3	1.1	0.6

DCO, death certificate only.

[a]Malignant cancer except non-melanoma skin cancer (ICD-10: C00–C43, C45–C66, C68–C96) and bladder cancer (C67, D09.0, D30.3, D41.4).

observed 28-year LDP and the same duration prevalence estimated starting from 15-year LDP using the completeness index calculated from Italian registries with a long-term period of incidence and survival (i.e., 1990–2018). Differences were more marked (+6.1% for colorectal cancer in men, +23.5% for thyroid in women) using only the completeness index based on shorter periods of incidence and survival (2003–2018) and also using the completeness index calculated on SEER data and provided with the ComPrev software (+3.5% for

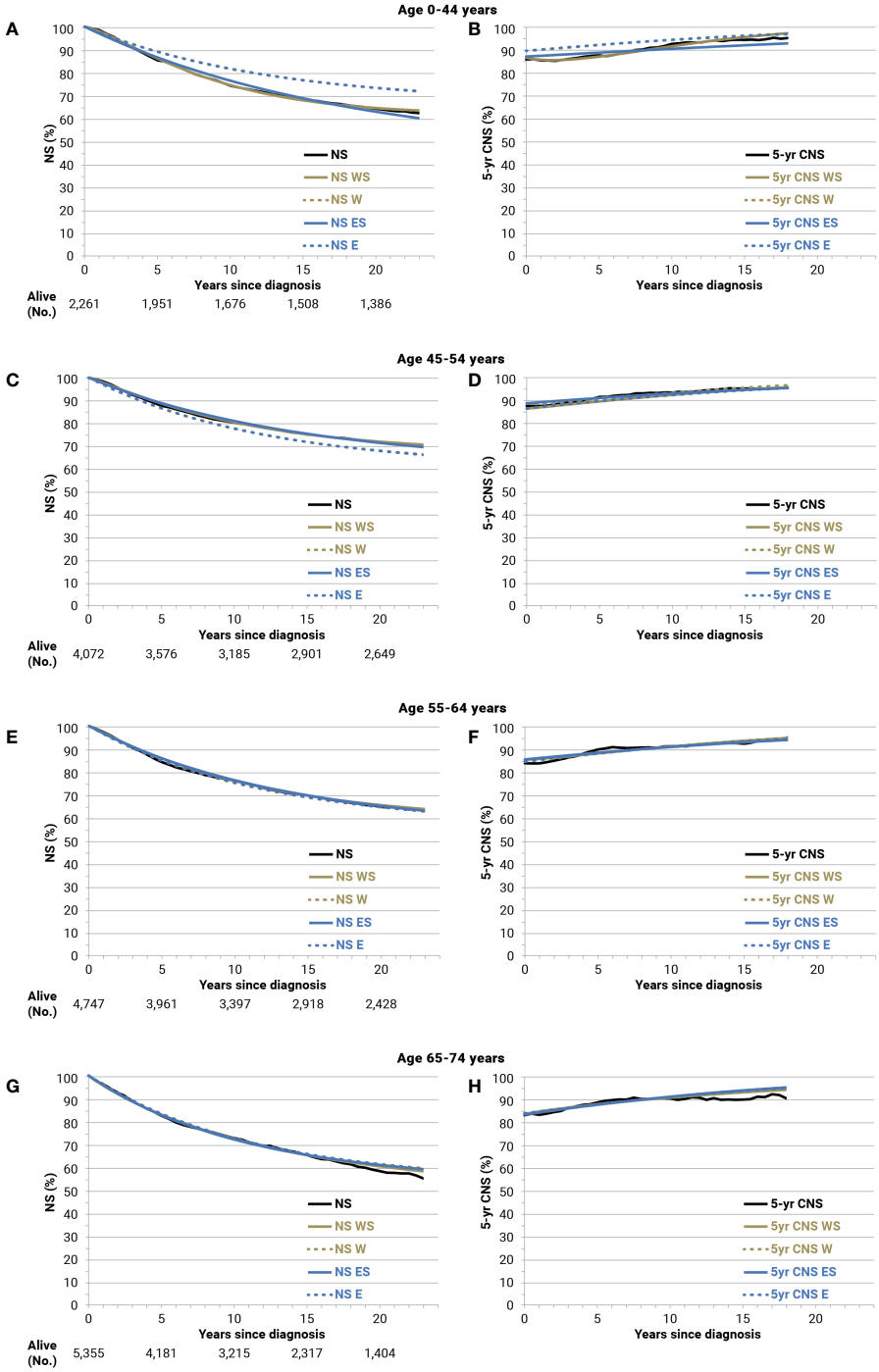


FIGURE 5
Net survival (NS), 5-year conditional NS (5-year CNS), and corresponding model-based estimates until 24 years of follow-up for breast cancer patients (all stages) diagnosed in 1994–1996 and followed up until 2018 by age group: Age 0-44 years **(A, B)**; 45-54 **(C, D)**; 55-64 **(E, F)**; 65-74 **(G, H)**. W, Weibull; WS, Weibull, age-stratified; E, exponential; ES, exponential, age-stratified.

melanoma in men and +9.2% for thyroid in women). In addition, a consistent overestimation emerged for the 37-year LDP completed by the 15-year LDP for the Ragusa registry, approximately +5% using the completeness index based on Italian data 1990–2018 but greater than 10% for some cancer types using both the completeness index based on short period or SEER data (Table 4).

3.5 Completeness index: comparisons

Values of R_L (i.e., completeness index for different lengths of observation) are presented in Table 5 for breast, colorectal, and prostate cancers and all cancer types. The R_L increases with lengths of follow-up and with decreasing age. For colorectal cancer, R_{20} (i.e.,

TABLE 4 Difference between the maximum duration prevalence calculated from 15-year limited duration prevalence (LDP), using different completeness indexes ($R_L(x)$)[a], and observed maximum LDP for selected cancer types.

Registry (maximum duration) Cancer type	Sex	Max LDP			
		Observed	Calculated (%)[b], using $R_L(x)$ from		
			Italy 1990–2018	Italy 2003–2018	SEER 1975–2005
Veneto (28 years)					
Colorectal	Men	8,184	8,143 (−0.5%)	8,680 (+6.1%)	8,196 (+0.2%)
Skin melanoma	Men	3,725	3,728 (+0.1%)	3,627 (−2.6%)	3,857 (+3.5%)
Breast	Women	30,792	31,135 (+1.1%)	32,004 (+3.9%)	30,115 (−2.2%)
Thyroid	Women	4,643	4,555 (−1.9%)	5,334 (+23.5%)	5,070 (+9.2%)
Ragusa (37 years)					
Colorectal	Men	886	931 (+5.1%)	1,004 (+13.3%)	949 (+7.1%)
Skin melanoma	Men	264	282 (+6.7%)	271 (+2.6%)	294 (+11.3%)
Breast	Women	2,763	2,947 (+6.7%)	3,089 (+11.8%)	2,849 (+3.1.%)
Thyroid	Women	698	723 (+3.5%)	895 (+28.1%)	858 (+22.8%)

[a]Calculated as described in Section 3.4.
[b]% represents the difference between calculated and observed.

for a 20-year duration) decreased from 97.2% at age 40–44 in men (96.0% in women) to 78.9% (75.0% in women) at 85+ years. R_{30} was approximately 100% until age 70 years and 10% higher than R_{20} for ages 70 years or more, while R_{40} was always above 98%. Values near 100% for a 20-year duration emerged for prostate cancers mainly diagnosed in older adults, while R_{20}<80% was estimated for breast cancer patients aged >70 years (61.6% for 85+ years). In other words, in CRs with a 20-year duration, the LDP underestimated complete prevalence, with a loss of >20% for women with a previous cancer diagnosis aged 70 years or more (>10% in men) (Table 5).

In Table 6, four estimates of the proportions of prevalent cases observed up to 20 years after diagnosis $R_{20}(x)$ have been compared: those according to estimates made in Italy for 2006 (27), 2010 (22), and 2018 (present estimates), as well as those estimated on SEER data (30).

R_{20} values estimated using the most recent Italian data (i.e., in 2018) were lower than those calculated in 2010, approximately −4% above age 40 years in men. In women, the gap gradually increased with age: −2% at 40 years, −3% at 50 years, and −6% at 75 years. R_{20} values based on SEER data (i.e., those provided by ComPrev) were consistently lower than those calculated from Italian data for women but higher in men above age 30 years (Table 6).

3.6 Cure fraction and long-term NS

In Table 7, CF estimated by mixture cure models until the asymptotical time after diagnosis (thus age) was compared with the estimated 50-year NS and with NS until the attained age of 100 or 90 years, by cancer type, sex, and age at diagnosis.

For pediatric cancer patients overall (age 0–14, Table 7), the difference between CF and 50-year NS is approximately 3%, suggesting a persistent excess risk of death throughout life, though limited. For the other patients, the difference was higher when diagnosed at ages 15–44 and 45–54 years (4%–5%). For older ages, both CF and 50-year NS go far beyond the maximum patient's life span, and their interpretation is fuzzy. For men diagnosed with cancer (all types) at age 65–74 years, CF (asymptotical) was 41%, while the estimated NS after 50 years (attained age over 115 years) was 48%, 49% at the reached age of 100 years, and 50% at the reached age of 90 years. Differences were similar in women aged 65–74 years after any cancer type and after breast cancer (i.e., CF was 61%, 50-year NS was 69%, NS until 100 years was 72%, and NS until 90 years was 76%) (Table 7). Notably, patients diagnosed with prostate cancer at age ≥75 years had a CF = 59%, but the 50-year NS = 68%. The NS until 100 years was even higher (73%) and was 80% until 90 years.

3.7 Cure prevalence (CurePrev): examples and interpretation

The number of patients with colorectal cancer alive in 2018 (January 1st) at age 55–64 years has been presented in Figure 6 (51,855 in the study area, sum of all bars). The green part of the bars included those expected to be cured, with the same mortality as the general population. CurePrev was 68.5% in those with diagnoses after ≤1 year (i.e., *CurePrev*(1) or the green area in the first vertical bar). CurePrev became 75.6% when diagnoses were >1 year and ≤2 years (i.e., the green area in the second vertical bar), and so on. The sum of CurePrev in all the annual intervals (vertical bars, overall CurePrev) was 89.0% and represented the proportion of colorectal cancer prevalent cases at age 55–64 years that will be cured (i.e., they will not die because of the neoplasm). Notably, the sum of *CurePrev*(x) for a duration longer than t years after diagnosis can be calculated as the sum of cases in green areas divided by all prevalent cases after a

10.3389/fonc.2023.1168325

TABLE 5 Completeness index (R_L, %)[a] by sex, age, length (L) of the observation period, and cancer type[b].

Age groups (years)	Cancer type, sex					
	Colorectal cancer, men			Colorectal cancer, women		
	L = 20	L = 30	L = 40	L = 20	L = 30	L = 40
40–44	97.2	99.9	100.0	96.0	99.5	100.0
45–49	96.6	99.7	100.0	96.1	99.3	100.0
50–54	96.4	99.4	100.0	96.0	99.2	99.9
55–59	96.1	99.3	99.9	95.4	99.1	99.8
60–64	95.4	99.1	99.9	94.0	98.8	99.8
65–69	93.9	98.8	99.8	91.8	98.4	99.7
70–74	91.6	98.3	99.7	88.6	97.6	99.6
75–79	88.2	97.4	99.5	84.5	96.3	99.3
80–84	83.8	95.9	99.2	79.5	94.3	98.8
85+	78.9	93.7	98.6	75.0	91.7	98.1
	Prostate cancer, men			Breast cancer, women		
	L = 20	L = 30	L = 40	L = 20	L = 30	L = 40
40–44	99.0	99.0	99.6	99.8	100.0	100.0
45–49	100.0	100.0	100.0	99.3	100.0	100.0
50–54	100.0	100.0	100.0	98.0	99.9	100.0
55–59	100.0	100.0	100.0	95.0	99.8	100.0
60–64	99.9	100.0	100.0	90.2	99.2	100.0
65–69	99.8	100.0	100.0	83.9	97.7	99.9
70–74	99.3	100.0	100.0	77.1	95.0	99.6
75–79	98.0	100.0	100.0	71.2	91.1	98.8
80–84	94.9	99.9	100.0	66.0	86.1	97.2
85+	89.4	99.8	100.0	61.6	81.6	94.7
	All cancers, men			All cancers, women		
	L = 20	L = 30	L = 40	L = 20	L = 30	L = 40
40–44	81.7	91.9	98.9	91.0	96.2	99.3
45–49	83.4	92.1	97.3	91.5	96.8	98.7
50–54	85.1	93.2	96.8	90.9	97.1	98.7
55–59	86.8	94.4	97.1	89.1	97.0	98.8
60–64	88.2	95.4	97.8	86.3	96.4	98.8
65–69	89.4	96.1	98.3	82.7	95.2	98.7
70–74	90.0	96.2	98.6	78.5	93.3	98.3
75–79	88.9	96.1	98.7	73.9	90.7	97.6
80–84	85.7	95.5	98.4	68.6	86.9	96.4
85+	80.9	94.5	98.2	64.2	83.1	94.6

[a]Completeness index calculated from Italian registries with a long-term period of incidence and survival (i.e., 1990–2018).
[b]The extended version is available upon request for the most frequent cancer types by the annual length of observation period from 7 to 40 years.

10.3389/fonc.2023.1168325

TABLE 6 Comparison of different completeness indexes for 20 years of length of the observation period (R_{20}, %) for all cancers combined by sex and age groups.

Age groups (years)	Men				Women			
	R_{20}, %				R_{20}, %			
	Italy at			USA[d]	Italy at			USA[d]
	2018[a]	2010[b]	2006[c]		2018[a]	2010[b]	2006[c]	
0–4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
05–09	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
10–14	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
15–19	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20–24	94.3	94. 6	94.0	92.3	92.7	91.9	89.5	91.5
25–29	85.5	86.1	85.7	85.4	85.9	85.3	85.9	84.3
30–34	81.7	83.0	83.3	85.1	87.1	87.4	83.7	85.4
35–39	80.6	83.4	84.0	85.5	89.4	90.4	86.5	87.2
40–44	81.7	85.8	86.5	85.9	91.0	92.6	89.8	88.2
45–49	83.4	88.2	89.3	86.6	91.5	93.4	91.7	88.2
50–54	85.1	89.8	91.7	87.9	90.9	93.1	92.3	87.4
55–59	86.8	91.1	93.4	89.6	89.1	91.8	91.8	85.9
60–64	88.2	91.7	94.4	91.1	86.3	89.5	90.3	83.9
65–69	89.4	92.4	94.8	92.0	82.7	86.7	88.1	81.4
70–74	90.0	93.0	94.5	91.7	78.5	83.3	85.3	78.5
75–79	88.9	92.4	93.7	90.2	73.9	79.7	82.5	75.1
80–84	85.7	90.9	92.2	87.1	68.6	76.1	80.0	71.4
85+	80.9	88.6	90.1	83.3	64.2	73.6	76.4	67.6

[a]Based on Italian incidence and survival trends in 2018.
[b]Based on Italian incidence and survival trends in 2010 (22).
[c]Based on Italian incidence and survival trends in 2006 (27).
[d]Based on the SEER incidence and survival trends in 2018, estimated from data in 2005 (Race: White) (30).

certain number of years (Figure 6). These *CurePrev* are the probabilities of being cured, conditioned to be already survive *t* years, and the complement of these quantities (i.e., 1 – *CurePrev*) can be read as the residual risk of death for cancer patients.

CurePrev for patients alive >5 years after diagnosis was 97.5% (i.e., 2.5% will die because of the neoplasms), 99.6% for patients alive after >10 years, and became 100.0% for those alive >15 years after diagnosis.

3.8 Already cured prevalence: examples

The same distribution of prevalent patients presented in Figure 6 allowed also the estimation of patients who were already cured, that is the sum of patients alive more than 6 years after diagnosis or 48% of all colorectal cancer patients alive in 2018 at age 55–64 years (Figure 4). Notably, using the TTC (i.e., 7 years) calculated in the same age group of prevalent cases (attained age) (4), the proportion of $Prev(>TTC)$ would be slightly underestimated, reaching only 42%.

4 Discussion

This study provides further insight into the models and procedures useful for estimating the number of people alive after a cancer diagnosis and several indicators of cancer cure. The validations presented describe reliable methods that can also be reproduced in different settings (i.e., countries).

According to our validations, some main observations deserve to be emphasized. The first one is on survival models, the basis for both the calculation of completeness indexes and cure indicators. Although the criteria for selecting the best model are still debated (25, 32), differences among the proposed parametric distributions to estimate long-term survival (e.g., non-mixture models, lognormal, flexible models with splines) (6, 14, 33) are limited

10.3389/fonc.2023.1168325

TABLE 7 Model-based estimates of cure fraction (CF, %) (centered at 2010 as the year of diagnosis), net survival (NS, %) 50 years after diagnosis, until 100 years of age, and until 90 years of age, for selected cancer types by sex and age at diagnosis.

Cancer type (sex) / Age at diagnosis (years)	CF	NS until		
		50 years after the diagnosis	Age 100 years	Age 90 years
All cancers (men)				
0–14	76%	79%	79%	79%
15–44	72%	77%	77%	77%
45–54	51%	56%	56%	57%
55–64	46%	53%	53%	53%
65–74	41%	48%	49%	50%
75+	35%	37%	38%	41%
All cancers (women)				
0–14	79%	81%	81%	81%
15–44	75%	79%	79%	79%
45–54	68%	72%	72%	72%
55–64	55%	60%	61%	62%
65–74	44%	48%	50%	52%
75+	34%	36%	36%	39%
Breast (women)				
0–44	72%	77%	77%	77%
45–54	77%	82%	82%	82%
55–64	71%	77%	77%	78%
65–74	61%	69%	72%	76%
75+	47%	51%	60%	72%
Prostate (men)				
0–44	68%	78%	78%	78%
45–54	86%	93%	93%	93%
55–64	89%	95%	95%	95%
65–74	81%	91%	93%	94%
75+	59%	68%	73%	80%

(32) when sufficient population size and long follow-up are available. In addition, model-based age-stratified estimates based on Weibull distribution of fatal cases showed a very good fit with "observed" net survival for common cancer types (i.e., breast or colorectal at any age and stage and prostate) (Supplementary Figures 2–4) and support their use to estimate completeness index and complete prevalence, as well as cure fraction and time to cure.

A second observation concerns our validation of the impact on the complete prevalence of using different completeness indexes. In principle, models should be built from complete and homogeneous registration periods (i.e., generally short) and, at the same time, should capture long-term survival and incidence trends (i.e., preferably long). Our validations show that the more accurate behavior of completeness indices was obtained using long-term incidence and survival data, although not all CRs provide data for all the years in the study period (Table 4). These results are explained by the assumptions of the completeness index method, calculated by including a back-estimation of incidence before the observed period through age-cohort models, assuming there is no period effect, although often very pronounced (e.g., for prostate after PSA diffusion, after breast cancer screening, for thyroid cancer). This observation may support similar choices in other countries (34) and suggests that more accurate complete prevalence estimations may be obtained using completeness indexes calculated from countries or regions with patterns (e.g., absolute values of incidence and survival and trends of incidence) similar to those of the registry or area to which they will be applied.

10.3389/fonc.2023.1168325

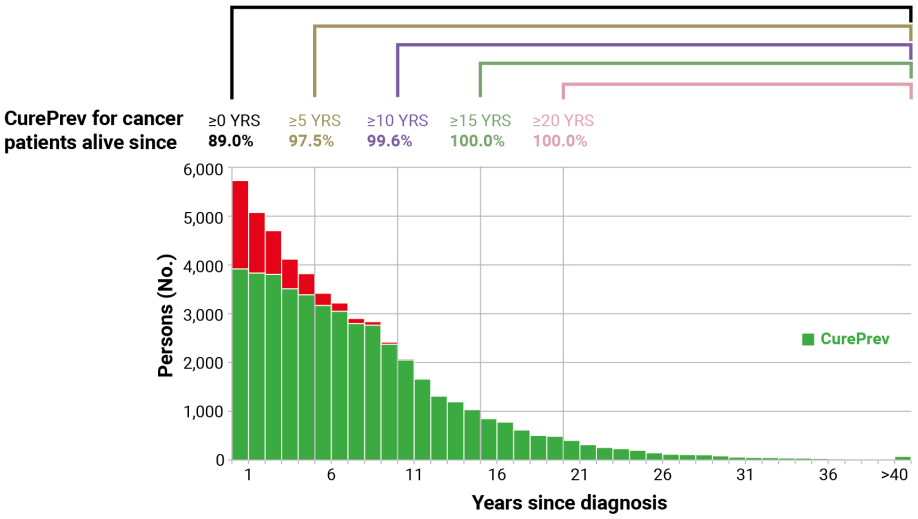


FIGURE 6
Cure prevalence (CurePrev) for Italian colorectal cancer patients (men and women), aged 55–64 years who were alive in 2018 (January 1st), overall and conditioned to be alive after more than 5, 10, 15, and 20 years. The red part of each bar includes cases who are expected to die because of their cancer.

A third point worthy of discussion concerns the assumptions and interpretations of the cure fraction, the estimation of which is also sensitive to the statistical model used. The population-level cure can be estimated by cure models assuming that there are two groups of patients: a group of individuals who experience no excess mortality, whose proportion is estimated by the cure fraction parameter, and a second group (i.e., uncured cases) who experience excess mortality that follows a survival function (35). Cure at the population level is a reasonable and widely accepted hypothesis when the net survival curves plateau and the excess mortality rate was negligible at some point within the follow-up interval (25). When excess mortality estimates (i.e., net survival) show a non-negligible decrease until the maximum follow-up time, the cure fraction should be read only as the proportion of diagnosed cancer patients that will die for causes other than their specific cancer (5), even if we know nothing about the time when those people will die. In the present study, we compared for the first time the estimates of the widely used "asymptotical" cure fraction (which are based on extrapolating very distant observations for periods beyond the end of available follow-up) and estimates of net survival until a reasonable maximum age that a patient may reach (i.e., until age 90 or 100 years, the long tail of the modeled NS curve). The difference between CF and 50-year NS in childhood cancer patients (3% in men and 2% in women), as well as in young adults (15–44 years, 5% and 4%, respectively), should be highlighted, in agreement with studies showing an excess risk of childhood cancer patients for many years after diagnosis (i.e., throughout life) due to treatment effects, second malignancies, or host features (36, 37). The same difference is still more marked for older patients. However, from the patient's point of view and to apply this information to clinical surveillance, it does not seem useful to consider a pediatric patient as uncured when they are alive several decades after diagnosis (38), or if she/he is still alive at age 100 years with a small excess risk of death.

In general, it should be noted that the assumption of only two groups of patients (i.e., cured and uncured), aside from being an extreme simplification, is very conservative. Some patients may have a risk of death higher than the general population associated with the same genetic background, lifestyle, and environmental factors associated with cancer diagnosis (39). The mixture cure models used in this paper did not include the patients' increased deaths from other causes that can be directly related (e.g., adverse effects of treatments) or not (e.g., independent second cancer) with the studied cancer, compared to the general population. Disregarding the presence of this factor leads to estimating a lower proportion of cures, given the definition of cures as those patients who will not die from relapse or disease progression (40). Younger patients, in particular, may be exposed to the detrimental effects of cancer treatments. To overcome these limitations, a more complex mixture model was proposed to capture not only cured and not cured but also the long-term risk of death in children diagnosed with cancer, due to the side effects of cancer treatments, second cancers, and risk factors associated with first cancer carrying an extra risk of death for patients (41). These models should be extended and validated also in adults.

A final point to be highlighted is the calculation and interpretation of cure prevalence, an indicator of the proportion of patients that have the same life expectancy as individuals in the general population of the same sex and age (4, 15). As the number of years since diagnosis increases (conditional on survival). This indicator can be read as the complement of the residual probability of dying from cancer (conditioned to be already survived) and can be helpful to overcome the difficulties of cancer survivors in accessing insurance for a home loan or a mortgage (42, 43).

4.1 Strengths and weaknesses

The major strengths of the presented study are the comprehensive description of the following issues: how the different completeness indices may impact the calculation of complete prevalence, the calculation of indicators of cure with the improvement of algorithms used, and the formal exposition of the links among the different indicators. In the estimation of already cured prevalence, we applied to prevalent cases at attained age the TTC calculated at the age of diagnosis, overcoming the simplified assumption used in the past, when TTC was applied to the complete prevalence of more advanced (reference) ages (4), an assumption that could lead to a slight underestimation of indicator since the TTC increased with age for most of the cancer types. The completeness and accuracy of the Italian CR incidence and survival data were deemed satisfactory (1, 44) and represent a major strength of the study, in particular for the estimation of long-term survival, cure, and prevalence. In addition, the size of the study population and the follow-up length (≥15 years for all CR used in the modelization) contributed also to maximize the reliability of the estimates of incidence and survival parameters, and indicators of cure. It should be noted that few CRs have the last available incidence year and LDP before 2017. For them, LDP and CP (not incidence or survival) were projected in 2018 and thereafter. In our medium-term projections, the hypothesis that CP can be predicted by a linear function of the calendar year as a regressor variable is supported by empirical evidence, at least for all cancer types combined and for most frequent cancer types, consistently showing an approximately linear trend in recent years (17, 22, 45).

Our study has some limitations. First, the probabilities of death for a cause (cancer *vs.* other causes) are estimated at the population level. Therefore, they reflect the overall behavior of a population, which may differ among individuals with cancer (i.e., an individual with comorbidities whose other cause of mortality might be greater or an individual who is compliant with cancer screening programs and whose high health awareness may result in lower other-cause mortality than the general population) (46). Second, in our study, we used an *a priori* threshold of 5% (of 5-year CNS) as a threshold of a low risk of death from cancer, which may be relatively unrestrictive for some groups and inevitably arbitrary. Sensitivity analyses were performed varying this threshold as well as different definitions were used (3, 6, 7, 10). A lower cutoff may be useful among younger individuals who are at low risk of death from other causes (10), and when years to reach 5- or 10-year CNS >90% or 95% were explored (4). It should be noted that the estimation of TTC is sensitive to the choice of the CNS threshold (i.e., 90% or 95% to fix a low risk of recurrence/death or the margin of clinical relevance) and the methodological approach used (3, 4, 7, 8, 10, 32), in particular for cancer types with a non-negligible long-term excess mortality rate (e.g., prostate or breast cancer). Nevertheless, the 5-year CNS >95% is not only clinically relevant and widely reproducible, but it also allows comparability between countries (5, 32, 47, 48).

In addition to the fact that estimates of cure indicators are sensitive to the different models used (whose choice has less impact on the calculation of the completeness index), a specific limit of the present study is that only mixture cure models parametrized according to Weibull or exponential distributions are allowed by the ComPrev software (30). Our mixture model was designed to capture only the long-term excess risk of death due to cancer. The advantages of alternative models include greater modeling flexibility as regards the shapes of the survival distributions and greater sensitivity to small excess risk (14, 33).

Another limitation of studies performing epidemiological indicator projections (17, 49) is the evolution of demographic trends (fertility, migration, and life expectancy) which have a strong impact on predictions of the future population at risk of cancer and profoundly affect the future burden of the cancer prevalence. For instance, the Italian population in 2020 observed in 2022 was 59.6 million, while the same population forecasted in 2015 (17) was 62.5 million (+5%), leading to an overestimation of the absolute number of prevalent cases.

Finally, it should be emphasized that net survival estimates, as cure models, are less reliable for older age groups (e.g., 75 years or more). It is, however, very useful to calculate prevalence (and related indicators) at all ages even if certain cure indicators (i.e., CF and TTC) are considerably less reliable (as well as possibly less useful) for older patients.

5 Conclusions

In the context of a population of cancer survivors expected to increase significantly in Europe and other high-income countries (45, 49, 50), this paper represents an important addition to the current knowledge on the topic providing a comprehensive picture of several available indicators of prevalence and cancer cure. They are unambiguously defined, measurable, and reproducible, e.g., the estimation of the same indicators can be performed in different countries and periods in areas with coverage by population-based cancer registries. Although cure fractions and time to cure are appealing in a clinical context and have widespread applicability, estimation relies on several choices, each associated with pitfalls, that the practitioner should be aware of (30, 43). Nevertheless, these indicators may help to better categorize cancer patients according to the risk of relapse or death many years after diagnosis (12, 51).

Data availability statement

Research data (aggregate) are available from the corresponding authors upon reasonable request.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

LDM and SG drafted the study protocol and the other authors revised the study protocol, collected the data, and prepared the cleaned data for the study database (SFr, RDA, DS, MZ, GM, EB, AR, FC, EM, AP, MFe, CG, MG, GCar, FS, MaM, RC, WM, MFu, PB, GS, SFe, LM, RR, MiM, GCas, LoB, RG, DP, MP, FB, PS, AF, and PP). FT, SG, ADP, and LDM designed the study and did the statistical analyses. SF, RDA, EC, LaB, SR, and SM contributed to the validation of statistical models and revised the statistical analyses. EC and DS specifically discussed the assumptions and clinical implications of the indicators of cancer cure. All authors contributed to the interpretation of the study results. All authors contributed to the article and approved the submitted version.

Members of AIRTUM Working Group

Fabiola Giudici, Ellina Evdokimova (CRO Aviano), Elena Demuru (ISS Roma), Gemma Gatta, Paolo Contiero, Giovanna Tagliabue (Fondazione IRCCS Istituto Nazionale Tumori Milano), Riccardo Capocaccia (E&P), Massimo Rugge (Veneto Cancer Registry–CR), Teresa Intrieri (Tuscany CR), Martina Taborelli (Friuli Venezia Giulia CR), Lucia Bisceglia (AReSS Puglia CR), Stefano Rosso (Piedmont Cancer Registry), Claudia Casella (Liguria CR), Antonietta Torrisi (Catania-Messina-Enna CR), Giovanni Maifredi (Brescia CR), Monica Lanzoni (ATS Insubria CR), Alessio Gili (Umbria CR), Sergio Mazzola (Palermo CR), Maria Francesca Vitale (Napoli 3 Sud CR), Erica Giacomazzi (Val Padana CR), Silvia Ghisleni (Bergamo CR), Maria Adalgisa Gentilini (Trento CR), Fabio Vitadello (SABES-ASDAA Cancer Registry; IRTS), Concetta Patrizia Rollo (Ragusa-Caltanissetta CR), Stefano Marguati (Pavia Cancer Registry), Luciana Del Riccio (Basilicata CR), Maria Rotella (Nord Sardegna CR), Alessandra Sessa (Caserta CR), Antonino Colanino Ziino (Siracusa CR), Ivan Cometti (Sondrio CR), Roberta Bosu (Nuoro CR).

Funding

This work was supported by the Italian Association for Cancer Research (AIRC) (Grant no. 21879). The funding sources had no involvement in the study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

Acknowledgments

The authors thank Dr. Angela B. Mariotto for her helpful comments and are grateful to Mrs. Luigina Mei and Ilaria Calderan for the editorial assistance. The authors would also like to thank the manuscript reviewers for their valuable suggestions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1168325/full#supplementary-material

References

1. Bray F, Colombet M, Mery L, Piñeros M, Znaor A, Zanetti R, et al. Cancer incidence in five continents, vol. XI. In: *IARC sci publ no. 166*. (Lyon, France: International Agency for Research on Cancer (2017).

2. Francisci S, Capocaccia R, Grande E, Santaquilani M, Simonetti A, Allemani C, et al. The cure of cancer: a European perspective. *Eur J Cancer* (2009) 45:1067–79. doi: 10.1016/j.ejca.2008.11.034

3. Janssen-Heijnen MLG, Gondos A, Bray F, Hakulinen T, Brewster DH, Brenner H, et al. Clinical relevance of conditional survival of cancer patients in Europe: age-specific analyses of 13 cancers. *J Clin Oncol* (2010) 28:2520–8. doi: 10.1200/JCO.2009.25.9697

4. Dal Maso L, Guzzinati S, Buzzoni C, Capocaccia R, Serraino D, Caldarella A, et al. Long-term survival, prevalence, and cure of cancer: a population-based estimation for 818902 Italian patients and 26 cancer types. *Ann Oncol* (2014) 25:2251–60. doi: 10.1093/annonc/mdu383

5. Dal Maso L, Panato C, Tavilla A, Guzzinati S, Serraino D, Mallone S, et al. Cancer cure for 32 cancer types: results from the EUROCARE-5 study. *Int J Epidemiol* (2020) 49:1517–25. doi: 10.1093/ije/dyaa128

6. Boussari O, Romain G, Remontet L, Bossard N, Mounier M, Bouvier A-M, et al. A new approach to estimate time-to-cure from cancer registries data. *Cancer Epidemiol* (2018) 53:72–80. doi: 10.1016/j.canep.2018.01.013

7. Dood RL, Zhao Y, Armbruster SD, Coleman RL, Tworoger S, Sood AK, et al. Defining survivorship trajectories across patients with solid tumors: an evidence-based approach. *JAMA Oncol* (2018) 4:1519–26. doi: 10.1001/jamaoncol.2018.2761

8. Romain G, Boussari O, Bossard N, Remontet L, Bouvier A-M, Monuier M, et al. Time-to-cure and cure proportion in solid cancers in France. *A Population-Based Study Cancer Epidemiol* (2019) 60:93–101. doi: 10.1016/j.canep.2019.02.006

9. Kou K, Dasgupta P, Cramb SM, Yu XQ, Baade PD. Temporal trends in population-level cure of cancer: the Australian context. *Cancer Epidemiol Biomarkers Prev* (2020) 29:625–35. doi: 10.1158/1055-9965.EPI-19-0693

10. Colonna M, Grosclaude P, Bouvier AM, Goungounga J, Jooste V. Health status of prevalent cancer cases as measured by mortality dynamics (cancer vs. noncancer): application to five major cancer sites. *Cancer* (2022) 128:3663–73. doi: 10.1002/cncr.34413

11. Prasad V. Use of the word "Cure" in the oncology literature. *Am J Hosp Palliat Care* (2015) 32:477–83. doi: 10.1177/1049909114524477

12. Dal Maso L, Santoro A, Iannelli E, De Paoli P, Minoia C, Pinto M, et al. Cancer cure and consequences on survivorship care: position paper from the Italian alliance against cancer (ACC) survivorship care working group. *Cancer Manage Res* (2022) 14:3105–18. doi: 10.2147/CMAR.S380390

13. De Angelis R, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A. Mixture models for cancer survival analysis: application to population-based data with covariates. *Stat Med* (1999) 18:441–54. doi: 10.1002/(SICI)1097-0258(19990228)18:4<441::AID-SIM23>3.0.CO;2-M

14. Andersson TML, Dickman PW, Eloranta S, Lambert PC. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Med Res Methodol* (2011) 11:96. doi: 10.1186/1471-2288-11-96

15. Gatta G, Capocaccia R, Berrino F, Ruzza MR, Contiero P, EUROPREVAL Working Group. Colon cancer prevalence and estimation of differing care needs of colon cancer patients. *Ann Oncol* (2004) 15:1136–42. doi: 10.1093/annonc/mdh234

16. AIRTUM Working Group. Italian Cancer figures, report 2016. *Survival Cancer Patients Italy Epidemiol Prev* (2017) 41(2 Suppl 1):1–244. doi: 10.19191/EP17.2S1.P001.017

17. Guzzinati S, Virdone S, De Angelis R, Panato C, Buzzoni C, Capocaccia R, et al. Characteristics of people living in Italy after a cancer diagnosis in 2010 and projections to 2020. *BMC Cancer* (2018) 18:169. doi: 10.1186/s12885-018-4053-y

18. Botta L, Gatta G, Trama A, Bernasconi A, Sharon E, Capocaccia R, et al. Incidence and survival of rare cancers in the US and Europe. *Cancer Med* (2020) 9:5632–42. doi: 10.1002/cam4.3137

19. *European Network of cancer registries (ENCR)* (2023). Available at: https://encr.eu/ENCR-Recommendations.

20. Johnson CJ, Weir HK, Yin D, Niu X. The impact of patient follow-up on population-based survival rates. *J Registry Manag* (2010) 37:86–103.

21. *SEER*Stat software, version 8.4.0* (2022). National Cancer Institute (Accessed December 31, 2022).

22. AIRTUM Working Group. Italian Cancer figures, report 2014. prevalence and cure of cancer in Italy. *Epidemiol Prev* (2014) 38:S1:1–144. doi: 10.19191/EP14.6.S1.113

23. Pohar Perme M, Stare J, Estève J. On estimation in relative survival. *Biometrics* (2012) 68:113–20. doi: 10.1111/j.1541-0420.2011.01640.x

24. ISTAT. *Demografia in cifre* (Accessed December 30, 2022).

25. Yu XQ, De Angelis R, Andersson TML, Lambert PC, O'Connell DL, Dickman PW. Estimating the proportion cured of cancer: some practical advice for users. *Cancer Epidemiol* (2013) 37:836–42. doi: 10.1016/j.canep.2013.08.014

26. Dal Maso L, Panato C, Guzzinati S, Serraino D, Francisci S, Botta L, et al. Prognosis of long-term cancer survivors: a population-based estimation. *Cancer Med* (2019) 8:4497–507. doi: 10.1002/cam4.2276

27. AIRTUM Working group. Italian Cancer figures–report 2010 cancer prevalence in Italy: persons living with cancer, long-term survivors, and cured patients. *Epidemiol Prev* (2010) 34(Suppl 2):1–188.

28. Capocaccia R, De Angelis R. Estimating the completeness of prevalence based on cancer registry data. *Stat Med* (1997) 16:425–40. doi: 10.1002/(SICI)1097-0258(19970228)16:4<425::AID-SIM414>3.0.CO;2-Z

29. Merrill RM, Capocaccia R, Feuer EJ, Mariotto A. Cancer prevalence estimates based on tumour registry data in the surveillance, epidemiology, and end results (SEER) program. *Int J Epidemiol* (2000) 29:197–207. doi: 10.1093/ije/29.2.197

30. COMPREV. *Complete prevalence program, version 3.0.29 (BETA)* (2019). Available at: https://surveillance.cancer.gov/help/comprev/technical-requirements/version-history.

31. Presidente del Consiglio dei Ministri. Decreto del presidente del consiglio dei ministri, in: *Identificazione dei sistemi di sorveglianza e dei registri di mortalità, di tumori e di altre patologie, 17A03142*. Available at: https://www.gazzettaufficiale.it/eli/id/2017/05/12/17A03142/sg (Accessed April 26, 2023). GU Serie Generale n.109 del 12-05-2017.

32. Jakobsen LH, Andersson TM, Biccler JL, Poulsen LØ, Severinsen MT, El-Galaly TC, et al. On estimating the time to statistical cure. *BMC Med Res Methodol* (2020) 20:71. doi: 10.1186/s12874-020-00946-8

33. Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics* (2007) 8:576–94. doi: 10.1093/biostatistics/kxl030

34. Demuru E, Rossi S, Ventura L, Guzzinati S, Dal Maso L, Jooste V, et al. Estimating complete cancer prevalence in Europe: validity of alternative vs standard completeness indexes. *Front Oncol* (2023) 13:1114701. doi: 10.3389/fonc.2023.1114701

35. Stedman MR, Feuer EJ, Mariotto AB. Current estimates of the cure fraction: a feasibility study of statistical cure for breast and colorectal cancer. *J Natl Cancer Inst Monogr* (2014) 2014:244–54. doi: 10.1093/jncimonographs/lgu015

36. Shah A, Stiller CA, Kenward MG, Vincent T, Eden TO, Coleman MP. Childhood leukaemia: long-term excess mortality and the proportion 'cured'. *Br J Cancer* (2008) 99:219–23. doi: 10.1038/sj.bjc.6604466

37. Byrne J, Schmidtmann I, Rashid H, Hagberg O, Bagnasco F, Bardi E, et al. Impact of era of diagnosis on cause-specific late mortality among 77 423 five-year European survivors of childhood and adolescent cancer: the PanCareSurFup consortium. *Int J Cancer* (2022) 150:406–19. doi: 10.1002/ijc.33817

38. Haupt R, Essiaf S, Dellacasa C, Ronckers CM, Caruso S, Sugden E, et al. The "Survivorship passport" for childhood cancer survivors. *Eur J Cancer* (2018) 102:69–81. doi: 10.1016/j.ejca.2018.07.006

39. Hinchliffe SR, Dickman PW, Lambert PC. Adjusting for the proportion of cancer deaths in the general population when using relative survival: a sensitivity analysis. *Cancer Epidemiol* (2012) 36:148–52. doi: 10.1016/j.canep.2011.09.007

40. Botta L, Gatta G, Trama A, Capocaccia R. Excess risk of dying of other causes of cured cancer patients. *Tumori* (2019) 105:199–204. doi: 10.1177/0300891619837896

41. Botta L, Gatta G, Capocaccia R, Stiller C, Cañete A, Dal Maso L, et al. Long-term survival and cure fraction estimates for childhood cancer in Europe (EUROCARE-6): results from a population-based study. *Lancet Oncol* (2022) 23:1525–36. doi: 10.1016/S1470-2045(22)00637-4

42. Dumas A, De Vathaire F, Vassal G. Access to loan-related insurance for French cancer survivors. *Lancet Oncol* (2016) 17:1354–6. doi: 10.1016/S1470-2045(16)30452-1

43. Scocca G, Meunier F. A right to be forgotten for cancer survivors: a legal development expected to reflect the medical progress in the fight against cancer. *J Cancer Policy* (2020) 25:1–4. doi: 10.1016/j.jcpo.2020.100246

44. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, et al. Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE–5-a population-based study. *Lancet Oncol* (2014) 15:23–34. doi: 10.1016/S1470-2045(13)70546-1

45. De Moor JS, Mariotto AB, Parry C, Alfano CM, Padgett L, Kent EE, et al. Cancer survivors in the united states: prevalence across the survivorship trajectory and implications for care. *Cancer Epidemiol biomark Prev* (2013) 22:561–70. doi: 10.1158/1055-9965.EPI-12-1356

46. Mariotto AB, Noone AM, Howlader N, Cho H, Keel GE, Garshell J, et al. Cancer survival: an overview of measures, uses, and interpretation. *J Natl Cancer Inst Monogr* (2014) 2014:145–86. doi: 10.1093/jncimonographs/lgu024

47. Yu XQ, Baade PD, O'Connell DL. Conditional survival of cancer patients: an Australian perspective. *BMC Cancer* (2012) 12:460. doi: 10.1186/1471-2407-12-460

48. Xia C, Yu XQ, Chen W. Measuring population-level cure patterns for cancer patients in the united states. *Int J Cancer* (2023) 152:738–48. doi: 10.1002/ijc.34291

49. Maddams J, Utley M, Møller H. Projections of cancer prevalence in the united kingdom, 2010-2040. *Br J Cancer* (2012) 107:1195–202. doi: 10.1038/bjc.2012.366

50. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin* (2019) 69:363–85. doi: 10.3322/caac.21565

51. Tralongo P, Mc Cabe MS, Surbone A. Challenge for cancer survivorship: improving care through categorization of risk. *J Clin Oncol* (2017) 35:3516–7. doi: 10.1200/JCO.2017.74.3450

frontiers | Frontiers in Oncology

TYPE Technology and Code
PUBLISHED 17 July 2023
DOI 10.3389/fonc.2023.1212434

OPEN ACCESS

EDITED BY
Hajo Zeeb,
Leibniz Institute for Prevention Research and Epidemiology (LG), Germany

REVIEWED BY
Renjun Peng,
Central South University, China
Zhirui Zeng,
Guizhou Medical University, China

*CORRESPONDENCE
Nicholas Nicholson
Nicholas.Nicholson@ec.europa.eu

RECEIVED 26 April 2023
ACCEPTED 27 June 2023
PUBLISHED 17 July 2023

CITATION
Nicholson N, Giusti F and Martos C (2023) An ontology design for validating childhood cancer registry data. *Front. Oncol.* 13:1212434. doi: 10.3389/fonc.2023.1212434

An ontology design for validating childhood cancer registry data

Nicholas Nicholson[1]*, Francesco Giusti[2] and Carmen Martos[3]

[1]European Commission, Joint Research Centre (JRC), Ispra, Italy, [2]Belgian Cancer Registry, Brussels, Belgium, [3]Rare Diseases Research Unit, Foundation for the Promotion of Health and Biomedical Research in the Valencian Region (FISABIO), Valencia, Spain

Ontologies can provide a valuable role in the work of cancer registration, particularly as a tool for managing and navigating the various classification systems and coding rules. Further advantages accrue from the ability to formalise the coding rule base using description logics and thereby benefit from the associated automatic reasoning functionality. Drawing from earlier work that showed the viability of applying ontologies in the data validation tasks of cancer registries, an ontology was created using a modular approach to handle the specific checks for childhood cancers. The ontology was able to handle successfully the various inter-variable checks using the axiomatic constructs of the web ontology language. Application of an ontological approach for data validation can greatly simplify the maintenance of the coding rules and facilitate the federation of any centralised validation process to the local level. It also provides an improved means of visualising the rule interdependencies from different perspectives. Performance of the automatic reasoning process can be a limiting issue for very large datasets and will be a focus for future work. Results are provided showing how the ontology is able to validate cancer case records typical for childhood tumours.

KEYWORDS

ontology design, description logic, data validation, data harmonization, cancer registration, childhood cancer

Introduction

A centralised process currently exists for collecting and validating data from the European cancer registries prior to the derivation of indicators that frame the information available on the European Cancer Information System (ECIS) website (1). Dedicated software is used for the validation task, the development of which is a labour-intensive process requiring frequent interactions between the development team and the domain experts. If the rules are updated, there is significant maintenance effort to refactor the code and release the new version. The centralised data collection process is itself facing increased challenges with stricter data-privacy rules and measures, especially for data related to minors. Both these issues impinge directly on the timely availability of cancer-burden indicators which in turn compromises their value in influencing policy-related actions. Computer ontologies provide a key for the provision of more efficient and verifiable data

validation processes as well as for the eventual federation of the processes to the local cancer registry level.

Ontologies are also a valuable tool in general for supporting the work of cancer registries. They provide a knowledge base able to describe entities and their relationships and consequently afford the means of capturing the semantics associated with any given domain. Moreover, the entities defined in one ontology can be linked to entities defined in another ontology. One immediate advantage is that the categorisation and linkage of entities can be made available in one application without the need of having to consult a wide set of different coding and classification standards; all the information is readily accessible. A further interesting feature is that the representation of knowledge in an ontology can be described formally using description logics (DLs). DLs constitute a branch of logics, with most DLs being decidable fragments of first-order logic (2). DLs also provide the possibility for some level of deductive reasoning and this is a useful feature for data validation, which is an essential task of cancer registries.

In order to ensure the necessary harmonisation of data-validation practices in Europe, the European Network of Cancer Registries (ENCR) agrees the rules that constrain the values and ranges cancer data variables can take. Many of these rules have multivariable dependencies and it is difficult to express them in unambiguous terms. Encoding the rules in an ontology allows them to be expressed in a formal sense *via* DL and can highlight inconsistencies in the rules that might otherwise have gone undetected (3).

The ease with which classes and their relationships can be created in an ontology editor such as Protégé (4) belies the difficulties of achieving a good ontology design. There are many ways in which the axioms can be constructed and the way in which they are formulated can have far-reaching implications on computational performance (especially where automatic reasoning is required) and on the ease of extracting information from the knowledge base. Guidelines, tools, and patterns are not widely available and ontology engineering is an emerging field. A key design principle is to achieve wide applicability of an ontology within the domain to avoid a multiplication of ontologies that cannot easily be integrated. This principle has been a driving factor in the design of the ontology for validating childhood cancer registry data.

An additional design aspect that has also to be kept in mind relates to the division of an ontology between pre and post coordination concepts. In pre coordination, knowledge about entities and their relations is asserted *a priori* in the ontology, whereas in post coordination (5), other relationships are inferred following an automatic reasoning process. Both mechanisms are useful and the degree to which one or other is used depends largely on the requirements of the application. Using a predominantly pre-coordinated ontology design would require an unmanageable set of axioms for the validation of cancer registry data. However, post coordination requires automated reasoning to make inferences based on the asserted axioms and can be computationally intensive depending on the expressivity of the DL in which the axioms are formulated.

The design of an ontology is here presented that can model the rules for validating childhood cancers. The ontology serves as the basis for developing a simple programme interface for the systematic validation of cancer registry records. The concept is notably different from the traditional approach of developing dedicated validation software. The validation conditions and machine intelligence are maintained within the ontology itself and the task of any programme interface is reduced solely to managing the data input process, invoking the standard machine reasoning tools and managing the output process. The ontology thereby provides a standalone resource that can be used for many different purposes resulting from its underlying knowledge base and can serve to reduce considerably software development and maintenance costs.

Method

An earlier tentative approach (6) showed the viability of using an ontology for validating cancer registry data and the associated advantages of expressing the rules in DL. The ontology had to be redesigned to allow a more scalable and comprehensive approach to the rules and to build on a number of shared core ontologies. Two validation modules dealing with cancer stage (7) and multiple primary tumours (8) have been developed according to this principle. Both these ontologies were developed as stand-alone applications since they are computationally quite demanding tasks and generally apply only to a subset of cancer registry case records, but they draw on the same shared core ontologies. The third application suite addressed in this article concerns the remainder of the ENCR validation checks, namely those relating to age constraints, tumour signatures, basis of diagnosis, grade, and sex. Figure 1 illustrates the ontology structure, in which the international classification of diseases for oncology, third edition, first revision (ICD-O-3.1) and the international classification of diseases for oncology, third edition, second revision (ICD-O-3.2) modules contain all the ICD-O third edition codes (ICD-O-3) and updates. The MorphologicalGroupChildhood ontology can be swapped out relatively seamlessly dependent on the requirements of the application. It has been designed for validating childhood cancer data which forms the focus of this article and draws from the grouping and subgrouping of the ICD-O-3 codes defined by the international classification of childhood cancer, third edition (ICCC-3) update 2017 (9). This module however can be replaced by any other grouping of ICD-O-3 codes and the resulting application used also for validating adult cancer records.

The ENCRTumourSignature ontology provides the permissible code couplets for topography and morphology values according to the ENCR rules (10) and can itself be used also as a standalone ontology if required. The modular approach to creating ontologies

Abbreviations: Computational complexity classes, PTIME, EXPTIME, N2EXPTIME; DL, description logics; DL Expressivities, ALC, EL, SHIQ, SROIQ; ECIS, European Cancer Information System; ENCR, European Network of Cancer Registries; ICD-O-3, International Classification of Diseases for Oncology; IRI, international resource identifier; LOD, linked open data; SNOMED CT, SNOMED Clinical Terminology; TNM, TNM classification of malignant tumours.

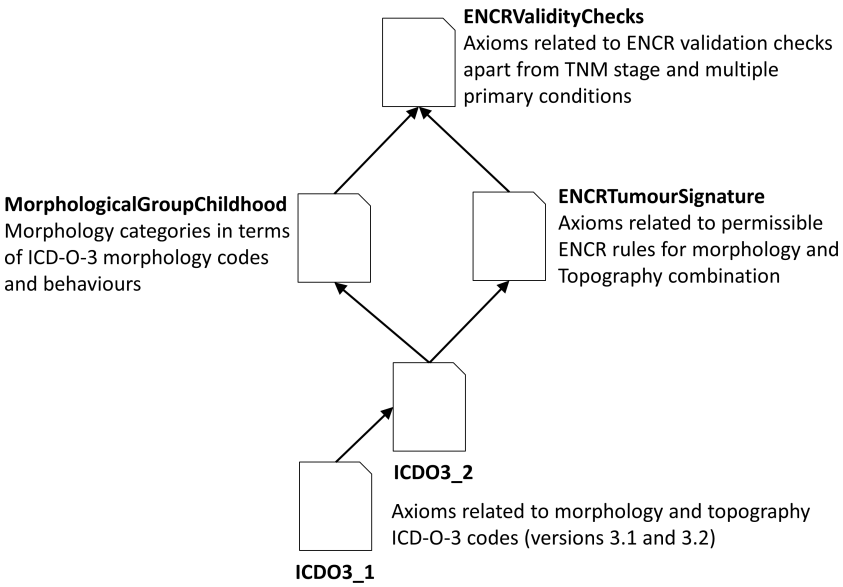


FIGURE 1
The ontology structure of the ENCR validation checks for childhood cancers, not including the cancer stage or multiple-primary checks.

via the mechanism to import ontologies into other ontologies is of great benefit to the scalability, reuse and maintenance of ontologies.

Computational performance accounts for one of the main current drawbacks to automatic reasoning in DL and requires further care in how the ontology is designed. DLs are classified by their expressivities, where expressivity describes the types of operators permitted. Higher expressivities are computationally more demanding but allow more complex reasoning. For example, the DL expressivity ELH (existential language with role hierarchy, allowing concept intersection, existential restrictions, and sub-properties) on which SNOMED CT (11) is modelled, can classify an ontology subsumption hierarchy in polynomial time (PTIME) (12). The higher expressivity of SHIQ has a worst-case complexity of EXPTIME (13) and SROIQ of N2EXPTIME (14). Whereas the introduction of optimised implementations of Tableau-based algorithms has enabled use of higher expressivities in practical applications even for complexities higher than PTIME (13), care has to be exercised to limit the expressivity as far as possible, especially with applications involving many thousands of axioms.

The ENCRValidityChecks ontology includes axioms relating to the constraints on morphology/topography combinations (or tumour signatures), basis of diagnosis, sex, grade, and age at diagnosis or incidence date. The tumour signature axioms (defined in the ENCRTumourSignature ontology) verify that the topography and morphology codes for each cancer case accord with the combinations considered permissible by the ENCR rules. The structure of the tumour signature ontology module passed through a number of design attempts to find an acceptable compromise between usefulness and efficiency. A major issue related to the very large number of morphology codes specified by ICD-O-3 (just under two thousand) and the combination of these codes with a substantial number of topography codes (330 codes).

In an initial design we subclassed the topography codes from the morphology codes, but this forced a coupling in the classification trees between morphologies and topographies. In other ontology modules we needed to specify existential relationships with morphology without automatically pulling in the associated topographies. Nor would it have helped to subclass the morphology codes from the topography codes since this would have resulted in the same problem when specifying existential relationships with topography. Moreover, the open world assumption of DLs meant that we were unable to specify the necessary class subsumption axioms required for automatic validation of the permitted morphology and topography combinations for a given tumour signature. Whereas this caused no difficulty in visualising the asserted topography-morphology relationships in the ontology's graphical user interface, it did mean that such information could not be inferred by the reasoner and therefore not optimal from the point of view of automating the validation checks themselves.

To overcome these issues, we had little option other than to duplicate the entire topography classification hierarchy (under a dummy name) and subclass the morphology codes under the dummy topography classification tree. This allowed a decoupling of the "real" topography codes from the morphology codes (since the morphology codes were then only associated with the dummy topography codes). Given that the real topography codes can be determined from the similarly named dummy topography codes, it is still possible from the graphical user interface to see which morphology codes are associated with a given topography code (and vice versa). This may be appreciated from the partial classification tree of the dummy topography code called "C323Morph" in Figure 2, where it is clear from the name that the associated real topography code is C323. All the morphologies associated with this code are visible in the classification tree under the dummy topography class.

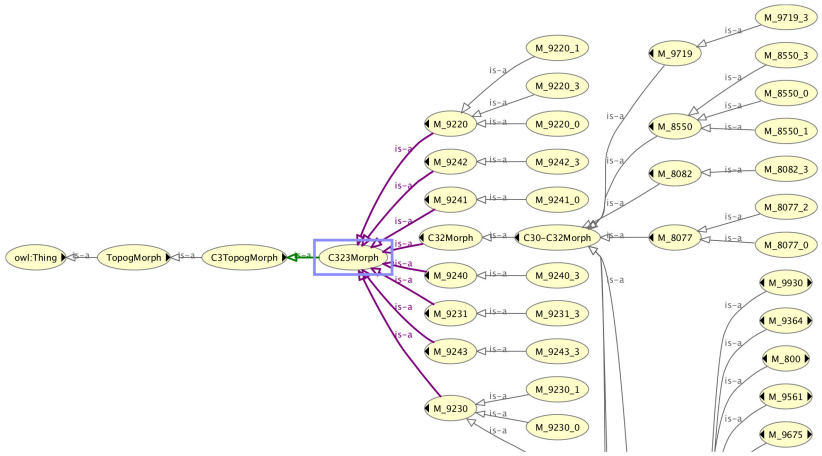


FIGURE 2
Part of the classification tree of the dummy topography code class C323Morph showing the associated morphologies.

Apart from the association of the morphology codes, there is one slight difference between the classification trees of the dummy topography codes and the real topography codes shown for the topography code C323 in Figure 3. The four-digit C323 code is subclassed from its three-digit code parent C32, in contrast to the dummy C323 code (C323Morph) that is the superclass of the three-digit dummy topography code C32Morph. The reasons for inverting the classification tree for the dummy code are firstly to avoid unnecessary duplication of the morphology codes under the dummy topography codes, and secondly to ensure that the existential relationships acting on the morphology codes are correctly specified. The four-digit dummy topography codes have more morphologies associated with them than the three-digit codes and specifying the three-digit codes should not pull in the morphology codes that are only associated with the more granular four-digit codes.

Ascertaining the dummy topography codes (and therefore the real topography codes) with which a given morphology is associated is also straightforward. Figure 4 shows the topography codes associated with the morphology code 8590/1 (namely C56 and C62).

The class subsumption axioms for a valid tumour signature can then be defined along the lines:

$$\exists hasMorphology(C000Morph) \sqcap$$

$$\exists hasTopography(C000) \sqsubset VALID\_TumourSignature$$

which states that the conjunction ($\sqcap$) of an existential relationship ($\exists$) of the topography code C000 and an existential relationship of the dummy topography code C000Morph (under which all the permitted morphologies for the topography code C000 are defined) is a valid tumour signature. This axiom clearly has to be duplicated for all topography codes and results in any valid combination of morphology and topography being subsumed under the class VALID_TumourSignature, allowing a simple test in a batch programme for validating compliant cancer case records.

The checks for basis of diagnosis and grade also raised an interesting challenge for handling them in description logic. DLs incorporate monotonic logic, meaning that a conclusion cannot vary with the addition of a new set of premises. In practical terms, this means that default values or exceptions cannot be attributed and cannot therefore be used to model the scenario in which a rule

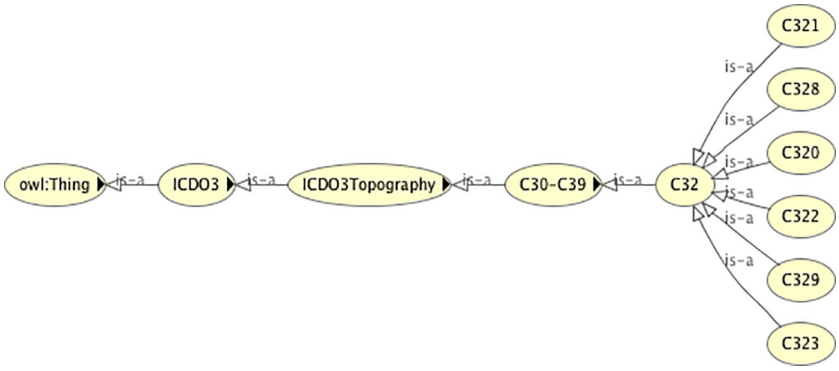


FIGURE 3
The classification tree of the real topography code class C323, showing its position in relation to the associated three-digit topography code C32.

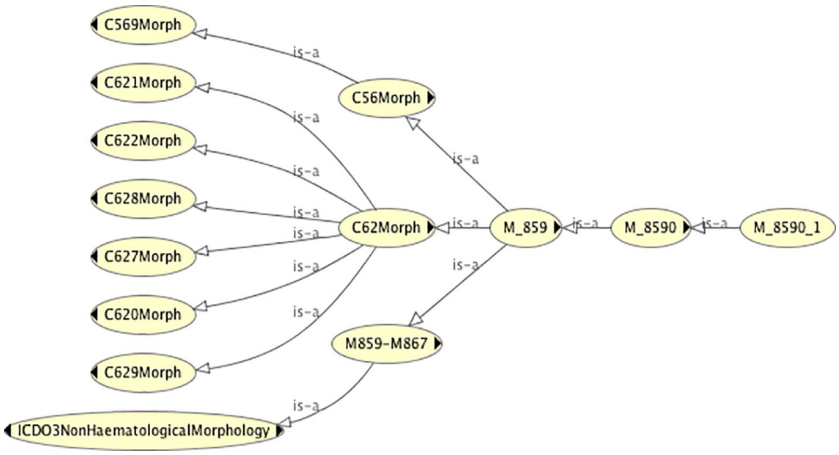


FIGURE 4
The classification tree of the morphology code class 8590/1, showing the dummy topography codes with which it is associated.

takes a default value that may then be overridden for a given condition. The rule tables for both basis of diagnosis and grade are in fact expressed in terms of exceptions to default values.

In order to circumvent this limitation we needed to introduce a default test value (for a given rule) and a violation flag. The default test value is defined either as default-valid or default-invalid. In the case of a default-valid test, if a given set of values violate the rule, then an invalid condition is flagged and vice versa. Thus, the user/application analysing the test results would need to look for any associated violation flag. In the absence of a violation flag, it can be concluded that the test result is valid or invalid depending on the default test value. An example of an axiom providing a default-invalid test value for the basis of diagnosis corresponding to *clinical investigation* is:

$$\exists prevalidatedBasisOfDiagnosis(BoDcode2\_Investigation) \sqsubset InvalidBoDDefaultCase$$

which states that any specified basis of diagnosis code 2 (clinical investigation) is an invalid basis of diagnosis default case. A rule for overriding this default value is:

$$\exists prevalidatedBasisOfDiagnosis(BoDcode2\_Investigation) \sqcap$$
$$\exists hasMorphology(M\_8960\_3) \sqcap$$
$$\exists hasTopography(C64) \sqsubset VALID\_BoD$$

Which, for a specified basis of diagnosis code 2, a morphology code 8960/3 and a topography code C64 (and all its four-digit subclasses), renders the check valid.

The axioms for validating age at diagnosis are less convoluted since they only require verification against minimum and maximum values. For combinations of topography and morphology that have an age restriction, the axioms for a minimum age limit take the form:

$$\exists hasMorphology(M801–M804) \sqcap$$
$$\exists hasTopography(C15) \sqsubset$$
$$\exists expectedAge(>14)$$

which states that the conjunction of morphology codes 801-804 (and all the associated subclasses) with topography code C15 (and associated subclasses) have an expected age greater than 14 years. The axioms for deriving the validation at post-coordination take the form:

$$\exists expectedAge(>14) \sqcap$$
$$\exists patientAgeAtDiagnosis(<15) \sqsubset WARNING\_age$$

for which a specified patient age at diagnosis less than 15 years when the expected age is greater than 14 years generates a warning condition.

The axioms for validating sex are simple, since they involve only a test on topography Thus:

$$\exists hasTopography(C60-C63) \sqsubset$$
$$\exists IsSexOf(Male)$$

ensures that topography codes C60-C63 are associated with the male sex, with the validation rule:

$$\exists IsSexOf(Male) \sqcap$$
$$\exists prevalidatedSex(Female) \sqsubset InvalidSexCombination$$

that states if the specified parameters require a male sex and a female sex is specified, then the cancer case will be subsumed under the class InvalidSexCombination.

Results

Examples are provided in Figures 5–12 of how the ontology handles the ENCR data-validation requirements *via* the post-coordination mechanism for a number of imaginary cancer-case scenarios. The yellow highlighted lines in the figures refer to the inferences made by the reasoner on the basis of the information passed to it (represented by the non-highlighted lines).

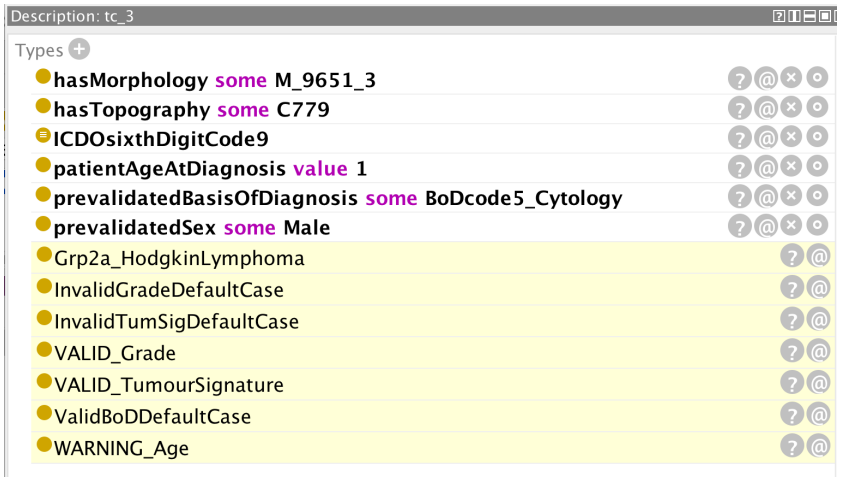


FIGURE 5
Inferences (highlighted lines) drawn by the reasoner on the basis of the specified parameters (non-highlighted lines). The values of grade, basis of diagnosis, and tumour signature are all valid but the reasoner has flagged a warning for the specified age.

In Figure 5, the pre-specified parameters are: morphology 9651 and behaviour (code 3 signifying malignancy in the primary site) – the composite code 9651/3 signifying Hodgkin lymphoma; patient age at diagnosis (one-year old); basis of diagnosis (code 5 signifying cytology); sex (male); and an ICD-O sixth digit code of 9. The ICD-O sixth digit code (grade of the tumour) can take the code values 0-9 and is used for histologic grading or differentiation. Codes 1–4 are only used for non-haematological or solid tumours (with the exception of morphology 9801), and codes 5–8 only for haematological tumours. Code 0 (not applicable) or code 9 (unknown) can be used for both classes of tumour. The highlighted yellow lines in the figure represent the inferences made by the reasoner on the basis of the pre-specified parameters. It can be seen from these inferences that the given parameters constitute invalid default cases for both tumour signature and grade, but that these default cases have been overridden by the respective "VALID" flags. Conversely, a default valid basis of diagnosis has been inferred and since this has not been overridden by an "INVALID" flag, it can be assumed that the basis of diagnosis is also valid. In addition, the reasoner has inferred an unlikely age for the input age parameter. The rationale for the inference of any given statement can be ascertained by clicking on the question mark next to the inferred statement. The explanation for the age warning (Figure 6) is that the expected age at diagnosis is greater than 2 for Hodgkin lymphomas (classified under the ICCC-3 group IIa, which the reasoner has deduced from the morphology code).

In Figure 7, the reasoner has inferred an ICCC-3 group V morphology (retinoblastoma) and an invalid grade code. The error results from the attempt to ascribe an immunophenotype grade code (codes 5–8) to a non-haematological tumour (Figure 8). Since this is an absolute rule that is triggered for all non-haematological morphologies (c.f. line 9 of Figure 8), there is no valid grade default case in this instance.

Figure 9 distinguishes between an invalid grade code inference and a grade code warning. Certain morphologies have an implied grade and these codes should be used instead of leaving the value unspecified (grade code 9). In order to determine the implied value (s) of a grade code, the reasoner is less informative and it is necessary to access the class description of the relevant morphology code (in this case 9511, c.f. Figure 10). The only grade code that is not invalid for this morphology (which is a non-haematological morphology) is 1 and thus it can be inferred that the implied grade is 1. Extra classes and rules could be added to the ontology to provide the implied grades directly but this is one of the compromises taken to avoid affecting performance further. An application programme interfacing with the ontology could determine the implied grade as easily as the user on the basis of the asserted axioms.

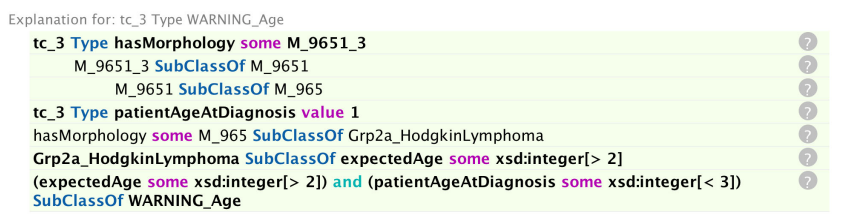


FIGURE 6
Explanation of why the reasoner generated a warning on the age specified in Figure 5.

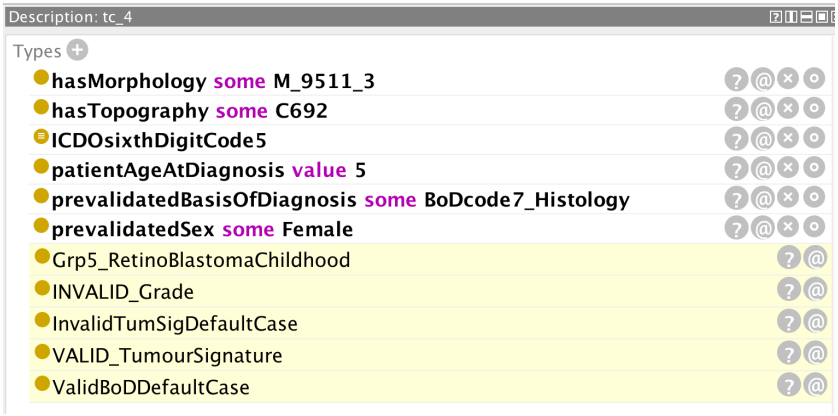


FIGURE 7
Inferences (highlighted lines) drawn by the reasoner on the basis of the specified parameters (non-highlighted lines). The reasoner has inferred an invalid grade on the basis of the specified morphology.

Figure 11 is an example of a cancer case with an erroneous basis of diagnosis code (non-microscopic clinical investigation) which is not admissible for this morphology type (juvenile myelomonocytic leukaemia). The rules for a basis of diagnosis code 2 are by default invalid, and valid cases are flagged as exceptions. The grade and tumour signature combinations also derive from invalid default conditions but these have been overridden by the valid flags (ultimate 2 lines of Figure 11).

Figure 12 is an example of a cancer case with an error in the encoding of the patient's sex. Topography code C569 (ovary) pertains solely to the female sex and the reasoner has inferred the error correctly,

Reasoning times are dependent on the specific reasoner used as may be appreciated from Table 1 which shows the time to classify the ENCR validation check ontology with and without the ENCR tumour signature checks for three reasoners (FaCT++, Hermit, and Pellet) on a 3GHz Intel Core i7 processor with 16 GB RAM. It is interesting to compare the performance of the Hermit reasoner with the other two reasoners in relation to the ENCR tumour signature ontology. It is not immediately clear why Hermit should take significantly longer to classify this particular ontology than the other reasoners (especially since they all use optimised Tableau-based techniques).

In terms of the domain knowledge encapsulated in the ontology, Table 2 shows a breakdown of the knowledge that can be ascertained *via* the pre-coordination or post-coordination processes. Pre-coordinated knowledge remains accessible also after post coordination.

A further useful functionality of ontologies comes from the ease of annotating a class with other information. Figure 13 shows the annotations associated with the ICD-O-3 morphology code 9651/3, from which it can be seen that all the descriptive text of ICD-O-3 can be captured for a given entity, as well as links to other resources (such as on-line data dictionaries, thesauri, and other ontologies). This allows access to a comprehensive set of knowledge describing the resource directly from a single application.

Discussion

The ontology described here for validating childhood cancer registry cases is a novel alternative approach for data cleaning processes that have traditionally been performed *via* dedicated application software. Using ontologies for these tasks brings a number of advantages. One key strength concerns the use of DL to describe the data validation rules in a formal sense. Formalising

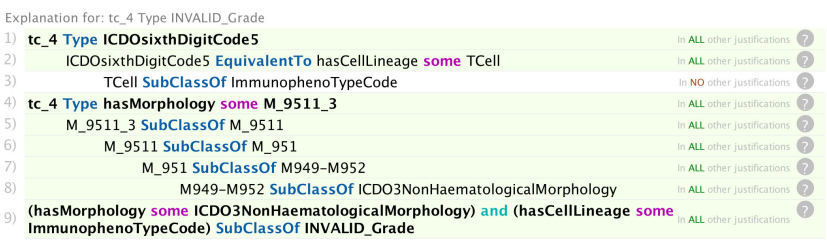


FIGURE 8
Explanation of why the reasoner inferred an invalid value of the grade specified in Figure 7.

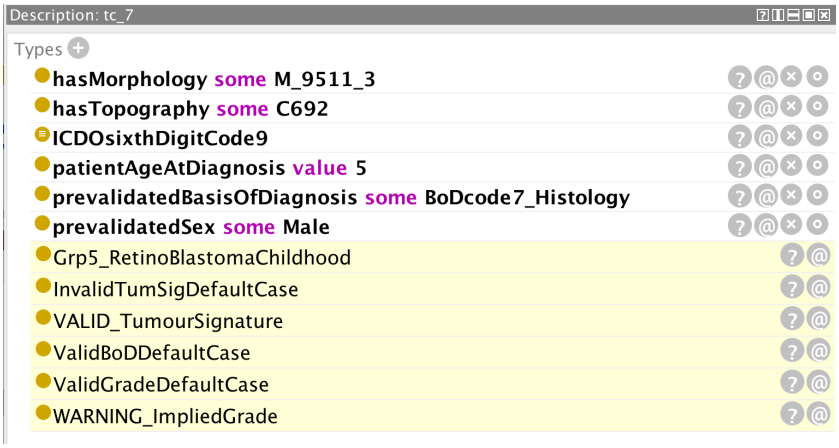


FIGURE 9
Inferences (highlighted lines) drawn by the reasoner on the basis of the specified parameters (non-highlighted lines). The reasoner has inferred an implied grade on the basis of the specified morphology.

the rules not only removes the inherent ambiguity of specifying them in natural language but can also detect inconsistencies within them. A further benefit of DL relates to its amenability to automatic machine reasoning tools that can pre-empt the need to handle complex validation conditions in dedicated software. Keeping the intelligence within the ontology allows a simpler programme interface and reduces software development and maintenance costs.

Ontologies also permit the integration of the data rules with the code classification systems. Cancer registries have to deal with many hundreds of codes from a variety of classification standards and from this standpoint alone, ontologies can structure the information to make it much more readily accessible. By expressing entities and relations in a comprehensive knowledge base, the task of ascertaining and verifying codes and their dependencies becomes a relatively straightforward task. This way of structuring information makes it considerably easier to verify data validation rules that otherwise require multiple table look-ups and also greatly facilitates maintenance issues by keeping the codes, rules, and variable values in a single application. An important corollary to this is the default functionality of OWL ontologies to maintain persistent metadata links *via* the international resource identifiers (IRIs) they assign to each entity as well as their ability to link to other metadata contexts. Access to a relevant set of comprehensive metadata is of fundamental importance to secondary data usage where data users need to understand the meaning of the data. For example, the cancer sites displayed on the ECIS data browser consist of groups of individual topography codes. Ontologies encode this information directly and moreover allow linkage *via* linked open data (LOD) principles to other metadata resources, such as thesauri and data dictionaries. Data users therefore have access *via* a single entry point to a wide source of information and reference material that extends far beyond the immediate classification needs of the ontology itself.

Furthermore, unifying the validation checks with the code classification systems ensures synchronisation of code classification editions with the data-validation rule base and a more thorough versioning control than can be assured *via* distributed software. These aspects are critical to expediting the

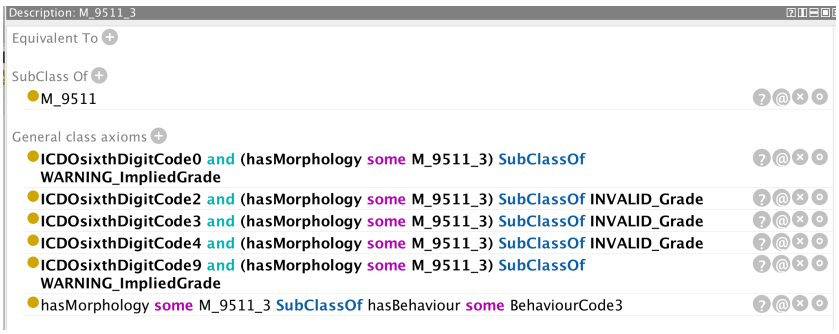


FIGURE 10
Class definition of the morphology code 9511 from which the implied value of grade can be determined.

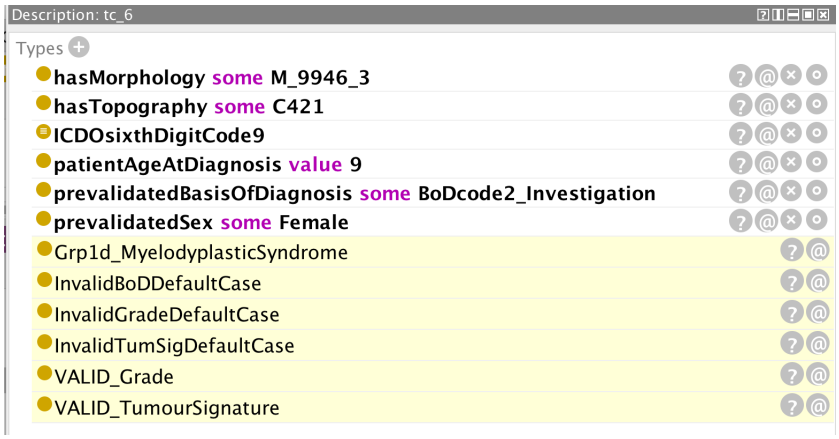


FIGURE 11
Example of a cancer case with an incorrect basis of diagnosis code (code 2, describing a non-microscopic clinical investigation) which is not admissible for this morphology type.

devolution of centralised data cleaning process to the local registry level and also to facilitating any eventual audit process for formally ensuring consistency of local data-cleaning processes.

An additional motivation that may perhaps be the most far-reaching is the potential stimulation of wider collaboration and development within the pan European cancer registry domain. It can justifiably be argued that code classification systems have been structured without the wider contexts in mind and lead to hierarchies that are not optimal to implementation in software. The way in which we had to group the morphology codes in the ontology design under many different class hierarchies dependent upon the particular rules points to the need for a more optimal code classification. This incidentally provides a useful example to show how the logic of ontologies can feed back into improving the representation and structuring of domain knowledge. Disease registry staff with knowledge of how ontologies work would provide a key input into future formulations of such classification systems.

It has to be emphasised however that the design of an ontology is a critical factor in its usefulness and performance. A modular structure such as the one described here helps limit complexity and aids maintenance and further development. It also allows the creation of specific dedicated ontologies using a "pick and mix" approach. For example, the childhood cancer ontology can be made equally applicable to adult cancers simply by swapping out the childhood cancer morphology grouping by an adult cancer morphology grouping. Likewise the most appropriate cancer stage ontology can be used, for example TNM for adult cancers or one that models a stage system more appropriate to childhood cancers, such as described by the Toronto childhood cancer stage guidelines (15). As long as the umbrella class names remain the same, no other changes need to be made in the other ontology applications that import the morphology grouping module. A modular structure is also useful for optimising performance for a particular set of checks and for deciding which reasoner may be best to use (c.f. Table 2).

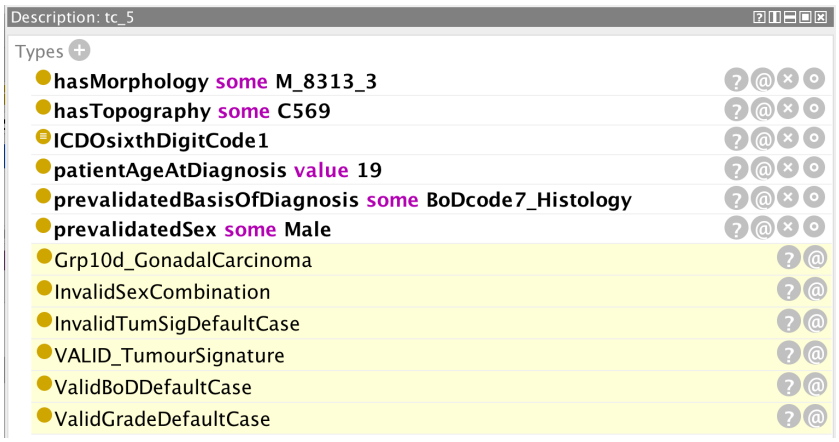


FIGURE 12
Example of a cancer case with an error in the encoding of the patient's sex.

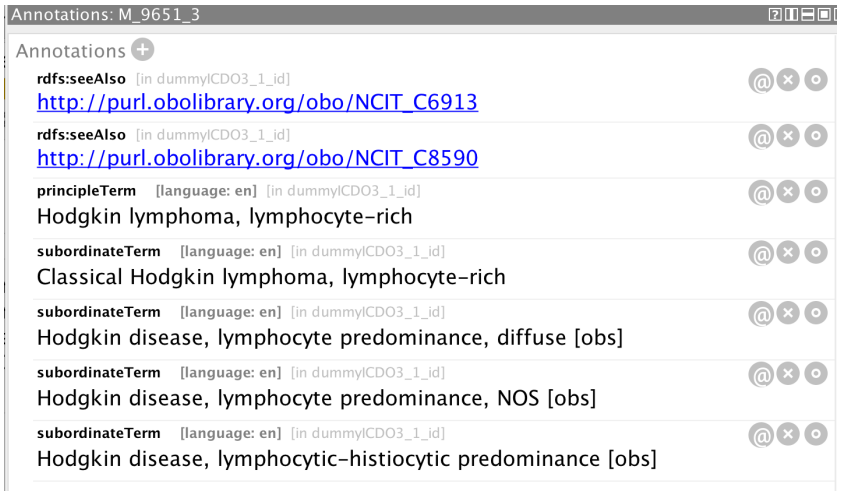


FIGURE 13
Annotations associated with the morphology class M_9651_3 representing the ICD-O-3 morphology code 9651/3.

Performance of automatic reasoning can provide limitations in the validation of large data sets although there are a number workarounds that do not negate the usefulness of ontologies for this task. Limiting DL expressivity to EL reasoning allows algorithms to complete in polynomial time and most cancer-registry validation rules can be handled within these constraints. Where higher expressivities are required, data sets can be ingested as a series of smaller sets and improve efficiency (since reasoning time is not linearly proportional to data-set size). There is also the possibility of exploiting the strengths of the various DL reasoners and future work will seek to understand the reason behind the performance differences observed in Table 2 in order to improve performance on the basis of the types of axioms. Whereas others have addressed comparisons of reasoners (16–22), work has generally focused on their accuracy, the types of operations and platforms they support, and overall performance rather than the strengths of the reasoners given a particular ontology structure. The OWL2Bench however provides a promising approach (23). Optimisation of reasoning processes and algorithms continues to be an active field of research.

A further consideration is that data validation is a highly parallelisable process and other semantic web tools are available for interfacing with an ontology apart from DL reasoning, such as SPARQL queries and direct access *via* a computer programme using the OWL-API application programme interface. The latter provides a solution superior to coding all the information in a dedicated computer programme. The OWL-API provides access to both pre- and post-coordinated information and where reasoning performance is a limiting factor, the computer application can swap out the reasoning functionality with its own dedicated logic on the basis of the ontology axioms without

TABLE 1 Summary of the expressivities and size of the various ontologies comprising the ENCR validation application, with comparison of reasoning performance between various reasoners.

Ontology	DL Expressivity	No. logical axioms	GCI count	Reasoner	Execution time (seconds)
ENCR validation (including ENCR tumour signature)	$ALC^{(D)}$	14,138	7,769	FaCT++	7
				Hermit	10
				Pellet	5
ENCR validation (excluding ENCR tumour signature)	$ALC^{(D)}$	10, 418	7,439	FaCT++	5
				Hermit	1
				Pellet	2.5
ENCR tumour signature (including the ICD-O-3 ontologies)	ALC	7,699	1,534	FaCT++	2
				Hermit	9
				Pellet	2

The GCI count refers to the number of general concept inclusion axioms, which Protégé defines as axioms whose subclass is a complex class expression (and more demanding in terms of reasoning). The DL expressivity ALC denotes attributive language (AL) with complex concept negation (C). The superscript (D) relates to the use of datatype properties..

TABLE 2 Summary of the type of information derivable from the knowledge base *via* the pre and post coordination mechanisms.

Item of knowledge	Mechanism for deriving the knowledge
All the codes for: ICD-O-3 (morphology, topography, behaviour, grade); basis of diagnosis; morphology groups	Pre coordination
Division of morphologies between haematological and non-haematological	Pre coordination
Permissible values of morphology for a given value of topography	Pre coordination (c.f. Figure 2)
Permissible values of topography for a given value of morphology	Pre coordination (c.f. Figure 4)
Morphologies with a specific given behaviour	Pre coordination
Topographies associated with male/female sex	Pre coordination
Whether a given morphology and topography code is a valid combination for a given basis of diagnosis	Post coordination
Whether a given age of diagnosis is unlikely for a given morphology or combination of morphology and topography	Post coordination (c.f. Figure 5)
Whether a given topography and morphology combination is a valid tumour signature	Post coordination (c.f. Figure 5)
The morphology group to which a topography and morphology combination is associated	Post coordination (c.f. Figure 5)
Whether a given grade is valid/invalid for a specific morphology or combination of morphology and topography	Post coordination (c.f. Figure 7)
Whether a morphology has an implied grade	Post coordination (c.f. Figure 9)

having to redefine all the rules and entity relationships. Thus, encoding domain knowledge in an ontology provides many advantages and flexibility in the way of handling information and deriving relationships beyond those explicitly expressed. For data validation purposes at least, this functionality is of considerable benefit.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: The datasets generated for this study are available from: http://data.europa.eu/89h/6f69e886-e5bc-4cd8-9b2d-8aaccf836789. License: European Commission Reuse and Copyright Notice.

Author contributions

Conceptualisation, NN; Data curation, FG, CM; Formal analysis, NN, FG, CM; Methodology, NN, FG and CM; Project administration, NN, CM; Software, NN; Supervision, CM; Validation, FG, CM; Writing—original draft, NN; Writing—review & editing, NN, FG, CM. All authors contributed to the article and approved the submitted version.

Acknowledgments

This work was conducted using Protégé.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. European Cancer Information System (ECIS). European Commission. Available at: https://ecis.jrc.ec.europa.eu/ (Accessed 08 June 2023).

2. Krötzsch M, Simančik F, Horrocks I. *A description logic primer* . Available at: https://arxiv.org/pdf/1201.4089.pdf (Accessed 27 February 2023).

3. Massicano F, Sasso A, Amaral-Silva H, Oleynik M, Nobrega C, Patrao DF. (2015). An ontology for TNM clinical stage inference, in: *Proceedings of the Brazilian Seminar on Ontologies (ONTOBRAS), São Paulo, Brazil*, Aachen, Germany. CEUR-WS. Available at: https://ceur-ws.org/Vol-1442/paper_6.pdf

4. Musen MAProtégé Team. The protégé project: a look back and a look forward. *AI Matters* (2015) 1(4):4–12. doi: 10.1145/2757001.2757003

5. Stevens R, Sattler U. Post-coordination: making things up as you go along (2013). Available at: http://ontogenesis.knowledgeblog.org/1305 (Accessed 26 December 2022).

6. Nicholson NC, Giusti F, Bettio M, Negrao Carvalho R, Dimitrova N, Dyba T, et al. An ontology-based approach for developing a harmonised data-validation tool for European cancer registration. *J BioMed Semant* (2021) 12:1. doi: 10.1186/s13326-020-00233-x

7. Nicholson NC, Giusti F, Bettio M, Negrao Carvalho R, Dimitrova N, Dyba T, et al. A multipurpose TNM stage ontology for cancer registries. *J BioMed Semant* (2022) 13:7. doi: 10.1186/s13326-022-00260-w

8. Nicholson NC, Giusti F, Bettio M, Negrao Carvalho R, Dimitrova N, Dyba T, et al. An ontology to model the international rules for multiple primary malignant tumours in cancer registration. *Appl Sci* (2021) 11(16):7233. doi: 10.3390/app11167233.3

9. International Agency for Research on Cancer. *International classification of childhood cancer* (2017). Available at: https://iicc.iarc.fr/ (Accessed 03 April 2023).

10. Martos C, Giusti F, Van Eycken E, Visser O. *A common data quality check procedure for European cancer registries* (2023). Ispra, Italy: European Commission. Available at: https://encr.eu/sites/default/files/Recommendations/JRC132486_cancer_data_quality_checks_procedure_report_2.0.pdf (Accessed 08 June 2023).

11. *SNOMED CT*. Available at: https://www.snomed.org/ (Accessed 18 April 2023).

12. Tongphu S, Suntisrivaraporn B. Algorithms for measuring similarity between ELH concept descriptions: a case study on SNOMED CT. *Computing Inf* (2017) 36 (4):733–64. doi: 10.4149/cai20174733

13. Baader F, Horrocks I, Sattler U. Description logics. In: Staab S, Studer R, editors. *Handbook on ontologies. international handbooks on information systems*. Berlin, Heidelberg: Springer (2009). doi: 10.1007/978-3-540-92673-3_1

14. Kazakov Y. RIQ and SROIQ are harder than SHOIQ(2008). Available at: https://aaai.org/papers/027-riq-and-sroiq-are-harder-than-shoiq/ (Accessed 18 April 2023).

15. Bhakta N, Rodriguez-Galindo C. The Toronto guidelines: a practical means for childhood cancer staging. *Lancet Child Adolesc Health* (2018) 2(3):158–9. doi: 10.1016/S2352-4642(18)30024-5

16. Weithöner T, Liebig T, Luther M, Böhm S. (2006). What's wrong with OWL benchmarks?, in: *Proc. of the Second Int. Workshop on scalable semantic web knowledgebase systems (SSWS 2006)*, Athens pp. 101–14. Citeseer. Available at: https://www.uni-ulm.de/fileadmin/website_uni_ulm/iui.inst.090/Publikationen/2006/weithoener-et-al-ssws06.pdf.

17. Gardiner T, Horrocks I, Tsarkov D. Automated benchmarking of description logic reasoners(2006). Available at: http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-189/submission_23.pdf (Accessed 18 April 2023).

18. Bock J, Haase P, Qiu J, Raphael V. Benchmarking OWL reasoners(2008). Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.143.1363&rep=rep1&type=pdf (Accessed 18 April 2023).

19. Dentler K, Cornet R, ten Teije A, de Keizer N. Comparison of reasoners for large ontologies in the OWL 2 EL profile. *Semantic Web* (2011) 2(2):71–87. doi: 10.3233/SW-2011-0034

20. Abburu S. A survey on ontology reasoners and comparison. *Int J Comput Appl* (2012) 57(17):33–39. doi: 10.5120/9208-3748

21. Kang Y-B, Li Y-F, Krishnaswamy S. (2012). Predicting reasoning performance using ontology metrics. the semantic web, ISWC 2012, in: *Proceedings of 11th International Semantic Web Conference*, Boston, MA, USA, pp. 198–214. Lecture Notes in Computer Science, vol 7649. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-35176-1

22. Khamparia A, Pandey B. Comprehensive analysis of semantic web reasoners and tools: a survey. *Educ Inf Technol* (2017) 22:3121–45. doi: 10.1007/s10639-017-9574-5

23. Singh G, Bhatia S, Mutharaju R. OWL2Bench: a benchmark for OWL 2 reasoners. In: *The semantic web – ISWC 2020. ISWC 2020. lecture notes in computer science*, vol. 12507. Athens: Springer, Cham (2020). doi: 10.1007/978-3-030-62466-8_6

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 25 July 2023
DOI 10.3389/fonc.2023.1197850

OPEN ACCESS

EDITED BY
Joachim Schüz,
International Agency For Research On Cancer (IARC), France

REVIEWED BY
Ana Xavier,
University of Alabama at Birmingham, United States
Maike Wellbrock,
Johannes Gutenberg University Mainz, Germany
Neimar De Paula Silva,
International Agency For Research On Cancer (IARC), France

*CORRESPONDENCE
Rafael Marcos-Gragera
rmarcos@iconcologia.net

RECEIVED 31 March 2023
ACCEPTED 07 July 2023
PUBLISHED 25 July 2023

CITATION
Trallero J, Sanvisens A, Almela Vich F, Jeghalef El Karoni N, Saez Lloret I, Díaz-del-Campo C, Marcos-Navarro AI, Aizpurua Atxega A, Sancho Uriarte P, De-la-Cruz Ortega M, Sánchez MJ, Perucha J, Franch P, Chirlaque MD, Guevara M, Ameijide A, Galceran J, Ramírez C, Camblor MR, Alemán MA, Gutiérrez P, Marcos-Gragera R and REDECAN (2023) Incidence and time trends of childhood hematological neoplasms: a 36-year population-based study in the southern European context, 1983–2018. *Front. Oncol.* 13:1197850. doi: 10.3389/fonc.2023.1197850

Incidence and time trends of childhood hematological neoplasms: a 36-year population-based study in the southern European context, 1983–2018

Jan Trallero[1,2], Arantza Sanvisens[1,2], Fernando Almela Vich[3,4], Noura Jeghalef El Karoni[3,4], Isabel Saez Lloret[4,5], Cristina Díaz-del-Campo[6], Ana Isabel Marcos-Navarro[7], Amaia Aizpurua Atxega[8], Patricia Sancho Uriarte[8], Marta De-la-Cruz Ortega[8], María José Sánchez[9,10,11,12], Josefina Perucha[13], Paula Franch[14,15], María Dolores Chirlaque[9,16], Marcela Guevara[9,17,18], Alberto Ameijide[19,20], Jaume Galceran[19,20], Cristina Ramírez[21], Marta Rodríguez Camblor[22], Maria Araceli Alemán[23], Pilar Gutiérrez[24], Rafael Marcos-Gragera[1,2,9,25]* and REDECAN

[1]Epidemiology Unit and Girona Cancer Registry, Oncology Coordination Plan, Catalan Institute of Oncology, Girona Biomedical Research Institute Dr. Josep Trueta (IDIBGI), Girona, Spain, [2]Josep Carreras Leukaemia Research Institute, Badalona, Spain, [3]Registry of Childhood and Adolescent Tumors of the Valencian Community, Valencian Community Department of Universal Health and Public Health, València, Spain, [4]Cancer Information System of the Valencian Community, Valencian Community Department of Universal Health and Public Health, València, Spain, [5]Castellón Cancer Registry, Directorate General of Public Health and Addictions, Valencian Government, Castellón, Spain, [6]Ciudad Real Cancer Registry, Health and Social Welfare Authority, Ciudad Real, Spain, [7]Cuenca Cancer Registry, Health and Social Welfare Authority, Cuenca, Spain, [8]Basque Country Cancer Registry, Basque Government, Vitoria-Gasteiz, Spain, [9]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain, [10]Granada Cancer Registry, Andalusian School of Public Health (EASP), Instituto de Investigación Biosanitaria Ibs. GRANADA, University of Granada, Granada, Spain, [11]Instituto de Investigación Biosanitaria Ibs.GRANADA, Granada, Spain, [12]Department of Preventive Medicine and Public Health, University of Granada, Granada, Spain, [13]La Rioja Cancer Registry, Epidemiology and Health Prevention Service, Logroño, Spain, [14]Mallorca Cancer Registry, Public Health and Participation Department, Palma de Mallorca, Spain, [15]Health Research Institute of the Balearic Islands (IdISBa), Palma de Mallorca, Spain, [16]Department of Epidemiology, Regional Health Authority, Instituto Murciano de Investigación Biosanitaria (IMIB)-Arrixaca, Murcia University, Murcia, Spain, [17]Navarra Cancer Registry, Navarra Public Health Institute, Pamplona, Spain, [18]Epidemiology and Public Health Area, Navarra Institute for Health Research (IdiSNA), Pamplona, Spain, [19]Tarragona Cancer Registry, Cancer Epidemiology and Prevention Service, Sant Joan de Reus University Hospital, Tarragona, Spain, [20]Institut d'Investigació Sanitària Pere Virgili (IISPV), Reus, Tarragona, Spain, [21]Albacete Cancer Registry, Health and Social Welfare Authority, Albacete, Spain, [22]Department of Health, Asturias Cancer Registry, Public Health Directorate, Asturias, Spain, [23]Canary Islands Cancer Registry, Public Health Directorate, Canary Islands Government, Tenerife, Spain, [24]Castilla y León Cancer Registry, Public Health Directorate, Castilla y León Government, Valladolid, Spain, [25]Nursing, University of Girona, Girona, Spain

Background: Hematological neoplasms (HNs) are the first and most common childhood cancers globally. Currently, there is a lack of updated population-based data on the incidence of these cancers in the Spanish pediatric population. This study aimed to describe the incidence and incidence trends of HNs in children (0–14 years) in Spain using data from the Spanish Network of Cancer Registries and to compare the results with other southern European countries.

Methods: Data were extracted from 15 Spanish population-based cancer registries between 1983 and 2018. Cases were coded according to the *International Classification of Diseases for Oncology*, third edition, first revision, and grouped according to the *International Classification of Childhood Cancer*, third edition. Crude rates (CRs), age-specific rates, and age-standardized incidence rates using the 2013 European population (ASR_E) were calculated and expressed as cases per 1,000,000 child-years. Incidence trends and annual percentage changes (APCs) were estimated.

Results: A total of 4,747 HNs were recorded (59.5% boys). Age distribution [n (%)] was as follows: <1 year, 266 (5.6%); 1–4 years, 1,726 (36.4%); 5–9 years, 1,442 (30.4%); and 10–14 years, 1,313 (27.6%). Leukemias were the most common group, with a CR and an ASR_E of 44.0 (95%CI: 42.5; 45.5) and 44.1 (95%CI: 42.6; 45.7), respectively. The CR and ASR_E of lymphomas were 20.1 (95%CI: 19.1; 21.1) and 20.0 (95%CI: 19.0; 21.1), respectively. The comparable incidence rates between our results and those of other southern European countries were similar for lymphomas, while some differences were observed for leukemias. From 1988 to 2016, the trend in leukemia incidence was stable for both sexes, with an APC of 0.0 (95%CI: −0.5; 0.7), whereas a constant overall increase was observed for lymphoma in both sexes, with an APC of 1.0 (95%CI: 0.4; 1.6).

Conclusion: Leukemias are the most common HNs in children, and their incidence has remained stable since 1988, whereas the incidence of lymphomas has increased every year. Lymphoma incidence is like that of other southern European countries, while leukemia incidence is similar only to that of southwestern European countries. Collaborative cancer registry projects allow for assessing epidemiological indicators for cancers such as HNs, which helps health authorities and clinicians provide more knowledge about these malignancies.

KEYWORDS

incidence, childhood, hematological neoplasms, incidence trends, leukemia-lymphoma, population-based registries

1 Introduction

Hematological neoplasms (HNs) are divided into leukemias and lymphomas. They account for one-third of all childhood cancers and are among the most common cancers in children (1). Leukemias are a group of diseases involving an uncontrolled proliferation of hematopoietic stem cells in the bone marrow caused by several risk factors such as genetic (e.g., chromosomal translocations, rare germline mutations, or epigenetics) and environmental factors (e.g., infections and exposure to chemicals or ionizing radiation) (2, 3). Meanwhile, lymphomas are a diverse group of diseases that arise from the clonal proliferation of lymphocytes (4). Due to the heterogeneity of these malignancies and the improvement of diagnostic methods based on genetic and pathological examinations (5), the *International Classification of Childhood Cancer* (ICCC), which classifies cancer histology codes in children, has been updated to the latest edition, the third edition (ICCC-3) of 2017 (1).

The most recent international collaborative study on childhood cancer (0–14 years), published in 2017 and covering the period 2001–2010, found an age-standardized rate (ASR) using the Segi world standard population (ASR_{SEGI}) of 61.6 cases per million

child-years, of which leukemias accounted for 46.4 and lymphomas for 15.2 (6). Slightly lower rates were estimated in a European study, covering the period 1988–1997, which compared childhood cancer incidence between geographical areas in Europe (7). Both studies reported higher incidence rates, particularly in southern Europe, compared with the rest of the world (6, 7).

The latest European study assessing incidence trends showed a statistically significant increase in the annual percentage change (APC) of 0.7% per year from 1991 to 2010 for childhood leukemias, whereas no increase was observed for lymphomas (8). In addition, some specific countries or regional population-based cancer registries have shown different results on childhood HN incidence trends (9, 10). Previously published data on childhood HNs in Spain covered the period 1983–2002 and reported an ASR_{SEGI} of 64.4 cases per million children, of which 49.9 cases were leukemias and 18.5 were lymphomas. The incidence trends for the same period were also discussed and showed an increase in the early years (1983–1991), followed by a stabilization in the second period (1992–2002) (11). Later studies covering a longer period up to 2013 for the incidence and up to 2007 for the incidence trends reported similar results for all HNs and leukemias (1, 12, 13); however, lymphomas showed stability throughout the period (13). The lack of updated results on the incidence and incidence trends of HNs in Spain requires an analysis of more recent years.

Previous studies have reported that developed countries tend to have higher incidence rates of HNs and southern European countries have higher incidence rates of lymphoma (6), in addition to a variation in trends (9, 10). Although stable incidence rates have previously been described in Spain (12, 13), we hypothesize that similar results to those reported in Europe will currently be observed, with higher incidence rates and changes in trends for both HN groups. Therefore, this study aims to provide an overview of the incidence and incidence trends of childhood HNs in Spain. This will be performed by adding more recent years and coverage areas to the previous studies, taking into account the most recent classification of childhood HNs, by age group, sex, and cancer type. Furthermore, we aim to compare the results with other population-based cancer registries in southern Europe.

2 Methods

2.1 Study population

Data on childhood HN cases were collected and harmonized from the 15 Spanish population-based cancer registries (PBCRs) belonging to the Spanish Network of Cancer Registries (REDECAN) (14) during the period 1983–2018. These PBCRs cover 17 provinces (Alacant, Albacete, Araba, Asturias, Bizkaia, Castelló, Ciudad Real, Cuenca, Girona, Gipuzkoa, Granada, La Rioja, Murcia, Navarra, Salamanca, Tarragona, and València) and three islands (Mallorca, Las Palmas, and Santa Cruz de Tenerife), representing ~35% of the total Spanish child population (Table 1). All the data provided by the registries share the same methodology of data collection, obtained through an active search in different data sources. The data meet the quality controls and follow the procedures and coding rules according to the standards of the International Agency for Research on Cancer (IARC) (15).

A case was defined as any child (0–14 years) diagnosed with an HN who resided in the areas covered by the cancer registries. Standard variables available for each tumor case included basic demographic data (age, sex, and province/island of residence) and tumor data (date of diagnosis, method of diagnosis, tumor histology, and tumor topography). All tumor cases were coded according to the *International Classification of Diseases for Oncology* (ICD-O), third edition, first revision (16). All cases were assembled into four age groups: <1 year, 1–4 years, 5–9 years, and 10–14 years. Diagnoses were grouped according to the most recent child-specific ICCC-3, which divides hematological cancers into two main groups: leukemias and lymphomas. Leukemias are subdivided into lymphoid leukemias (LLs), acute myeloid leukemias (AMLs), chronic myeloproliferative diseases (CMDs), myelodysplastic syndromes and other myeloproliferative diseases (MSs), and unspecified and other specified leukemias. Similarly, lymphomas are grouped into Hodgkin's lymphoma (HL), non-Hodgkin's lymphoma (except Burkitt's lymphoma) (NHL), Burkitt's lymphoma (BL), miscellaneous lymphoreticular neoplasms (MLNs), and unspecified or other specified lymphomas (1).

2.2 Ethics statement

This study was conducted using anonymized data from the participating PBCRs that make up REDECAN. For their part, the cancer registries comply with European and Spanish legislation on the protection of personal data. No intervention was performed on human or animal subjects. Informed consent of the patients is not required for this type of study.

2.3 Statistical analysis

Absolute and relative frequencies of all hematological cancers and subgroups were analyzed by age group and sex. HN incidence rates, expressed as crude rates (CRs), age-specific rates, and ASR were estimated by age and sex, expressed per million child-years. The population at risk used was obtained from the National Statistics Institute (INE) (17). ASRs were calculated by the direct method from the summation of the age-specific rates for each age group using the weights of the European population of 2013 (ASR_E) (18), the ASR_{SEGI} (19), and the new world standard population of the World Health Organization (WHO) (2000–2025) (ASR_{WHO}) (20). Incidence sex ratios (ISRs) were calculated as the ratio of the ASR in boys to ASR in girls. In addition, to compare the ASR_{SEGI} from our study with other southern European PBCRs, the ASR_{SEGI} values were obtained from the ICCC, Volume III, published by IARC (1). The countries and time periods included in the comparison were the following: Portugal (1991–2012), Greece (1996–2009), Cyprus (1998–2012), Croatia (2001–2014), France (1993–2012), Bulgaria (1990–2013), Italy (1992–2013), and Malta (1994–2013).

TABLE 1 Spanish provinces and islands covered by the cancer registries included in the childhood (0–14 years) hematological neoplasms analysis, period of participation, person-years, number of cases contributing to the incidence analysis, and data quality indicators.

				Quality indicator		
Province/island	Period	Person-years	Number of cases	MV %	NOS %	DCO %
Albacete*	1991–2012	1,378,731	93	97.9	1.1	1.1
Asturias	1991–2013	2,832,009	185	99.5	0.5	0.0
Las Palmas	1993–2015	2,944,080	210	99.5	0.5	0.0
Santa Cruz de Tenerife	1993–2015	2,779,698	168	98.8	0.0	1.2
Ciudad Real*	2004–2012	700,649	35	100.0	0.0	0.0
Cuenca*	1993–2010	528,351	36	97.2	2.9	0.0
Araba	1987–2015	1,243,597	88	97.7	1.1	1.1
Gipuzkoa	1986–2016	3,121,467	235	97.9	1.3	0.9
Bizkaia	1986–2015	4,794,719	342	98.5	0.0	1.5
Girona	1983–2018	3,665,590	246	97.2	1.6	1.2
Granada	1985–2016	4,997,941	281	99.3	0.7	0.0
La Rioja*	1993–2014	906,089	45	91.1	4.4	4.4
Mallorca	1988–2013	2,976,006	181	98.9	1.1	0.0
Murcia	1983–2015	7,872,968	539	97.8	1.5	0.7
Navarra	1983–2015	3,002,901	192	97.9	0.5	1.6
Tarragona	1983–2015	3,567,874	230	99.1	0.0	0.9
Salamanca*	2011–2016	247,071	22	100.0	0.0	0.0
València	1983–2018	13,936,208	915	99.9	0.0	0.1
Castelló	1983–2018	3,016,416	188	100.0	0.0	0.0
Alacant	1983–2018	9,534,776	516	99.8	0.0	0.2
All		74,047,141	4,747	98.4	0.9	0.7

MV, microscopically verified cases; NOS, not otherwise specified cases; DCO, death certificate only cases.
*Provinces/islands excluded in the incidence trend analysis.

Data normality and homoscedasticity assumptions were checked in order to analyze the incidence trends; to meet these assumptions, trend analyses were restricted to the period 1985–2016. In addition, provinces/islands with shorter time periods, especially those in the first quartile in terms of number of years (<23 years), were excluded from the trend analysis to ensure that most of the provinces/islands contributed to each year of the time trend. Incidence trends were modeled using a simple logarithmic regression model with the ASR_E as the dependent variable and time (years) as the independent variable. Changes in the trend were estimated using segmented models, and the APC was calculated for each of the segment trends (21, 22). All the statistical analyses were performed using R version 4.1.3 (23).

3 Results

3.1 Description of cases

A total of 4,747 childhood HN cases were included in the study of a total population at risk of 74 million child-years. The proportion of the overall microscopically verified cases was 98.4%, with 0.7% of cases based on death certificates only (DCO) (Table 1).

Figure 1 shows the distribution of HNs. Leukemias accounted for two-thirds of HNs diagnosed, while lymphomas accounted for the remaining third. The proportion of leukemias was higher in girls and younger age groups, whereas the proportion of lymphomas was higher in boys and older age groups. A total of 59.5% of patients were boys (N = 2,823), and the age distribution of the groups [n (%)] was as follows: <1 year, 266 (5.6%); 1–4 years, 1,726 (36.4%); 5–9 years, 1,442 (30.4%); and 10–14 years, 1,313 (27.6%).

3.2 Incidence

Supplementary Table S1 shows HN incidence rates for all provinces/islands as a whole and separately for each province/island. Overall, CR and ASR_E were 64.1 (95%CI: 62.3; 65.9) and 64.2 (95%CI: 62.4; 66.0) cases of HNs per million child-years, respectively. Higher ASR_E values were found in boys than in girls

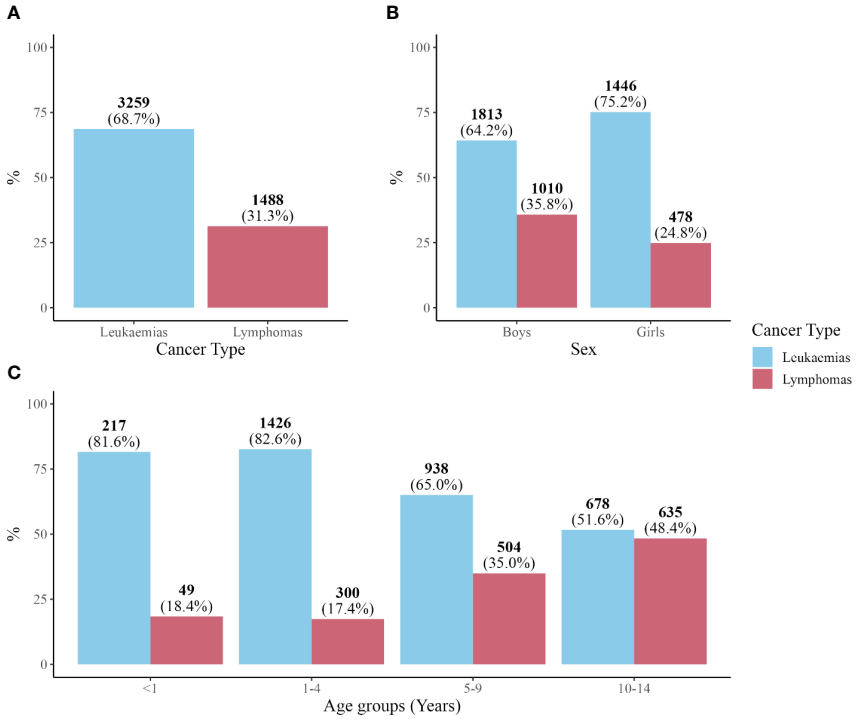


FIGURE 1
Number of cases and percentage. **(A)** Leukemias and lymphomas. **(B)** Leukemias and lymphomas by sex. **(C)** Leukemias and lymphomas by age group.

(74.2 (95%CI: 71.5; 77.0) and 53.5 (95%CI: 51.1; 55.9), respectively) with an ISR of 1.4. Differences were found by age group, with the highest age-specific rate of 92.8 cases per million child-years in the 1–4 year age group. The lowest age-specific rate of 49.8 cases per million child-years was observed in the oldest age group (10–14 years).

3.2.1 Leukemias

Table 2 shows the CR and ASR_E for leukemia with 44.0 (95%CI: 42.5; 45.5) and 44.1 (95%CI: 42.6; 45.7) cases per million child-years, respectively. Sex differences in leukemia were smaller compared to those in all HNs, with an ISR of 1.2. Boys had an ASR_E of 47.8 (95%CI: 45.6; 50.0), while girls had an ASR_E of 40.3 (95%CI: 38.2; 42.4). Age-specific rates revealed a peak incidence of 76.6 (81.9 in boys and 71.0 in girls) cases per million child-years in the age group 1–4 years, while the lowest rate of 25.7 (27.7 in boys and 23.6 in girls) was observed in the age group of 10–14 years.

Differences between leukemia subgroups were observed, with LL having the highest ASR_E of 33.1 (95%CI: 31.8; 34.5) cases per million child-years, representing 75.0% of all leukemia cases. AML was the second most common subtype with an ASR_E of 7.6 (95%CI: 7.0; 8.2) cases per million child-years, accounting for 17.2% of all leukemia cases. The remaining leukemia subtypes accounted for less than 8.0% of all leukemia cases, as shown in Supplementary Figure S1 and Table 2.

Sex differences were also observed between leukemia subtypes, with LL reporting higher differences with an ISR of 1.2, compared to AML with an ISR of 1.1. Furthermore, age group differences within the leukemia subtypes revealed a peak at the age of 1–4 years with an age-specific rate of 63.5 corresponding to LL. In contrast, the remaining subtypes showed a peak in the age group of <1 year with age-specific rates of 16.0, 2.2, 3.8, and 4.0 per million child-years, corresponding to AML, CMDs, MSs, and unspecified and other specified leukemias, respectively.

Table 3 shows the differences in leukemia incidence rates between southern European countries, using the latest ICCC-3 from 2017. The results revealed that Western countries (Portugal, Spain, and France) had an ASR_{SEGI} below 50 cases per million child-years. Conversely, Eastern countries (Greece, Malta, Italy, Cyprus, and Croatia) exhibited an ASR_{SEGI} above 50 cases, with the exception of Bulgaria, which reported 43.4 cases. For further reference, the ASR_{WHO} values for Spain are depicted in Supplementary Table S2.

3.2.2 Lymphomas

Table 4 shows the CR and ASR_E of lymphomas with 20.1 (95% CI: 19.1; 21.1) and 20.0 (95%CI: 19.0; 21.1) cases per million child-years, respectively. Higher sex differences were observed in lymphomas with an ISR of 2.0. Boys had an ASR_E of 26.5 (95% CI: 24.9; 28.1), while girls had an ASR_E of 13.2 (95%CI: 12.0; 14.4) cases per million child-years. Age-specific rates revealed an increase in lymphoma incidence with age, with the highest rate of 24.1 (30.4 in boys and 17.4 in girls) cases per million child-years in the age group 10–14 years and the lowest rate of 10.9 (11.7 in boys and 10.1 in girls) in the age group <1 year.

Similar incidence rates were estimated for HL, NHL, and BL, reporting ASR_E values of 6.0 (95%CI: 5.4; 6.5), 6.0 (95%CI: 5.4; 6.5),

TABLE 2 Age-specific rates, crude rates, European age-standardized rates, and incidence sex ratios of all the leukemias and leukemia subgroups by sex and age group.

		Age-specific rate								
Sex	N	<1	1–4	5–9	10–14	CR	(CR 95%CI)	ASR_E	(ASR_E 95%CI)	ISR
I. Leukemias, myeloproliferative diseases, and myelodysplastic diseases										
Both	3,259	48.3	76.6	38.2	25.7	44.0	(42.5; 45.5)	44.1	(42.6; 45.7)	1.2
Boys	1,813	44.0	81.9	43.6	27.7	47.6	(45.4; 49.8)	47.8	(45.6; 50.0)	
Girls	1,446	52.9	71.0	32.4	23.6	40.2	(38.1; 42.3)	40.3	(38.2;42.4)	
Ia. Lymphoid leukemias										
Both	2,445	22.3	63.5	30.2	16.0	33.0	(31.7; 34.3)	33.1	(31.8; 34.5)	1.2
Boys	1,368	19.0	68.1	34.6	17.3	35.9	(34.0; 37.8)	36.1	(34.2; 38.0)	
Girls	1,077	25.8	58.6	25.6	14.6	29.9	(28.2; 31.7)	30.1	(28.3; 31.9)	
Ib. Acute myeloid leukemias										
Both	561	16.0	9.5	5.5	6.7	7.6	(7.0; 8.2)	7.6	(7.0; 8.2)	1.1
Boys	306	14.2	10.8	6.3	6.6	8.0	(7.1; 8.9)	8.0	(7.1; 8.9)	
Girls	255	17.9	8.2	4.7	6.7	7.1	(6.2; 8.0)	7.1	(6.2; 8.0)	
Ic. Chronic myeloproliferative diseases										
Both	82	2.2	0.9	0.9	1.2	1.1	(0.9; 1.4)	1.1	(0.9; 1.4)	1.4
Boys	49	1.7	0.9	1.0	1.7	1.3	(0.9; 1.7)	1.3	(0.9; 1.6)	
Girls	33	2.8	0.9	0.8	0.7	0.9	(0.6; 1.2)	0.9	(0.6; 1.2)	
Id. Myelodysplastic syndromes										
Both	63	3.8	1.2	0.5	0.5	0.9	(0.6; 1.1)	0.9	(0.6; 1.1)	1.1
Boys	34	5.2	0.8	0.6	0.5	0.9	(0.6; 1.2)	0.9	(0.6; 1.2)	
Girls	29	2.3	1.7	0.3	0.4	0.8	(0.5; 1.1)	0.8	(0.5; 1.1)	
Ie. Unspecified and other specified leukemias										
Both	108	4.0	1.6	1.1	1.4	1.5	(1.2; 1.7)	1.5	(1.2; 1.7)	1.0
Boys	56	3.9	1.3	1.1	1.6	1.5	(1.1; 1.9)	1.5	(1.1; 1.9)	
Girls	52	4.1	1.7	1.0	1.3	1.5	(1.1; 1.8)	1.5	(1.1; 1.8)	

CR, crude rate; ASR_E, age-standardized rate using the 2013 European population; CI, confidence interval; ISR, incidence sex ratio.

and 5.3 (95%CI: 4.8; 5.9), respectively, representing 86.0% of all lymphoma cases. The remaining 14.0% of cases were predominantly MLN, as shown in Supplementary Figure S2 and Table 4.

Differences in incidence by sex and age group were observed between lymphoma subtypes. The incidence of MLN decreased with age, and girls showed similar incidence rates to boys with an ISR of 1.1. In contrast, BL had the highest sex difference of all HNs, with an ISR of 3.6. In addition, HL incidence rates increased dramatically with age, with no cases in the age group of <1 year and 11.5 cases in the age group of 10–14 years.

Table 3 shows lymphoma ASR_{SEGI} among southern European countries. The majority of these countries exhibited incidence rates of approximately 20 cases per million child-years. In particular, Italy and Cyprus demonstrated the highest rates with 25.7 and 23.8 cases, respectively. In contrast, Greece displayed the lowest at 15.9 cases per million child-years. In addition, Supplementary Table S3 provides the ASR_{WHO} for Spain for further reference.

3.3 Incidence trends

ASR_E by diagnosis period (1983–1994, 1995–2006, and 2007–2018) was calculated as shown in Supplementary Table S4 and revealed statistically significant differences between the first and last periods for all HNs and lymphomas. Figure 2 depicts the differences in trend patterns between childhood leukemias and lymphomas. Leukemias showed a statistically significant increase in the first 3 years (1985–1988) with an APC of 15.3% (95%CI: 5.9; 24.7), followed by a stable period between 1988 and 2016 (APC: 0.0% (95%CI: −0.5; 0.7). In contrast, childhood lymphomas showed no

10.3389/fonc.2023.1197850

TABLE 3 Comparison of world Segi 1960 age-standardized rates between southern European countries (1).

ASR_{SEGI}								
Spain	Portugal	Greece	Cyprus (south-west)	Croatia	France	Bulgaria	Italy	Malta
I. Leukemias, myeloproliferative diseases, and myelodysplastic diseases								
47.3	44.6	51.9	58.0	60.3	45.2	43.4	58.4	60.7
Ia. Lymphoid leukemias								
35.8	32.3	45.2	44.5	50.5	34.5	32.5	47.1	46.1
Ib. Acute myeloid leukemias								
7.9	8.2	5.5	10.1	7.7	6.9	4.8	7.7	7.5
Ic. Chronic myeloproliferative diseases								
1.1	1.3	0.5	0.4	0.8	1.0	1.0	1.6	3.8
Id. Myelodysplastic syndrome								
1.0	1.5	0.1	1.3	0.3	1.6	0.4	0.7	2.7
Ie. Unspecified and other specified leukemias								
1.5	1.2	0.5	1.6	1.0	1.2	4.7	1.2	0.7
II. Lymphomas and reticuloendothelial neoplasms								
19.4	22.8	15.9	23.8	23.1	20.0	18.1	25.7	21.8
IIa. Hodgkin lymphomas								
5.3	8.7	6.1	12.1	7.2	7.1	8.2	10.3	6.8
IIb. Non-Hodgkin lymphomas (except Burkitt's lymphoma)								
5.8	6.7	5.3	9.2	7.8	6.2	6.0	7.4	6.0
IIc. Burkitt's lymphoma								
5.4	4.6	4.5	1.8	2.9	4.9	1.7	4.6	3.9
IId. Miscellaneous lymphoreticular neoplasms								
2.3	2.1	0.0	0.0	5.1	1.5	1.4	2.6	5.1
IIe. Unspecified and other specified lymphomas								
0.4	0.7	0.0	0.8	0.2	0.2	0.8	0.8	0.0

ASR_{SEGI}, age-standardized rate using the 1960 Segi world population.

change in trend over the period, with a statistically significant steady increase in the APC of 1.0% (95%CI: 0.4; 1.6). Additional incidence trend analyses were performed by sex and age group for HNs, leukemia, and lymphoma, as shown in Supplementary Table S5.

4 Discussion

Information on the epidemiology of childhood HNs is scarce, as most studies focus on adults due to their higher incidence compared to children (6). Therefore, this study, using data from 15 PBCRs, allows us to provide a broad descriptive analysis of the incidence and incidence trends of all childhood HNs over 36 years, from 1983 to 2018 in Spain. In addition, the long time period of this study and the increased population coverage allow us to update the epidemiological results of these malignancies since previous publications in Spain covered a smaller population area and a shorter time period until 2007 (12, 13).

New histology codes created after the 2005 ICCC-3, due to multiple new diagnostic methods in progress using molecular techniques, were accurately grouped according to the updated version, the 2017 ICCC-3 (5). The latest classification is not widely used in other studies due to its recent publication. At the national level in Spain, previous publications on HNs in children have used the ICCC-3 of 2005 (12, 13, 24). Therefore, the analysis of incidence and incidence trends not only provides insight over a long period of time but also provides more recent results in terms of histological aggregation. At the international level, only one study was found that used the ICCC-3 of 2017 (25), since the most recent studies used the ICCC-3 of 2005 (26–28). Therefore, we suggest the use of this more recent classification for future studies in European PBCRs. Furthermore, ASR exhibited variability between the

10.3389/fonc.2023.1197850

TABLE 4 Age-specific rates, crude rates, European age-standardized rates, and incidence sex ratio of all lymphomas and lymphoma subgroups by sex and age group.

		Age-specific rate								
Sex	N	<1	1–4	5–9	10–14	CR	(CR 95%CI)	ASR_E	(ASR_E 95%CI)	ISR
II. Lymphomas and reticuloendothelial neoplasms										
Both	1,488	10.9	16.1	20.5	24.1	20.1	(19.1; 21.1)	20.0	(19.0; 21.1)	2.0
Boys	1,010	11.7	22.2	28.4	30.4	26.5	(24.9; 28.2)	26.5	(24.9; 28.1)	
Girls	478	10.1	9.6	12.2	17.4	13.3	(12.1; 14.5)	13.2	(12.0; 14.4)	
IIa. Hodgkin lymphomas										
Both	449	0.0	1.0	5.2	11.5	6.1	(5.5; 6.6)	6.0	(5.4; 6.5)	1.5
Boys	276	0.0	1.8	7.2	12.4	7.3	(6.4; 8.1)	7.2	(6.3; 8.0)	
Girls	173	0.0	0.2	3.1	10.4	4.8	(4.1; 5.5)	4.7	(4.0; 5.4)	
IIb. Non-Hodgkin lymphomas (except Burkitt's lymphoma)										
Both	443	1.6	4.9	6.8	6.8	6.0	(5.4; 6.5)	6.0	(5.4; 6.5)	2.3
Boys	312	1.7	6.5	9.5	9.3	8.2	(7.3; 9.1)	8.2	(7.3; 9.1)	
Girls	131	1.4	3.2	3.9	4.1	3.6	(3.0; 4.3)	3.6	(3.0; 4.3)	
IIc. Burkitt's lymphoma										
Both	393	0.2	6.6	6.3	4.4	5.3	(4.8; 5.8)	5.3	(4.8; 5.9)	3.6
Boys	312	0.4	10.8	9.3	6.7	8.2	(7.3; 9.1)	8.2	(7.3; 9.1)	
Girls	81	0.0	2.1	3.1	2.0	2.3	(1.8; 2.7)	2.3	(1.8; 2.8)	
IId. Miscellaneous lymphoreticular neoplasms										
Both	172	8.9	3.4	1.7	1.0	2.3	(2.0; 2.7)	2.3	(2.0; 2.7)	1.1
Boys	92	9.5	3.1	1.7	1.4	2.4	(1.9; 2.9)	2.4	(1.9; 2.9)	
Girls	80	8.3	3.8	1.7	0.6	2.2	(1.7; 2.7)	2.3	(1.8; 2.7)	
IIe. Unspecified and other specified lymphomas										
Both	31	0.2	0.2	0.6	0.4	0.4	(0.3; 0.6)	0.4	(0.3; 0.6)	1.3
Boys	18	0.0	0.1	0.7	0.6	0.5	(0.3; 0.7)	0.5	(0.3; 0.7)	
Girls	13	0.5	0.3	0.5	0.2	0.4	(0.2; 0.6)	0.4	(0.2; 0.6)	

CR, crude rate; ASR_E, age-standardized rate using the 2013 European population; CI, confidence interval; ISR, incidence sex ratio.

different standard populations used. In particular, the use of the Segi 1960 standard population yielded the highest incidence rates due to its emphasis on younger age groups (20). Conversely, the ASR_E and the ASR_{WHO} demonstrated comparable rates, suggesting that the new WHO standard reflected the study population structure more accurately. In light of these findings, we recommend adopting the new WHO standard to be used for international comparisons.

Overall, we reported that two-thirds of all childhood HNs were leukemias and the remaining third were lymphomas. Similar proportions have been reported in southern and eastern Europe (29). An ASR_E of 64 cases per million child-years of HNs was observed during the study period, with a statistically significant increase in incidence during the first 3 years, followed by stable rates for the remaining years. To the best of our knowledge, no publication in the literature provides an overview of the incidence and incidence trends of HNs as a whole. Most of the authors report the incidence and incidence trends of HNs divided into two large groups: leukemias and lymphomas (6, 7, 30). Differences between sexes and age groups have been observed, with boys and children aged 1–4 years tending to have higher incidence rates. Furthermore, the sex differences in HNs are greater than in other cancers, but the endogenous causes are largely unknown despite all the research that has been performed (31, 32).

Lymphoid leukemias were the main subgroup of all leukemias, followed by a smaller proportion of AML. An ASR_{SEGI} of 47.3 cases of childhood leukemia per million child-years was observed, with similar results to those previously reported in Spain and other European regions (12, 24, 30). These results contrast with the lower values reported in developing countries, such as sub-Saharan Africa and South Asia (6). The peak in leukemia incidence in the age group 1–4 years, specifically observed in LL, is a pattern that has been reported in previous studies (12, 24, 33). At present, this peak was

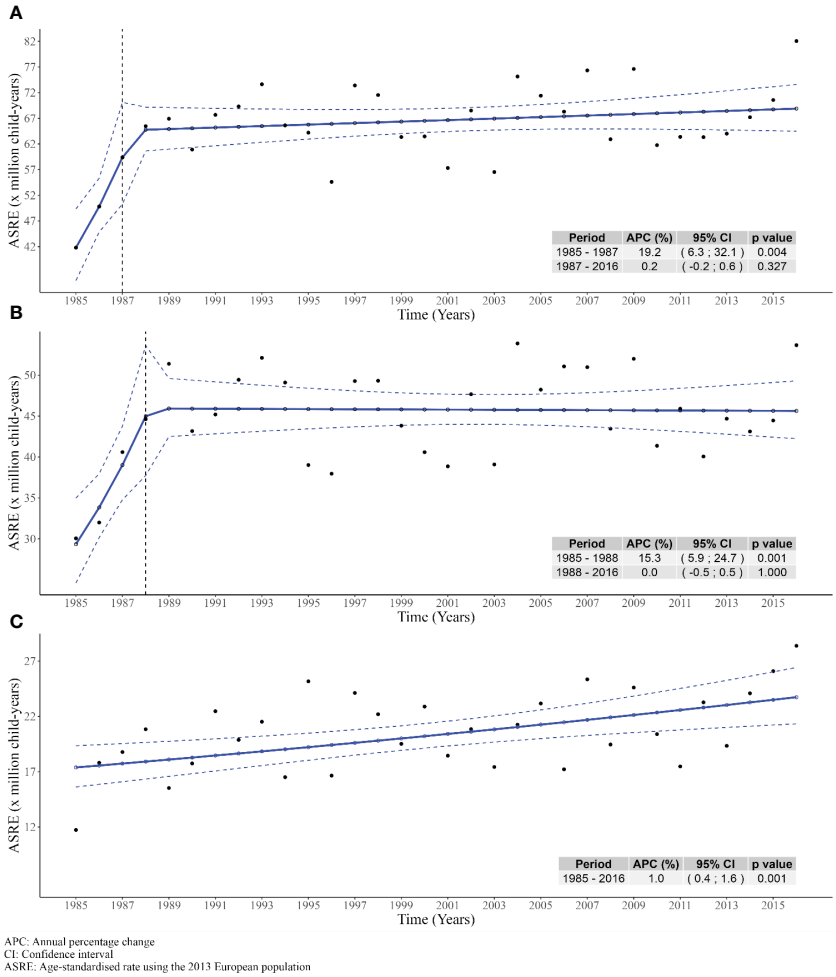


FIGURE 2
Trends of age-standardized rates using the 2013 European standard population during the period 1985-2016 in Spain. (A) All hematological neoplasms. (B) Leukemias. (C) Lymphomas.

first identified in the early 20th century, and it was suggested that the risk factors triggering this increase were unknown (34). However, subsequent studies have suggested that benzene, agricultural exposures, smoking and alcohol consumption, cigarettes, and illicit drugs during pregnancy are predisposing factors for childhood LL (35). Meanwhile, the reported incidence peak at the age of <1 year, corresponding to AML, is mainly attributed to hereditary conditions such as Down syndrome and Fanconi anemia (36). From an international perspective, the leukemia incidence patterns observed in this study by age group, sex, and subtype were similar to those in France and Portugal (1, 30, 33). In addition, Surveillance, Epidemiology, and End Results (SEER) leukemia incidence rates were similar to those observed in this study, whereas Cancer Research UK incidence rates were higher (37, 38).

Similar proportions were observed for HL, NHL, and BL, with ASR_E values of 6.0, 5.4, and 6.0 cases per million child-years, respectively, out of a total of 20 cases per million child-years for all lymphomas. The most recent study published in Spain reported lower incidence values for lymphomas as a whole, but this can be attributed to the exclusion of MLN and unspecified and other specified lymphomas subtypes from the incidence analysis (13). Internationally, southern Europe, and therefore the region included in this study, has one of the highest lymphoma incidence rates in the world, with the exception of other Mediterranean regions such as North African countries (6, 7). It has been suggested that the elevated incidence rates observed in the Mediterranean region may be due to environmental factors, as different ethnic groups living in the region, including Caucasians, Arabs, and Jews, exhibit higher incidence rates of these malignancies (39). Aside from this region, Cancer Research UK lymphoma incidence rates were similar to those observed in this study, while SEER results were higher, specifically for HL (37, 38).

Sex and age differences were also observed in our results, which have been widely reported by other authors (6, 12, 13). Although the main risk factors for these sex differences are currently unknown, some authors suggest that boys have an innate susceptibility associated with immune surveillance that makes them more vulnerable to proto-oncogenic mutations (6, 31, 40). The association between lymphoma incidence and age is due to the

fact that these malignancies are more common in adolescents, with the age group of 10–14 years having the highest age-specific incidence rate of all age groups (13, 41).

Our findings showed that the overall incidence trends of HNs were similar to those previously reported for leukemias in Spain, with a statistically significant increase in the early years of the period, followed by a stabilization since 1989 (12). We suggest that this steep increase in leukemia incidence during 1985–1988 is probably an artifact due to the lack of cases registered by the provinces/islands during the first years of the study. In our study, the incidence rates were found to be stable in most recent years added. Meanwhile, other studies have shown different patterns of leukemia incidence trends, with some reporting a steady statistically significant increase without a stabilization (8, 10, 26, 27). However, others reported a stabilization over the whole period of the study (9, 42, 43), especially for LL and AML in Germany for the most recent period available in the study (28).

This heterogeneity in incidence trends was also observed for lymphomas, as our results revealed a statistically significant steady increase of 1% per year, in contrast to previous findings in southern Europe and Spain, where no increase was detected (8, 13). However, a similar increase was observed in Denmark and Australia for the periods 1977–2014 and 1983–2015, respectively (9, 26). We hypothesized that the increase in lymphoma incidence could be due to an increase in environmental risk factors in the Mediterranean region, as these have been suggested to be the main risk factors for these malignancies in this area (39). However, other countries, such as Switzerland, Italy, Japan, and the England, reported a stabilization since the beginning of the period analyzed (10, 42, 43). Many factors may contribute to the differences in incidence trend results, despite all the standardized definitions of quality criteria by IARC. One factor that may contribute to these differences is the use of different versions of the ICCC in the histological grouping, as this may complicate the interpretation of incidence trends (44).

Despite all the studies discussed on incidence trends, the number of collaborative studies is scarce. One of the reasons contributing to this may be the implementation of the General Data Protection Regulation in Europe since 2018, which, among other things, may affect researchers and make data sharing more difficult than before (45).

5 Strengths and limitations

Our results are strengthened by the use of population-based incidence data from REDECAN, in which more than 95% of cases were verified microscopically. In addition, the large area covered in Spain and the long time period allowed us to perform the main analysis with high statistical power. Furthermore, few studies grouped the histology codes according to the latest ICCC-3 of 2017 and provided the ASR using the new WHO world standard population (2000–2025). However, the sample size was limited when performing all sub-analyses involving less common pathologies, such as calculating the incidence of specific sex, age group, and HN subgroup. Another limitation of the study was the lack of period homogeneity between cancer registries, but this was addressed by ensuring that most of the provinces/islands contributed to each year of the trend. In addition, this study does not represent the entire Spanish child population, as incidence rates may differ between the covered and non-covered areas, as a previous study has revealed differences in incidence rates between urban and rural areas in Spain (46). Finally, changes in histological grouping and diagnostic criteria throughout the period could have also affected the results obtained and the comparisons with previous studies.

6 Conclusions

This study presents an updated population-based analysis of the incidence and incidence trends of childhood HNs in Spain covering a long period from 1983 to 2018 and a large area. Our results showed that leukemias are the most common HNs in children, and their incidence has remained stable since 1988, while lymphomas are less common but their incidence is increasing every year. The observed incidence rates of lymphomas are similar to those reported in other southern European countries, while the similarities in leukemia incidence rates were limited to the southwestern European countries. Furthermore, we recommend the use of the new ICCC-3 from 2017 and the use of the new WHO world standard population (2000–2025) in future studies conducted by European PBCRs. Collaborative cancer registry projects, such as the REDECAN, provide the opportunity to assess epidemiological indicators of less common cancers, such as pediatric HNs. Consequently, all these reported epidemiological findings could help health authorities and clinicians to have updated results of these malignancies for the more recent years in Spain. Furthermore, the results suggest the need for more studies focusing on the risk factors of childhood lymphomas, as their incidence is increasing every year in Spain.

REDECAN working group

REDECAN Working Group: **Girona Cancer Registry** – Coordination of hematology neoplasms in REDECAN: Rafael Marcos Gragera, Montse Puigdemont, Anna Vidal Vila, Arantza Sanvisens, and Jan Trallero; **Tarragona Cancer Registry** – Data coordination in REDECAN: Marià Carulla, Alberto Ameijide, Clàudia Pla, and Jaume Galceran; **Álava Cancer Registry**: Arantza López de Munain, Patricia Sancho, and María Luisa Iruretagoyena; **Albacete Cancer Registry**: Cristina Ramírez; **Asturias Cancer Registry**: Susana Merino Perera, Virginia Menéndez García, and Marta Rodríguez Camblor; **Bizkaia Cancer Registry**: Visitación de Castro, Marta De La Cruz, and Joseba Bidaurrazaga; **Castellón Cancer Registry**: Consol Sabater Gregori, Isabel Sáez Lloret, Ana Vizcaíno Batllés, and Xavier Peñalver Herrero; **Ciudad Real Cancer Registry**: Cristina Díaz del Campo; **Cuenca Cancer Registry**: Ana Isabel Marcos and Rosario Jimenez Chillarón; **Gipuzkoa Cancer Registry**: Amaia Aizpurua; **Canary Islands Cancer Registry**: María Dolores Rojas Martín, María Araceli Alemán Herrera, Emilio De Miguel, María Magdalena Ramos Marrero, María Carmen Gábas, Beatriz Hidalgo

Martín, and Olga Leonor Velázquez Castilla; **Granada Cancer Registry**: María José Sánchez, Daysi Yoe-Ling Chang-Chan, and Miguel Rodríguez Barranco; **La Rioja Cancer Registry**: María Isabel Palacios and Enrique Ramalle; **Mallorca Cancer Registry**: Paula Franch, Patricia Ruiz Armengol, and Carmen Sánchez Contador; **Murcia Cancer Registry**: María Dolores Chirlaque, Antonia Sánchez Gil, and Ricardo José Vaamonde; **Navarra Cancer Registry**: Marcela Guevara and Eva Ardanaz; **Castilla y León Cancer Registry**: Pilar Gutiérrez, Rufino Álamo, and Enrique Cabrera; **Registro Español de Tumores Infantiles**: Adela Cañete, Elena Pardo, and Rafael Peris Bonet; **Registro de Tumores Infantiles y Adolescentes de la Comunitat Valenciana**: Ana Vizcaino, Fernando Almela Vich, and Noura Jeghalef El Karoni.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by CEIM Girona Institut d'Investigació Biomèdica de Girona Dr. Josep Trueta (IDIBGI). Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and institutional requirements.

Author contributions

JT, AS, and RM-G contributed to the study conception and design. Data collection was performed by FA, NJ, IS, CD-D-C, AM-N, AAA, PS, MO, MS, JP, PF, MC, MG, CR, MC, MA, PG, and RM-G. Data analysis was performed by JT, AS, and AA. The first draft of the manuscript was written by JT. All authors contributed to the article and approved the submitted version.

Funding

This work was partially funded by the Josep Carreras Leukaemia Research Institute (Grant No. FIJC1100).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1197850/full#supplementary-material

References

1. Steliarova-Foucher E, Colombet M, Ries LAG, Hesseling P, Moreno F, Shin HY, et al. *International Incidence of Childhood Cancer* Vol. III. . Lyon, France: International Agency for Research on Cancer (2017).

2. Davis AS, Viera AJ, Mead MD. Leukemia: an overview for primary care. *Am Fam Physician* (2014) 89:731–8.

3. Schmidt JA, Hornhardt S, Erdmann F, Sánchez-García I, Fischer U, Schüz J, et al. Risk factors for childhood leukemia: radiation and beyond. *Front Public Heal* (2021) 9:805757. doi: 10.3389/FPUBH.2021.805757

4. Matasar MJ, Zelenetz AD. Overview of lymphoma diagnosis and management. *Radiol Clin North Am* (2008) 46:175–98. doi: 10.1016/J.RCL.2008.03.005

5. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. *Cancer* (2005) 103:1457–67. doi: 10.1002/cncr.20910

6. Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Bray F, et al. International incidence of childhood cancer, 2001–10: a population-based registry study. *Lancet Oncol* (2017) 18:719–31. doi: 10.1016/S1470-2045(17)30186-9

7. Stiller CA, Marcos-Gragera R, Ardanaz E, Pannelli F, Almar Marqués E, Cañada Martinez A, et al. Geographical patterns of childhood cancer incidence in Europe, 1988-1997. Report from the Automated Childhood Cancer Information System project. *Eur J Cancer* (2006) 42:1952–60. doi: 10.1016/j.ejca.2006.05.017

8. Steliarova-Foucher E, Fidler MM, Colombet M, Lacour B, Kaatsch P, Piñeros M, et al. Changing geographical patterns and trends in cancer incidence in children and adolescents in Europe, 1991–2010 (Automated Childhood Cancer Information System): a population-based study. *Lancet Oncol* (2018) 19:1159–69. doi: 10.1016/S1470-2045(18)30423-6

9. Grabas MR, Kjaer SK, Frederiksen MH, Winther JF, Erdmann F, Dehlendorff C, et al. Incidence and time trends of childhood cancer in Denmark, 1943-2014. *Acta Oncol* (2020) 59:588–95. doi: 10.1080/0284186X.2020.1725239

10. Sommer G, Schindler M, Redmond S, Pfeiffer V, Konstantinoudis G, Ammann RA, et al. Temporal trends in incidence of childhood cancer in Switzerland, 1985-2014. *Cancer Epidemiol* (2019) 61:157–64. doi: 10.1016/J.CANEP.2019.06.002

11. Peris-Bonet R, Salmerón D, Martínez-Beneito MA, Galceran J, Marcos-Gragera R, Felipe S, et al. Childhood cancer incidence and survival in Spain. *Ann Oncol Off J Eur Soc Med Oncol* (2010) 21:103–10. doi: 10.1093/ANNONC/MDQ092

12. Marcos-Gragera R, Galceran J, Martos C, de Munain AL, Vicente-Raneda M, Navarro C, et al. Incidence and survival time trends for Spanish children and adolescents with leukaemia from 1983 to 2007. *Clin Transl Oncol* (2017) 19:301–16. doi: 10.1007/s12094-016-1531-z

13. Marcos-Gragera R, Solans M, Galceran J, Fernández-Delgado R, Fernández-Teijeiro A, Mateos A, et al. Childhood and adolescent lymphoma in Spain: incidence and survival trends over 20 years. *Clin Transl Oncol* (2018) 20:1289–301. doi: 10.1007/s12094-018-1860-1

14. *Home | RED ESPAÑOLA DE REGISTROS DE CÁNCER (REDECAN)* (2010). Available at: https://redecan.org/es/redecan (Accessed October 13, 2022).

15. Navarro C, Martos C, Ardanaz E, Galceran J, Izarzugaza I, Peris-Bonet R, et al. Spanish Cancer Registries Working Group. Population-based cancer registries in Spain and their role in cancer control. *Ann Oncol* (2010) 21 Suppl 3:iii3–13. doi: 10.1093/annonc/mdq094

16. World Health Organization. *International classification of diseases for oncology (ICD-O), 3rd ed., 1st revision. 3rd ed.* Geneva: World Health Organization ([amp]] lrm;2013). Available at: https://apps.who.int/iris/handle/10665/96612.

17. Instituto Nacional Estadística (INE). *Población residente por fecha, sexo y edad (31304)* . Available at: https://www.ine.es/jaxiT3/Tabla.htm?t=31304.

18. Commission E, Eurostat, Pace M, Gissler M, Lanzieri G, Grande E, Wojtyniak B, et al. *Revision of the European Standard Population : report of Eurostat's task force : 2013 edition*. Luxembourg: Publications Office (2013). doi: 10.2785/11470

19. Segi M. "Cancer mortality for selected sites in 24 countries (1950-1957)". In: *Department of Public Health, Tohoku University School of Medicine, editor. Cancer mortality for selected sites in 24 countries*. Nagoya, Japan (1960).

20. Ben AO, Boschi Pinto C, Lopez AD. Age standardization of rates: A new WHO standard. *GPE Discuss Pap Ser* (2001) 31:10–2.

21. Muggeo VMR. Estimating regression models with unknown break-points. *Stat Med* (2003) 22:3055–71. doi: 10.1002/SIM.1545

22. Muggeo V. Segmented: an R package to fit regression models with broken-line relationships. *R News* (2008) 8:20–5.

23. RStudio Team. *RStudio: Integrated Development Environment for R* (2022). Available at: http://www.rstudio.com/.

24. Marcos-Gragera R, Cervantes-Amat M, Vicente ML, De Sanjosé S, Guallar E, Godoy C, et al. Population-based incidence of childhood leukaemias and lymphomas in Spain (1993-2002). *Eur J Cancer Prev* (2010) 19:247–55. doi: 10.1097/CEJ.0B013E328339E2F3

25. Sun K, Zheng R, Zhang S, Zeng H, Wang S, Chen R, et al. Patterns and trends of cancer incidence in children and adolescents in China, 2011-2015: A population-based cancer registry study. *Cancer Med* (2021) 10:4575–86. doi: 10.1002/CAM4.4014

26. Youlden DR, Baade PD, Green AC, Valery PC, Moore AS, Aitken JF. The incidence of childhood cancer in Australia, 1983-2015, and projections to 2035. *Med J Aust* (2020) 212:113–20. doi: 10.5694/MJA2.50456

27. Erdmann F, Li T, Luta G, Giddings BM, Torres Alvarado G, Steliarova-Foucher E, et al. Incidence of childhood cancer in Costa Rica, 2000-2014: An international perspective. *Cancer Epidemiol* (2018) 56:21–30. doi: 10.1016/J.CANEP.2018.07.004

28. Wellbrock M, Spix C, Grabow D, Borkhardt A, Zeeb H, Erdmann F. 28-year incidence and time trends of childhood leukaemia in former East Germany compared to West Germany after German reunification: A study from the German Childhood Cancer Registry. *Cancer Epidemiol* (2021) 73:101968. doi: 10.1016/J.CANEP.2021.101968

29. Petridou ET, Dimitrova N, Eser S, Kachanov D, Karakilinc H, Varfolomeeva S, et al. Childhood leukemia and lymphoma: time trends and factors affecting survival in five Southern and Eastern European Cancer Registries. *Cancer Causes Control* (2013) 24:1111–8. doi: 10.1007/S10552-013-0188-Y

30. Lacour B, Guyot-Goubin A, Guissou S, Bellec S, Désandes E, Clavel J. Incidence of childhood cancer in France: National Children Cancer Registries, 2000-2004. *Eur J Cancer Prev* (2010) 19:173–81. doi: 10.1097/CEJ.0B013E32833876C0

31. Tevfik Dorak M, Karpuzoglu E. Gender differences in cancer susceptibility: an inadequately addressed issue. *Front Genet* (2012) 3:268. doi: 10.3389/FGENE.2012.00268

32. Liu Z, Yang Q, Cai N, Jin L, Zhang T, Chen X. Enigmatic differences by sex in cancer incidence: evidence from childhood cancers. *Am J Epidemiol* (2019) 188:1130–5. doi: 10.1093/AJE/KWZ058

33. Desandes E, Clavel J, Berger C, Bernard JL, Blouin P, De Lumley L, et al. Cancer incidence among children in France, 1990-1999. *Pediatr Blood Cancer* (2004) 43:749–57. doi: 10.1002/PBC.20148

34. Hrusák O, Trka J, Zuna J, Poloucková A, Kalina T, Stáry J. Acute lymphoblastic leukemia incidence during socioeconomic transition: selective increase in children from 1 to 4 years. *Leukemia* (2002) 16:720–5. doi: 10.1038/SJ.LEU.2402422

35. Fujita TC, Sousa-Pereira N, Amarante MK, Watanabe MAE. Acute lymphoid leukemia etiopathogenesis. *Mol Biol Rep* (2021) 48:817–22. doi: 10.1007/S11033-020-06073-3

36. Rubnitz JE, Gibson B, Smith FO. Acute myeloid leukemia. *Hematol Oncol Clin North Am* (2010) 24:35–63. doi: 10.1016/J.HOC.2009.11.008

37. Howlander N, Noone A, Krapcho M, Miller D, Brest A, Yu M, et al. *SEER Cancer Statistics Review (CSR) 1975-2018*. Bethesda, MD: National Cancer Institute (2021). Available at: https://seer.cancer.gov/csr/1975_2018/.

38. Welham C, Charles S, Irvine L, Luke H, Martin M, Gavin A, et al. *Children, teenagers and young adults UK cancer statistics report 2021*. London: Public Health England (2021). Available at: http://www.ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/cancer_in_children_teenagers_and_young_adults/.

39. Stiller CA, Parkin DM. International variations in the incidence of childhood lymphomas. *Paediatr Perinat Epidemiol* (1990) 4:303–24. doi: 10.1111/J.1365-3016.1990.TB00654.X

40. Radkiewicz C, Bruchfeld JB, Weibull CE, Jeppesen ML, Frederiksen H, Lambe M, et al. Sex differences in lymphoma incidence and mortality by subtype: A population-based study. *Am J Hematol* (2023) 98:23–30. doi: 10.1002/AJH.26744

41. Loizou L, Demetriou A, Erdmann F, Borkhardt A, Brozou T, Sharp L, et al. Patterns and temporal trends in the incidence of childhood and adolescence cancer in Cyprus 1998-2017: A population-based study from the Cyprus Paediatric Oncology Registry. *Cancer Epidemiol* (2022) 80:102239. doi: 10.1016/J.CANEP.2022.102239

42. Nakata K, Ito Y, Magadi W, Bonaventure A, Stiller CA, Katanoda K, et al. Childhood cancer incidence and survival in Japan and England: A population-based study (1993-2010). *Cancer Sci* (2018) 109:422–34. doi: 10.1111/CAS.13457

43. Isaevska E, Manasievska M, Alessi D, Mosso ML, Magnani C, Sacerdote C, et al. Cancer incidence rates and trends among children and adolescents in Piedmont, 1967–2011. *PloS One* (2017) 2(7):e0181805. doi: 10.1371/journal.pone.0181805

44. Kaatsch P. Epidemiology of childhood cancer. *Cancer Treat Rev* (2010) 36:277–85. doi: 10.1016/J.CTRV.2010.02.003

45. Cornock M. General Data Protection Regulation (GDPR) and implications for research. *Maturitas* (2018) 111:A1–2. doi: 10.1016/J.MATURITAS.2018.01.017

46. González García H, Garrote Molpeceres R, Urbaneja Rodríguez E, Gutiérrez Meléndez P, Herráiz Cristóbal R, Pino Vázquez MA. Differences in incidence and survival to childhood cancer between rural and urban areas in Castilla y León, Spain (2003-2014): A Strobe-compliant study. *Med (Baltimore)* (2018) 97:e12797. doi: 10.1097/MD.0000000000012797

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 28 July 2023
DOI 10.3389/fonc.2023.1219128

OPEN ACCESS

EDITED BY
Paulo S. Pinheiro,
University of Miami, United States

REVIEWED BY
Gonçalo Forjaz,
Westat, United States
Emanuele Crocetti,
Independent Researcher, Florence, Italy

*CORRESPONDENCE
Francesco Giusti
Francesco.Giusti@registreducancer.org;
francescogiusti@hotmail.com
Manola Bettio
JRC-ENCR@ec.europa.eu

RECEIVED 08 May 2023
ACCEPTED 06 July 2023
PUBLISHED 28 July 2023

CITATION
Giusti F, Martos C, Negrão Carvalho R, Van Eycken L, Visser O and Bettio M (2023) Quality indicators: completeness, validity and timeliness of cancer registry data contributing to the European Cancer Information System.
Front. Oncol. 13:1219128.
doi: 10.3389/fonc.2023.1219128

Quality indicators: completeness, validity and timeliness of cancer registry data contributing to the European Cancer Information System

Francesco Giusti[1,2]*, Carmen Martos[1,3], Raquel Negrão Carvalho[1], Liesbet Van Eycken[2], Otto Visser[4] and Manola Bettio[1]*

[1]European Commission, Joint Research Centre (JRC), Ispra, Italy, [2]Belgian Cancer Registry, Brussels, Belgium, [3]Rare Diseases Research Unit, Foundation for the Promotion of Health and Biomedical Research in the Valencian Region (FISABIO), Valencia, Spain, [4]Department of Registration, Netherlands Comprehensive Cancer Organisation (IKNL), Utrecht, Netherlands

Population-based Cancer Registries (PBCRs) are tasked with collecting high-quality data, important for monitoring cancer burden and its trends, planning and evaluating cancer control activities, clinical and epidemiological research and development of health policies. The main indicators to measure data quality are validity, completeness, comparability and timeliness. The aim of this article is to evaluate the quality of PBCRs data collected in the first ENCR-JRC data call, dated 2015.

Methods: All malignant tumours, except skin non-melanoma, and *in situ* and uncertain behaviour of bladder were obtained from 130 European general PBCRs for patients older than 19 years. Proportion of cases with death certificate only (DCO%), proportion of cases with unknown primary site (PSU%), proportion of microscopically verified cases (MV%), mortality to incidence (M:I) ratio, proportion of cases with unspecified morphology (UM%) and the median of the difference between the registration date and the incidence date were computed by sex, age group, cancer site, period and PBCR.

Results: A total of 28,776,562 cases from 130 PBCRs, operating in 30 European countries were included in the analysis. The quality of incidence data reported by PBCRs has been improving across the study period. Data quality is worse for the oldest age groups and for cancer sites with poor survival. No differences were found between males and females. High variability in data quality was detected across European PBCRs.

Conclusion: the results reported in this paper are to be interpreted as the baseline for monitoring PBCRs data quality indicators in Europe along time.

KEYWORDS

cancer registry, data quality, completeness, validity, timeliness, Europe

1 Introduction

Population-based Cancer Registries (PBCRs) are tasked with collecting high-quality data, important for monitoring cancer burden and its trends, planning and evaluating cancer control activities, clinical and epidemiological research and developing of health policies (1). Therefore, the value of a PBCR is inherent in the quality of its data and the related quality control measures. The main indicators to measure data quality are validity, completeness, comparability and timeliness (2, 3).

Validity or accuracy refers to the proportion of cases with specific characteristics that actually have such attribute. Completeness indicates the extent of which all incident cancer cases in the area covered by the PBCR are indeed recorded by the PBCR. Comparability is the adherence to common international guidelines. Timeliness refers to how quickly cancer incidence data is collected, processed and reported. There is usually a trade-off between timeliness and both completeness and validity. Cancer data quality indicators include proportion of cases with death certificate only (DCO%), the proportion of microscopically verified cases (MV%) and the mortality to incidence (M:I) ratio (2–4).

The European Network of Cancer Registries (ENCR) has been operating since 1990 to support the collaboration among European PBCRs. One of the ENCR main aims is the improvement of the quality and comparability of cancer incidence data. The ENCR Secretariat has been hosted in Ispra, Italy, since 2012 by the Directorate-General Joint Research Centre (JRC), the science and knowledge centre of the European Commission. The JRC supports the ENCR in the harmonisation of PBCR data, with the goal of accurately comparing data between European areas (5).

In 2015 a first ENCR-JRC data call was launched by the ENCR Steering Committee and the JRC to the European PBCRs (6). After harmonisation, EU-wide statistics on incidence and mortality by cancer site, sex, age group and PBCR have been computed, feeding the European Cancer Information System (ECIS) as the web tool developed and maintained by the JRC to report on the burden of cancer in EU and Europe (7).

The goal of this study is to evaluate the quality of PBCRs data collected in the first ENCR-JRC data call, dated 2015, and is based on indicators evaluating completeness, validity and timeliness as data quality dimensions.

2 Methodology

2.1 Data sources

Incidence and mortality data from 130 European general PBCRs (collecting data for all ages and all tumours), contributing to the ECIS through the 2015 ENCR-JRC data call (Figure 1; Supplementary Table 1) were selected for patients older than 19 years. Data quality in children and adolescents will be analysed in a separate publication, since for this age group tumours are grouped taking into account morphology and topography combinations according the International Classification of Childhood Cancer

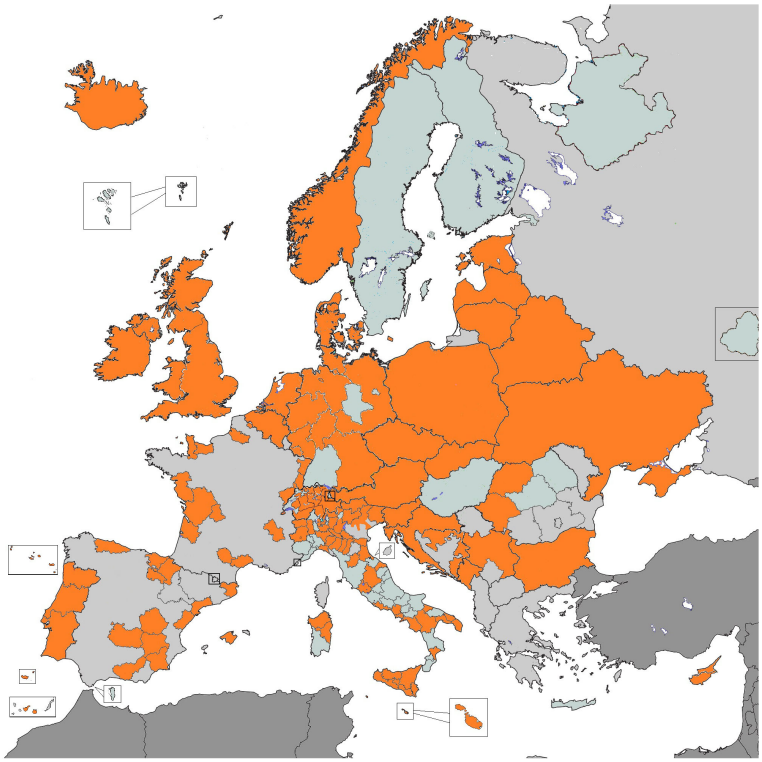


FIGURE 1
Population-based cancer registries contributing data for the analysis (in orange).

and also have different definitions of unspecified morphology compared to adults (8).

All malignant tumours (ICD-O-3.1 behaviour = 3), except skin non-melanoma, and *in situ* and uncertain behaviour (ICD-O-3.1 behaviour 2 and 1 respectively) of bladder were included in the analysis.

Among others, the 2015 data call protocol (9) included the following variables: topography, morphology and behaviour, coded according to the International Classification of Diseases for Oncology, Third Edition, (ICD-O-3) (10), as well as basis of diagnosis.

Patients with the same patient identification code and tumour identification code were checked, and excluded from the analysis if other variables such as topography, morphology and behaviour were also duplicated.

Cancer sites were defined with ICD-O-3 topography and morphology combinations reported in Supplementary Table 2.

2.2 Quality indicators

Validity, completeness, and timeliness of the PBCRs datasets were evaluated. The following indicators were calculated, with type of indicator specified in italics between parentheses (2, 3):

- DCO% (*validity*).
- Proportion of cases with unknown primary site (PSU%, *validity*) (ICD-O-3 topography = C80.9).
- MV% (*validity* and *completeness*). Tumours with basis of diagnosis as cytology, histology of a primary tumour or histology of a metastasis were considered as MV cases.
- M:I ratio (*completeness*), computed dividing the number of deaths by the number of incident cases.
- Proportion of cases with unspecified morphology (UM%, *validity*). The ICD-O-3.1 morphology codes considered as unspecified morphologies were 8000-8005 for solid tumours and 9590-9591, 9596, 9727, 9760, 9800-9801, 9805-9809, 9820, 9832, 9835, 9860, 9960, 9970-9971, 9975, 9989 for haematological malignancies.
- Median of the difference between the registration date and the incidence date (*timeliness*). Date of registration was defined in the 2015 data call protocol as the date in which a cancer case was first recorded in the registry database (9).

When applicable, all indicators were disaggregated by sex, age group (20-59, 60-69, 70-79, 80+ years), cancer site, period (1995-1999, 2000-2004, 2005-2009, 2010-2014) and PBCR.

Benchmarks for the latest available period (2010-2014) were computed for the first tertile (30%) of PBCRs with the higher performance for each indicator. Two-sided 95% confidence intervals were calculated using the Clopper–Pearson method for DCO%, MV%, UM%, PSU%, with a ratio paired t-test for M:I ratio and with the normal approximation method for the timeliness indicator.

3 Results

A total of 28,776,562 cases from 130 PBCRs, 21 National and 109 regional PBCRs, operating in 30 European countries were included in the analysis (Figure 1).

MV%, PSU% and UM% were computed for all 130 PBCRs, DCO % for 102 PBCRs which had access to death certificate information, M:I ratio for 92 PBCRs with available mortality data, and timeliness for the 49 PBCRs which provided date of registration.

Table 1 includes DCO%, MV%, UM%, M:I ratio by age at diagnosis and cancer site for the period 1995-2014 and timeliness for the period 2000-2014.

Table 2 includes reference values based for the best performing tertile of PBCRs for DCO%, MV%, UM%, PSU%, M:I ratio and timeliness, by age at diagnosis and cancer site for the period 2010-2014.

Results by period (Figures 2–6; Table 1; Supplementary Figures 1, 3, 5, 8) included PBCRs with available incidence data at least in period 1998-2011.

Results by period for timeliness (Figure 7; Table 1; Supplementary Figure 10) included PBCRs with available incidence data at least in period 2003-2011, with at least 2 incidence years in each considered period: 2000-2006 and 2007-2014.

3.1 Proportion of cases with death certificate only (DCO%)

The highest DCO% was recorded for liver, pancreas cancer and unknown primary site cases, followed by other haematological malignancies, stomach cancer, brain and central nervous system tumours and lung cancer. The lowest DCO% occurred for testicular cancer, skin melanoma and cervical cancer (Figure 2).

When comparing different time periods, a decrease in DCO% was observed over time for all cancer sites, except PSU cases, changing on average from 4.9% in the period 1995-1999 to 3.0% in the period 2010-2014 (Figure 2). In particular, between 1995-1999 and 2010-2014 DCO cases decreased on average from 15.1% to 8.7% for liver, from 10.9% to 7.8% for pancreatic cancer and from 7.9% to 4.5% for stomach respectively (Figure 2).

The DCO% for all PBCRs and all cancer sites combined did not show any difference between males (3.8%) and females (4.0% - data not shown).

Considering the whole analysed period, an increase in DCO% was observed with increasing age, from 1.4% in patients aged 20-59 years at diagnosis, up to 9.4% for those aged 80 and more. Differences by age group were found for most cancer sites. In particular, age group 20-59 and 80+ had a respective DCO% of 8.1% vs 17.2% for liver, 5.3% vs 15.1% for pancreas, 3.3% vs 14.9% for central nervous system and 1.3% vs 12.1% for ovary (Table 1; Supplementary Figure 1). There was a high variability among PBCRs for this indicator. Whereas the majority of PBCRs had less than 5% DCO cases between 1995 and 2014, 25 out of 102 PBCRs had more than 5% DCOs in at least one of the considered 5-year periods. However, the latter group of PBCRs showed a general improvement for this indicator between 1995 and 2014 (Supplementary Figure 2).

TABLE 1 Proportion of cases: with death certificate only (DCO%), microscopically verified (MV%), with unspecified morphology (UM%), by age at diagnosis and cancer site, 1995-2014.

Cancer site/age	DCO%					MV%					UM%					M:I ratio					Timeliness				
	20-59	60-69	70-79	80+	Total	20-59	60-69	70-79	80+	Total	20-59	60-69	70-79	80+	Total	20-59	60-69	70-79	80+	Total	20-59	60-69	70-79	80+	Total
Lip, Oral cavity and Pharynx	1.3	1.7	2.4	5.2	2.0	96.6	95.7	94.1	88.1	95.0	2.7	3.4	4.5	8.9	3.8	0.32	0.38	0.42	0.55	0.38	610	623	723	730	650
Oesophagus	2.6	2.7	3.1	5.1	3.3	92.8	91.8	89.0	80.5	88.9	5.2	5.5	6.5	10.5	6.7	0.78	0.84	0.93	1.07	0.90	422	392	371	396	394
Stomach	3.2	4.3	5.8	10.7	6.3	92.6	89.9	86.7	76.3	86.0	7.3	9.2	11.1	17.6	11.5	0.60	0.65	0.72	0.89	0.73	671	681	686	709	690
Colon and Rectum	1.2	1.7	2.9	7.4	3.4	95.9	94.4	91.1	78.6	89.9	3.0	4.0	6.1	13.7	6.8	0.33	0.35	0.43	0.65	0.46	628	634	665	668	650
Liver	8.1	9.3	11.3	17.2	11.4	62.1	58.1	51.0	31.3	51.2	19.4	21.0	23.7	31.6	23.7	0.81	0.90	1.01	1.19	0.98	966	1053	1032	1006	1021
Pancreas	5.3	6.7	9.2	15.1	9.4	73.8	65.8	51.8	26.6	52.9	21.2	25.7	34.7	51.0	34.1	0.84	0.92	0.98	1.07	0.97	720	712	761	786	750
Larynx	1.8	2.1	3.2	7.1	2.7	96.0	94.9	93.0	84.4	94.0	3.4	4.0	5.1	10.6	4.6	0.34	0.37	0.46	0.72	0.42	760	790	836	791	790
Lung	3.3	3.9	5.1	9.3	5.2	86.2	81.4	72.9	49.2	74.2	10.3	12.8	17.0	29.7	16.5	0.75	0.81	0.88	1.01	0.86	713	656	690	724	693
Melanoma of the Skin	0.3	0.5	0.8	2.2	0.7	98.2	98.0	97.1	95.1	97.6	–	–	–	–	–	0.13	0.17	0.22	0.32	0.20	418	377	373	347	392
Breast (Female)	0.4	0.9	2.0	7.5	1.8	98.0	97.0	93.9	82.0	94.8	1.6	2.4	4.7	12.2	3.8	0.17	0.23	0.34	0.57	0.28	525	523	578	617	547
Cervix Uteri	0.5	1.7	3.5	8.6	1.6	98.1	95.6	92.4	82.5	95.9	1.7	3.6	5.9	12.1	3.2	0.22	0.39	0.56	0.80	0.34	384	506	534	582	425
Uterus: Corpus and Unspecified	0.5	0.7	1.9	8.2	2.0	98.3	97.9	95.5	82.9	95.4	1.3	1.5	3.1	11.8	3.2	0.13	0.19	0.31	0.60	0.26	670	633	671	722	668
Ovary	1.3	2.5	4.7	12.1	4.1	94.7	91.0	83.5	59.7	85.6	4.4	7.0	12.2	27.7	10.6	0.44	0.64	0.80	0.98	0.70	590	549	579	653	590
Prostate	0.3	0.6	1.9	9.2	2.6	97.7	96.6	91.9	67.8	89.9	1.6	2.2	5.1	19.6	6.3	0.08	0.11	0.24	0.70	0.26	448	560	714	792	633
Testis	0.2	2.0	7.6	17.3	0.5	97.8	94.0	84.9	59.6	97.2	1.4	5.2	13.1	31.0	1.9	0.05	0.16	0.33	0.55	0.06	500	564	707	941	508
Bladder	0.6	0.9	1.8	5.4	2.3	96.7	95.9	94.0	84.9	92.7	2.7	3.2	4.4	10.5	5.3	0.15	0.21	0.31	0.57	0.32	890	877	866	828	866
Kidney, Renal Pelvis, Ureter	1.2	2.0	3.6	9.7	3.4	91.6	86.9	78.1	48.3	79.9	6.1	8.9	14.6	33.1	13.4	0.26	0.34	0.43	0.67	0.41	673	654	721	799	707
Central Nervous System	3.3	5.1	8.5	14.9	6.1	86.0	77.9	56.3	20.5	70.8	9.8	15.1	27.8	52.5	19.4	0.70	0.71	0.72	0.67	0.82	659	669	755	834	712
Thyroid	0.2	1.1	3.1	9.2	1.3	98.1	96.2	92.0	77.6	95.9	1.6	3.0	6.1	16.7	3.2	0.03	0.16	0.36	0.79	0.14	866	837	803	671	851
Hodgkin Lymphoma	0.6	2.4	3.9	8.3	1.6	97.3	94.2	92.1	85.8	95.8	–	–	–	–	–	0.13	0.38	0.55	0.88	0.25	694	618	670	582	672
Non-Hodgkin Lymphoma	1.0	1.6	2.9	5.8	2.5	95.7	94.4	91.6	85.3	92.4	22.3	24.0	28.3	35.3	26.7	0.24	0.36	0.49	0.71	0.42	648	635	669	626	646
Other haematological malignancies (HM)	2.3	3.7	6.4	13.2	6.4	91.5	88.8	84.1	72.9	84.6	12.4	8.7	10.2	14.5	11.3	0.41	0.51	0.67	0.89	0.62	784	827	867	860	834
Mesothelioma	1.5	2.0	2.9	5.7	2.9	93.7	92.6	88.5	78.1	88.6	–	–	–	–	–	0.67	0.79	0.85	0.96	0.84	701	737	716	641	701
Primary site unknown (C80)	4.7	6.6	9.5	16.7	10.5	75.1	64.1	50.9	31.5	51.7	21.9	28.4	35.3	47.4	35.4	0.70	0.83	0.91	1.03	0.90	565	580	650	706	649
Other	1.8	3.4	5.9	12.8	6.0	91.1	87.6	81.7	66.3	81.9	7.5	10.3	15.2	26.7	14.8	0.39	0.56	0.68	0.92	0.64	709	754	812	800	773
Total	1.4	2.3	4.0	9.4	3.9	93.9	90.3	84.2	67.6	85.4	5.8	7.8	11.7	22.3	11.1	0.32	0.44	0.55	0.79	0.51	625	647	715	730	678

Mortality to incidence (M:I) ratio by age at diagnosis and cancer site, 1995-2014. Timeliness by age at diagnosis and cancer site, 2000-2014.

10.3389/fonc.2023.1219128

10.3389/fonc.2023.1219128

TABLE 2 Quality indicators benchmarks, 2010-2014.

	DCO%	MV%	UM%	PSU%	M:I ratio	Timeliness
Total	0.3 (0.3-0.3)	95.6 (95.5-95.6)	5.5 (5.5-5.5)	1.3 (1.3-1.3)	0.41 (0.40-0.41)	336 (237-435)
Sex						
Males	0.3 (0.2-0.3)	95.4 (95.3-95.4)	5.6 (5.6-5.7)	1.3 (1.2-1.3)	0.42 (0.41-0.43)	354 (254-454)
Females	0.3 (0.3-0.3)	95.8 (95.8-95.8)	5.4 (5.4-5.4)	1.4 (1.4-1.4)	0.39 (0.38-0.40)	324 (224-424)
Age group						
20-59	0.1 (0.1-0.1)	98.6 (98.6-98.6)	2.2 (2.1-2.2)	0.8 (0.8-0.9)	0.22 (0.21-0.23)	321 (219-422)
60-69	0.1 (0.1-0.1)	97.8 (97.7-97.8)	2.8 (2.8-2.8)	0.9 (0.9-1.0)	0.32 (0.32-0.33)	325 (225-425)
70-79	0.2 (0.2-0.2)	95.6 (95.5-95.6)	5.2 (5.1-5.2)	1.3 (1.3-1.4)	0.44 (0.43-0.45)	347 (244-449)
80+	1.0 (1.0-1.1)	86.7 (86.6-86.8)	13.8 (13.7-13.9)	2.6 (2.5-2.7)	0.73 (0.71-0.76)	496 (279-714)
Cancer site						
Lip, Oral cavity and Pharynx	0.1 (0.1-0.2)	99.1 (99.0-99.2)	1.2 (1.1-1.2)	–	0.35 (0.32-0.37)	327 (222-433)
Oesophagus	0.3 (0.2-0.4)	98.4 (98.2-98.5)	2.2 (2.1-2.3)	–	0.78 (0.73-0.84)	307 (210-403)
Stomach	0.3 (0.3-0.4)	98.1 (98.0-98.2)	2.6 (2.5-2.8)	–	0.66 (0.63-0.69)	323 (221-426)
Colon and Rectum	0.2 (0.2-0.3)	97.5 (97.4-97.5)	2.9 (2.9-3.0)	–	0.40 (0.39-0.42)	312 (208-416)
Liver	0.8 (0.7-0.9)	65.2 (64.6-65.8)	15.2 (14.8-15.6)	–	0.86 (0.82-0.91)	463 (248-678)
Pancreas	0.8 (0.7-0.9)	76.8 (76.4-77.2)	20.0 (19.8-20.3)	–	0.91 (0.85-0.96)	494 (264-724)
Larynx	0.1 (0.0-0.1)	98.9 (98.8-99.1)	1.3 (1.1-1.4)	–	0.35 (0.32-0.38)	568 (344-791)
Lung	0.4 (0.3-0.4)	91.4 (91.3-91.5)	8.3 (8.2-8.4)	–	0.78 (0.76-0.79)	346 (249-442)
Melanoma of the Skin	0.0 (0.0-0.0)	99.9 (99.8-99.9)	–	–	0.16 (0.14-0.19)	304 (197-411)
Breast (Female)	0.1 (0.1-0.1)	99.4 (99.4-99.5)	0.8 (0.8-0.8)	–	0.21 (0.20-0.22)	296 (195-397)
Cervix Uteri	0.1 (0.0-0.1)	99.5 (99.4-99.6)	0.8 (0.7-0.9)	–	0.26 (0.23-0.29)	423 (196-650)
Uterus: Corpus and Unspecified	0.1 (0.1-0.2)	99.0 (98.9-99.1)	1.2 (1.1-1.3)	–	0.28 (0.25-0.31)	410 (232-587)
Ovary	0.3 (0.3-0.4)	95.7 (95.5-96.0)	5.5 (5.3-5.7)	–	0.69 (0.66-0.73)	339 (236-443)
Prostate	0.2 (0.2-0.2)	97.7 (97.6-97.7)	3.3 (3.2-3.4)	–	0.19 (0.18-0.20)	402 (271-533)
Testis	0.0 (0.0-0.1)	99.5 (99.4-99.7)	1.0 (0.8-1.1)	–	0.03 (0.03-0.04)	430 (211-649)
Bladder	0.1 (0.1-0.2)	98.5 (98.4-98.6)	2.3 (2.2-2.4)	–	0.25 (0.22-0.27)	344 (240-448)
Kidney, Renal Pelvis, Ureter	0.2 (0.2-0.3)	90.1 (89.8-90.3)	8.3 (8.1-8.4)	–	0.33 (0.31-0.35)	436 (262-610)
Central Nervous System	0.3 (0.2-0.4)	84.6 (84.1-85.1)	9.5 (9.2-9.8)	–	0.77 (0.73-0.82)	347 (258-436)
Thyroid	0.1 (0.0-0.1)	99.6 (99.5-99.6)	1.1 (1.0-1.2)	–	0.07 (0.06-0.09)	357 (252-463)
Hodgkin Lymphoma	0.0 (0.0-0.1)	99.8 (99.7-99.9)	–	–	0.15 (0.12-0.18)	463 (248-678)
Non-Hodgkin Lymphoma	0.2 (0.2-0.3)	99.4 (99.3-99.4)	12.8 (12.6-13.0)	–	0.32 (0.30-0.34)	444 (251-636)
Other haematological malignancies (HM)	0.4 (0.4-0.5)	99.1 (99.0-99.2)	6.4 (6.2-6.6)	–	0.62 (0.58-0.65)	552 (343-761)
Mesothelioma	0.2 (0.1-0.4)	99.1 (98.8-99.3)	–	–	0.88 (0.80-0.96)	327 (234-420)
Primary site unknown (C80)	1.4 (1.2-1.5)	76.6 (76.1-77.0)	28.4 (28.0-28.7)	–	0.84 (0.73-0.95)	529 (304-753)
Other	0.5 (0.5-0.6)	94.5 (94.3-94.6)	8.6 (8.5-8.7)	–	0.48 (0.44-0.51)	517 (289-745)

3.2 Proportion of microscopically verified cases (MV%)

The lowest MV% occurred for hepatic and pancreatic cancer, followed by lung and central nervous system. The highest MV% was observed for lip and oral cancers, larynx, melanoma, female breast cancer, cancer of the cervix and uterus, testis, thyroid and Hodgkin and non-Hodgkin lymphoma (Figure 3).

MV% increased over time across cancer sites, from an average 81% in the period 1995-1999, to 88% for the period 2010-2014.

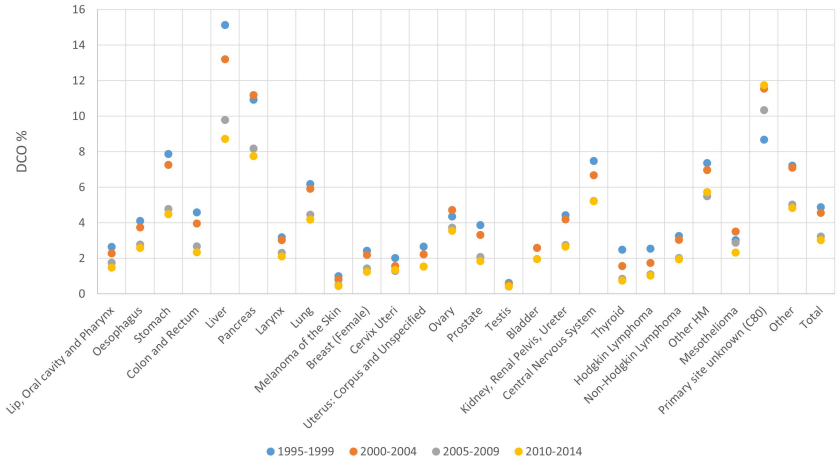


FIGURE 2
Proportion of cases with death certificate only (DCO%) by period of diagnosis and cancer site, 1995-2014.

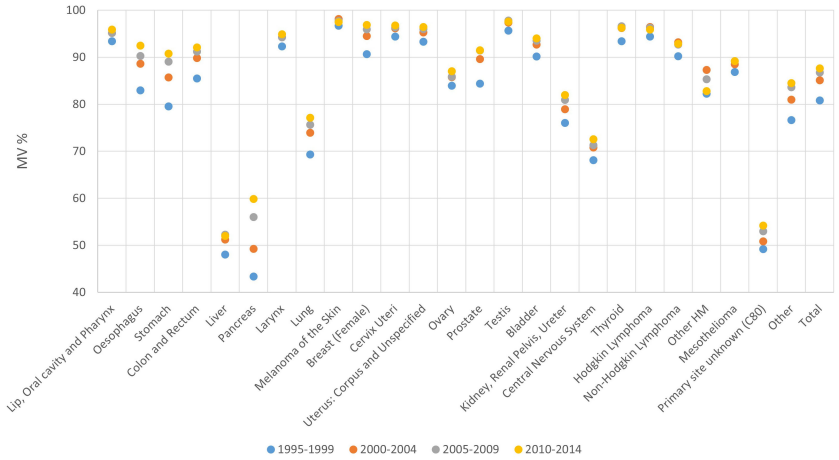


FIGURE 3
Proportion of microscopically verified cases (MV%) by period of diagnosis and cancer site, 1995-2014.

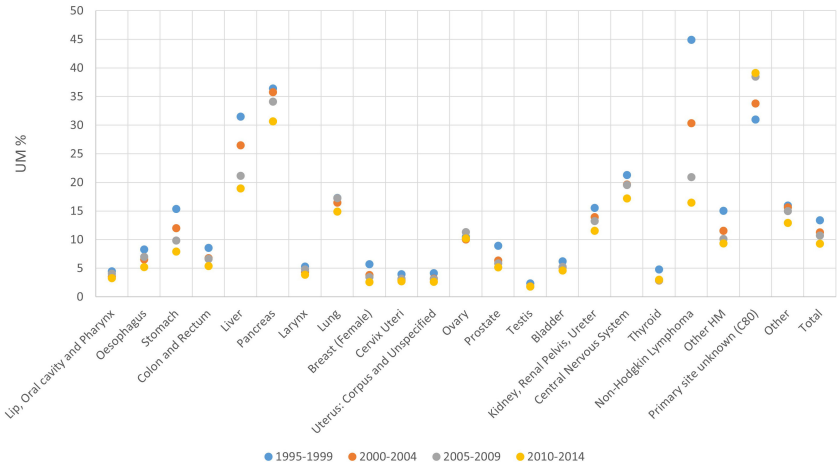


FIGURE 4
Proportion of cases with unspecified morphology (UM%) by period of diagnosis and cancer site, 1995-2014.

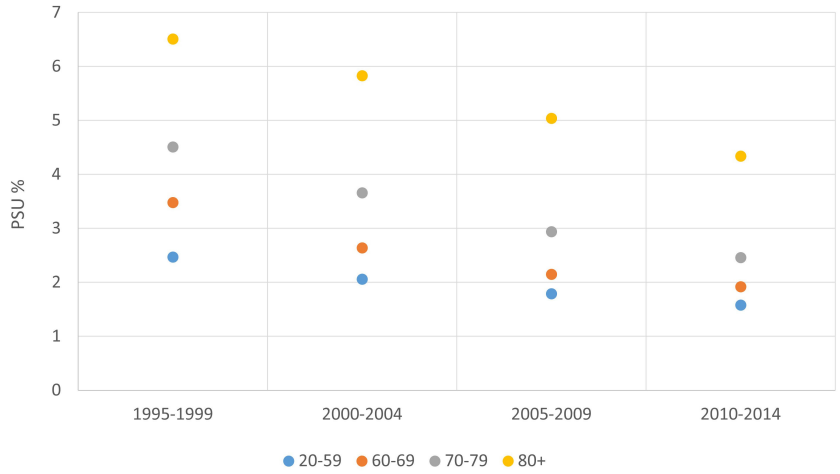


FIGURE 5
Proportion of cases with unknown primary site/primary site uncertain (PSU%) by age group and period of diagnosis, 1995-2014.

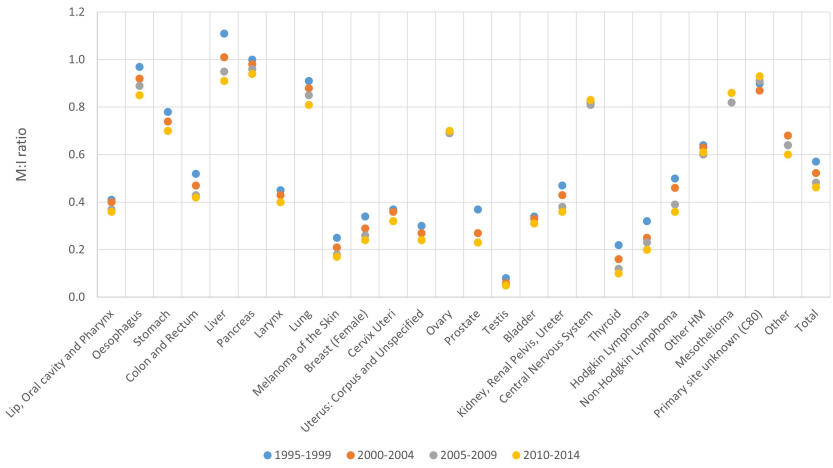


FIGURE 6
Mortality to incidence (M:I) ratio by period of diagnosis and cancer site, 1995-2014.

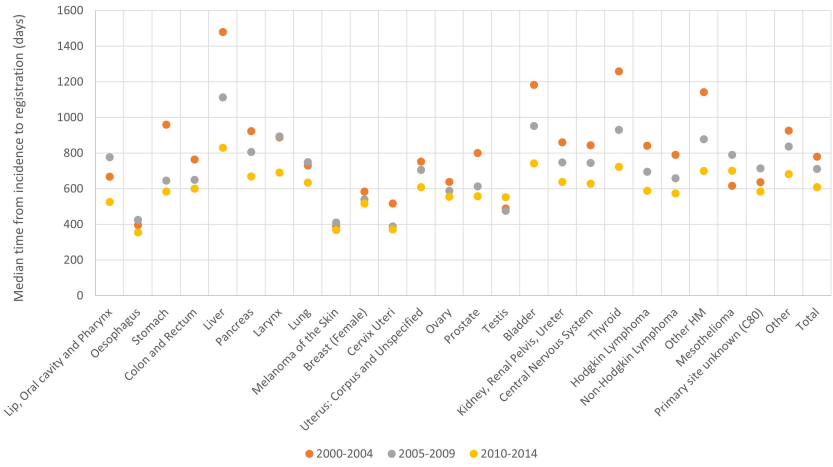


FIGURE 7
Timeliness by period of diagnosis and cancer site, 2000-2014.

The biggest improvement between 1995-1999 and 2010-2014 was observed for pancreas (43% and 60% respectively), stomach (80% and 91% respectively) and oesophagus (83% and 92% respectively) (Figure 3).

The MV% was similar for males and females (85% and 86% respectively - data not shown).

When comparing different age groups, the highest MV% (94%) occurred in the younger age group (20-59 years), followed by age groups 60-69 (90%), 70-79 (84%) and 80+ (68%). MV% decreased by age group for all cancer sites. In particular, age groups 20-59 and 80+ had respective MV%s of 86% vs 20% for central nervous system, 74% vs 27% for pancreas and 86% vs 49% for lung (Table 1; Supplementary Figure 3).

As for DCO%, a high variability among PBCRs was found, although 117 out of 128 PBCRs had an overall MV% of at least 80% in the latest available period of incidence, MV% increased for most PBCRs between 1995-1999 and 2010-2014 (Supplementary Figure 4).

3.3 Proportion of cases with unspecified morphology (UM%)

The highest UM% was found for non-Hodgkin lymphoma, mainly in period 1995-2004, primary site unknown, pancreas and liver and the lowest was found for testis, thyroid, uterus and lip, oral cavity and pharynx.

The UM% decreased over time, from an average of 13% in the period 1995-1999 to 9% in the period 2010-2014. The highest decrease was observed for non-Hodgkin lymphoma (45% and 16% in the periods 1995-1999 and 2010-2014 respectively), liver (31% vs 19%) and stomach (15% vs 8%) (Figure 4).

The UM% was 11% for both males and females (data not shown).

As for the previous indicators, UM% was lower with increasing age for all cancer sites (on average, 6% and 22% for ages 20-59 and 80+ respectively). In particular, age groups 20-59 and 80+ had a respective UM% of 10% vs 53% for central nervous system, 21% vs 51% for pancreas, 6% vs 33% for kidney, renal pelvis and ureter, and 4% vs 28% for ovary (Table 1; Supplementary Figure 5).

A high variability in UM% was observed among PBCRs, although 112 out of 130 PBCRs had an overall UM% below 20% in the latest available period of incidence. As for previously considered indicators, an improvement occurred for most PBCRs between incidence years 1995-1999 and 2010-2014 (Supplementary Figure 6).

3.4 Proportion of cases with unknown primary site/primary site uncertain (PSU%)

The PSU% was 3% for both males and females (data not shown). As far as this dimension is considered, data quality decreased with increasing age (Figure 5).

Similarly to the other indicators presented above, PSU% improved over time for all age groups (Figure 5).

All 130 PBCRs had less than 5% PSU cases in the latest available period, and the indicator decreased for the majority of PBCRs (Supplementary Figure 7).

3.5 Mortality to incidence (M:I) ratio

The highest M:I ratio was observed for hepatic and pancreatic cancer, followed by cancer of the oesophagus, lung and stomach. The lowest ratio was observed for testicular cancer, followed by thyroid and melanoma of the skin (Figure 6).

Overall M:I ratio was 0.53 for males and 0.49 for females in the analysed period (data not shown).

The overall M:I ratio declined over time, from 0.57 in 1995-1999 to 0.46 in 2010-2014. The biggest improvement was observed for liver, which decreased from 1.11 between 1995-1999 to 0.91 between 2010-2014. A decrease in M:I ratio between 1995-1999 and 2010-2014 was observed also for other cancer sites such as oesophagus (0.97 and 0.85 respectively) and prostate (0.37 and 0.23 respectively) (Figure 6).

M:I ratio increased with increasing age, from 0.32 in patients aged 20-59 years to 0.79 in patients aged 80+. M:I ratio increased by age group in all cancer sites except central nervous system. In particular, age groups 20-59 and 80+ had a respective M:I ratio of 0.81 vs 1.19 for liver, 0.78 vs 1.07 for oesophagus, 0.84 vs 1.07 for pancreas and 0.44 vs 0.98 for ovary (Table 1; Supplementary Figure 8).

Out of 92 PBCRs with available mortality data, 76 had an overall M:I ratio between 0.4 and 0.5 in the latest available period of incidence (Supplementary Figure 9).

3.6 Timeliness

For the 49 PBCRs with available data, the median time from incidence to registration decreased from 781 to 610 days between incidence years 2000-2004 and 2010-2014 (Figure 7). This indicator improved particularly for liver (from 1479 to 830 days respectively), thyroid (from 1259 to 723 days) and bladder (from 1184 to 743 days) and remained relatively low throughout incidence years 2000-2014 for oesophagus, melanoma of the skin and cervix uteri (Figure 7).

The median time to registration was lower for younger patients for the majority of cancer sites, for instance, for cervix uteri (384 vs 582 days respectively for age groups 20-59 and 80+ years) and prostate (448 vs 792 days) (Table 1; Supplementary Figure 10).

A huge variability in timeliness was observed among PBCRs, although 31 out of 49 considered PBCRs had a median time from incidence date to registration date between one and four years in the latest available period. For most PBCRs the indicator improved between incidence years 2000-2004 and 2010-2014 (Supplementary Figure 11).

4 Discussion

This article gives an overview of data quality among the European PBCRs contributing to the ECIS in the 2015 ENCR-

JRC data call. Reference values were computed for the most recently available incidence period (2010-2014) in order to evaluate data quality for future submission to the ECIS (Table 2).

Most of the indicators computed in this study have been used at international level for comparing and interpreting cancer data among different PBCRs (2–4, 11). In addition, PBCRs are using them for data quality evaluation (12–21). UM% and PSU% were computed for all 130 PBCRs included in the analysis. UM% and PSU% indicators are based on topography and morphology variables, considered as core variables and available for all PBCRs.

A limitation of this first evaluation is the delay after the latest submissions, in 2018, to the previous ENCR-JRC data call and the present analysis. The benchmarks that were calculated and the experience with the previous data call will help reducing such delay in future data quality assessments in ECIS.

MV% and DCO% were computed on the basis of diagnosis variable, which is also considered a core variable and also available for all PBCRs. Nevertheless, the "death certificate only" category (i.e. basis of diagnosis = 0) of this variable is available only for PBCRs with access to death certificate.

Mortality data by cause of death, sex and age group were not available for 38 PBCRs and M:I ratio could therefore not be computed for these PBCRs.

Only 49 PBCRs submitted registration date for at least two years in each of the two considered periods (2000-2006 and 2007-2014). Therefore, timeliness, median of the difference between the registration date and the incidence date, was computed for the 49 PBCRs.

It will be not possible to compute timeliness at European level in the near future, because date of registration is among the variables not included in the 2022 Call for Data Protocol for European Population-Based Cancer Registries (22) due to the low number of the PBCR that submitted this variable in the 2015 ENCR-JRC data call. Nevertheless, this indicator could be useful at PBCR level for improving the efficiency of PBCR procedures (2).

The use of death certificates as information source is a mean for PBCRs of finding cases not captured by other registration procedures (23). A higher DCO% is often linked to poor cancer prognosis. A high percentage of DCOs can point out incompleteness, as well as low validity.

Liver and pancreas were the cancer sites with the highest proportion of DCO%. This observation is consistent with data from other PBCRs (12, 16, 18). In any case, the DCO% varies highly across cancer sites. The Finnish PBCR reported an overall DCO% (all sites) of 2.6%, also with high differences between cancers. The highest values were reported for unspecified topographies such as respiratory tract NOS (C37 and C39), other digestive organs (C26) and uterus NOS (C55, C58) with values 39%, 23% and 20% respectively. The DCO% for pancreas was 9.5% and for liver 4.8% (16).

A decrease in DCO% was observed between 1995-1999 and 2010-2014. This is in line with what reported for similar periods in Cancer Incidence in Five Continents volumes IX and X (24, 25) and as reported also in selected PBCRs' studies, namely Zurich and Zug PBCRs, where the proportion of DCO cases declined between 1997 (6.4%) and 2014 (0.8%) (18). As a matter of fact, a declining DCO% trend is a natural consequence of increasing attention and efforts over time to improve data quality. An important activity aimed at improving PBCRs data quality is carried out by the JRC and the ENCR, in the form of training opportunities, the set-up of working groups to draft guidelines provided for data coding, registry visits and most importantly validation of cancer registry data itself (26–30).

Although death certificates are available for the majority of the European PBCRs, there is still a consistent percentage (22%) with problems in accessing death certificates. This issue could have an impact on cancer incidence computation and also survival estimations (31). Nevertheless, DCO% is low for the majority of cancer sites and for the European PBCRs contributing to the ECIS. Therefore, it is unlikely to have significant impact in data comparability among PBCRs, in particular in the latest period of incidence. Lastly, it should be noted that the proportion of death certificate initiated cases (DCI%) is presently not available in ECIS. This indicator can be an important complement to evaluate DCO% but is still not routinely reported by many European PBCRs (3).

The MV% was overall high, with an average value of 85%. The lowest MV% was observed for liver and pancreatic cancer. High overall MV% are consistent with what was reported in selected countries, namely Finland (93%) (16), Norway (94%) (12) and Iceland (96%) (13). Lower MV% values were observed for few PBCRs, consistent with what was reported for instance in Ukraine (78%) and Hungary (58%) (19, 21).

Opposite to the overall value, Iceland reported a low MV% (67%) for liver (13), and Finland reported a MV% of 63% for pancreatic cancer (16).

The highest MV% occurred in the youngest age group and declined with increasing age. This could be explained, at least partially, by a lower diagnostic activity in elderly patients.

An increase in MV% over time was observed, in line with what reported for similar periods in Cancer Incidence in Five Continents volumes IX and X (24, 25). MV% is mainly considered as a measure of validity, but a very high proportion of cases diagnosed by histology or cytology may also suggest that a PBCR is over-reliant on pathology as a source of information and might not detect part of the cases normally diagnosed by other means (2). As an example, the Swiss PBCRs of Zurich and Zug reported a MV% of 62% for 1997, which increased to 81% for 2014 (19).

The UM% was 11% in the observed periods, with a decrease from 13% in 1995-1999 to 9% in 2010-2014. This decrease was highest for non-Hodgkin lymphoma, liver and stomach, at least in part explained by the improvement of the diagnosis techniques for these tumours.

The PSU% was overall around 3%. This indicator decreased for the majority of PBCRs over time.

The PSU% reported by the Iceland PBCR was 1.9% for men and 3.1% for women, while it was 2.2% for both sexes in Norway in 2001-2005. In both countries there was an increase of this percentage with advancing age (12, 13). Differences in the age

distribution of men and women populations could partially explain these differences. Nevertheless, differences by sex were not found in our study when all PBCRs were considered together.

Since rare tumours are defined by topography and specific morphology (32), UM% and PSU% could have an important impact in rare cancer incidence computation and data comparability.

The M:I ratio declined over time (from 0.57 in 1995-1999 to 0.46 in 2010-2014), confirming the findings from Cancer Incidence in Five Continents volumes IX and X (24, 25) and reported by selected PBCRs studies: in two Swiss PBCRs, the M:I ratio declined from 0.58 in 1980 to 0.37 in 2014 (18). Bulgaria reported an M:I ratio of 0.5 for males, and 0.4 for females (15). The higher M:I ratio for males observed also in our study (0.53 vs 0.49 for females) is also in line with the usual inverse relationship between this indicator and survival, which is higher in females (3, 25).

M:I ratio can help interpreting cancer incidence in PBCRs, by comparing the indicator with cancer incidence rates. A higher M:I ratio could be associated with lower completeness and incidence rates, which should be interpreted with caution (see for instance the example in Supplementary Figure 12). Other factors, such as the quality of death certificates, should be also taken into account into the interpretation of M:I ratio.

As a limitation, mortality data was not available for 38 PBCRs at the moment of the analysis; these were mostly regional registries. In some cases, data was provided by PBCRs in a different format from the one required in the ECIS data call protocol (e.g. less than 18 age classes). Following the analysis most of the problems related to such data were solved, and updated mortality figures can be found in the ECIS web application (7).

Timeliness was evaluated computing the median time from incidence to case registration, which ranged between one and four years for the majority of PBCRs recording this information. This is in line with what reported in a survey performed in 2011, where European PBCRs stated a median time from incidence to data publication (which is related with data registration) of 18 months, with a range between 4 months and 5 years (11). Timeliness indicators have not been frequently reported by PBCRs; however the reduction in time to registration observed in our analysis (with an average decrease of 171 days between 2000-2004 and 2010-2014) has a similar trend to what reported by Norway (from over 525 days in 2001 to 261 days in 2005), whereas Iceland reported a median time from date of diagnosis to registration of 238 days (with a range between 49 and 1445 days) (12, 13). Lastly, an increase in time to registration was observed for 3 PBCRs between 2000-2004 and 2010-2014; this could possibly be due to resource constraints, which have been common for smaller regional PBCRs throughout Europe in recent years.

Indicators for European PBCRs such as MV%, DCO% and M:I ratio were found to be similar to those reported for other developed areas worldwide, in particular to North America, Australia and New Zealand (24).

5 Conclusion and way forward

The quality of incidence data reported by PBCRs has been improving across the study period. Data quality is worse for the oldest age groups and for cancer sites with poor survival. No differences were found between males and females. High variability in data quality could be detected across European PBCRs.

The harmonisation of PBCR' data as the input source for the assessment of cancer burden is one of the main aims of the support provided by the JRC to the ENCR to strengthen the basis for monitoring the cancer burden. In order to improve data quality and harmonisation, the JRC and the ENCR have been carrying out several activities along the years, namely the set-up of yearly training agendas and organisation of trainings, the coordination of thematic Working Groups to draft guidelines and recommendations on data coding, the development and provision of common rules and related validation software to check data compliance to agreed EU-wide standards.

In this context, the results reported in this paper are to be interpreted as the baseline for monitoring PBCRs data quality indicators in Europe along time.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Requests to access the datasets should be directed to francescogiusti@hotmail.com.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

The first draft of the manuscript was written by FG and CM. MB supervised data acquisition. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors acknowledge Carine Dochez (Belgian Cancer Registry) for her editorial assistance in the preparation of this manuscript. Grateful acknowledgement is given to the European cancer registries for contributing to the initiatives and studies included in ECIS.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1219128/full#supplementary-material

References

1. White MC, Babcock F, Hayes NS, Mariotto AB, Wong FL, Kohler BA, et al. The history and use of cancer registry data by public health cancer control programs in the United States. *Cancer* (2017) 123(suppl 24):4969–76. doi: 10.1002/cncr.30905

2. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: Comparability, validity and timeliness. *EJC* (2009) 45:747–55. doi: 10.1016/j.ejca.2008.11.032

3. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods. Part II: Completeness. *EJC* (2009) 45:756–64. doi: 10.1016/j.ejca.2008.11.033

4. Bray F, Colombet M, Mery L, Piñeros M, Znaor A, Zanetti R, et al. *Cancer Incidence in Five Continents, Vol. XI. IARC Scientific Publication No. 166*. Lyon: International Agency for Research on Cancer (2021). Available at: https://publications.iarc.fr/597.

5. *European Network of Cancer Registries* (2022). Available at: https://encr.eu/ (Accessed 24 March 2023).

6. *ENCR-JRC Project 2015 - Updating population-based cancer incidence and mortality in Europe* (2015). Available at: https://www.encr.eu/sites/default/files/ENCR-JRC2015_Protocol.pdf (Accessed 24 March 2023).

7. *ECIS - European Cancer Information System* (2023). European Union. Available at: https://ecis.jrc.ec.europa.eu (Accessed 24 March 2023).

8. *International Classification of Childhood Cancer* (2017). Available at: https://iicc.iarc.fr/classification/diagnostic-groups-and-subgroups/iccc-3-2017-main-sub-groups.pdf (Accessed 05 July 2023).

9. *European Network of Cancer Registries, ENCR-JRC Call for Data* (2015). Available at: https://encr.eu/sites/default/files/Data_call/2015_ENCR_JRC_Call_for_Data_Version_1_1.pdf (Accessed 12 March 2023).

10. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al. *International Classification of Diseases for Oncology. 3rd ed*. Geneva: World Health Organization (2013).

11. Zanetti R, Schmidtmann I, Sacchetto L, Binder-Foucard F, Bordoni A, Coza D, et al. Completeness and timeliness: Cancer registries could/should improve their performance. *Eur J Cancer* (2015) 51(9):1091–8. doi: 10.1016/j.ejca.2013.11.040

12. Larsen IK, Smastuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. *Eur J Cancer* (2009) 45(7):1218–31. doi: 10.1016/j.ejca.2008.10.037

13. Sigurdardottir LR, Jonasson JG, Stefansdottir S, Jonsdottir A, Olafsdottir GH, Olafsdottir EJ, et al. Data quality at the Icelandic cancer registry: comparability, validity, timeliness and completeness. *Acta Oncol* (2012) 51(7):880–9. doi: 10.3109/0284186X.2012.698751

14. Dehler S, Tonev S, Korol D, Rohrmann S, Dimitrova N. Recent trends in cancer incidence: impact of risk factors, diagnostic activities and data quality of registration. *Tumori* (2014) 100(4):399–405. doi: 10.1700/1636.17894

15. Dimitrova N, Parkin DM. Data quality at the Bulgarian National Cancer Registry: An overview of comparability, completeness, validity and timeliness. *Cancer Epidemiol* (2015) 39(3):405–13. doi: 10.1016/j.canep.2015.03.015

16. Leinonen MK, Miettinen J, Heikkinen S, Pitkäniemi J, Malila N. Quality measures of the population-based Finnish Cancer Registry indicate sound data quality for solid malignant tumours. *Eur J Cancer* (2017) 77:31–9. doi: 10.1016/j.ejca.2017.02.017

17. Lorez M, Bordoni A, Bouchardy C, Bulliard JL, Camey B, Dehler S, et al. Evaluation of completeness of case ascertainment in Swiss cancer registration. *Eur J Cancer Prev* (2017) 26:S139–46. doi: 10.1097/CEJ.0000000000000380

18. Wanner M, Matthes KL, Korol D, Dehler S, Rohrmann S. Indicators of data quality at the cancer registry Zurich and Zug in Switzerland. *BioMed Res Int* (2018) 2018:7656197. doi: 10.1155/2018/7656197

19. Ryzhov A, Bray F, Ferlay J, Fedorenko Z, Goulak L, Gorokh Y, et al. Evaluation of data quality at the National Cancer Registry of Ukraine. *Cancer Epidemiol* (2018) 53:156–65. doi: 10.1016/j.canep.2018.02.002

20. Barchuk A, Tursun-Zade R, Belayev A, Moore M, Komarov Y, Moshina N, et al. Comparability and validity of cancer registry data in the northwest of Russia. *Acta Oncol* (2021) 60(10):1264–71. doi: 10.1080/0284186X.2021.1967443

21. Wéber A, Mery L, Nagy P, Polgár C, Bray F, Kenessey I. Evaluation of data quality at the Hungarian National Cancer Registry, 2000-2019. *Cancer Epidemiol* (2023) 82:102306. doi: 10.1016/j.canep.2022.102306

22. *Call for Data Protocol for European Population-Based Cancer Registries* (2022). Available at: https://www.encr.eu/sites/default/files/Data_call/ECIS%20call%20for%20data%20protocol_20221124.pdf (Accessed 18 January 2023).

23. Bray F, Ferlay J. (2021) Chapter 5: data comparability and quality. In: *Cancer Incidence in Five Continents volume XI. IARC Scientific Publication No. 166*. Lyon: International Agency for Research on Cancer.

24. Curado MP, Edwards B, Shin HR, Storm H, Ferlay J, Heanue M, et al. *Cancer Incidence in Five Continents, Vol. IX. IARC Scientific Publication No. 160*. Lyon: International Agency for Research on Cancer (2007).

25. Forman D, Bray F, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M, et al. *Cancer Incidence in Five Continents, Vol. X. IARC Scientific Publication No. 164*. Lyon: International Agency for Research on Cancer (2014).

26. Martos C, Crocetti E, Visser O, Rous B, Giusti F and the Cancer Data Quality Checks Working Group. *A proposal on cancer data quality checks: one common procedure for European cancer registries – version 1.1*. Luxembourg: Publications Office of the European Union (2018). doi: 10.2760/429053

27. Giusti F, Martos C, Adriani S, Flego M, Brunetto A, Dyba T, et al. *The JRC-ENCR Quality Check Software (QCS) for the validation of cancer registry data: user compendium – version 2.0*. Ispra: European Commission (2022).

28. ENCR. *Materials from trainings and workshops*. Available at: https://encr.eu/node/535 (Accessed 12 March 2023).

29. ENCR. *e-learning modules on Cancer Coding*. Available at: https://encr.eu/news/e-learning (Accessed 12 March 2023).

30. ENCR. *Registries Reviews*. Available at: https://encr.eu/registries-reviews (Accessed 12 March 2023).

31. Andersen MR, Storm HH. Cancer registration, public health and the reform of the European data protection framework: Abandoning or improving European public health research? *Eur J Cancer* (2015) 51:1028–38. doi: 10.1016/j.ejca.2013.09.005

32. Trama A, Marcos-Gragera R, Sánchez Pérez MJ, van der Zwan JM, Ardanaz E, Bouchardy C, et al. Data quality in rare cancers registration: the report of the RARECARE data quality study. *Tumori* (2017) 103:22–32. doi: 10.5301/tj.5000559

frontiers | Frontiers in Oncology

TYPE Technology and Code
PUBLISHED 08 August 2023
DOI 10.3389/fonc.2023.1098342

OPEN ACCESS

EDITED BY
Liesbet Van Eycken,
Belgian Cancer Registry, Belgium

REVIEWED BY
Stephanie G. C. Kroeze,
Aarau Cantonal Hospital, Switzerland
Roberta De Angelis,
National Institute of Health (ISS), Italy
Emanuele Crocetti,
Independent researcher, Florence, Italy
Maximilian Knoll,
German Cancer Research Center (DKFZ), Germany

*CORRESPONDENCE
Siri Larønningen
sla@kreftregisteret.no

RECEIVED 14 November 2022
ACCEPTED 04 July 2023
PUBLISHED 08 August 2023

CITATION
Larønningen S, Skog A, Engholm G, Ferlay J, Johannesen TB, Kristiansen MF, Knoors D, Kønig SM, Olafsdottir EJ, Pejicic S, Pettersson D, Skovlund CW, Storm HH, Tian H, Aagnes B and Miettinen J (2023) Nordcan.R: a new tool for federated analysis and quality assurance of cancer registry data. *Front. Oncol.* 13:1098342. doi: 10.3389/fonc.2023.1098342

Nordcan.R: a new tool for federated analysis and quality assurance of cancer registry data

Siri Larønningen[1]*, Anna Skog[1], Gerda Engholm[2], Jacques Ferlay[3], Tom Børge Johannesen[1], Marnar Fríðheim Kristiansen[4], Daan Knoors[5], Simon Mathis Kønig[2], Elinborg J. Olafsdottir[6], Sasha Pejicic[7], David Pettersson[7], Charlotte Wessel Skovlund[2], Hans H. Storm[2], Huidong Tian[1], Bjarte Aagnes[1] and Joonas Miettinen[8]

[1]Department of Registration, Cancer Registry of Norway, Oslo, Norway, [2]Cancer Surveillance and Pharmacoepidemiology, Danish Cancer Society Research Center, Copenhagen, Denmark, [3]Cancer Surveillance Branch, International Agency for Research on Cancer, Lyon, France, [4]University of the Faroe Islands, Faculty of Health Sciences, Tórshavn, Faroe Islands, [5]Research and Development, Netherlands Comprehensive Cancer Organisation, Eindhoven, Netherlands, [6]The Icelandic Cancer Registry, Reykjavik, Iceland, [7]The National Board of Health and Welfare, Stockholm, Sweden, [8]Finnish Cancer Registry, Helsinki, Finland

Aim of the article: We present our new GDPR-compliant federated analysis programme (nordcan.R), how it is used to compute statistics for the Nordic cancer statistics web platform NORDCAN, and demonstrate that it works also with non-Nordic data.

Materials and methods: We chose R and Stata programming languages for writing nordcan.R. Additionally, the internationally used CRG Tools programme by International Agency for Research on Cancer (IARC/WHO) was employed. A formal assessment of (GDPR-compliant) anonymity of all nordcan.R outputs was performed. In order to demonstrate that nordcan.R also works with non-Nordic data, we used data from the Netherlands Cancer Registry.

Results: nordcan.R, publicly available on Github, takes as input cancer and general population data and produces tables of statistics. Each NORDCAN participant runs nordcan.R locally and delivers its results to IARC for publication. According to our anonymity assessment the data can be shared with international organizations, including IARC. nordcan.R incidence results on Norwegian and Dutch data are highly similar to those produced by two other independent methods.

Conclusion: nordcan.R produces accurate cancer statistics where all personal and sensitive data are kept within each cancer registry. In the age of strict data

protection policies, we have shown that international collaboration in cancer registry research and statistics reporting is achievable with the federated analysis approach. Undertakings similar to NORDCAN should consider using nordcan.R.

KEYWORDS

NORDCAN, quality, harmonization, cancer, epidemiology, software, GDPR-compliance, federated analysis

1 Introduction

The Nordic cancer registries have collaborated closely since the first registries were established in the 1940s and early 1950s. The collaboration was formalized in 1966 as the Association of Nordic Cancer Registries (ANCR) and has resulted in numerous projects and collaborations to describe incidence and mortality trends in the Nordic countries and to develop statistics to support cancer surveillance, decision making and research in the Nordic countries (1–7). A crucial part of these projects is to prove the quality of the Nordic cancer registry data through quality estimates like completeness, validity, timeliness and comparability (8, 9).

NORDCAN, the Nordic cancer database and webtool for cancer incidence, mortality, prevalence and survival (10) was created in the mid-90's in a collaboration between the International Agency for Research on Cancer (IARC) and driving forces in some of the Nordic cancer registries. NORDCAN on the web was established in 2002. A work group including representatives from the cancer registries in all the Nordic countries was established to maintain and develop NORDCAN. The history, organization and use of NORDCAN and the comparability of the Nordic cancer registries have been described earlier (11–13).

The European General Data Protection Regulation (GDPR) (14) in effect since 2018, strengthened standards for sharing personal data. Although the GDPR allows for sharing personal data within the European Union and EEA, the stricter statutory interpretation of GDPR in most Nordic countries creates additional barriers. This includes the definition of personal data, individual consent to research use and international collaboration. Sharing personal data with UN agencies such as the IARC (WHO), which are exempted from national legislation and thus not subject to the GDPR regulations, is not allowed. It became apparent that the established NORDCAN procedures on data extraction, data preparation and data processing should be changed. The old and new data flow are shown in Figure 1.

The counts and statistics produced by nordcan.R are sent to IARC through a file transfer and are stored in the NORDCAN database to be visually represented in the NORDCAN webtool. Incorporated in the webtool is the logic to perform calculation of rates, predictions, estimated annual percentage change and survival improvement, based on the output files from nordcan.R. All statistics are based on the counts and analyses done in nordcan.R, and rates are available using the standard populations World, Europe and Nordic.

The aim of this article is to present the new GDPR compliant federated analysis programme (nordcan.R), how it is used to compute statistics for NORDCAN and how it can be used also for non-Nordic data.

2 Materials and methods

In 2018 we started rebuilding the data management, quality control, data flow and webtool of NORDCAN on research grants

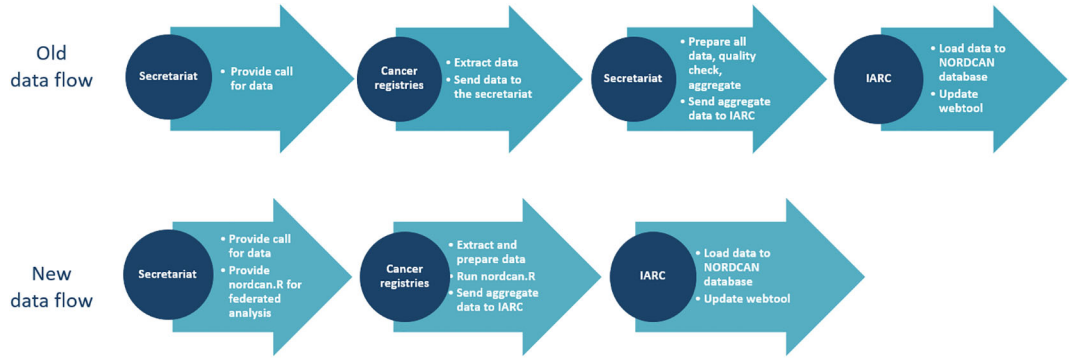


FIGURE 1
Old and new data flow for NORDCAN.

from the Danish Cancer Society, the Nordic Cancer Union (NCU) and in kind-contributions from the Nordic cancer registries.

There were two key issues to address. First, we had to ensure that personal data on cancer patients, in line with the national implementation of GDPR, would not be released from the cancer registries. Second, we had to make sure that released (aggregated) data to the NORDCAN database in IARC were comparable between countries and to previous releases.

To keep personal data on cancer patients within each cancer registry, we developed a simple federated system; a program that could be run separately in each country, producing all counts and statistics necessary for NORDCAN. The program was named nordcan.R. We established an IT-group with representatives from each of the Nordic cancer registries. The IT-group decided on architecture, technology, software and development guidelines to be used for the federated analysis application, developed the application itself, and tested the tool. Prior experience in designing and programming a system for data preprocessing and aggregation for the Finnish Cancer Registry was taken as the model.

We decided to use R (15) and Stata (16) for the development of the new tool. In addition, we chose to continue using the IARC check and conversion program (17) for conversions from International Classification of Diseases for Oncology, third edition (ICD-O-3) (18) to International Classification of Diseases, tenth edition (ICD-10) (19) and for implementation of the multiple primary rules (20).

We specified four common input datasets to be used in nordcan.R: a cancer case dataset (21), a cancer death dataset (22), population datasets (including population projections) (23) and life tables to be used in survival analysis (24). Specifications on cancer entities (the cancer dictionary) used in NORDCAN already existed, but were revised to better meet the current needs of clinicians, cancer unions, patient organizations, researchers, and politicians (25, 26). We also developed a conversion table for mortality to be used for the conversion of old ICD-codes (ICD6-9) to ICD-10 (27).

Before starting the development of nordcan.R, we identified the core functionality of the program (Table 1). The NORDCAN secretariat also provided all participating countries with an overview of the necessary tools to be used, with links to installation files and sources, and additional information on licensing where necessary (Table 2).

We did a risk assessment on the aggregate data of the output files (31). We drew knowledge from other sources on statistical disclosure control and data anonymization (32–34) to evaluate the risk of disclosure from the datasets created by nordcan.R and to identify possible risk-reducing measures.

We established a proof-of-concept project between the NORDCAN secretariat and the Netherlands Comprehensive Cancer Organization (IKNL) to test the usability of nordcan.R outside of the Nordic countries. We used data from the Netherlands Cancer Registry (NCR) hosted by IKNL for the period 2000-2019. For the comparison of the output, we calculated all rates in Stata, using the World standard population (35) for age-adjusted rates.

3 Results

The first version of nordcan.R (nordcan_9.0_1.0) was published to the NORDCAN-participants on November 27th, 2020. The most recent, nordcan_9.3_1.3, was released May 4th, 2023. The application has so far been used to update data in NORDCAN from 2016 to 2021.

The main result of revamping the software was that all Nordic countries were able to deliver updated, anonymous, tabular data to the NORDCAN database and webtool, hosted by IARC, without being subject to legal discussions and hindrance due to differences in the interpretation of GDPR. Through nordcan.R, all data were

TABLE 1 Identified core functionality of nordcan.R.

Core functionality	Specification
User-specified global settings	The user should be allowed to set: - Country - First and last year of national incidence data - First and last year of national mortality data - Last year of follow-up for survival data - First year of regional data
Check of input files	nordcan.R should check that the input-files adhere to the call for data
Integration with the IARC check and conversion program	nordcan.R should create the necessary input files for the IARC check and conversion program, and be able to read the output files from this program back into nordcan.R.
Enrichment of original data	nordcan.R should enrich the original data using output-files from the IARC check and conversion program and specifications from the NORDCAN group. This includes grouping data into NORDCAN entities and documenting reasons for exclusion of cases. nordcan.R should do recoding and calculations necessary for the final computation of data.
Create necessary output files	nordcan.R should create the following output: A zip-file with all counts and analyses necessary for the NORDCAN database - Counts for cancer cases, cancer deaths and prevalent cases by year, sex, region/country and entity (28) - Age standardized relative survival using the Pohar Perme estimator via stnet in Stata (29, 30). A folder with maintainer summary files - Technical information - Comparison summary - Graphics comparing counts between two different runs

TABLE 2 Tools necessary for nordcan.R users.

Tool	
R	https://www.r-project.org/ We recommend using R-studio (https://rstudio.com), but it is not necessary to run nordcan.R.
Stata (Stata/IC is sufficient)	https://www.stata.com/
IARCcrgTools	http://www.iacr.com.fr/index.php?option=com_content&view=article&id=72:iarccrgtools&catid=68&Itemid=445

quality assured, converted, analyzed and counted the same way, and the same inclusions and exclusions were applied to all data, regardless of country and region.

Nordcan.R produces a "maintainer summary" with plots and tables comparing the newly created counts to those created in a previous call for data. The user is required to inspect these comparisons for unexpected discrepancies between two versions as an additional quality assurance step. Figure 2 shows a comparison of counts of cancer cases in Norway between version 9.1 and version 9.2 of NORDCAN for selected entities. The reasons for changes in entity 980 (all sites) is clearer when looking at visualizations for each specific entity. We see an increase in entity 430 (malignant hematopoietic diseases), and changes in entity 280 (urinary tract cancers). Both these changes are explained by a thorough process of quality assurance and corrections in the Cancer registry of Norway. Smaller changes to the cancer counts, mainly caused by regular day-to-day corrections in cancer registries, are expected. The maintainer summary serves to acknowledge that each nordcan.R-user has seen the results for their own country and the comparison with previous results and accepts the results as valid. The maintainer summary is also an indication that the technical requirements of nordcan.R have been fulfilled.

Figure 3 shows the comparisons of adjusted incidence rates (World) between The Global Cancer Observatory (GCO) (36), official cancer statistics in the Netherlands (37) and Norway (38) and results from nordcan.R for colorectal cancer, men and women shown separately. There are only minor differences.

4 Discussion

As shown in the results, we achieved our main goal. We have produced the necessary data and we have been able to share the data with IARC for updating the NORDCAN database and webtool.

As GCO and nordcan.R both use the IARCcrgTool for conversions and exclusions of multiple primary cancers, and both cancer registries adhere to international rules on cancer registration, we expected that the trend lines would be quite similar. It is still valuable to make this comparison to see that nordcan.R did not introduce any unforeseen errors. Comparison with official national statistics produced by the national cancer registries is of additional value to see that the numbers and rates in NORDCAN do not deviate even if the IARCcrgTool is not used. The slight differences we see might be due to changes in the original data caused by updates, differences in implementation of multiple primary rules or corrections or small differences in exclusions and inclusions of cases.

We have a common call for data to be adhered to by all participants. We aimed to make this call for data as similar as possible to calls for data issued by IARC for Cancer Incidence in Five Continents (CI5) – Volume XII (39) and by The European Network of Cancer Registries (ENCR) in collaboration with the Joint Research Centre of the European Commission (JRC) for the update of the European Cancer Information System (ECIS) (40). NORDCAN has, however, focused mainly on the traditional cancer registry-variables which are needed to calculate incidence, mortality, prevalence and survival on country and regional level, whereas the JRC/ENCR call for data also includes variables on stage, treatment and more detailed geographical area.

Incorporating quality checks, conversions and inclusions/exclusions into the data management process when running nordcan.R, and using transparent and unified specifications were discussed in the NORDCAN work group as important steps towards comparability. This is a well-known recipe, used by both IARC and ENCR-JRC, and was also the same process used previously in NORDCAN. In the current NORDCAN pipeline, each country is responsible for completing the entire process

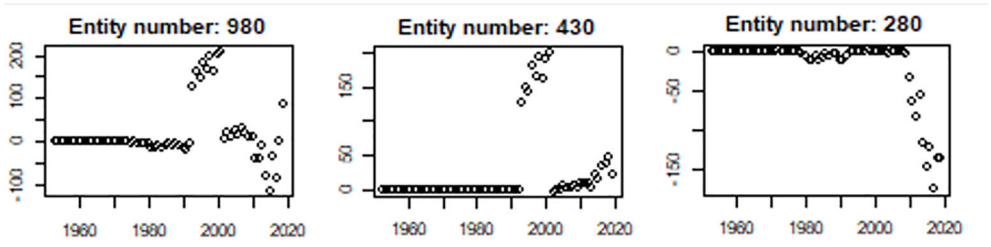


FIGURE 2
Comparison between previous and current number of incident cases for selected entities produced as a standard output by nordcan.R. The Y-axis represents change in number of cases, the X-axis represents year of cancer counts. The changes in all sites (980) are mainly driven by changes in malignant hematopoietic diseases (430) and urinary tract cancers (280).

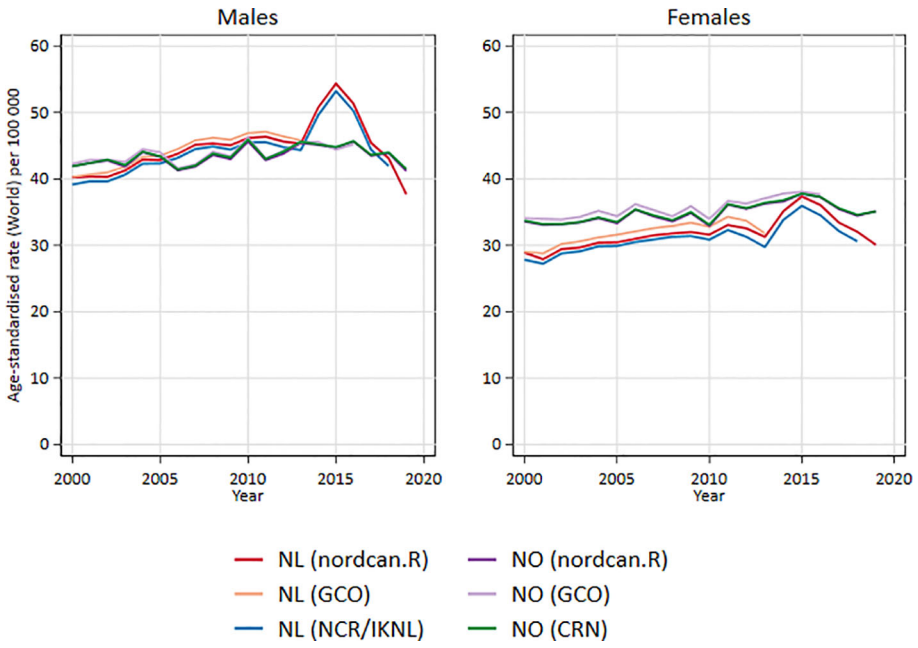


FIGURE 3
Comparison of adjusted incidence rates (World) between GCO, official national statistics and nordcan.R for colorectal cancer in Netherland and Norway, men and women separately.

themselves, whereas in the old, centralized approach, each participant sent data on patients to one central node for checks, conversions and aggregation. A big advantage of the new procedure is that the cancer registries run their own data through the entire process and see all errors and warnings. Another big advantage – and a prerequisite for the development of nordcan.R – is that all personal and sensitive health data is kept within each cancer registry. The disadvantage of this data flow is that there is a possibility that the countries select and map their original data in a slightly different way to fit the call for data, and thereby introduce a few minor differences. However, taking all pros and cons into consideration, our view is that the cost of introducing a few differences has less impact than the benefit of users with first-hand knowledge of the data doing all data processing, aided by a tailor-made tool.

Using a common check and conversion program like the one from IARC is also an advantage, as it ensures that all data in NORDCAN have been subject to the same conversion and check rules. A disadvantage of the current version of the IARC check and conversion program is that it is not updated with the most recent morphology codes from the ICD-O-3 second revision. Also, although nordcan.R includes written instructions on what selections to use in the tool, there is still room for manual errors that can lead to data being less comparable.

In the current version of nordcan.R, all entities or cancer groups are predefined through a table in the application. The user cannot change the groups or run different specifications without reprogramming parts of nordcan.R. This is an advantage, as it ensures comparability as to which ICD10-codes are included in the different entities, and which morphologies might be included or excluded in addition to the ICD10-codes, but makes nordcan.R less useful if the user wishes to prepare statistics for other cancer sites or wishes to compare to international statistics with different cancer groups.

When running nordcan.R, the user gets a visualization and an output showing how similar the current counts are to previous counts. Although this does not ensure comparability between countries, it ensures that data within one country are comparable over time, and that any bigger changes affecting the comparability of data can be documented and explained. Most cancer registries are "living organisms", where data are changed and updated over time, also for previously published data. It is therefore not unusual to see changes in counts from one version of NORDCAN to the next, but the reasons for changes should be clear and well documented, and any bigger changes should be accompanied by a warning on the webpage.

The usability of nordcan.R outside of NORDCAN was tested in our proof-of-concept project with Dutch cancer data. The first main hindrance was the necessity of installing three different types of software to run nordcan.R: R and the R-packages, Stata, and the IARC check and conversion tool. As nordcan.R operates on sensitive data, all tools needed to be installed in a secure environment which allowed for this kind of data processing. For Stata, there is an additional hindrance that installation and use require a license which the institution might not be willing to purchase. Not every potential user can fulfill the requirement of an on-site Stata installation. However, this only excludes the computation of survival statistics while other statistics are unaffected. Preparing a dataset according to the specifications given for nordcan.R did not reveal any major problems for IKNL. An upcoming inclusion of quality tables will also make it easier to compare reasons for exclusions and other quality measures of the data (41).

4.1 Future perspectives

The steps taken so far are only an introduction to future challenges and possibilities. Although nordcan.R fulfills its purpose for the current requirements of NORDCAN, there are still missing parts. For instance: we can currently show survival in Norway and Sweden compared to each other, but not survival for Norway and Sweden together.

With the current problems we have in sharing data with IARC, we see the need for a better alternative. This is why we look to federated analysis as one possible solution for sharing data and knowledge across borders without the current hindrances we meet in laws and regulations. A future incarnation of NORDCAN should allow for more complex analyses than are currently possible and retain the successful federated analysis approach. Establishing common data models in each cancer registry, for instance using the OMOP Common Data Model (42), might make it easier to implement different federated learning infrastructures to run analysis on local data and only share the results. However, getting approval to install software on-site which enables external access to sensitive data is not straightforward.

5 Conclusion

The Nordic cancer registries successfully developed a new software tool (nordcan.R) which enables them to share aggregated cancer statistics, comparable to previous releases and in a GDPR compliant manner, with the NORDCAN database and webtool. Nordcan.R was also successfully used in a non-Nordic country.

Accessibility to software and specifications

Documentation for nordcan.R, the application itself and the packages it consists of can be downloaded from the Cancer Registry of Norway GitHub repositories NORDCAN, basicepistats, nordcanepistats, nordcansurvival, nordcancore and nordcanpreprocessing (43). The code and packages are free to use as is under the given license. The NORDCAN work group and IT-group maintains and further develops nordcan.R for NORDCAN purposes and needs. Conversion tables, cancer dictionary and other documentation used in NORDCAN and nordcan.R which is not available through the GitHub-pages can be shared from the secretariat upon request. Questions can be directed to head of the NORDCAN secretariat, Siri Larønningen (sla@kreftregisteret.no).

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/CancerRegistryOfNorway/NORDCAN, https://nordcan.iarc.fr/en.

Author contributions

Main writers of manuscript: SL, HS, JM. Corrections, improvements and approval of final version: All authors. Developers of nordcan.R: JM, SP, HT, BA. Statistical analysis: AS, DK. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors would like to acknowledge Henrik Mulvad Hansen, Johanne Jerven Hestad, Sheila Kristensen, Lina Steinrud Mørch, Tomas Tanskanen and Anni Virtanen from the NORDCAN work group, Gijs Geleijnse in IKNL, Freddie Bray, Morten Ervik, Frederic Lam and Mathieu Laversanne in IARC, the ANCR board, NCU and JRC-ENCR, as well as all colleagues in the Nordic cancer registries. The process of replacing the NORDCAN database and webtool in IARC with a modernized platform was financially supported by The Danish Cancer Society, and the rebuild of the pipeline and the development of a new tool was financially supported by the Nordic Cancer Union (NCU) and in kind-contributions from the Nordic cancer registries.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor LV declared a past co-authorship with the author TB.

Publisher's note

Author disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

References

1. Ringertz N. Epidemiology of gastrointestinal cancers in Scandinavia. I. Report on Denmark, Finland, Norway, and Sweden. *Natl Cancer Institute Monograph* (1967) 25:219–39. Available at: https://europepmc.org/article/med/3465196.

2. Hakulinen T, Andersen A, Malker B, Pukkala E, Schou G, Tulinius H. Trends in cancer incidence in the Nordic countries. A collaborative study of the five Nordic Cancer Registries. *Acta Pathol Microbiol Immunol Scand Suppl* (1986) 288:1–151.

3. Jensen OM. *Atlas of cancer incidence in the Nordic countries: a collaborative study of the five Nordic cancer registries*. Pura Musta, Helsinki: Nordic Cancer Union (1988).

4. Tulinius H, Storm H, Pukkala E, Andersen A, Ericsson J. Cancer in the Nordic countries, 1981-86. *APMIS* (1992) 100(31). Acta pathologica, microbiologica et immunologica scandinavica. Supplementum. p. 194. Available at: https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&lang=en&idt=730737.

5. Engeland A, Haldorsen T, Tulinius H, Vaittinen P, Tretli S, Hakulinen T. Prediction of cancer incidence in the Nordic cancer up to the years 2000 and 2010. *APMIS* (1993) 101(38):1–124. Acta pathologica, microbiologica et immunologica scandinavica. Supplementum. Available at: https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&lang=en&idt=589545.

6. Møller B, Fekjær H, Hakulinen T, Tryggvadóttir L, Storm HH, Talback M, et al. Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Cancer Prev* (2002) 11. Available at: https://www.researchgate.net/profile/Hans-Storm/publication/11025416_Prediction_of_cancer_incidence_in_the_Nordic_countries_up_to_the_year_2020/links/0046351d66737db360000000/Prediction-of-cancer-incidence-in-the-Nordic-countries-up-to-the-year-2020.pdf.

7. Storm HH. Nordic cancer registration, a review of an invaluable source and example for surveillance, research and public health for more than 70 years. *Norsk Epidemiol* (2022) 30(1-2):25–30. doi: 10.5324/nje.v30i1-2.4973

8. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *Eur J Cancer* (2009) 45 (5):747–55. doi: 10.1016/j.ejca.2008.11.032

9. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. *Eur J Cancer* (2009) 45(5):756–64. doi: 10.1016/j.ejca.2008.11.033

10. Larønningen S, Ferlay J, Beydogan H, Bray F, Engholm G, Ervik M, et al. *NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 9.2*. Oslo, Norway: Association of the Nordic Cancer Registries. Cancer Registry of Norway (2022).

11. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint Å, et al. NORDCAN–a Nordic tool for cancer information, planning, quality control and research. *Acta Oncol* (2010) 49(5):725–36. doi: 10.3109/02841861003782017

12. Larønningen S, Larsen IK, Møller B, Engholm G, Storm HH, Johannesen TB. *NORDCAN - Cancer data from the Nordic countries*. Oslo: Cancer Registry of Norway (2013).

13. Pukkala E, Engholm G, Højsgaard Schmidt LK, Storm H, Khan S, Lambe M, et al. Nordic Cancer Registries–an overview of their procedures and data comparability. *Acta Oncol* (2018) 57(4):440–55. doi: 10.1080/0284186X.2017.1407039

14. European Union. *General Data Protection Regulation - GDPR* (2018). Available at: https://gdpr-info.eu/.

15. The R Foundation. *The R Project for Statistical Computing* (2022). Available at: https://www.r-project.org/.

16. StataCorp. *Stata* (2022). Available at: https://www.stata.com/.

17. International Agency for Research on Cancer (IARC). *Check and conversion Program* (2018). Available at: http://www.iacr.com.fr/index.php?option=com_content&view=article&id=72:iarccrgtools&catid=68&Itemid=445.

18. World Health Organization. *International Classification of Diseases for Oncology. 3rd ed*. Malta: WHO Press, World Health Organization (2013). First Revision (ICD-O-3).

19. World Health Organization. *International Classification of Diseases* (2019). Available at: https://icd.who.int/browse10/2019/en.

20. International Agency for Research on Cancer (IARC), World Health Organization (WHO) and International Association of Cancer Registries (IACR) European Network of Cancer Registries (ENCR). *International rules for multiple primary cancers. (ICD-O Third Edition)*. Lyon: IARC (2004). Available at: http://www.iacr.com.fr/images/doc/MPrules_july2004.pdf.

21. NORDCAN work group. *Call for data - Incidence* (2022). Available at: https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/Call-for-data—Incidence.

22. NORDCAN work group. *Call for data - Mortality* (2022). Available at: https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/Call-for-data—Mortality.

23. NORDCAN work group. *Call for data - Population* (2022). Available at: https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/Call-for-data-Population.

24. NORDCAN work group. *Call for data - Survival* (2022). Available at: https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/Call-for-data-Survival.

25. NORDCAN work group. *NORDCAN - The cancer dictionary* (2022). Available at: https://nordcan.iarc.fr/en/database#bloc1.

26. NORDCAN work group. *The Cancer Registry of Norway metadatabase: Entity* (2022). Available at: https://metadata.kreftregisteret.no/variables/detail/814?tabIndex=4.

27. NORDCAN work group. *ICD10 - ICD6-7/ICD8/ICD9*. The Cancer Registry of Norway metadatabase (2022). Available at: https://metadata.kreftregisteret.no/tables/view/7.

28. NORDCAN work group. *nordcan.R output* (2022). Available at: https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/nordcan.R-output.

29. NORDCAN work group. *Nordcan survival methods* (2022). Available at: https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/nordcansurvival.

30. Coviello E, Dickman P, Seppa K, Pokhrel A. *STNET: Stata module to calculate net survival*. Boston College Department of Economics: EconPapers (2020). Available at: https://econpapers.repec.org/software/bocbocode/S457533.htm.

31. NORDCAN work group. *Protecting confidentiality in NORDCAN* (2021). Available at: https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/Protecting-confidentiality-in-NORDCAN.

32. Hundepool A, Domingo-Ferrer J, Franconi L, Giessing S, Nordholt ES, Spicer K, et al. *Statistical disclosure control*. Wiley New York (2012), vol 2.

33. *Policy on protecting confidentiality in tables of birth and death statistics*. United Kingdom: Office for National Statistics (2018). Available at: https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/policyonprotectingconfidentialityintablesofbirthanddeathstatistics.

34. The Norwegian Data Protection Authority. *A guide to the anonymisation of personal data*. Oslo: The Norwegian Data Protection Authority (2015).

35. NORDCAN work group. *Glossary of statistical terms* (2022). Available at: https://nordcan.iarc.fr/en/additional-information?tab=2.

36. Ervik M, et al. *Global Cancer Observatory: Cancer Over Time* (2021). Available at: https://gco.iarc.fr/overtime/en (Accessed 2022 20.09).

37. Netherlands Cancer Registry and Netherlands Comprehensive Cancer Organisation (IKNL). *NKR cijfers* (2022). Available at: https://iknl.nl/nkr-cijfers (Accessed 2022 10.10).

38. Cancer Registry of Norway. *STINE - Cancer Registry of Norway databank* (2022). Available at: https://sb.kreftregisteret.no/.

39. International Agency for Research on Cancer (IARC). *Cancer Incidence in Five Continents - Volume XII* (2022). Available at: http://www.iacr.com.fr/index.php?option=com_content&view=article&id=90&Itemid=566.

40. European Network of Cancer Registries. *Call for data* (2022). Available at: https://www.encr.eu/call-for-data.

41. NORDCAN work group. *Quality tables* (2022). Available at: https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/Specification-quality-tables.

42. European Health Data & Evidence Network (EHDEN). *EHDEN: Becoming the trusted open science community built with standardised health data via a European federated network*. (2022). Available at: ehden.eu/consortium-partners/.

43. Miettinen J, Aagnes B, Pejcic S, Tian H. *nordcan.R*. (2022). Available at: https://github.com/CancerRegistryOfNorway/NORDCAN/tree/master/releases.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 22 August 2023
DOI 10.3389/fonc.2023.1232451

OPEN ACCESS

EDITED BY
Otto Visser,
Netherlands Comprehensive Cancer Organisation (IKNL), Netherlands

REVIEWED BY
Wafaa M. Rashed,
Ahram Canadian University, Egypt
Raoull Hoogendijk,
Princess Maxima Center for Pediatric Oncology, Netherlands

*CORRESPONDENCE
Angela Lopez-Cortes
angela.lopez@ucl.ac.uk;
benchista@istitutotumori.mi.it

†These authors share first authorship

‡These authors share last authorship

RECEIVED 31 May 2023
ACCEPTED 24 July 2023
PUBLISHED 22 August 2023

CITATION
Lopez-Cortes A, Didonè F, Botta L, Hjalgrim LL, Jakab Z, Cañete Nieto A, Stiller C, Zeller B, Gatta G, Pritchard-Jones K and The BENCHISTA Project Working Group (2023) Cancer data quality and harmonization in Europe: the experience of the BENCHISTA Project – international benchmarking of childhood cancer survival by stage. *Front. Oncol.* 13:1232451. doi: 10.3389/fonc.2023.1232451

Cancer data quality and harmonization in Europe: the experience of the BENCHISTA Project – international benchmarking of childhood cancer survival by stage

Angela Lopez-Cortes[1*†], Fabio Didonè[2†], Laura Botta[2†], Lisa L. Hjalgrim[3], Zsuzsanna Jakab[3], Adela Cañete Nieto[3], Charles Stiller[3], Bernward Zeller[3], Gemma Gatta[2‡], Kathy Pritchard-Jones[1‡] and The BENCHISTA Project Working Group

[1]University College London (UCL) Great Ormond Street Institute of Child Health, Developmental Biology & Cancer Research Department, London, United Kingdom, [2]Fondazione IRCCS "Istituto Nazionale dei Tumori di Milano" (INT), Department of Evaluative Epidemiology, Milan, Italy, [3]BENCHISTA Project Management Team, London, United Kingdom

Introduction: Variation in stage at diagnosis of childhood cancers (CC) may explain differences in survival rates observed across geographical regions. The BENCHISTA project aims to understand these differences and to encourage the application of the Toronto Staging Guidelines (TG) by Population-Based Cancer Registries (PBCRs) to the most common solid paediatric cancers.

Methods: PBCRs within and outside Europe were invited to participate and identify all cases of Neuroblastoma, Wilms Tumour, Medulloblastoma, Ewing Sarcoma, Rhabdomyosarcoma and Osteosarcoma diagnosed in a consecutive three-year period (2014-2017) and apply TG at diagnosis. Other non-stage prognostic factors, treatment, progression/recurrence, and cause of death information were collected as optional variables. A minimum of three-year follow-up was required. To standardise TG application by PBCRs, on-line workshops led by six tumour-specific clinical experts were held. To understand the role of data availability and quality, a survey focused on data collection/sharing processes and a quality assurance exercise were generated. To support data harmonization and query resolution a dedicated email and a question-and-answers bank were created.

Results: 67 PBCRs from 28 countries participated and provided a maximally de-personalized, patient-level dataset. For 26 PBCRs, data format and ethical approval obtained by the two sponsoring institutions (UCL and INT) was sufficient for data sharing. 41 participating PBCRs required a Data Transfer Agreement (DTA) to comply with data protection regulations. Due to heterogeneity found in legal aspects, 18 months were spent on finalizing the DTA. The data collection survey was answered by 68 respondents from 63

PBCRs; 44% of them confirmed the ability to re-consult a clinician in cases where stage ascertainment was difficult/uncertain. Of the total participating PBCRs, 75% completed the staging quality assurance exercise, with a median correct answer proportion of 92% [range: 70% (rhabdomyosarcoma) to 100% (Wilms tumour)].

Conclusion: Differences in interpretation and processes required to harmonize general data protection regulations across countries were encountered causing delays in data transfer. Despite challenges, the BENCHISTA Project has established a large collaboration between PBCRs and clinicians to collect detailed and standardised TG at a population-level enhancing the understanding of the reasons for variation in overall survival rates for CC, stimulate research and improve national/regional child health plans.

KEYWORDS

childhood cancer, population-based, cancer registry, Toronto staging, diagnosis, survival, data quality, data harmonization

1 Introduction

According to the International Agency for Research of Cancer (IARC) and estimates from 2020, nearly 280.000 children and teenagers (0-19 years old) were diagnosed with cancer around the world and almost 110.000 died of this cause (1). When considering estimates of total childhood cancer incidence accounting for underdiagnosis, a simulation-based analysis found that there were 397.000 incident cases of childhood cancer for 200 territories worldwide and 43% of these were undiagnosed with substantial variation by region (range:3%-57%). Furthermore, considering population projections for 2015-2030 it is estimated there will be 6.7 million cases of CC worldwide, from which 2.9 million of cases will be missed (2). In addition to these estimates, and due to delay in diagnosis, variation in treatment and rates of relapse, paediatric oncology patients in low-and middle-income countries (LMICs) are five times as likely to die from a cancer diagnosis compared with patients in high-income countries (HICs) (3).

Population-based cancer registries (PBCRs) are key organizations that generate estimates of incidence and survival essential for cancer research (4). When considering paediatric patients, data completeness and accuracy represent a challenge due to the rarity and heterogeneity of childhood cancer. It has also been noted that stage data is not consistently recorded for paediatric patients. The tumour/node/metastasis (TNM) system is the standard staging system for most adult cancers; however, it is inappropriate for documenting the extent of disease in most childhood cancers (5).

Disease-specific staging systems have been developed for childhood cancers within the context of broader risk-stratification schemes used by various clinical trial groups. This means that for many diagnostic groups, two or more systems are in clinical use and there was no international standard suitable for global use by population-based cancer registries. Thus, in 2014 and through a collaborative effort between epidemiologists, clinical trial groups and registration experts, a consensus definition of tumour stage was agreed for most childhood cancer types - the Toronto Paediatric Cancer Stage Guidelines (TG) for population cancer registries (4, 6). These are endorsed by the European Network of Cancer Registries (ENCR), the Group for Cancer Epidemiology and Registration in Latin Language Countries (GRELL) the African Network of Cancer Registries (ANCR) and published in the UICC TNM Classification of Malignant Tumours 8th Edition (5, 7).

Different childhood cancer population-based studies have demonstrated survival disparities between countries and European regions. Several factors may explain this variation including late diagnosis, delayed treatment, variation in quality of diagnostic and treatment services, management of acute complications, lack of resources, limited access to health services, abandonment to treatment, among others (8–10). Further understanding of the international variation in childhood cancer survival may be explained by the distribution of stage at diagnosis and stage-specific survival (4, 5, 11–13).

Other research studies have assessed the feasibility and validity of the TG demonstrating that PBCRs can reconstruct stage according to TG (11). This standardised framework supports PBCRs to assign cancer stage using data that can be found routinely in clinical records for most childhood cancers. The success of the pilot study emphasised the importance of a larger number of cancer registries in different countries applying the TG so that the paediatric staging system can be further improved (11, 14).

The International Benchmarking of Childhood Cancer Survival by Stage, also called BENCHISTA Project, is a research collaboration between multiple population-based cancer registries from European and non-European countries. It aims to stimulate the application of Toronto Stage Guidelines by participating PBCRs for six of the most common paediatric solid cancers (15) to lead to a better understanding of the reasons for variation in childhood cancer survival between countries and to highlight areas for

improvement. The research sponsors are University College London (UCL) and the Fondazione IRCCS Istituto Nazionale dei Tumori di Milano (INT).

Due to the large number of participating PBCRs, data quality, harmonization and standardization are essential. The aim of this paper is to present the resources that the project established to ensure high-quality, standardised data, comparable across participating PBCR. The resources used have provided understanding on current procedures at cancer registry level and highlighted strengths and limitations when gathering stage at diagnosis, other prognostic, and non-stage prognostic factors to understand childhood cancer survival and its variation.

2 Materials and methods

All European population-based cancer registries (PBCRs) included in the EUROCARE studies were invited to participate in the BENCHISTA Project. Additionally, other non-European PBCRs from Australia, Canada, Brazil, and Japan confirmed their contribution to the project. A great number of PBCRs are checked for quality indicators by the International Agency for Research on Cancer (IARC) based on four dimensions of quality: comparability, validity, timeliness, and completeness (16, 17). Not all PBCRs have a government mandate. Some are coordinated by the National Society for Paediatric Haematology-Oncology and/or register all cases diagnosed at all hospitals authorised for childhood cancer treatment in the relevant country, with the aim of achieving population coverage.

Participating PBCRs were required to assign stage at diagnosis at a population-level using the Internationally recognized Toronto Stage Guidelines (TG) to six paediatric solid tumours (Ewing sarcoma, osteosarcoma, rhabdomyosarcoma – for ages 0 to 19 years old – and neuroblastoma, medulloblastoma and Wilms Tumour – for ages 0 to 14 years old) diagnosed in a three-year period (in the window 2014-2017). The selection of these tumour diagnoses was based on several factors, including their unambiguous diagnosis by histological code, previous studies showing geographical disparities in outcomes, limited improvements in survival rates over a prolonged period, and their significant representation among childhood solid tumours (4, 9, 10). For the three sarcomas, we included cases in the adolescent range (15 to 19 years old) as many of the participating registries collected data in this age range, where bone and soft tissues sarcomas peak in incidence. The process of determining the stage of diagnosis was performed by cancer registrars or relevant staff, who use various available data sources, such as clinical records, histopathology and imaging reports, and other administrative files. Clinical personnel could also be consulted in cases where uncertain or inconclusive information was encountered.

The staging classification used the TG to enable registries to derive the best estimate of stage at diagnosis in a standardised fashion. The guidelines endorse a two-tier approach, Tier 1 focuses on registries with limited resources and/or restricted data access and requires less detailed criteria and stage categories; Tier 2 involves more detailed criteria for cancer registries with further access to medical information or well-resourced (4, 5, 11). All Tier 2 can be converted into Tier 1. For the BENCHISTA Project, TG is defined as extent of disease at the time of diagnosis and based on detailed evidence before receiving treatment with two exceptions: staging of localized (non-metastatic) Wilms Tumour after neoadjuvant chemotherapy, since stage is based on surgical and histopathological examination of the nephrectomy specimen; and tumours in which investigations to exclude distant metastases may be performed shortly after surgery to the primary tumour but before systemic therapy is commenced.

The variables collected included depersonalized patient demographic data plus information on clinical investigations and types of data sources used by the registrars for applying Toronto staging for each of the six solid tumours (e.g., imaging/examinations performed and their results, when available). PBCRs were requested to use the International Classification of Childhood Cancers (ICCC-3) and to assign tumour stage according to the Toronto consensus staging guidelines. Moreover, to review all available data sources and to seek advice from clinicians when further explanation or clarification was required to ensure consistency in data collection and accuracy. Follow-up for life status was requested for a minimum of three years from diagnosis.

The BENCHISTA Project also assessed the availability to PBCRs of optional but clinically relevant variables for understanding any variation in treatment and survival for the six solid tumours. These optional variables included the more recently agreed 'Toronto non-stage prognostic factors' (NSP), and primary treatment modalities, relapse/recurrence/progression, and cause of death (6). To avoid limitations due to language barriers, the TG provided detailed guidance (5) translated in different languages (Italian, Spanish, Japanese, French, Bulgarian and Portuguese) and an electronic tool available to facilitate its use by different audiences (18, 19).

Data gathered from each participating PBCR were merged in a maximally anonymized dataset created by and stored within the secure environment of the data controller at INT. Comparative analysis of distribution of tumour stage at diagnosis at a population-level and analysis of survival estimates by stage for each tumour type between large geographical regions with similar groupings to previous EUROCARE studies is in progress. Validation is being conducted by the project analytical team to verify the coverage, number of submitted cases and national/local reported incidence.

Several factors were considered for standardization and harmonization parameters. Data files were checked with *ad hoc* developed procedures in regular use by the data controller (INT). Likewise, the validity of each variable and variable combinations for each tumour record were checked to detect unlikely or incorrect values. Records that were flagged during the data checking process were sent to the registries for revision and amendment. Furthermore, cases ascertained only by death certificate (DCO), number of cases diagnosed by cytology and those with unspecified morphology codes (NOS) were considered as data quality indicators for the completeness and accuracy of population-level data. Additional assessments to define the accuracy of sub-typing definitions in the six solid tumours of interest were also conducted.

To support the TG staging by PBCRs and ensure standardized processes, a series of three online training workshops led by clinical experts in the six solid tumours of interest and generated in collaboration with the Belgian Cancer registry were held. Moreover, to understand the modalities of data collection and staging processes in each PBCR a survey was designed to verify local/national processes and understand current practices and the possibility to seek advice from clinicians when clarification is required. This survey was addressed to registrars, clinical and

non-clinical staff gathering data and completing stage at diagnosis for the BENCHISTA Project.

To assess data comparability, a quality assurance tool including a set of twelve fictitious cases (two per each tumour type of interest) was developed and completed by selected representatives of each participating cancer registry. Both surveys were developed using the platform SurveyMonkey® and the results were gathered and analysed by the team at INT.

2.1 Project's governance

The Project Management Team (PMT) comprises the two principal investigators in the UK and Italy, and four representatives from participating cancer registry staff (Norway, Denmark, Spain and Hungary). The Project Working Group (PWG) involves one or two representatives from each contributing cancer registry, six tumour-specific oncology experts nominated by the relevant European clinical study group, representatives from parent/survivor groups and communication and dissemination partners. The BENCHISTA team, PMT and PWG meet regularly to review the project's advances and overview preliminary results helping to ensure a broader assessment of tasks and upcoming plans to guarantee the achievement of the project's goals.

Moreover, the project has established an Independent Advisory Board (IAB) that includes a cancer registry director not directly involved in the day-to-day project, parent and survivor representatives, clinical executive level members of a national paediatric oncology society, a clinical trial study group and a medical director-level clinician involved in organisation of childhood cancer services. Importantly, there is also representation by patient/public involvement and engagement (PPIE) structures to ensure the perspective of parents and survivors is included in the different stages of the project's development and upcoming results.

3 Results

3.1 PBCR participation and database status

80 PBCRs within and outside of Europe were interested to participate in the project at outset, however 16% of them could not participate due to different reasons, including limited or no availability of population-based data, anticipated restrictions in sharing patient-level datasets beyond national boundaries and limited access to clinical data to apply TG. 67 PBCRs from 24 European countries, Australia, Brazil, Japan, and Canada committed to participate in the project, which commenced in January 2021 (Figure 1). This process entailed close work between the data controller, cancer registry leaders and in some cases legal representatives from the PBCRs to achieve research collaboration.

Seeking for standardised TG collection by PBCRs, three on-line workshops were held in October-November 2021. A total of 60 PBCRs, both within and outside Europe, actively participated in real-time; each session attracted an attendance ranging from 70 to 80 individuals and was centred in two specific tumour types. The sessions covered various topics, including the fundamental principles of Toronto Staging, introducing and discussing clinical aspects, diagnosis, therapy, and non-stage prognostic factors for each tumour type and exemplar staging exercises based on pre-created cases. The training workshops were recorded and are publicly available on the BENCHISTA Project website, together with other supporting materials.

The content and format of compulsory and optional data variables were agreed by the Project Working Group members to ensure almost complete anonymisation whilst retaining patient-level information (Table 1). Although the process to agree the content of the required datafile submitted by each registry was finalised by March 2021, heterogeneity in the approach and legal requirements from participating countries led to a lengthy process to finalise the format and content of the data sharing agreement and hence delays in data submission.

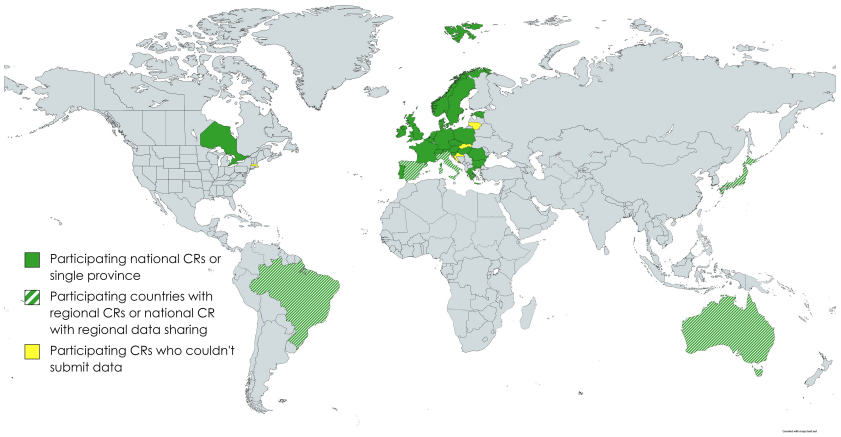


FIGURE 1
Participating countries.

10.3389/fonc.2023.1232451

TABLE 1 Variables and dataset structure.

Variable	No. of characters	Notes and encoding
Basic variables		
Registry	10	alphabetic
Registry Patient Identification code	10	assigned by the registry, it is a project-specific pseudonymised code
Year of birth	4	yyyy
Age at diagnosis	3	Numeric (in months)
Year of diagnosis	4	yyyy
Sex	1	boy/girl/unknown 1/2/9
Base of diagnosis (as coded in the ENCR protocol)	1	DCO/Clinical/Clinical investigation/Specific tumour markers/Cytology/Histology of a metastasis/ Histology of a primary tumour/Unknown 0/1/2/4/5/6/7/9
ICDO-3-Topography	3	Only the numeric part of the ICD-O-3 topography code will be reported (the "C" and "." will not be included)
ICDO-3-Morphology	4	Malignant, only, behaviour=3
First previous cancer	1	Y/N/unknown 1/0/9
First previous cancer definition		International Classification of Childhood Cancers (ICCC) 3rd edition
Year of diagnosis of the first previous cancer	4	yyyy/9
Second previous cancer	1	Y/N/unknown 1/0/9
Second previous cancer definition		International Classification of Childhood Cancers (ICCC) 3rd edition
Year of diagnosis of the second previous cancer	4	yyyy/9
Imaging/examination used for staging before any treatment		
CT/MRI primary site	1	Y/N/unknown 1/0/9
MRI whole neuraxis	1	Y/N/unknown 1/0/9
MRI whole neuraxis outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9
CT thorax	1	Y/N/unknown 1/0/9
CT thorax outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9
Imaging of regional lymph nodes	1	Y/N/unknown 1/0/9
Imaging of regional lymph nodes outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9
CSF	1	Y/N/unknown 1/0/9
CSF outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9
MIBG scan	1	Y/N/unknown 1/0/9
MIBG scan outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9
Abdominal ultrasound	1	Y/N/unknown 1/0/9
Abdominal ultrasound outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9
Bone scan	1	Y/N/unknown 1/0/9
Bone scan outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9
Bone marrow aspirate or biopsy	1	Y/N/unknown 1/0/9
Bone marrow aspirate or biopsy outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9
X-Ray thorax	1	Y/N/unknown 1/0/9
X-Ray thorax outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9

(Continued)

TABLE 1 Continued

Variable	No. of characters	Notes and encoding
PET	1	Y/N/unknown 1/0/9
PET outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9
Tissue biopsy	1	Y/N/unknown 1/0/9
Tissue biopsy outcome		Negative/Positive/Suspicious/Unknown 0/1/2/9
Source used for staging		
Clinical report (hospital clinical records)	1	Y/N/unknown 1/0/9
Pathological report	1	Y/N/unknown 1/0/9
Administrative files (hospital discharge, etc.)	1	Y/N/unknown 1/0/9
Clinical study group	1	Y/N/unknown 1/0/9
Others (string)	10	alphabetic
Toronto staging, Neuroblastoma		
Stage Tier 1	2	L/LR/M/MS/X 1/2/3/4/9
Stage Tier 2	2	L1/L2/M/MS/X 1/2/3/4/9
Laterality	1	Not applicable/Right/Left/Unilateral NOS/Bilateral//unknown 0/1/2/3/4/9
* NSP: N-Myc	1	Amplified Y/N (exact definitions to be discussed)
Toronto staging, Wilms tumour		
Stage Tier 1 after pre-surgery chemotherapy	1	L/M/X 1/2/9
Stage Tier 2 after pre-surgery chemotherapy	1	y-I/y-II/y-III/IV/9 1/2/3/4/9
Stage Tier 1 after immediate surgery (i.e., surgery first)	1	L/M/X 1/2/9
Stage Tier 2 after immediate surgery	1	I/II/III/IV/X 1/2/3/4/9
Laterality	1	R/L/B 1/2/3
O_NSP: Wilms Presence of anaplasia	1	No/Yes, but unknown if focal or diffuse/Yes, focal/Yes, diffuse/Anaplasia unknown 0/1/2/3/9
Toronto staging, Medulloblastoma		
Stage Tier 1	1	L/M/X 1/2/9
Stage Tier 2	2	M0/M1/M2/M3/M4/X 0/1/2/3/4/9
*Evaluation of postoperative residual disease		R0/R1/R2/R+/unknown 0/1/2/3/9
*_NSP: Wingless (WNT) medulloblastoma	1	Y/N/unknown 1/0/9
*_NSP: Sonic Hedgehog (SHH) medulloblastoma	1	Y/N/unknown 1/0/9
Toronto staging, Osteosarcoma, Ewing sarcoma		
Stage Tier 1	1	L/M/X 1/2/9
Stage Tier 2	1	L/M/X 1/2/9
Toronto staging, Rhabdomyosarcoma		
Stage Tier 1	1	L/M/X 1/2/9
Stage Tier 2	1	I/II/III/IV/X 1/2/3/4/9
*_NSP: FKR-PAX3 rhabdomyosarcoma	1	Y/N/unknown 1/0/9

(Continued)

TABLE 1 Continued

Variable	No. of characters	Notes and encoding
*_NSP : FKR-PAX7 rhabdomyosarcoma	1	Y/N/unknown 1/0/9
Primary Treatment defined as given within 1 year from diagnosis		
*_Surgery	1	Y/N/unknown 1/0/9
*_Chemotherapy	1	Y/N/unknown 1/0/9
*_Chemotherapy type	1	Preoperative chemo/Postoperative chemo/Both, preoperative and postoperative chemo/Chemotherapy only/Unknown 1/2/3/4/9
*_Radiotherapy	1	Y/N/unknown 1/0/9
*_Relapse/recurrence/progression		
*_Relapse/recurrence/progression	1	Y/N/unknown 1/0/9
*_Time in days from diagnosis to relapse/ recurrence/progression		numeric
Follow-up		
Status of life alive/dead	1	alive/dead/unknown 1/2/9
*_Causes of death (CoD)	1	Toxicity of treatment, Tumour, Comorbidity previously present in the child, Others, unknown 1/2/3/4/9
Time in days from diagnosis to death or last follow up		numeric

*Optional Variables.

Data started to flow to the data controller from March 2022 and up to now 58 databases including nearly 11000 cases have been successfully submitted and are under final stages of quality assessment. Specific queries on data or further requests to ensure high-quality are discussed among the Project Management Team and the PBCR if required. Model answers to each query received were collated and published on the project website as a series of Frequently Asked Questions.

3.2 Data privacy process and challenges

While data sharing and data transfer involve movement of data from one institution to other, there are several differences between these two concepts. Data sharing involves making data available more broadly, enabling reuse and access in ways that allow control and management from one or several parties. Data transfer involves moving specific data from one entity to another with a purpose and typically to solve a specific research question; it tends to be more targeted and involves providing the information to another entity for analysis, storage but without giving full control over data itself. Both, data sharing, and data transfer are subjected to legal and ethical considerations that need to be considered to comply with General Data Protection Regulation principles including purpose minimization, lawfulness of processing, accuracy, storage limitation and accountability (20).

In compliance with institutional and legal requirements, the project was granted with ethical approvals from UCL and INT. Minor amendments to the protocol and appendix were submitted and approved to ensure information is clear and adequate. Individual parent consent is not required in general as the information is collected under existing permissions for cancer registration in each jurisdiction.

Each participating cancer registry was approached to understand their individual requirements to proceed with data sharing. For 26 PBCRs, the project-specific ethical approval obtained by the research sponsors (UCL and INT) was sufficient to confirm their participation and submit data for analysis. The rest of the PBCRs required a specific legal document that allowed collaboration and further research.

Considering requirements and aims of the project, a Data Transfer Agreement (DTA) was developed by the legal officer and data controller's team at INT. It contains general information about the project and legal considerations to share patient-level data in a highly de-personalised format. Its aim is to meet specific legislations to adhere to General Data Protection Regulations (GDPR) and relevant laws from each participating country; in total, 44 PBCRs required the DTA and 3 signed it in a second phase. Discussions between legal officers from PBCRs and the BENCHISTA team regarding the acceptable wording of the DTA continued over a period of 18 months before finalisation of the DTA, leading to a delay in the project's timeline. For one participating centre a country-specific transparency statement was generated and made publicly available in the project's website. Another participating centre required the Data Protection Impact Assessment (DPIA) in a format that complied with their specific requirements. The DPIA, evaluates the impact of the processing activity generated within the BENCHISTA Project focusing on the rights and freedoms of the data subjects. The outcome of this assessment was categorized as data processing with a low risk level.

Noticeable differences in interpretation of legislations, laws and required processes were encountered whilst confirming the requirements for participation by each PBCR. For some PBCRs, interpretation of their national laws meant they were not allowed to share patient-level data. These registries reluctantly dropped out of the project as the anticipated work to change this interpretation for sharing this standard dataset was felt to be too complex or not possible. The requirement for local ethical approval, above and beyond sharing of the sponsors' ethical approval documents, also varied between countries and sometimes between regions within the same country.

After multiple interactions from legal representatives, the DTA was fully executed on 14th November 2022. After final signatures, other PBCRs expressed their interest in participating in the project and submitting data and some others that did not need the signature in the first place but then required it. For these cases the 'accession document' included in the Appendix of the DTA was created by the legal officer at INT. This latter document did not permit any further changes to the wording of the DTA.

3.3 Survey on data collection/data sharing processes

The BENCHISTA Project conducted a survey among all participating PBCRs to gather information on available data sources, as well as approaches used for data collection and interpretation. The survey received responses from 63 out of the 67 cancer registries (94%) involved in the project, representing 31 countries (Table 2 and Figure 2).

Among the respondents, 83% participated in the online training workshops. Additionally, 49 out of 63 respondents (78%) confirmed their ability to collect external data if necessary. In cases where stage ascertainment diagnosis posed difficulties or uncertainties, 28 out of 63 respondents (44%) reported the ability to seek consultation with a clinician. In addition, 20 out of 63 registries (32%) stated that re-consultation was only possible under specific circumstances, such as the availability of clinicians, limited access to clinical records, or depending on the anatomical location of the tumour. For 15 out of 63 registries (24%), clinical re-consultation was not available.

Additionally, 44% of respondents reported having access to individual-patient imaging results for staging purposes, while 13% did not have access to such resources. 43% indicated that they had some access, but not for all cases.

3.4 Data standardization and quality assurance

To maximize the efficiency of cancer registry staff time, the project's quality assurance tool was limited to twelve fictitious cases (two for each tumour type). These cases were designed by the project leaders, discussed among the members of the Project Management Team and the relevant tumour-specific clinical experts, and piloted with a limited number of PBCR staff who had been involved in the design of the training workshops. The cases were then refined and revised for clarity, readability and correctness.

All participating PBCRs were invited to stage these 12 fictitious cases blind to any answers by others. The Project Working Group agreed in advance that a concordance rate of 90% or greater should be aimed for to demonstrate sufficient standardisation.

Of the total participating PBCRs (24 out of 29 countries) 75% completed this exercise. The correct answer proportion for the Toronto stage ranged from 70% (rhabdomyosarcoma, Tier 2) to 100% (Wilms tumour), with a median score of 92% (Figure 3).

The average correct score varied across registries, ranging from 67% to 100%. The score for rhabdomyosarcoma was lower due to limited correctness in one of the fictitious case exercises (50% correct answers) where assignment of the 'paranasal' sinus anatomical site to either favourable or unfavourable category led to a change in the assigned Toronto stage from I to III. For neuroblastoma, the discrepancies were mainly related to variable interpretation of 'image-defined' risk factors. For medulloblastoma, there was variable interpretation of whether Tier 2 staging could be applied in the absence of a cerebrospinal fluid (CSF) cytology result. These specific cases highlighted the importance of training and clear definitions in support information such as the CanStaging+ Tool (18, 19) when staging complex cases. The 'artificial' nature of the fictitious cases also contributed to discrepancies as registry staff would have access to multiple data sources for cross-verification and to advice from senior colleagues for 'real' clinical cases.

As of May 2023, out of the total expected cases, ~98% or 10,504 cases were collected. Information on the stage is complete for 94% of the cases at Tier 1 level and 88% at Tier 2 level for all six tumours combined. Regarding optional variables, completeness is 73% for relapse or progression, while treatment variables such as surgery, chemotherapy, and radiotherapy have a completeness of 83%, 86%, and 81% respectively. NSP have a completeness of 44%. Additionally, cause of death data was reported in 69% of cases.

4 Discussion

The BENCHISTA Project has enabled PBCRs from different geographical areas within and beyond Europe to share patient-level data to better understand the factors underlying variation in childhood cancer survival rates. Participation in the project has stimulated their efforts to access the data required to apply the Toronto Guidelines to stage their cases in a standardized way that allows international comparisons. In addition, the feasibility to collect other non-stage prognostic factors and summary information on treatments given, relapse and cause of death has been demonstrated.

The project has focused on achieving consistent participation and compilation of information in line with local or national laws despite legal heterogeneity that led to delays in finalising requirements such as the DTA impacting the project's timeline. Despite this, the BENCHISTA Project has achieved participation from across most of Europe and with several key international partners to compile detailed information on nearly 11,000 cases of six childhood solid tumours diagnosed in a recent period at a population-level. This is a tremendous 'proof of principle' to

10.3389/fonc.2023.1232451

TABLE 2 Answers to Survey on data collection per country.

Participating Cancer Registry	Online training session attendance	External Data Collection	Possibility to re-consult a clinician	Imaging access for staging
Australia	NO	NO	NO	YES
Belgium	YES	YES	YES	SOMETIMES
Brazil	YES	YES	SOMETIMES	SOMETIMES
Bulgaria	YES	YES	YES	YES
Ontario	NO	YES	YES	YES
Croatia	YES	YES	SOMETIMES	SOMETIMES
Czech Republic (Hospital Brno)	YES	NO	YES	YES
Czech Republic	YES	YES	YES	YES
Denmark	YES	YES	YES	YES
England	YES	YES	NO	SOMETIMES
Estonia	YES	YES	YES	SOMETIMES
France	YES	NO	YES	YES
Germany	YES	NO	SOMETIMES	NO
Greece	YES	YES	SOMETIMES	SOMETIMES
Hungary	YES	YES	YES	SOMETIMES
Ireland	YES	YES	NO	YES
Osaka	YES	YES	YES	NO
Malta	NO	YES	SOMETIMES	YES
The Netherlands	NO	NO	YES	NO
Northern Ireland	YES	YES	SOMETIMES	YES
Norway	YES	YES	SOMETIMES	SOMETIMES
Poland	YES	YES	YES	SOMETIMES
Portugal	YES	YES	YES	YES
Romania	YES	YES	YES	SOMETIMES
Scotland	YES	YES	YES	YES
Slovakia	NO	YES	YES	SOMETIMES
Slovenia	YES	YES	SOMETIMES	YES
Sweden	YES	YES	YES	SOMETIMES
Switzerland	YES	YES	SOMETIMES	SOMETIMES
Wales	YES	YES	YES	YES

catalyse continued outcomes research that uses routine healthcare data available to PBCRs. The project has revealed aspects of data access and staging definitions that require further attention if we are to achieve the ultimate aim of truly harmonized data to ensure reliable estimates and survival comparisons.

Several challenges were observed during the DTA generation and sign off by PBCRs. Initially differences in legislation and laws on data sharing/transfer and processing were encountered. Multiple interactions across legal and cancer registry staff were required to reach a consensus on the requirements for data transfer. Additional limiting factors included, different understanding and application of the terms "anonymous" versus "maximally de-personalised" in relation to general data protection regulations, introduction of new data systems that led to delays in finalising cancer incidence data, limited access to clinical data sources and advice for applying TGs, limited registry workforce capacity for data collection/staging tasks in specific timelines, administrative changes or high turn-over leading to delays. Some of these difficulties were more noticeable in countries with regional rather than national coverage, where the interpretation of General Data Protection Regulations varied between regions.

Nevertheless, data available to PBCRs has generally increased over time, with direct data feeds from clinical reporting

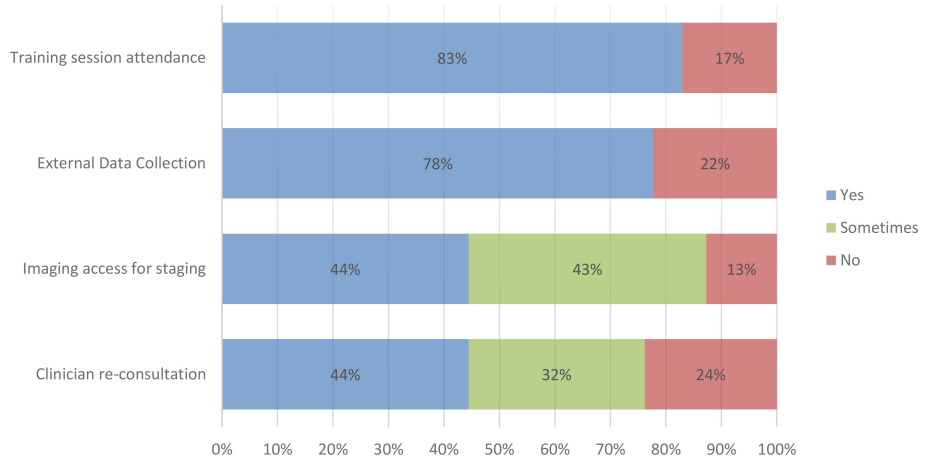


FIGURE 2
Results of survey on data collection/sharing processes.

(histopathology, imaging, treatments etc) permissible in many jurisdictions. In spite of this, access to the detailed information and clinical support required to apply Toronto staging at Tier 2, together with non-stage prognostic variables, remains very variable.

Considering the results from the survey on data collection, there is noticeable variation in the access to data sources at an individual patient-level and available clinical support. This may impact the application of TG in some cases and highlights the importance of understanding current modalities for data collection and standardised parameters for staging by PBCRs.

The quality assurance tool aimed to assess how standardised the collection of TG is across the PBCRs. This exercise highlighted challenges related to differences in terminology, risk group definitions and access to required clinical information to complete TG accurately. Some examples include the availability of CSF cytology results for medulloblastoma, image-defined risk factors for neuroblastoma and differences in interpretation/classification of favourable and unfavourable anatomical sites for rhabdomyosarcoma. These were discussed among the Independent Advisory Board and other members of the project leading to further conversations with TG leaders, clinicians, cancer registry staff and researchers to improve the understanding of key clinical parameters to improve staging and therefore healthcare data research. Recommendations from this project's experience are already being considered for inclusion in the next revision of the CanStaging+ Tool (5, 18).

Additionally, a key point discussed among the project's team focused on the importance of standardization of the definition of *metastasis*, particularly in relation to lung nodules (for Wilms Tumour and all three sarcomas) and how it requires further attention. PBCR staff rely on the interpretation provided in the imaging reports or by the clinician providing the stage information. These are inherently variable (in their definitions of metastasis) and could benefit from a move towards standardised structured reporting.

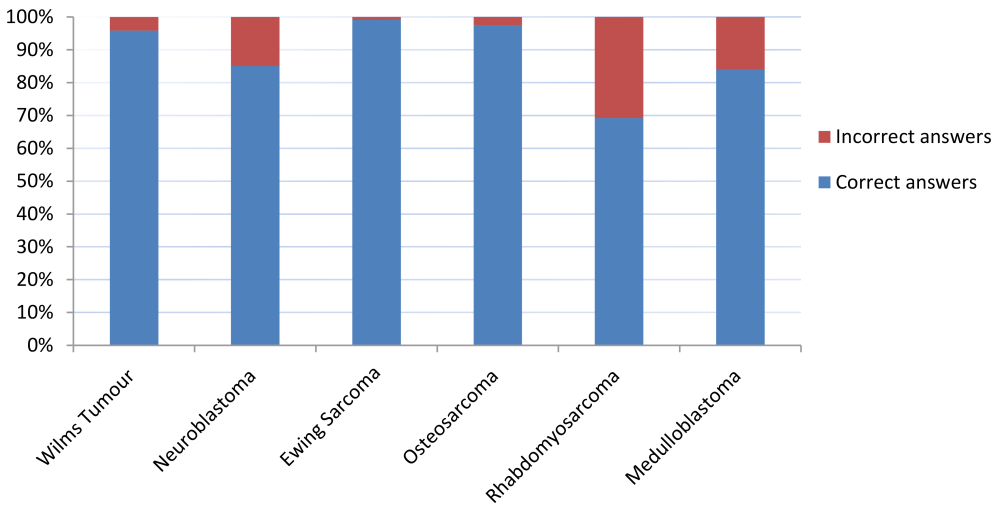


FIGURE 3
Results of quality assurance tool.

Considering previous challenges and to enhance communication channels, project-specific resources were created to ensure consistent interaction with PBCRs; these include newsletters, social media and web platforms, on-line meetings and active participation on scientific conferences. However, this might not be sustainable beyond research-funded activities. In spite of this, the project has provided valuable insights to demonstrate the viability of the general approach of PBCRs collecting and sharing patient-level routine health care data. This approach not only paves the way for continuous benchmarking of stage distribution at diagnosis and survival by stage, but also serves as the foundation for population-level outcomes research in cancers with different prognosis. For example, in tumour-types and subgroups with overall survival rates in excess of 90%, implementation and funding of prospective clinical trials is increasingly challenging. Hence, the ability to design prospective studies that can use the capabilities of PBCRs to collect additional non-stage prognostic variables, offers an efficient mechanism to monitor population-level survival rates for clinically defined subgroups for whom there are no open interventional clinical trials but who are treated according to nationally agreed clinical practice guidelines.

The BENCHISTA Project represents an opportunity to understand reasons for international variation in overall survival for childhood cancer at a population-level by enhancing the collaboration with PBCRs and stimulating their ability to use TG in childhood cancer cases in a standardized way. Data harmonization also requires strengthened relationships with clinicians, medical sources, the European Network of Cancer Registries and other stakeholders to ensure cancer data recorded in registries are high-quality, comprehensive and accurate empowering the PBCRs to routinely collect TG for future benchmarking research leading to outcome improvement.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

In compliance with institutional and legal requirements, the project was granted with ethical approvals from UCL and INT.

Author contributions

KP-J is the principal investigator of the BENCHISTA Project, and GG is a co-investigator, both with equal contribution and senior authorship. AL-C, LB and FD have equal contribution as joint first authors and contributed to writing – original draft and editing. FD and LB contributed also to formal analysis. KP-J, GG, AL-C, LB, and FD contributed also to conceptualization, project administration and methodology. The Project Management Team (PMT) members supported in the discussion of the methods and interpretation of the results. The Project Working Group (PWG) contributed to data collection, interpretation and discussion of the results. All authors approved the submitted version.

BENCHISTA Project Working Group

For a full list of the international members of the BENCHISTA Project Working Group, see Appendix 3.

Funding

The BENCHISTA Project is funded by Children with Cancer UK (Grant Reference: 20-329) and the Associazione Italiana per la Ricerca sul Cancro (AIRC) Italy (Grant reference: IG 2020 - ID. 24933).

Acknowledgments

The data used in this project have been generated during the treatment and care of patients. The BENCHISTA Project is sponsored by UCL and INT. We would like to thank the clinical experts (Dr Sandra Strauss, Dr Natalie Gaspar, Dr Adela Canete, Prof Simon Bailey, Dr Filippo Spreafico and Dr Lisa L Hjalgrim), the Belgian Cancer Registry and Meric Klein for the leadership and support provided during the on-line training sessions. We highly appreciate the support provided by cancer registry leaders and legal parties ensuring the data transfer is compliant and working along with the data controller legal officer to finalise the DTA.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. IARC. *International Childhood Cancer Day* (2022). Available at: https://www.iarc.who.int/featured-news/iccd-2022/.

2. Ward ZJ, Yeh JM, Bhakta N, Frazier AL, Atun R. Estimating the total incidence of global childhood cancer: a simulation-based analysis. *Lancet Oncol* (2019) 20(4):483–93. doi: 10.1016/S1470-2045(18)30909-4

3. Lam CG, Howard SC, Bouffet E, Pritchard-Jones K. Science and health for all children with cancer. *Science* (2019) 363(6432):1182–6. doi: 10.1126/science.aaw4892

4. Gupta S, Aitken JF, Bartels U, Brierley J, Dolendo M, Friedrich P, et al. Paediatric cancer stage in population-based cancer registries: the Toronto consensus principles and guidelines. *Lancet Oncol* (2016) 17(4):e163–72. doi: 10.1016/S1470-2045(15)00539-2

5. Aitken JF, Youlden D, O'Neill L, Gupta S, Frazier AL. *Childhood cancer staging for population registries according to the Toronto Childhood Cancer Stage Guidelines – Version 2*. Brisbane A: Cancer Council Queensland and Cancer Australia (2021).

6. Gupta S, Aitken J, Bartels U, Bhakta N, Bucurenci M, Brierley JD, et al. Development of paediatric non-stage prognosticator guidelines for population-based cancer registries and updates to the 2014 Toronto Paediatric Cancer Stage Guidelines. *Lancet Oncol* (2020) 21(9):e444–51. doi: 10.1016/S1470-2045(20)30320-X

7. Aitken JF, Youlden DR, Moore AS, Baade PD, Ward LJ, Thursfield VJ, et al. Assessing the feasibility and validity of the Toronto Childhood Cancer Stage Guidelines: a population-based registry study. *Lancet Child Adolesc Health* (2018) 2(3):173–9. doi: 10.1016/S2352-4642(18)30023-3

8. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* (2018) 391(10125):1023–75. doi: 10.1016/S0140-6736(17)33326-3

9. Botta L, Gatta G, Capocaccia R, Stiller C, Cañete A, Dal Maso L, et al. Long-term survival and cure fraction estimates for childhood cancer in Europe (EUROCARE-6): results from a population-based study. *Lancet Oncol* (2022) 23(12):1525–36. doi: 10.1016/S1470-2045(22)00637-4

10. Gatta G, Botta L, Rossi S, Aareleid T, Bielska-Lasota M, Clavel J, et al. Childhood cancer survival in Europe 1999-2007: results of EUROCARE-5–a population-based study. *Lancet Oncol* (2014) 15(1):35–47. doi: 10.1016/S1470-2045(13)70548-5

11. Gatta G, Botta L, Capocaccia R, Cañete A, Pritchard-Jones K. Staging childhood cancers in Europe: Application of the Toronto stage principles for neuroblastoma and Wilms tumour. The JARC pilot study. *Pediatr Blood Cancer* (2021) 68(9):e29020. doi: 10.1002/pbc.29020

12. Pritchard-Jones K, Graf N, van Tinteren H, Craft A. Evidence for a delay in diagnosis of Wilms' tumour in the UK compared with Germany: implications for primary care for children. *Arch Dis Child* (2016) 101(5):417–20. doi: 10.1136/archdischild-2015-309212

13. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. *CA Cancer J Clin* (2014) 64(2):83–103. doi: 10.3322/caac.21219

14. Sacerdote C, Mosso ML, Alessi D, Merletti F, Tagliabue G, D'Agostino A, et al. An application of the Toronto Childhood Cancer Stage Guidelines in three population-based cancer registries: The case of central nervous tumors. *Pediatr Blood Cancer* (2020) 67(6):e28303. doi: 10.1002/pbc.28303

15. Botta L, Gatta G, Didonè F, Lopez Cortes A, Pritchard-Jones K. International benchmarking of childhood cancer survival by stage at diagnosis: The BENCHISTA project protocol. *PloS One* (2022) 17(11):e0276997. doi: 10.1371/journal.pone.0276997

16. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *Eur J Cancer* (2009) 45(5):747–55. doi: 10.1016/j.ejca.2008.11.032

17. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. *Completeness Eur J Cancer* (2009) 45(5):756–64. doi: 10.1016/j.ejca.2008.11.033

18. *CanStaging*. Available at: https://www.canstaging.org/.

19. Soerjomataram I, Ervik M, Fox C, Hawkins S, Yeung K, Napolitano G, et al. CanStaging+: an electronic staging tool for population-based cancer registries. *Lancet Oncol* (2021) 22(8):1069. doi: 10.1016/S1470-2045(21)00188-1

20. EU. *General Data Protection Regulation*. Available at: https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504 [accessed Feb 2023].

Appendix 3 The BENCHISTA Project Working Group:

Australia: Australian Childhood Cancer Registry: Joanne Aitken, Leisa O'Neill;

Austria: Austrian Cancer Registry: Monika Hackl, Ruth Ladenstein;

Belgium: Belgian Cancer Registry: Elizabeth Van Eycken, Nancy Van Damme;

Boston, MA (USA): Dana-Farber Cancer Institute: Lindsay Frazier;

Brasil: Cancer Registry Representatives: Beatriz De Camargo, Marceli de Oliveira Santos;

Bulgaria: Bulgarian Cancer Registry: Zdravka Valerianova, Dobrin Konstantinov;

Canada: OOntario Children Cancer Registry_POGO: Sumit Gupta, Jason D Pole;

Croatia: Croatian Cancer Registry: Mario Sekerija;

Czech Republic: Czech National Cancer Registry: Jan Stary, Jaroslav Sterba;

Denmark: Danish Childhood Cancer Registry and Department of Pediatric Oncology: Lisa L. Hjalgrim (PMT Member); Childhood Cancer Research Group, Danish Cancer Institute, Copenhagen, Denmark & Department of Clinical Medicine, Faculty of Health, Aarhus University and University Hospital, Aarhus, Denmark: Jeanette Falck Winther;

Estonia: Estonia National Institute for Health Development: Keiu Paapsi;

France: French National Registry of Childhood Cancer - *Solid tumours:* Brigitte Lacour, Emmanuel Desandes; *Hematopoietic Malignancies:* Jacqueline Clavel, Claire Poulalhon;

Germany: German Childhood Cancer Registry, Mainz: Claudia Spix, Meike Ressing;

Greece: Greek Nationwide Registry for Childhood Hematological Malignancies and Solid Tumours (NARECHEMST): Eleni T Petridou, Evdoxia Bouka;

Hungary: Hungarian Child Cancer Registry: Zsuzsanna Jakab (PMT Member), Miklos Garami;

Italy:

Basilicata Cancer Registry: Rocco Galasso;

Bergamo Cancer Registry: Giuseppe Sampietro;

Campania Childhood Cancer Registry: Francesco Vetrano, Marcella Sessa;

Childhood Cancer Registry of Piedmont: Milena M Maule, Carlotta Sacerdote;

Cremona & Mantova Cancer Registry: Paola Ballotari;

Friuli Venezia Giulia Cancer Registry, CRO Aviano National Cancer Institute: Emilia De Santis;

Integrated Cancer Registry CT-ME-EN: Margherita Ferrante, Rosalia Ragusa;

Innovation, Research and Teaching Service (SABES-ASDAA), Teaching Hospi tal of the Paracelsus Medizinischen Privatuniversität, Bolzano-Bozen, Italy: Michael Mian;

Liguria Cancer Registry, Ospedale Policlinico San Martino IRCCS: Luca Boni;

Monza-Brianza Cancer Registry: Magda Rognoni;

Palermo Province Cancer Registry: Rosalba Amodio;

Pavia Cancer Registry: Lorenza Boschetti;

Puglia Cancer Registry: Francesco Cuccaro, Danila Bruno;

Registro tumori ATS della Città metropolitana di Milano: Antonio G Russo, Federico Gervasi;

Registro Tumori ATS Insubria: Maria L Gambino;

Registro tumori dell'Emilia-Romagna, Unità di Piacenza: Elisabetta Borciani;

Registro tumori dell'Emilia-Romagna, Unità di Parma: Maria Michiara;

Registro tumori dell'Emilia-Romagna, Unità di Reggio Emilia: Lucia Mangone;

Registro tumori dell'Emilia-Romagna, Unità di Modena: Gianbattista Spagnoli;

Registro tumori dell'Emilia-Romagna, Unità di Ferrara: Stefano Ferretti;

Registro tumori dell'Emilia-Romagna, Unità della Romagna, IRCCS IRST Meldola: Fabio Falcini;

Registro Tumori di Ragusa e Caltanissetta: Eugenia Spata;

Registro Tumori Regione Marche: Sonia Manasse, Paola Coccia;

Siracusa Province Cancer Registry: Francesca Bella;

Toscana Cancer Registry: Adele Caldarella, Teresa Intrieri;

Trapani Cancer Registry: Tiziana Scuderi;

Trento Cancer Registry (Trento CR), Servizio Epidemiologia Clinica e Valutativa, APSS Trento: Roberto V Rizzello;

Veneto Cancer Registry, Regional Epidemiological Service, Azienda Zero, Padova, Italy: Massimo Rugge, Stefano Guzzinati;

Ireland: Ireland Cancer Registry: Deirdre Murray;

Japan: NNational Cancer Center: Tomohiro Matsuda; Osaka CR: Kayo Nakata;

Malta: Malta National Cancer Registry, Health Information and Research: Miriam J Azzopardi;

Norway: Norwegian Cancer Registry: Tom Børge Johannesen, Aina H Dahlen; Bernward Zeller (PMT Member);

Poland: Polish Childhood Cancer Clinical Database provided by the Polish Society of Paediatric Oncology and Haematology (PSPOH) - Prof Jerzy Kowalczyk, Medical University of Lublin and Prof Anna Raciborska, Department of Oncology and Surgical Oncology for Children and Youth, Institute of Mother and Child.

Portugal: Portuguese Pediatric Cancer Registry: Ana M Ferreira, Gabriela Caldas;

Romania: Romanian Child Cancer Registry: Mihaela Bucurenci, Dana Coza;

Slovenia: Cancer Registry of Republic of Slovenia: Vesna Zadnik;

Spain:

Basque Country, Euskadi-CIBERESP Cancer Registry: Arantza Lopez de Munain;

Childhood and Adolescents CR - CISCV: Fernando Almela-Vich, Nieves Fuster-Camarena;

Girona CR, CIBERESP, ICO, IDIBGI: Rafael Marcos-Gragera;

Granada CR, EASP, CIBERESP, ibs.GRANADA, UGR: Maria José Sanchez;

Madrid Childhood Cancer Registry: Aragonés, Raquel López;

Murcia Cancer Registry, CIBERESP, IMIB-Arrixaca: Maria Dolores Chirlaque;

Navarra Cancer Registry, ISPLN, CIBERESP, IdiSNA: Marcela Guevara; Peris-Bonet, Adela Cañete Nieto (PMT Member);

Tarragona Cancer Registry, HUSJR, IISPV: Marià Carulla;

Sweden: Swedish Childhood Cancer Registry – SCCR, Karolinska Institute, Stockholm, Sweden: Päivi Lähteenmäki;

Switzerland: Childhood Cancer Registry, Institute of Social and Preventive Medicine, University of Bern, Switzerland: Claudia E Kuehni, Shelagh M Redmond;

The Netherlands: The Netherlands Cancer Registry: Otto Visser, Henrike Karim-Kos;

United Kingdom:

England: National Disease Registration Service, Transformation Directorate, NHS England: Lucy Irvine, Paul Stacey, Charles Stiller (PMT Member);

Northern Ireland Cancer Registry : Anna Gavin, Deirdre Fitzpatrick;

Scotland: Scottish Cancer Registry: David S Morrison; Karen Smith;

Welsh Cancer Intelligence and Surveillance Unit, Public Health Wales: Dyfed Wyn Huws, Janet Warlow;

Clinical Lead Experts:

Ewing Sarcoma: Sandra Strauss - University College London Hospital (UCLH), London, England, UK.

Medulloblastoma: Simon Bailey - Great North Children's Hospital, Newcastle upon Tyne, England, UK.

Neuroblastoma: Adela Cañete Nieto - Hospital UiP La Fe, Paediatric Oncology and Hematology Unit, Valencia, Spain.

Osteosarcoma: Nathalie Gaspar - Gustave Roussy Cancer Campus, Villejuif, France.

Rhabdomyosarcoma: Lisa L. Hjalgrim - University of Copenhagen, Rigshospitalet, Department of Pediatrics and Adolescent Medicine, Copenhagen, Denmark.

Wilms Tumour: Filippo Spreafico - IRCCS National Cancer Institute, Milan, Italy.

Angela Polanco (Patient and Public Involvement and Engagement – PPIE Representative);

Riccardo Capocaccia (Epidemiologia & Prevenzione, Italy);

Andrea Di Cataldo (Università degli Studi di Catania, Italy);

Meric Klein (Belgian Cancer Registry – Training Workshops Lead).

frontiers | Frontiers in Oncology

TYPE Correction
PUBLISHED 28 March 2024
DOI 10.3389/fonc.2024.1397101

OPEN ACCESS

EDITED AND REVIEWED BY
Otto Visser,
Netherlands Comprehensive Cancer Organisation (IKNL), Netherlands

*CORRESPONDENCE
Angela Lopez-Cortes
angela.lopez@ucl.ac.uk;
benchista@istitutotumori.mi.it

†These authors share first authorship

‡These authors share last authorship

RECEIVED 06 March 2024
ACCEPTED 18 March 2024
PUBLISHED 28 March 2024

CITATION
Lopez-Cortes A, Didonè F, Botta L, Hjalgrim LL, Jakab Z, Cañete Nieto A, Stiller C, Zeller B, Gatta G, Pritchard-Jones K and The BENCHISTA Project Working Group (2024) Corrigendum: Cancer data quality and harmonization in Europe: The experience of the BENCHISTA Project – International Benchmarking of Childhood Cancer Survival by Stage.
Front. Oncol. 14:1397101.
doi: 10.3389/fonc.2024.1397101

Corrigendum: Cancer data quality and harmonization in Europe: The experience of the BENCHISTA Project – International Benchmarking of Childhood Cancer Survival by Stage

Angela Lopez-Cortes[1*†], Fabio Didonè[2†], Laura Botta[2†], Lisa L. Hjalgrim[3], Zsuzsanna Jakab[3], Adela Cañete Nieto[3], Charles Stiller[3], Bernward Zeller[3], Gemma Gatta[2‡], Kathy Pritchard-Jones[1‡] and The BENCHISTA Project Working Group

[1]University College London (UCL) Great Ormond Street Institute of Child Health, Developmental Biology & Cancer Research Department, London, United Kingdom, [2]Fondazione IRCCS "Istituto Nazionale dei Tumori di Milano" (INT), Department of Evaluative Epidemiology, Milan, Italy, [3]BENCHISTA Project Management Team, London, United Kingdom

KEYWORDS

childhood cancer, population-based, cancer registry, Toronto staging, diagnosis, survival, data quality, data harmonization

A Corrigendum on

Cancer data quality and harmonization in Europe: the experience of the BENCHISTA Project – international benchmarking of childhood cancer survival by stage

By Lopez-Cortes A, Didonè F, Botta L, Hjalgrim LL, Jakab Z, Cañete Nieto A, Stiller C, Zeller B, Gatta G, Pritchard-Jones K and The BENCHISTA Project Working Group (2023). *Front. Oncol.* 13:1232451. doi: 10.3389/fonc.2023.1232451

In the published article, there was an error in Supplementary **Appendix (Number 3)**. The Project Working Group was displayed according to initial known information. A concern related to the Polish representative affiliation was raised after submission. After internal review and discussion with the relevant involved parties, an agreement on the amendment to the name/affiliation of the Polish representative and clarification to the source of data collection was reached. Furthermore, other changes to the Appendix 3 were incorporated as these were communicated by other representatives from the Project Working Group after the article deadline/submission.

The appendix in the published article was displayed as:

Appendix 3

The BENCHISTA Project Working GroupŁ:

Australia: Australian Cancer Registry (CR): Joanne Aitken, Leisa O'Neill;

Austria: Austrian CR: Monika Hackl, Ruth Ladenstein;

Belgium: Belgian CR: Elizabeth Van Eycken, Nancy Van Damme;

Boston, MA (USA): Dana-Farber Cancer Institute: Lindsay Frazier;

Brasil: CR Representatives: Beatriz De Camargo, Marceli de Oliveira Santos;

Bulgaria: Bulgarian CR: Zdravka Valerianova, Dobrin Konstantinov;

Canada: Ontario Children CR_POGO: Sumit Gupta, Jason D Pole;

Croatia: Croatian CR: Mario Sekerija;

Czech Republic: Czech National CR: Jan Stary, Jaroslav Sterba;

Denmark: Danish Childhood Cancer Registry and Department of Pediatric Oncology: Lisa L. Hjalgrim (PMT Member); Danish Cancer Society: Jeanette F Winther;

Estonia: Estonia National Institute for Health Development: Keiu Paapsi;

France: French National Registry of Childhood Cancer - Solid tumours: Brigitte Lacour, Emmanuel Desandes; Hematopoietic Malignancies: Jacqueline Clavel, Claire Poulalhon;

Germany: German Childhood CR, Mainz: Friederike Erdmann, Claudia Spix;

Greece: Greek Nationwide Registry for Childhood Hematological Malignancies and Solid Tumours (NARECHEM-ST): Eleni T Petridou, Evdoxia Bouka;

Hungary: Hungarian Child CR: Zsuzsanna Jakab (PMT Member);

Italy:

Alto Adige CR: Michael Mian;

Basilicata CR: Rocco Galasso;

Bergamo CR: Giuseppe Sampietro;

Campania Childhood CR: Francesco Vetrano, Marcella Sessa;

Childhood Cancer Registry of Piedmont: Milena M Maule, Carlotta Sacerdote;

Cremona & Mantova CR: Paola Ballotari;

Friuli Venezia Giulia CR, CRO Aviano National Cancer Institute: Luigino Dal Maso; Integrated Cancer Registry CT-ME-EN: Margherita Ferrante, Rosalia Ragusa;

Liguria CR, Ospedale Policlinico San Martino IRCCS: Luca Boni;

Monza-Brianza CR: Magda Rognoni;

Palermo CR: Rosalba Amodio;

Pavia CR: Lorenza Boschetti;

Puglia CR: Francesco Cuccaro, Danila Bruno;

Registro tumori ATS della Città metropolitana di Milano: Antonio G Russo, Federico Gervasi; Registro Tumori ATS Insubria: Maria L Gambino;

Registro tumori dell'Emilia-Romagna, Unità di Piacenza: Elisabetta Borciani;

Registro tumori dell'Emilia-Romagna, Unità di Parma: Maria Michiara;

Registro tumori dell'Emilia-Romagna, Unità di Reggio Emilia: Lucia Mangone;

Registro tumori dell'Emilia-Romagna, Unità di Modena: Gianbattista Spagnoli;

Registro tumori dell'Emilia-Romagna, Unità di Ferrara: Stefano Ferretti;

Registro tumori dell'Emilia-Romagna, Unità della Romagna, IRCCS IRST Meldola: Fabio Falcini;

Registro Tumori di Ragusa e Caltanissetta: Eugenia Spata;

Registro Tumori Regione Marche: Sonia Manasse, Paola Coccia;

Siracusa Province CR: Francesca Bella;

Toscana CR: Adele Caldarella, Teresa Intrieri;

Trapani CR: Tiziana Scuderi;

Trento Cancer Registry (Trento CR), Servizio Epidemiologia Clinica e Valutativa, APSS Trento: Roberto V Rizzello;

Veneto CR: Massimo Rugge, Stefano Guzzinati;

Ireland: Ireland CR: Deirdre Murray;

Japan: National Cancer Center: Tomohiro Matsuda; Osaka CR: Kayo Nakata;

Malta: Malta National Cancer Registry, Health Information and Research: Miriam J Azzopardi;

Norway: Norwegian CR: Tom Børge Johannesen, Aina H Dahlen; Bernward Zeller (PMT Member);

Poland: Polish Childhood Cancer Registry, Medical University of Lublin: Jerzy Kowalczyk; Polish Institute of Mother and Child: Monika Jedrzejczyk;

Portugal: Portuguese Pediatric CR: Ana M Ferreira, Gabriela Caldas;

Romania: Romanian Child CR: Mihaela Bucurenci, Dana Coza;

Slovenia: Cancer Registry of Republic of Slovenia: Vesna Zadnik;

Spain:

Basque Country, Euskadi-CIBERESP CR: Arantza Lopez de Munain;

Childhood and Adolescents CR - CISCV: Fernando Almela-Vich, Nieves Fuster-Camarena; Girona CR, CIBERESP, ICO, IDIBGI: Rafael Marcos-Gragera;

Granada CR, EASP, CIBERESP, ibs.GRANADA, UGR: Maria José Sanchez;

Madrid Childhood CR: Nuria Aragones, Raquel Lopez;

Murcia CR, CIBERESP, IMIB-Arrixaca: Maria Dolores Chirlaque;

Navarra CR, CIBERESP, IdiSNA: Marcela Guevara Eslava;

Spain RETI-SEHOP: Elena Pardo, Rafael Peris-Bonet, Adela Cañete Nieto (PMT Member); Tarragona CR: Jaume Galceran;

Sweden: Swedish Childhood Cancer Registry - SCCR: Päivi Lähteenmäki;

Switzerland: Childhood Switzerland CR: Claudia E Kuehni, Shelagh M Redmond;

The **Netherlands:** The Netherlands CR: Otto Visser, Henrike Karim-Kos;

United Kingdom:

England: National Disease Registration Service, Transformation Directorate, NHS England: Lucy Irvine, Paul Stacey, Charles Stiller (PMT Member);

Northern Ireland CR: Anna Gavin, Deirdre Fitzpatrick;

Scotland: Scottish CR: David S Morrison; Karen Smith;

Welsh Cancer Intelligence and Surveillance Unit - WCISU: Dyfed Wyn Huws, Janet Warlow;

Clinical Lead Experts:

Ewing Sarcoma: Sandra Strauss - University College London Hospital (UCLH), London, England, UK.

Medulloblastoma: Simon Bailey - Great North Childrens Hospital, Newcastle upon Tyne, England, UK.

Neuroblastoma: Adela Canete Nieto -Hospital UiPLaFe, Paediatric Oncology and Hematology Unit, Valencia, Spain.

Osteosarcoma: Nathalie Gaspar - Gustave Roussy Cancer Campus, Villejuif, France. Rhabdomyosarcoma: Lisa L. Hjalgrim - University of Copenhagen, Rigshospitalet, Department of Pediatrics and Adolescent Medicine, Copenhagen, Denmark.

Wilms Tumour: Filippo Spreafico - IRCCS National Cancer Institute, Milan, Italy.

Patient and Public Involvement and Engagement – PPIE representative: Angela Polanco;

Epidemiologia & Prevenzione, Italy: Riccardo Capocaccia;

Università degli Studi di Catania, Italy: Andrea Di Cataldo;

Belgian Cancer Registry: Meric Klein.

The correct material statement appears below:

Appendix 3

The BENCHISTA Project Working Groupł

Australia: Australian Childhood Cancer Registry: Joanne Aitken, Leisa O'Neill;

Austria: Austrian Cancer Registry: Monika Hackl, Ruth Ladenstein;

Belgium: Belgian Cancer Registry: Elizabeth Van Eycken, Nancy Van Damme;

Boston, MA (USA): Dana-Farber Cancer Institute: Lindsay Frazier;

Brasil: Cancer Registry Representatives: Beatriz De Camargo, Marceli de Oliveira Santos;

Bulgaria: Bulgarian Cancer Registry: Zdravka Valerianova, Dobrin Konstantinov;

Canada: Ontario Children Cancer Registry _ POGO: Sumit Gupta, Jason D Pole;

Croatia: Croatian Cancer Registry: Mario Sekerija;

Czech Republic: Czech National Cancer Registry: Jan Stary, Jaroslav Sterba;

Denmark: Danish Childhood Cancer Registry and Department of Pediatric Oncology: Lisa L. Hjalgrim (PMT Member); Childhood Cancer Research Group, Danish Cancer Institute, Copenhagen, Denmark & Department of Clinical Medicine, Faculty of Health, Aarhus University and University Hospital, Aarhus, Denmark: Jeanette Falck Winther;

Estonia: Estonia National Institute for Health Development: Keiu Paapsi;

France: French National Registry of Childhood Cancer - *Solid tumours:* Brigitte Lacour, Emmanuel Desandes; *Hematopoietic Malignancies:* Jacqueline Clavel, Claire Poulalhon;

Germany: German Childhood Cancer Registry, Mainz: Claudia Spix, Meike Ressing;

Greece: Greek Nationwide Registry for Childhood Hematological Malignancies and Solid Tumours (NARECHEM-ST): Eleni T Petridou, Evdoxia Bouka;

Hungary: Hungarian Child Cancer Registry: Zsuzsanna Jakab (PMT Member), Miklos Garami;

Italy:

Basilicata Cancer Registry: Rocco Galasso;

Bergamo Cancer Registry: Giuseppe Sampietro;

Campania Childhood Cancer Registry: Francesco Vetrano, Marcella Sessa;

Childhood Cancer Registry of Piedmont: Milena M Maule, Carlotta Sacerdote;

Cremona & Mantova Cancer Registry: Paola Ballotari;

Friuli Venezia Giulia Cancer Registry, CRO Aviano National Cancer Institute: Emilia De Santis;

Integrated Cancer Registry CT-ME-EN: Margherita Ferrante, Rosalia Ragusa;

Innovation, Research and Teaching Service (SABES-ASDAA), Teaching Hospital of the Paracelsus Medizinischen Privatuniversität, Bolzano-Bozen, Italy: Michael Mian;

Liguria Cancer Registry, Ospedale Policlinico San Martino IRCCS: Luca Boni;

Monza-Brianza Cancer Registry: Magda Rognoni;

Palermo Province Cancer Registry: Rosalba Amodio;

Pavia Cancer Registry: Lorenza Boschetti;

Puglia Cancer Registry: Francesco Cuccaro, Danila Bruno;

Registro tumori ATS della Città metropolitana di Milano: Antonio G Russo, Federico Gervasi;

Registro Tumori ATS Insubria: Maria L Gambino;

Registro tumori dell'Emilia-Romagna, Unità di Piacenza: Elisabetta Borciani;

Registro tumori dell'Emilia-Romagna, Unità di Parma: Maria Michiara;

Registro tumori dell'Emilia-Romagna, Unità di Reggio Emilia: Lucia Mangone;

Registro tumori dell'Emilia-Romagna, Unità di Modena: Gianbattista Spagnoli;

Registro tumori dell'Emilia-Romagna, Unità di Ferrara: Stefano Ferretti;

Registro tumori dell'Emilia-Romagna, Unità della Romagna, IRCCS IRST Meldola: Fabio Falcini;

Registro Tumori di Ragusa e Caltanissetta: Eugenia Spata;

Registro Tumori Regione Marche: Sonia Manasse, Paola Coccia;

Siracusa Province Cancer Registry: Francesca Bella;

Toscana Cancer Registry: Adele Caldarella, Teresa Intrieri;

Trapani Cancer Registry: Tiziana Scuderi;

Trento Cancer Registry (Trento CR), Servizio Epidemiologia Clinica e Valutativa, APSS Trento: Roberto V Rizzello;

Veneto Cancer Registry, Regional Epidemiological Service, Azienda Zero, Padova, Italy: Massimo Rugge, Stefano Guzzinati;

Ireland: Ireland Cancer Registry: Deirdre Murray;

Japan: National Cancer Center: Tomohiro Matsuda; Osaka CR: Kayo Nakata;

Malta: Malta National Cancer Registry, Health Information and Research: Miriam J Azzopardi;

Norway: Norwegian Cancer Registry: Tom Børge Johannesen, Aina H Dahlen; Bernward Zeller (PMT Member);

Poland: Polish Childhood Cancer Clinical Database provided by the Polish Society of Paediatric Oncology and Haematology

(PSPOH) - Prof Jerzy Kowalczyk, Medical University of Lublin and Prof Anna Raciborska, Department of Oncology and Surgical Oncology for Children and Youth, Institute of Mother and Child.

Portugal: Portuguese Pediatric Cancer Registry: Ana M Ferreira, Gabriela Caldas;

Romania: Romanian Child Cancer Registry: Mihaela Bucurenci, Dana Coza;

Slovenia: Cancer Registry of Republic of Slovenia: Vesna Zadnik;

Spain:

Basque Country, Euskadi-CIBERESP Cancer Registry: Arantza Lopez de Munain;

Childhood and Adolescents Cancer Registry - CISCV: Fernando Almela-Vich, Nieves Fuster-Camarena;

Girona Cancer Registry, CIBERESP, ICO, IDIBGI: Rafael Marcos-Gragera;

Granada Cancer Registry, EASP, CIBERESP, ibs.GRANADA, UGR: Maria José Sanchez;

Madrid Childhood Cancer Registry: Nuria Aragonés, Raquel López;

Murcia Cancer Registry, CIBERESP, IMIB-Arrixaca: Maria Dolores Chirlaque;

Navarra Cancer Registry, ISPLN, CIBERESP, IdiSNA: Marcela Guevara;

Spanish Registry of Childhood Tumours (RETI SEHOP), University of Valencia, Faculty of Medicine: Elena Pardo, Rafael Peris-Bonet, Adela Cañete Nieto (PMT Member);

Tarragona Cancer Registry, HUSJR, IISPV: Marià Carulla;

Sweden: Swedish Childhood Cancer Registry – SCCR, Karolinska Institute, Stockholm, Sweden: Päivi Lähteenmäki;

Switzerland: Childhood Cancer Registry, Institute of Social and Preventive Medicine, University of Bern, Switzerland: Claudia E Kuehni, Shelagh M Redmond;

The Netherlands: The Netherlands Cancer Registry: Otto Visser, Henrike Karim-Kos;

United Kingdom:

England: National Disease Registration Service, Transformation Directorate, NHS England: Lucy Irvine, Paul Stacey, Charles Stiller (PMT Member);

Northern Ireland Cancer Registry: Anna Gavin, Deirdre Fitzpatrick;

Scotland: Scottish Cancer Registry: David S Morrison; Karen Smith;

Welsh Cancer Intelligence and Surveillance Unit, Public Health Wales: Dyfed Wyn Huws, Janet Warlow;

Clinical Lead Experts:

Ewing Sarcoma: Sandra Strauss - University College London Hospital (UCLH), London, England, UK.

Medulloblastoma: Simon Bailey - Great North Children's Hospital, Newcastle upon Tyne, England, UK.

Neuroblastoma: Adela Cañete Nieto - Hospital UiP La Fe, Paediatric Oncology and Hematology Unit, Valencia, Spain.

Osteosarcoma: Nathalie Gaspar - Gustave Roussy Cancer Campus, Villejuif, France.

Rhabdomyosarcoma: Lisa L. Hjalgrim - University of Copenhagen, Rigshospitalet, Department of Pediatrics and Adolescent Medicine, Copenhagen, Denmark.

Wilms Tumour: Filippo Spreafico - IRCCS National Cancer Institute, Milan, Italy.

Angela Polanco (Patient and Public Involvement and Engagement – PPIE Representative);

Riccardo Capocaccia (Epidemiologia & Prevenzione, Italy);

Andrea Di Cataldo (Università degli Studi di Catania, Italy);

Meric Klein (Belgian Cancer Registry – Training Workshops Lead).

In the published article, the first paragraph is correct. A phrase is required to be added in the first paragraph of the material and methods section to clarify that not all of the population-based cancer registries that contributed data to the project are constituted in the same way.

A correction has been made to **Materials and Methods**, at end of first paragraph.

This sentence previously stated:

"All European population-based cancer registries (PBCRs) included in the EUROCARE studies were invited to participate in the BENCHISTA Project. Additionally other non-European PBCRs from Australia, Canada, Brazil, and Japan confirmed their contribution to the project. A great number of PBCRs are checked for quality indicators by the International Agency for Research on Cancer (IARC) based in four dimensions of quality: comparability, validity, timeliness, and completeness (16, 17)."

The corrected sentence appears below:

"All European population-based cancer registries (PBCRs) included in the EUROCARE studies were invited to participate in the BENCHISTA Project. Additionally, other non-European PBCRs from Australia, Canada, Brazil, and Japan confirmed their contribution to the project. A great number of PBCRs are checked for quality indicators by the International Agency for Research on Cancer (IARC) based on four dimensions of quality: comparability, validity, timeliness, and completeness (16, 17). Not all PBCRs have a government mandate. Some are coordinated by the National Society for Paediatric Haematology-Oncology and/or register all cases diagnosed at all hospitals authorised for childhood cancer treatment in the relevant country, with the aim of achieving population coverage."

The authors apologize for these errors and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 30 August 2023
DOI 10.3389/fonc.2023.1201464

OPEN ACCESS

EDITED BY
Paulo S. Pinheiro,
University of Miami, United States

REVIEWED BY
Jean-Michel Billette,
Statistics Canada, Canada
Theresa Devasia,
National Cancer Institute (NIH),
United States

*CORRESPONDENCE
Silvia Francisci
silvia.francisci@iss.it
Luigino Dal Maso
epidemiology@cro.it

RECEIVED 06 April 2023
ACCEPTED 07 August 2023
PUBLISHED 30 August 2023

CITATION
Francisci S, Tursini F, Dal Maso L, Gigli A and Guzzinati S (2023) Projecting cancer prevalence by phase of care: a methodological approach for health service planning. *Front. Oncol.* 13:1201464. doi: 10.3389/fonc.2023.1201464

Projecting cancer prevalence by phase of care: a methodological approach for health service planning

Silvia Francisci [1*], Francesco Tursini [2], Luigino Dal Maso [3*], Anna Gigli [2] and Stefano Guzzinati [4]

[1]National Center for Disease Prevention and Health Promotion, National Health Institute, Rome, Italy, [2]Institute for Research on Population and Social Policies, National Research Council, Rome, Italy, [3]Cancer Epidemiology Unit, Centro di Riferimento Oncologico di Aviano (CRO) Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Aviano, Italy, [4]Regional Epidemiological Service, Veneto Cancer Registry (RTV), Azienda Zero, Padova, Italy

Background: In most developed countries, the number of cancer survivors is expected to increase in the coming decades because of rising incidence and survival rates and an aging population. These patients are heterogeneous in terms of health service demands: from recently diagnosed patients requiring first-course therapy to patients with extensive care needs and severe disabilities to long-term survivors who only need minimal care. Therefore, in terms of providing healthcare planners and policymakers with useful indicators for addressing policies according to health service demands, it is worth supplying updated measures of prevalence for groups of patients based on the level of care they require. The aim of this paper is to illustrate a new method for estimating short-term projections of cancer prevalence by phase of care that applies to areas covered by cancer registration.

Methods: The proposed method combines linear regression models to project limited duration prevalence derived from cancer registry data and a session of the freely available software COMPREV to estimate the projected complete prevalence into three distinct clinically relevant phases of care: initial, continuing, and final. The method is illustrated and validated using data from the Veneto region in Italy for breast, colorectal, and lung cancers.

Results: Prevalence is expected to increase in 2015-2026 for all considered cancer sites and sexes, with average annual variations spanning from 2.6% for women with lung cancer to 0.5% for men with colorectal cancer. The only exception is lung cancer prevalence in men, which shows an average annual decrease of 1.9%. The majority of patients are in the continuing phase of care, followed by the initial and final phases, except for lung cancer, where the final phase of care prevails over the initial one.

Discussion: The paper proposes a method for estimating (short-term) future cancer healthcare needs that is based on user-friendly and freely available software and linear regression models. Validation results confirm the applicability of our method to the most frequent cancer types, provided that cancer registry data with at least 15 years of registration are available. Evidence

from this method is addressed to policymakers for planning future cancer care, thus improving the cancer survivorship experience for patients and caregivers.

KEYWORDS

cancer prevalence, phase of care, cancer registry, projections, survivorship, health service planning

1 Introduction

In most European countries, prevalent cases make up an important share of the whole population; according to recent estimates, prevalent cases of cancer in Italy accounted for 5.7% of the national population in 2020 (1), corresponding to 3.6 million inhabitants. These absolute numbers are forecast based on an increasing trend of 3.2% per year in the first decade of the 2000s, which is consistent with estimates reported for the USA (2.8% per year) (2), Switzerland (3), and the UK (4). Cancer prevalence is a function of incidence and survival. It increases when new cases are diagnosed and decreases when cancer patients die. Moreover, population growth and changes in the age structure of the population have a relevant impact on prevalence since the risk of cancer increases with age (5).

Cancer survivors represent a growing population because of increases in cancer survival, due to advances in treatment and early diagnosis, and the aging of the population, and the impact of these trends is exceeding the declining incidence observed for some neoplasms (6, 7). Cancer survivors have complex health problems and are heterogeneous in their needs for medical care, psychosocial support, and practical assistance (1, 3). They are generally classified according to the length of survival time and disease outcome, and the vast majority of cancer survivors diagnosed with the most common cancer types survive more than 5 years after diagnosis (8). Most of them receive cancer-related medical care at diagnosis, and some will receive cancer care throughout the rest of their lives. Therefore, medical care expenditures associated with cancer are substantial and are projected to increase dramatically in the near future (9).

Cancer prevalence represents a fundamental measure of cancer burden and cancer survivorship (10). It includes all survivors, irrespective of their patterns of care, and is therefore not suitable to inform healthcare planning, resource allocation, or cost estimation. To overcome this limitation and to better understand the burden of cancer on the healthcare system, several studies have proposed and implemented a breakdown of prevalence into phases of care, i.e., clinically relevant periods related to diagnosis and death (11–13). Different stakeholders are interested in estimating and forecasting cancer prevalence by phase of care: policymakers, to plan sustainable healthcare policies and resource allocation according to the needs of cancer survivors; epidemiologists, to describe the impact of cancer in the population, taking into account the combined effect of incidence, survival, and demographic changes; clinicians, to develop guidelines to improve standardized medium- and long-term follow-up of cancer survivors; and patients, to find support for a complete social recovery and to better meet their rehabilitation needs (1, 13).

Estimates of prevalence are commonly based on limited duration prevalence (LDP) derived from population-based cancer registry data. However, LDP only includes cancer survivors who were diagnosed during the period of activity of the cancer registry, and the shorter this period, the lower the LDP measure (14). Moreover, data collection is retrospective, and the delay between the present time and the time of registration is at least three years (15). To overcome these drawbacks, there are well-consolidated statistical models to estimate complete prevalence, which includes all persons diagnosed with cancer in a given population who are alive at a given prevalence date, regardless of how long ago they were diagnosed (16, 17). However, it is necessary to have more updated prevalence figures than those derived from cancer registry data and to be able to break down complete prevalence by phase of care to account for the heterogeneity of cancer survivors with respect to their healthcare needs.

These needs are addressed in this study, which aims at presenting a methodological approach to project the complete prevalent population by phase of care in the near future. This approach combines methods specifically developed for deriving LDP from population-based cancer registries, using the counting method implemented in the SEER*Stat software (18); implementing short-term projections of LDP; estimating projected complete prevalence in three distinct clinically relevant phases of care - the initial phase following diagnosis, the last year of life, and the continuing phase in between - using the completeness index method - as implemented in the COMPREV software (19).

The method is illustrated and validated using cancer registry data from the Veneto region (Italy), which have been collected in the framework of the Epicost-2 study (20). The method was applied to forecast prevalence by phase of care in 2025 for the following cancer sites: breast (female subjects), colon and rectum (male and female subjects), and lung (male and female subjects).

2 Materials and methods

2.1 Definitions

Limited-duration prevalence (LDP) is the number of people who are alive on a certain date X and have had a cancer diagnosis in

a limited period. The maximum duration of this period depends on the number of years the registry has been collecting incidence cases. LDP is calculated from cancer registry data using the SEER*Stat software (18). When running the limited-duration session in SEER*Stat, the option "All Tumors Matching Selection Criteria/ One Tumor Per Statistic" was used. According to that option, LDP refers to person prevalence: that is, a person will not contribute to a single prevalence estimate with more than one tumor diagnosis.

Complete prevalence (CP) is the number of people who are alive on a given date X with a prior diagnosis of cancer, regardless of when the diagnosis occurred. CP is estimated using LDP and the completeness index method to estimate survivors diagnosed before cancer registration (16).

P is the proportion of LDP per 100,000, i.e., the ratio of the number of cases in a specific population to the population itself.

2.2 Data sources

We used data from the Veneto Cancer Registry (VCR), a population-based cancer registry that covers approximately 2.1 million inhabitants (43% of the whole region) in northeastern Italy. Patients diagnosed with colon and rectum, lung, and breast (female subjects only) cancer between 1990 and 2018 were selected and followed up for vital status until 31/12/2019.

2.3 Input data

LDP and P matrices were stratified by single year of age at prevalence date (t= 0,…, 84, 85+) and by single-year duration d, intended as the distance in years from diagnosis to prevalence date. LDP and P matrices are derived for the more recent five years of incidence: from 01/01/2015 to 01/01/2019. These matrices are the input data for projections; each LDP corresponds to a different maximum duration, and the maximum common duration is 25 years.

Completeness indices were obtained from parameter estimation of survival and incidence models from eight historical Italian cancer registries in the period 1985-2009 (1).

2.4 Projecting limited duration prevalence

We assumed that the prevalence proportion P follows a linear trend in time based on the trend of the last five calendar years. The assumption of a linear trend in P is reasonable for short- or medium-term (e.g., 10-year) projections (1). The steps below were applied to the five LDP matrices from 01/01/2015 to 01/01/2019 to derive the CP by phase of care projected to 01/01/2025 in the population covered by the VCR.

The projection algorithm is made up of the following steps:

i. Compute the LDP proportion (P) (summed for all ages and durations) in the last 5 years of observation (from 01/01/2015 to 01/01/2019 in our example):

$$P(x) = \frac{\sum_{t=1}^{T}\sum_{d=1}^{D} LDP(t,d,x)}{\sum_{t=1}^{T} Pop(t,x)} \times 100,000 \quad (1)$$

where LDP (t,d,x) is the number of prevalent cases of age t and duration d alive on prevalence date x (=2015,…, 2019), Pop(t,x) is the population of the area covered by the VCR on prevalence date x, stratified by age t, maximum age is T=85+ years, and the maximum common duration is D=25 (incidence data from 1990 to 2018).

ii. Fit a linear regression to the LDP proportion for all ages and durations combined

$$P(x) = \alpha + \beta x \quad (2)$$

where the dependent variable is the prevalence proportion P and the covariate is the prevalence date x (=01/01/2015,…, 01/01/2019) and obtain the estimates of the two parameters: $\hat{\alpha}$ and $\hat{\beta}$.

iii. Project the linear regression in year X (in our example, 01/01/2025) to obtain the projected prevalence proportion

$$\hat{P}(X) = \hat{\alpha} + \hat{\beta} \times X \quad (3)$$

The 95% prediction intervals of the projected prevalence proportion were calculated using the "predict" function in R software (21).

iv. Compute the distribution of prevalent cases in the last available year (01/01/2019 in our example). For each age t and duration d we have:

$$w(t,d,2019) = \frac{LDP(t,d,2019)}{\sum_{t=1}^{T}\sum_{d=1}^{D} LDP(t,d,2019)} \quad (4)$$

Where LDP (t, d, 01/01/2019) is the number of cases of age t and duration d alive on 01/01/2019

v. For each annual age t and annual duration d, compute the projected prevalent cases in year X:

$$\widehat{LDP}(t,d,X) = \hat{P}(X) \times Pop(X) \times w(t,d,2019) \quad (5)$$

where Pop(X) is the projected population of the region (computed by the Italian National Institute of Statistics ISTAT (22)) in year X, and w(t,d,2019) are the weights computed in (4) reflecting the distribution of prevalent cases by age and duration in the latest prevalence date of available observations (01/01/2019).

vi. Repeat steps iii and v for year X+1 (in our example, 01/01/2026); notice that the same weights computed in iv are used for the calculation of projected LDP in year X+1.

The projected LDP matrices $\hat{LDP}$ (t, d, X) and $\hat{LDP}$ (t, d, X+1) will be used to decompose the projected Complete Prevalence in year X by phase of care.

2.5 Decomposing the projected complete prevalence by phase of care

The COMPREV software (23) allows estimating the complete prevalence by phase of care, i.e., to break down the complete prevalence into three mutually exclusive phases: the initial phase (Ini, the first 12 months after diagnosis), the end-of-life phase (EOL,

or final, i.e., 12 months before death), and the continuing phase (Cont), defined as the time in between initial and EOL. At the prevalence date, each patient belongs to one of these phases, according to the date of diagnosis and life status: a patient diagnosed within 12 months before the prevalence date and alive 12 months after the prevalence date belongs to the initial phase (Ini); a patient diagnosed more than 12 months before the prevalence date and alive 12 months after the prevalence date belongs to the continuing phase (Cont); and a patient who died within 12 months after the prevalence date, regardless of when they were diagnosed, belongs to the EOL phase (EOL). The EOL phase can be further subdivided into EOL cancer (prevalent cases whose death is due to cancer) and EOL other cause (prevalent cases whose death is due to causes other than cancer), according to the cause of death. This breakdown of the final phase is feasible when information on the cause of death is available.

COMPREV requires the input of two LDP data files: the first one refers to year X, and the second one must refer to the successive year X+1; these files must be identical in their settings except for the year of prevalence to which they refer and must be stratified by single ages at prevalence and single year durations (19). COMPREV also requires completeness indices, specific to cancer type and sex, obtained by statistical regression models of incidence and survival data from cancer registries. A survival matrix containing a crude probability of death is also required to break down the EOL phase into EOL cancer and EOL other causes.

We applied COMPREV to the LDP projected matrices (5), which were computed at prevalence dates X and X+1, to obtain an estimate of the projected complete prevalence by phase of care in year X, stratified by age at prevalence:

CP(t,X,Ini), CP(t,X,Cont),CP(t,X,EOL), where CP(t,X,Ini)+ CP (t,X,Cont)+CP(t,x,EOL)= CP(t,X).

2.6 Validation of the projected complete prevalence by phase of care

In order to validate the method, we applied the above-illustrated algorithm to a subset of the VCR data, comprising patients diagnosed with colon and rectum, lung, and breast (female subjects only) cancer in 1990-2011 and followed for vital status until 01/01/2012:

i. derive LDP in five consecutive years, 2008-2012, with a maximum common duration of 18 years;
ii. project LDP proportions in the years 2018 and 2019, as described in Section 2.4;
iii. compute the projected complete prevalence by phase of care on 01/01/2018 via COMPREV, as described in Section 2.5;
iv. directly estimate the complete prevalence by phase of care by applying completeness indices to the LDP in years 2018 and 2019 derived from the complete set of VCR data (i.e., patients diagnosed over the entire period of data availability 1990-2018 and followed up to 01/01/2019);
v. compare the projected and estimated complete prevalence by cancer site and phase of care.

The results of this validation are illustrated in Supplementary Table 1 and Supplementary Figure 1.

We also investigated the minimum length of cancer registry data required for the projections by comparing the projected complete prevalence by phase of care in 2025 using 25-year LDP data (incidence data period 1990-2018, follow-up 01/01/2019) with that obtained using 15-year LDP data (incidence data period 2000-2018, follow-up 01/01/2019.

The results of this validation are illustrated in Supplementary Figure 2.

3 Results

Figure 1 shows time trends of 25-year LDP proportions P by cancer site and sex in the Veneto region. From 2015 to 2019, the proportions are based on VCR data; from 2020 to 2026, the proportions are projected via linear regression; the lower and upper bounds of the projections are also included in the figure.

P increases in the seven-year projection period 2020-2026 for all combinations of cancer site and sex, except for lung cancer in the male population (-10.6%, corresponding to a -1.9% average annual variation). The largest increases are in women with lung cancer (16.5%, corresponding to a 2.6% average annual variation), breast cancer (14.7%, corresponding to a 2.3% average annual variation), and colorectal cancer (5.9%, corresponding to a 1% average annual variation). In men, there is a 3% increase in P for colorectal cancer (corresponding to a 0.5% average annual variation).

The increasing trends in LDP proportions derive from increasing incidence (as is the case for lung cancer in women) or stable incidence (as is the case for breast cancer) combined with population aging. For colorectal cancer, the reduction in the risk of developing the disease, which led to a decrease in incidence in 2007-2008 for both men and women (24), does not yet compensate for the combined effect of aging and increasing survival, thus resulting in a slight but positive trend until 2026.

The decreasing trend of LDP proportions in men diagnosed with lung cancer is due to a decrease in incidence: in the Veneto region, the APC (Annual Percent Change) incidence spans from -1.3% in the 1990s to -3.7% at the beginning of the 2000s (25). Despite the aging of the population and the increase in survival, this decrease in incidence determines the reduction of prevalence: from 201 prevalent cases per 100,000 in 2015 (corresponding to 4,809 patients overall in the Veneto region) to 166 prevalent cases per 100,000 in 2026 (corresponding to 3,927 patients overall in the Veneto region).

The projection of prevalent cases decomposed by phase of care is the main application of the methodology, and the results are illustrated in Table 1. The total number of prevalent cases (complete prevalence) estimated in 2018 and projected in 2025 in the Veneto region are reported by phase of care and age group at prevalence (<50, 50-69, 70+). During the 7 years, the complete prevalence shows an increase in the percent variation between 6% for colorectal

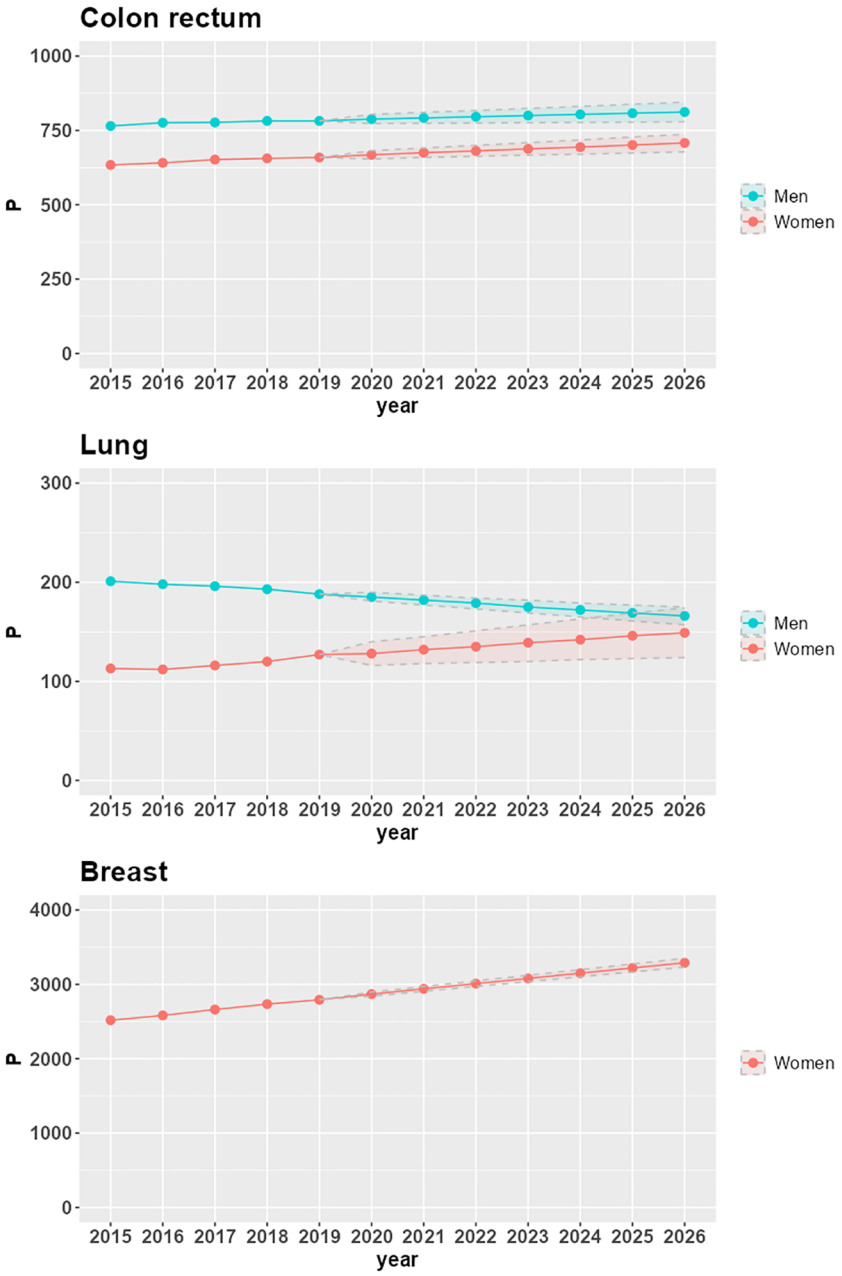


FIGURE 1
Time trends of 25-year LDP proportions per 100,000 (P), with lower and upper bounds of 95% confidence intervals, by cancer site and sex in the Veneto region.

cancer in men, where the number of survivors increases from 19,342 to 20,436, and 25% for lung cancer in women, where the number of survivors increases from 3,010 to 3,750. The only exception is lung cancer in men, showing a 12% decrease in the percent variation with the number of survivors decreasing from 4,848 to 4,263.

For women with breast cancer and men with lung cancer, variations are evenly distributed by phase of care, with an increasing trend by age. For men with colorectal cancer, the complete prevalence increases between 6% and 7% in the initial and continuing phases of care, respectively, and decreases by about 6% in the final phase of care. For women, most of the variation is due to the increase in survivors in the continuing phase of care (10% percent variation, from 15327 to 16843 patients). For women with lung cancer, most of the variation is due to the increasing number of survivors in the final phase of care (from 406 to 635 women), thus representing an increasing share of the prevalence cohort (from 14% in 2018 to 17% in 2025).

Major variations concern the elderly population, aged 70 years and over. Time trends and patterns by age at prevalence are due to the aging of the population and the consequent increased risk of developing cancer.

In the initial phase of care, the increase in prevalence for colorectal cancer is less pronounced among patients aged 50 to

10.3389/fonc.2023.1201464

TABLE 1 Complete prevalence (counts) estimated in 2018 (CP 2018) and projected in 2025 (CP 2025) in the Veneto region by cancer site, age group at prevalence, and phase of care.

CP 2018		Phase of Care				CP 2025		Phase of Care			
Cancer Site - Sex	Age Group	Initial	Continuing	Final	Total	Cancer Site - Sex	Age Group	Initial	Continuing	Final	Total
Colon Rectum-M	<50	53	341	7	400	Colon Rectum-M	<50	53	347	6	407
	50-69	553	4753	251	5557		50-69	578	4972	222	5773
	70 +	752	11566	1067	13385		70 +	810	12435	1011	14257
	all ages	1358	16660	1324	19342		all ages	1441	17755	1239	20436
Colon Rectum-F	<50	69	366	19	453	Colon Rectum-F	<50	71	384	17	472
	50-69	390	4079	161	4630		50-69	415	4386	133	4935
	70 +	706	10882	690	12278		70 +	754	12073	753	13580
	all ages	1164	15327	869	17361		all ages	1241	16843	903	18987
Breast-F	<50	811	4647	104	5562	Breast-F	<50	928	5302	110	6341
	50-69	2182	27706	467	30355		50-69	2548	32592	493	35632
	70 +	1658	36312	1637	39606		70 +	2049	44078	2013	48139
	all ages	4651	68665	2208	75523		all ages	5525	81971	2616	90112
Lung-M	<50	17	58	18	93	Lung-M	<50	16	54	14	84
	50-69	206	806	257	1269		50-69	180	731	218	1129
	70 +	342	2306	839	3487		70 +	305	2011	733	3049
	all ages	565	3170	1114	4848		all ages	501	2796	965	4263
Lung-F	<50	21	76	15	112	Lung-F	<50	24	103	21	148
	50-69	184	720	168	1072		50-69	214	867	238	1319
	70 +	223	1378	224	1825		70 +	274	1633	377	2284
	all ages	429	2174	406	3010		all ages	513	2602	635	3750

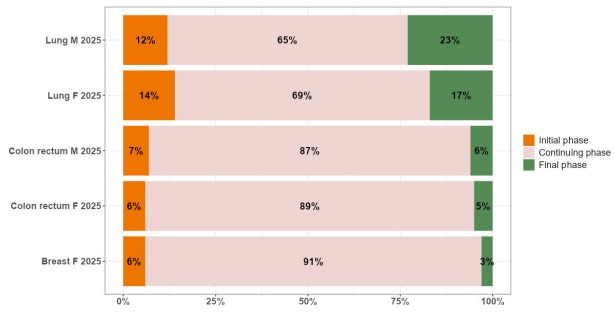


FIGURE 2
Distribution by phase of care (% values) of complete prevalence in 2025, by cancer site and sex in the Veneto region.

69, possibly as a consequence of screening programs that allow the detection of pre-cancerous lesions, thus reducing the number of newly diagnosed patients; in all phases of care, the increase in the number of lung cancer survivors among the female population is higher for women aged 15 to 49, consistent with the increasing prevalence of smoking among young women (26).

The bar plot in Figure 2 presents the breakdown of complete prevalence by phase of care in 2025 by cancer site and sex.

The dynamics of prevalence during the projection period from 2018 to 2025 slightly affect the distribution of cancer survivors in the three phases of care: most patients are in the continuing phase of care, followed by the initial and final phases, except for lung cancer, where the percentage of patients in the final phase is higher than in the initial one. There are two patterns according to survival: among cancer patients with a better prognosis, as is the case for women with breast cancer, and all patients with colorectal cancer, 87%-91% are in the continuing phase, 6%-7% are in the initial phase, and 3%-6% are end-of-life patients. Among cancer patients with poorer prognoses, as is the case for all patients with lung cancer, 66%-69% are in the continuing phase, 12%-14% are in the initial phase, and 17-23% are end-of-life patients.

4 Discussion

Estimating and projecting cancer prevalence according to different phases of care is a prerequisite for estimating the economic impact of cancer.

The estimation of cancer prevalence by phase of care was mentioned in 2001 by Brown et al. in their seminal paper on the economic burden of cancer (27). Since then, many researchers have contributed to the field (11, 13, 28, 29). Estimation of cancer prevalence by phase of care is feasible when longitudinal data, identified at the level of individual-incidence cancer cases, are available. Cancer registries typically collect these data.

Prevalence is a complex indicator that depends on incidence, survival, and population dynamics. These determinants are to be taken into account in the projection of prevalence. A step forward in this direction was the projection of cancer prevalence based on a deterministic relationship between cancer mortality, incidence, and survival: the PIAMOD approach (5), which estimates and projects cancer prevalence as a function of incidence and survival models, with minor *ad hoc* hypotheses on the population evolution patterns. In PIAMOD, a linear period trend is assumed for incidence projections. For survival projections, two hypotheses are proposed: a conservative one, which assumes that cancer patient survival will remain stable for the projected years, and an optimistic one, which assumes that cancer patient survival will continue to improve at the same rate as observed in recent years (5).

PIAMOD was used as a basis for projecting prevalence by phase of care by Mariotto et al. (30) and later by Yu et al. (31). It is used for purposes similar to our method, but it requires more data (incidence and survival) and modeling than our approach. On the other hand, PIAMOD is more flexible as it allows one to distinguish the contribution of incidence from that determined by survival in the prevalence projections.

According to Yu, this approach has some drawbacks: *"The process involves many decisions to be made, such as selecting a high-dimensional polynomial incidence model and mixture cure model for relative survival based on different assumptions for future trends. All of these decisions must be informed by a high level of cancer epidemiological and statistical knowledge, and the resulting prevalence estimates are highly dependent on these modelling decisions and assumptions."*

In this paper, we propose an alternative semi-parametric approach that combines the projection of LDP data from CR (1) and the decomposition of the projected prevalence into phases of care (19).

This approach is quite straightforward and does not require complex modeling, as completeness indices are externally estimated from other studies. It provides good results for the most frequent cancer types, which are the most interesting cases from the perspective of estimating the economic burden. To decompose the projected prevalence by phase of care, we used the software COMPREV, which is freely available and easy-to-use; the software contains a set of default parameter estimates obtained from SEER (Surveillance, Epidemiology, and End Results) data. Further, population forecasts can easily be embedded; for Italy, these were provided by the Italian Institute of Statistics (ISTAT). Finally, the method works equally well on shorter incidence data series (15 years); thus, it can also be applied to more recently established CRs and allows one to incorporate more recent trends in the projections, as has been shown in the case of colorectal cancer.

There are some limitations:

The method does not allow one to project prevalence according to different scenarios of incidence and survival dynamics. However, according to a sensitivity analysis presented in previous studies (31), approximately 71% of the prevalence dynamics reflect the impact of population growth and aging, while the remaining 29% are attributable to incidence and survival changes.

The phase of care decomposition does not function well when the number of cases is too small, especially when the scarce numbers are concentrated in only one of the phases of care, such as for thyroid cancer, where there are virtually no cases in the end-of-life phase. We must bear in mind, however, that less frequent cancers have a smaller economic impact. Therefore, the applicability of the proposed methodology is limited to the most frequent cancers that have a substantial economic impact on the healthcare system, and to a short- to medium-term forecast horizon, which is typically considered for planning healthcare intervention policies. Within this context, the proposed complete prevalence projections by phase of care have been validated and produce reliable results. The continuing phase includes patients who may be highly heterogeneous in terms of healthcare: some of them have recently completed their initial therapy and require follow-up, some others require treatment for cancer recurrence or second primary cancers and, finally, some have survived for a long period since their initial treatment and can be considered cured.

Further developments of this method can be considered:

Data on specific treatments and procedures collected in the framework of the Epicost study (32) could be used to disentangle patterns of patients with homogeneous care needs and to decompose the continuing phase accordingly.

As also highlighted by Mariotto (6), since cancer incidence is highest in the elderly, the impact of population changes on cancer prevalence may exceed the impact of declining cancer incidence rates for some cancers. We are considering the possibility of incorporating the dynamics of the age structure in addition to the population changes.

For the purposes of our method, it would be worthwhile to project the initial phase prevalence stratified by stage at diagnosis. To implement this methodological enhancement, we need to retrieve information on the stage at diagnosis for initial phase patients in the last five years used as the basis for the projections.

5 Conclusions

Complete cancer prevalence is a fundamental but crude indicator of health service needs, as it covers all steps of the clinical pathway and includes patients with a wide range of health service requirements. Here, we presented a method, applicable where cancer registry data are available, to monitor the size of the cancer burden in a given population to define care requirements concerning the prevalence breakdown across the three phases of

care, to establish priorities, and to project, in combination with average individual cost profiles, expenditures directly related to cancer care (20). For these purposes, 7 to 10 years is the time span usually considered by policymakers, and the focus is on the most frequent cancer sites that have a major economic impact on the healthcare system. Evidence from this methodology will be useful in facing the challenge of planning and developing a healthcare system that is able to respond in the short- to medium term to the increasing needs of people living with cancer.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SF, AG, and SG designed the study, drafted the study protocol, collected the data, and prepared the cleaned data for the study database. FT and SG performed the statistical analyses. SF, FT, LDM, AG, and SG contributed to the validation of the statistical models and revised the statistical analyses. SF, LDM, AG, and SG discussed modeling assumptions and applicability. All authors contributed to the interpretation of the study results and reviewed and approved the final version.

Funding

This study was funded by the Italian Ministry of Health (RICERCA FINALIZZATA 2018, grant number RF-2018-12365530) and by the European Commission (Work Programme 2017, grant number 801520 HP-JA-2017 "Innovative Partnership for Action Against Cancer") and by the Italian Association for Cancer Research (AIRC, grant no. 21879).

Acknowledgments

The authors thank Francesca Vicari for the preparation of the figures.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1201464/full#supplementary-material

References

1. Guzzinati S, Virdone S, De Angelis R, Panato C, Buzzoni C, Capocaccia R, et al. Characteristics of people living in Italy after a cancer diagnosis in 2010 and projections to 2020. *BMC Cancer* (2018) 18:169. doi: 10.1186/s12885-018-4053-y

2. De Moor JS, Mariotto AB, Parry C, Alfano CM, Padgett L, Kent EE, et al. Cancer survivors in the United States: prevalence across the survivorship trajectory and implications for care. *Cancer Epidemiol Biomarkers Prev* (2013) 22(4):561–70. doi: 10.1158/1055-9965.EPI-12-1356

3. Herrmann C, Cerny T, Savidan A, Vounatsou P, Konzelmann I, Bouchardy C, et al. Cancer survivors in Switzerland: a rapidly growing population to care for. *BMC Cancer* (2013) 13:287. doi: 10.1186/1471-2407-13-287

4. Maddams J, Utley M, Møller H. Projections of cancer prevalence in the United Kingdom, 2010–2040. *Br J Cancer* (2012) 107:1195–202. doi: 10.1038/bjc.2012.366

5. Verdecchia A, De Angelis G, Capocaccia R. Estimation and projections of cancer prevalence from cancer registry data. *Statist. Med* (2002) 21:3511–26. doi: 10.1002/sim.1304

6. Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML. Projections of the cost of cancer care in the United States: 2010–2020. *J Natl Cancer Inst* (2011) 103:117–28. doi: 10.1093/jnci/djq495

7. Allemani C, Weir HK, Carreira H, Harewood R, Spika D, Wang X-S, et al. Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* (2015) 385:977–1010. doi: 10.1016/S0140-6736(14)62038-9

8. Dal Maso L, Santoro A, Iannelli E, De Paoli P, Minoia C, Pinto M, et al. Cancer cure and consequences on survivorship care: position paper from the italian alliance against cancer (ACC) survivorship care working group. *Cancer Manage Res* (2022) 14:3105–18. doi: 10.2147/CMAR.S380390

9. Mariotto AB, Enewold L, Zhao J, Zeruto CA, Yabroff KR. Medical care costs associated with cancer survivorship in the United States. *Cancer Epidemiol Biomarkers Prev* (2020) 29:1304–12. doi: 10.1158/1055-9965.EPI-19-1534

10. Toffolutti F, Guzzinati S, De Paoli A, Francisci S, De Angelis R, Crocetti E, et al. Complete prevalence and indicators of cancer cure in Italy: Enhanced 1 methods and validation. *Front Oncol* (2023) 733. submitted. doi: 10.3389/fonc.2023.1168325

11. Laudicella M, Walsh B, Burns E. Cost of care for cancer patients in England: evidence from population-based patient-level data. *Br J Cancer* (2016) 114:1286–92. doi: 10.1038/bjc.2016.77

12. Francisci S, Guzzinati S, Capodaglio G, Pierannunzio D, Mallone S, Tavilla A, 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) 21(7):1003–13. doi: 10.1007/s10198-020-01190-z

13. Yabroff KR, Lund J, Kepka D, Mariotto A. Economic burden of cancer in the United States: estimates, projections, and future research. *Cancer Epidemiol Biomarkers Prev* (2011) 20(10):2006–14. doi: 10.1158/1055-9965.EPI-11-0650

14. Corazziari I, Mariotto A, Capocaccia R. Correcting the completeness bias of observed prevalence. *Tumori J* (1999) 85:370–81. doi: 10.1177/030089169908500503

15. AIRTUM Working Group. Italian cancer figures, report 2014: Prevalence and cure of cancer in Italy. *Epidemiol Prev* (2014) 38(6 Suppl 1):1–122. doi: 10.19191/EP14.6.S1.113

16. Capocaccia R, De Angelis R. Estimating the completeness of prevalence based on cancer registry data. *Stat Med* (1997) 16(4):425–40. doi: 10.1002/(SICI)1097-0258(19970228)16:4<425::AID-SIM414>3.0.CO;2-Z

17. Merrill RM, Capocaccia R, Feuer EJ, Mariotto A. Cancer prevalence estimates based on tumour registry data in the Surveillance, Epidemiology, and End Results (SEER) Program. *Int J Epidemiol* (2000) 29(2):197–207. doi: 10.1093/ije/29.2.197

18. National Cancer Institute. *SEER*Stat Software, version 8.4.0.1*. Available at: https://seer.cancer.gov/seerstat/ (Accessed March 21, 2023).

19. Gigli A, Francisci S, Guzzinati S, Hall A, Hachey M, Scoppa S, et al. Cancer prevalence by phase of care: an indicator for assessing health service needs. *Tumori J* (2020). doi: 10.1177/0300891620961839

20. Francisci S, Capodaglio G, Gigli A, Mollica C, Guzzinati S. Cancer cost profiles: The Epicost estimation approach. *Front Public Health* (2022) 10:974505. doi: 10.3389/fpubh.2022.974505

21. Chambers JM, Hastie TJ. *Statistical models in S*. Department of Statistics and Actuarial Science, University of Waterloo Waterloo, Ontario, Canada :Wadsworth & Brooks/Cole (1992).

22. *Demography in figures*. Available at: https://demo.istat.it/?l=en (Accessed March 21, 2023).

23. *COMPREV, complete prevalence program, version 3.0.31 (Beta)*. Available at: https://surveillance.cancer.gov/comprev/ (Accessed March 21, 2023). released August 2022.

24. Bucchi L, Mancini S, Baldacchini F, Ravaioli A, Giuliani O, Vattiato R. How a faecal immunochemical test screening programme changes annual colorectal cancer incidence rates: an Italian intention-to-screen study. *Br J Cancer* (2022) 127:541–548. doi: 10.1038/s41416-022-01813-7

25. *The tumours in Veneto*. Available at: https://gecoopendata.registrotumoriveneto.it/incidenza.php?sede=polmone&codSede=C33-C34.9&lang=EN (Accessed March 21, 2023).

26. Trama A, Boffi R, Contiero P, Buzzoni C, Pacifici R, Mangone L, et al. Trends in lung cancer and smoking behaviour in Italy: an alarm bell for women. *Tumori J* (2017) 103(6):543–50. doi: 10.5301/tj.5000684

27. Brown ML, Lipscomb J, Snyder C. The burden of illness of cancer: economic cost and quality of life. *Annu Rev Public Health* (2001) 22:91–113. doi: 10.1146/annurev.publhealth.22.1.91

28. Mariotto A, Warren JL, Knopf KB, Feuer EJ. The prevalence of patients with colorectal carcinoma under care in the US. *Cancer: Interdiscip Int J Am Cancer Soc* (2003) 98(6):1253–61. doi: 10.1002/cncr.11631

29. de Oliveira C, Weir S, Rangrej J, Krahn MD, Mittmann N, Hoch JS, et al. The economic burden of cancer care in Canada: A population-based cost study. *CMAJ Open* (2018) 6:E1–E10. doi: 10.9778/cmajo.20170144

30. Mariotto AB, Yabroff KR, Feuer EJ, De Angelis R, Brown M. Projecting the number of patients with colorectal carcinoma by phases of care in the US: 2000–2020. *Cancer Causes Control* (2006) 17:1215–26. doi: 10.1007/s10552-006-0072-0

31. Yu XQ, Clements M, O'Connell D. Projections of cancer prevalence by phase of care: a potential tool for planning future health service needs. *J Cancer Surviv* (2013) 7:641–51. doi: 10.1007/s11764-013-0303-9

32. Corti MC, Avossa F, Schievano E, Gallina P, Ferroni E, Alba N. A case-mix classification system for explaining healthcare costs using administrative data in Italy. *Eur J Int Med* (2018) 54:13–6. doi: 10.1016/j.ejim.2018.02.035

frontiers | Frontiers in Oncology

TYPE Technology and Code
PUBLISHED 26 October 2023
DOI 10.3389/fonc.2023.1250195

OPEN ACCESS

EDITED BY
Jerry Polesel,
Aviano Oncology Reference Center (IRCCS), Italy

REVIEWED BY
Paolo Contiero,
National Cancer Institute Foundation (IRCCS), Italy
Maria Paula Curado,
A.C.Camargo Cancer Center, Brazil

*CORRESPONDENCE
Francesco Giusti
Francesco.Giusti@registreducancer.org;
francescogiusti@hotmail.com
Manola Bettio
JRC-ENCR@ec.europa.eu

†PRESENT ADDRESSES
Francesco Giusti,
Belgian Cancer Registry, Brussels, Belgium
Carmen Martos,
Foundation for the Promotion of Health and Biomedical Research in the Valencian Region (FISABIO), Valencia, Spain

RECEIVED 29 June 2023
ACCEPTED 26 September 2023
PUBLISHED 26 October 2023

CITATION
Giusti F, Martos C, Adriani S, Flego M, Carvalho RN, Bettio M and Ben E (2023) The Joint Research Centre-European Network of Cancer Registries Quality Check Software (JRC-ENCR QCS). *Front. Oncol.* 13:1250195. doi: 10.3389/fonc.2023.1250195

The Joint Research Centre-European Network of Cancer Registries Quality Check Software (JRC-ENCR QCS)

Francesco Giusti[1,2*†], Carmen Martos[1,3†], Stefano Adriani[1], Manuela Flego[1], Raquel Negrão Carvalho[1], Manola Bettio[1*] and Enrico Ben[1]

[1]European Commission, Joint Research Centre (JRC), Ispra, Italy, [2]Belgian Cancer Registry, Brussels, Belgium, [3]Foundation for the Promotion of Health and Biomedical Research in the Valencian Region (FISABIO), Valencia, Spain

The core activity of population-based cancer registries (PBCRs) is to gather information from all new cancer cases in a defined geographic area, in order to measure the magnitude of cancer burden and to provide a basis for cancer research. The Joint Research Centre-European Network of Cancer Registries Quality Check Software (JRC-ENCR QCS) is a Java standalone desktop application, under development since 2015, created to support PBCRs in the validation of the collected data. The JRC-ENCR QCS performs internal consistency checks on the cancer registry dataset, to detect impossible or unlikely codes or combination of codes, and is thereby an important tool to support the validation efforts by registries and improve data quality and European-wide harmonisation. The software package also includes the JRC CSV Data layout converter, a complementary tool for transforming PBCR incidence files into a format compatible with the JRC-ENCR QCS. This paper gives an overview of the JRC-ENCR QCS, describing the role of the software in processing data files submitted by PBCRs contributing to the European Cancer Information System (ECIS) as well as its functionalities. The development of the JRC-ENCR QCS is an evolving process, with regular updates implementing new and revised European and International recommendations and classifications.

KEYWORDS

cancer registry, validation, harmonisation, data quality, software, Europe

1 Introduction

Population-based cancer registries (PBCRs) systematically collect data from all new reportable cancer cases occurring in a defined geographic area (1). In Europe, PBCRs are organised within the European Network of Cancer Registries (ENCR), established in 1989 in the framework of the Europe Against Cancer Programme of the European Commission. The ENCR is a professional, non-profit society dedicated to promoting collaboration

between PBCRs, defining data collection standards and providing training to PBCR personnel. It aims to strengthen the basis for monitoring cancer burden in the EU and the rest of Europe, through the provision of regular and timely information from European PBCRs.

The European Cancer Information System (ECIS) was developed by the European Commission's Joint Research Centre (JRC) in collaboration with the Directorate-General for Health and Food Safety (DG SANTE), following the 2009 "*Communication on Action Against Cancer: European Partnership*" (2). Launched in 2018, the ECIS provides indicators (incidence, mortality and survival) to quantify cancer burden across Europe thanks to the contribution of cancer registries data (3) through periodic data calls. A dedicated data submission portal (4) was developed to collect registry data files for incidence, mortality, population and life tables, submitted by ENCR registries according to a well-defined protocol (5) that details the list of variables and allowed range of values required for the calculation of cancer burden indicators and publication in ECIS.

The reliability and use of the information provided by PBCRs depend on data quality, measured through its different dimensions of comparability, completeness, validity and timeliness (6–8).

Adherence to protocol, data standardisation and internal consistency checks are the core steps of the data validation process carried out by the JRC to ensure harmonisation and comparability of European PBCRs data.

In support to this process, an ENCR expert working group (WG) published a comprehensive and standardised list of data quality checks to be adopted by European PBCRs and European projects.

The WG addressed case and variables definition, format for data collection and related internal consistency rules. The results of the initiative were ENCR-endorsed reports (9–11) which serve as guidelines for the data acquisition and further validation of PBCRs data.

This work was also the basis for the development of the JRC-ENCR Data Quality Check Software (JRC-ENCR QCS) described here, an open-access software to facilitate standardisation and validation of PBCR data (12). The aim of this paper is to give an overview of the JRC-ENCR QCS by describing its role in processing data files submitted by PBCRs contributing to the ECIS, and list its main validation routines.

2 Method

2.1 Overview of the JRC-ENCR QCS

The JRC-ENCR QCS has been designed as a standalone desktop application that can run locally by PBCRs, without the need of internet connection.

Processing cancer data locally is a common precautionary practice to protect sensitive patient health data from external access, thus allowing PBCRs to directly check and correct their data files while avoiding the stringent General Data Protection Regulation (GDPR) rules that must be applied when sharing individual data, even after pseudonymisation. Incidence files are considered sensitive even if pseudonymised, as they contain patient's sensitive data, such as date of birth, date of cancer diagnosis, sex, geographical location, therefore, in order to be compliant with the GDPR regulation, incidence data must be handled with the proper precautions (13).

The software was developed using an almost pure Java design pattern in which strictly necessary libraries are used either to reduce the number of dependencies or to avoid vendor's lock-ins.

The JRC-ENCR QCS has a flexible architecture and can perform checks on different data collection protocols, such as the 2015 and 2022 ECIS protocols (5, 14).

The standard execution of the software is the Graphical User Interface (GUI) mode, which opens the GUI window and waits for actions from the user. Alternatively, the JRC-ENCR QCS may be run in the command-line mode, validating the dataset passed as an argument.

Regardless of the execution mode, output reports are produced in the output subfolder within the directory where the software is installed.

The JRC-ENCR QCS has been initially developed for the Windows operating system, requiring Java (Windows 8 and above). Starting from JRC-ENCR QCS version 1.7 it can also run on MacOS and Linux operating systems.

Since the first release, the software has been upgraded and improved based on the JRC Technical reports "*A common data quality check procedure for European cancer registries*" (9–11), on the new ENCR recommendations (15), on the experience acquired in data validation for ECIS and on the feedback received from the PBCR users.

2.2 The dataset

The inputs required by the JRC-ENCR QCS are text files, with default data fields (variables) delimited by semicolon (";"). The software configuration allows also for comma-separated variables.

The software is able to check four different data files (incidence, mortality, population and life tables), required to update all the incidence, mortality and survival indicators published in ECIS, and the calculation of different quality indicators.

The cancer incidence file contains different groups of variables:

i) variables related to the patient,
ii) variables related to the tumour,
iii) variables related to the follow-up,
iv) variables related to stage and
v) variables related to treatment.

When PBCR contribute with their incidence file to European projects, they are responsible for the prior pseudonymisation of their data. This is the case of the contribution to ECIS, which requires the upload of pseudonymised data in the JRC data submission portal.

Information about the processing of data is published at the European Commission's Register of the Data Protection Officer

(DPO) and available at the following link: https://ec.europa.eu/dpo-register/detail/DPR-EC-00417.

There is correspondence between the population and cancer incidence files with respect to registration area, period, sex, age-range and geographical reference codes. The variables included in the population file are: calendar year, sex, age, geographical area code and label, as well as number of residents. The information included in the population file should be obtained from official censuses, from intercensal/postcensal estimates provided by Vital Statistics Departments, or equivalent, or other official sources.

The mortality data corresponds to the cancer incidence file with respect to registration area, period, sex and age-range. The mortality information is obtained from the Vital Statistics Department, or equivalent, and based on certificates/death records. The file should contain the following variables: calendar year, sex, age, cause of death and number of deaths.

Life tables file must be provided by registries covering their entire period of incidence or the period in which the follow-up is available. Life tables have the same geographical and temporal reference as for the cases of the incidence file and contain the following variables: calendar year, sex, age, geographical area code and label, as well as all causes of death probability.

2.3 The data check process

Data checks are performed in consecutive cycles, each comprised of different steps (Figure 1). Generally, data checks fall within three main categories:

- Checks on single variables (univariate checks)
- Checks between variables of the same record (multivariate checks)
- Checks between variables of different records belonging to the same patient (perfect duplicates and multiple primary tumours checks)

Population, life tables and mortality files have a simple structure that can be validated by univariate checks only. On the other hand, incidence files are more complex, as they contain more variables (the ECIS protocol requested 56 and 39 variables for the 2015 and 2022 data calls respectively). This data can be recorded/codified in different ways, and can be more prone to errors. For this reason, in addition to the checks on single variables, the checks on the incidence file should include as well checks between variables of the same record.

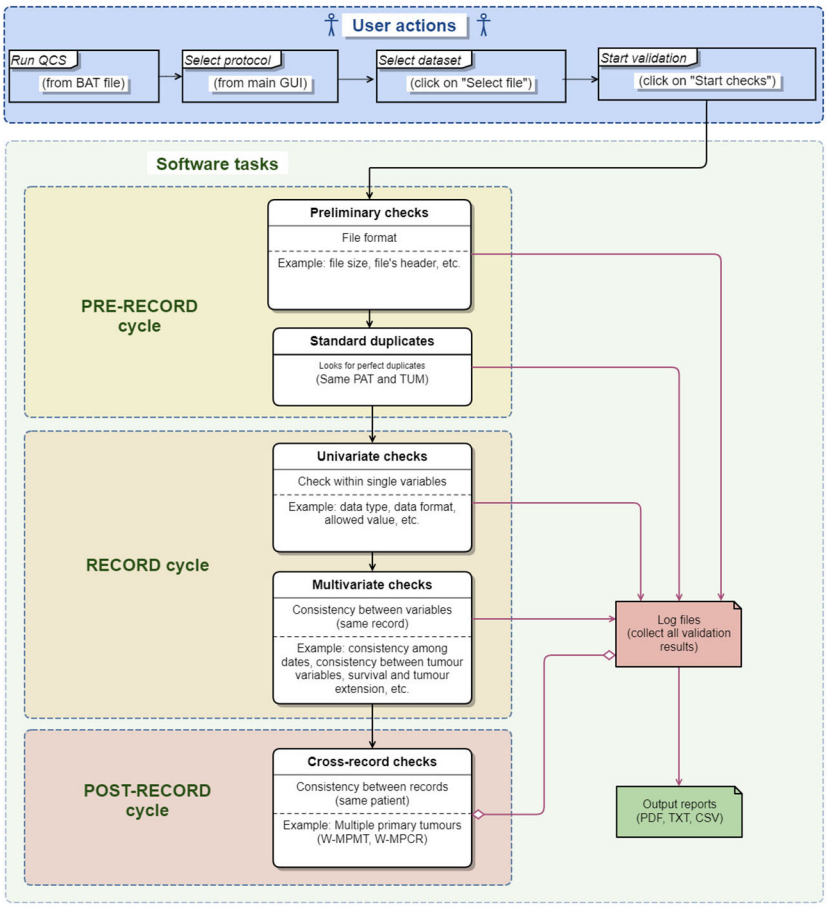


FIGURE 1
Data quality check process performed by the JRC-ENCR QCS.

Additionally, the JRC-ENCR QCS also checks for the presence of multiple primary tumours in the incidence file, which is an important validation step for the correct ascertainment of number of cancer cases. This is because the same patient can be associated to multiple tumours, recorded either at the same time or at different times, and that can represent the same primary tumour or completely independent tumours.

The time needed to perform all the checks and create the output reports depends mainly on the computer's memory; the performance of the process increases significantly starting from 4GB of RAM. A second factor affecting the process is the number of errors/warnings. There is an approximately linear relation between the number of errors/warnings and the time spent to finalise the validation of data. For instance, with 4GB of RAM, 100,000 records and circa one error/warning per record the total time of the process is around 2 minutes, and around 20 minutes with one million records.

The minimum number of records checked by the software is one, while the maximum is related to the number of errors/warnings, at around 10 million messages (e.g. 1 million records with 10 errors/warnings per line).

2.4 Univariate checks

Univariate checks on single variables are the simplest checks, used to verify that the value of each variable is compliant with the required format and is within the range specified by the protocol. An example is the variable "sex" which must be codified as numeric and has four allowed values: 1=male, 2=female, 3=other or 9=unknown.

Sometimes the software has to adapt the theoretical range of values allowed by the protocol to the values that are used (and submitted) by PBCRs. This is the case of the "*pN*" variable, i.e. the pathological assessment of the regional lymph nodes in the TNM classification system (16). In 2021, the JRC-ENCR QCS expanded the set of allowed values for "*pN*" in order to include also the TNM notation "*1biv*", coming from previous TNM editions. Such TNM notation was still in use by some PBCRs, and resulted in an excessive amount of "out of range" errors identified in the software output, shadowing the actual critical errors in the dataset.

2.5 Multivariate checks of the same record

Multivariate checks identify inconsistencies among values in different variables of the same record. Some checks, such as comparing topography and morphology codes, are straightforward and are performed according to a well-defined table of allowed/refused combinations of cancer morphologies and topographies.

By contrast, the consistency between age/topography/morphology, which is required because some cancer types occur almost exclusively in certain age groups, is one of the most complex checks. This is the case of retinoblastoma (tumour of young children) or prostate cancer (in older men). Therefore, some combinations of age/topography/morphology are unlikely or very rare and should produce a warning message according to the list of unlikely/rare combinations of age/topography/morphology. Additionally, several morphology codes may be related to multiple topography values, thus increasing the complexity in the checks for age/topography/morphology combinations (11). For example, in the age group 0-14 years, about 50 different morphology codes for the gonadal carcinoma type can be combined with the two topography values C56 (ovary) and C62 (testis), producing about 100 different unlikely combinations.

2.6 Multivariate checks between different records of the same patient

The implementation of the multivariate checks between different records of the same patient is rather complex. Firstly, not all records in the dataset need to be checked, only records that meet certain requirements and pass the univariate or multivariate checks (Sections 2.4 and 2.5) must be compared against each other. This implies that the software has to store the results of all the basic checks before deciding which records should be evaluated by the rules for the multivariate checks between different records of the same patient.

Secondly, some complex filtering criteria must be applied on the dataset, including those on the behaviour of multiple primary tumours. These are quite critical and at the same time, have been subject to several changes in the latest years due to updates of the international cancer coding rules, new ENCR recommendations for coding and reporting tumours, new requirements from the PBCRs, results of data checks, etc.

Finally, while in the dataset each patient is identified by a patient ID, records are usually not ordered by patient ID and two or more records addressing the same patient may not be in consecutive rows of the data file. The JRC-ENCR QCS is not designed to store in memory the entire dataset, which could contain millions of rows and would require a computer with a large amount of memory (the procedure for these checks is detailed in Section 2.7.2).

The software performs two checks between records of the same patient: the first identifies perfect duplicates (records with same patient and tumour IDs) in the pre-record cycle (Figure 1) and the second addresses Multiple Primary Tumours (MPT) (records with same patient ID and different tumour ID). The rules used for checking MPT take place in the Post-record cycle (Figure 1) since they are usually performed after the basic univariate and multivariate checks described in the previous sections.

This could be the case of a patient with multiple metastases originating from a single primary tumour but recorded separately at different times. The multiple primaries rule can identify multiple records of the same patient that refer to the same primary tumour, reducing the risk of counting the same tumour twice (or more). This check, and the subsequent elimination of these records from the dataset must be performed before data aggregation, and is crucial to correctly calculate the cancer indicators published on the ECIS website.

The pre-record cycle, record cycle and post-record cycle, each consisting of several steps are described in more detail in the following sections.

2.7 Validation workflow

2.7.1 File structure validation (pre-record cycle)

In this validation phase, only preliminary rules (e.g. checking the file format, file size, file header, looking for perfect duplicates) are applied. Depending on the JRC-ENCR QCS configuration, some errors found during these checks, such those occurring when a dataset is empty or too small, can stop the execution of the program and interrupt the validation process (blocking errors).

Log files (e.g. *qcs_rule_output.csv*) collect all errors/warnings identified at any stage of the validation process and are used to produce the final output reports.

2.7.2 Record cycle

If the whole dataset passes the checks of the previous validation phase, further checks are performed as described below.

All records of the input dataset are read and processed one by one, and all basic rules (univariate and multivariate checks) are applied to each record independently. This approach optimises the memory management and allows the JRC-ENCR QCS to analyse big datasets (with millions of rows) efficiently as only one record at time is kept in memory to be processed. During the record cycle, issues can be detected in each record for single variables (e.g. data format, data type, data range) and/or between variables (e.g. coherence between topography and morphology, consistency between age and tumour type).

After a record has been checked, but before moving on to the next record, a special notification is sent according to all the rules (e.g. MPT rule) applied in the next post-record cycle. This notification contains the list of errors/warnings found in the processed record, including critical errors on some core variables (e.g. date of incidence or topography) that, according to specific acceptance criteria, prevent the software from applying such rules. The acceptance criteria and the related list of critical errors are specific for each rule.

For example, if the JRC-ENCR QCS identifies an error E-MISS (value missing) on the variable *YoI* (Year of Incidence) this is considered critical and the MPT rule of the post-record cycle (described in the following section) cannot be applied to the record with this type of error. The same error on the variable "*Stage*" is not critical and the record could be accepted and processed.

All acceptable records are stored in a temporary file. For example, if two rules 1 and 2 are defined in the post-record cycle, two log files will be produced (e.g. *qcs_acceptable_by_rule_1.csv*, *qcs_acceptable_by_rule_2.csv*), each with the list of records accepted by the specific rule according to its respective criteria.

These log files usually contain fewer records than the original dataset, depending on the acceptance criteria of the corresponding rule. For example, the list of records acceptable for the MPT rule will not include records with critical errors on core variables.

At the end of the record cycle, all errors and warnings produced for each record are added to the log files created in the first validation phase: these log files will be later used to produce the final output files.

2.7.3 Sorting of temporary files

The process of sorting temporary files produced in the record cycle occurs between the record and post-record cycles. Acceptable records stored in the temporary files are sorted according to some criteria, specific for each rule. For example, for the MPT rule all records are sorted by patient ID, so that consecutive records with the same patient ID can be easily analysed without scanning and saving into the computer memory the whole dataset for each patient ID.

The list of sorted acceptable records produced during this step is saved in a new temporary file. For example, file *qcs_acceptable_by_rule_1-by-PAT.csv* will include the list of records which are "acceptable" for the MPT rule and sorted by patient ID. These records are ready to be processed in the next validation cycle.

2.7.4 Post-record cycle

In the post-record cycle, a set of rules which perform specific checks between records is applied to the temporary files produced in the previous cycle. All records are compared to each other according to specific criteria defined for each rule. During this process, a set of specific errors and warnings is produced.

The new MPT rules have been defined and implemented in 2022 and will be available in the next version of the JRC-ENCR QCS.

When these rules are applied, records in the temporary files are read in batches, meaning that all records having the same patient ID are loaded and handled at the same time. At this point, some specific filtering criteria (e.g. on the tumour's behaviour) can be applied to exclude some records from the batch. For example, a record with topography code C50 (breast) and behaviour 0 (benign) will be ignored, while all tumours with behaviour 3 (malignant) will be processed, in line with the current call for data protocol (5).

For each batch of records with the same patient ID, some equivalence criteria on the morphology and topography codes are then applied to check if two tumours can be considered the same primary tumour according to the rules for checking solid MPTs defined in the quality check report "*A common data quality check procedure for European cancer registries*" (11).

2.8 Producing output reports

When all validation steps have been completed, the log files, which include all errors and warnings collected during the validation process, are used to produce the final output.

Each log file is sorted with respect to the original ordering. This is necessary to print all errors and warnings for each record exactly in the same order of the original dataset.

2.9 Outcome of the validation process

All the checks performed by the software are following the International Classification of Diseases for Oncology, Third Edition (ICD-O-3) topography and morphology codes (17).

After all the checks described above (univariate, multivariate and between records checks) have been executed, the JRC-ENCR QCS produces some output reports listing all messages collected during the validation process. Output messages can be warnings (i.e. the record should be reviewed by the PBCR) or errors (the record cannot be accepted as it is and should be corrected by the PBCR).

In some rare cases, the software can produce also critical messages, meaning that something went fatally wrong. This could be the case of a "broken rule" (e.g. the user tampered with the configuration files and removed a resource used by a specific check) or it could be the case of a dataset that cannot be read correctly (e.g. wrong number of columns, wrong separator between variables).

Two types of messages are printed in the output reports: W-YYYY (warning code) and E-YYYY (error code), where the code YYYY identifies the specific type of message.

Below some examples:

- W-AGMT: Unlikely Age + tumour type
- W-BDMO: Morphology too specific taking into account the basis of diagnosis
- W-MOTO: Morphology + Topography not valid
- E-FORM: Format error
- E-MISS: Value missing
- E-OUTR: Value out of range

The list of all error and warning codes used by the JRC-ENCR QCS is available in the header of the PDF and TXT output reports.

2.10 The output reports

Each validation run generates three output files in PDF, TXT and CSV format. The PDF file contains a summary of the execution process (date and time, name of the processed file, number of rows, total number of errors and warnings) and the detailed list of all errors and warnings detected for each record of the input data file. The TXT output file has the same content of the PDF, but in text format.

The CSV file contains all errors and warnings in a format easily readable by automated procedures. This file can be used by users to load the results of the validation process in a database, or to perform detailed statistics and analyses. This format can be particularly useful if the input file generates a large number of errors.

2.11 Protocol and application configurations

Starting from version 2.1 the software can perform the validation of the input data set according to different versions of the protocol, 2014 and 2020, corresponding respectively to the protocols for the JRC-ENCR calls for data in 2015 and 2022. These documents detail the guidelines for submitting to the JRC four types of data files: incidence, mortality, population and life tables. The software can run in 10 different modalities, according to the specific data call protocol. These modalities are referred to as "protocols":

- Incidence 2014: Incidence protocol 2014 (56 variables)
- Incidence 2020: Incidence protocol 2020 (39 variables)
- Mortality 2014: Mortality protocol 2014 (5 variables)
- Mortality 2020 - Age Unit: Mortality protocol 2020 "Age Unit" (5 variables)
- Mortality 2020 - Age Range: Mortality protocol 2020 "Age Range" (5 variables)
- Population 2014: Population protocol 2014 (4 variables)
- Population 2020 - Age Unit: Population protocol 2020 "Age Unit" (6 variables)
- Population 2020 - Age Range: Population protocol 2020 "Age Range" (6 variables)
- Life Table 2014: Life Table protocol 2014 (4 variables)
- Life Table 2020: Life Table protocol 2020 (6 variables)

A summary of the number and type of files needed to configure the JRC-ENCR QCS application is specified below:

- About 10 configuration files for the general configuration
- About 70 configuration files for defining the list of variables and the allowed range of values
- About 10 configuration files for defining the protocols and the list of rules
- About 40 configuration files for defining the internal logic of each validation rule

3 Results

3.1 The role of the JRC-ENCR QCS in the data processing workflow

As a first step, the PBCR should extract data from its database to create the incidence file, following the requirements of the ENCR call for data protocol (4). This step is not always straightforward, since data might be extracted with a slightly different format or structure, or the PBCR might use a different coding for the variables than the one specified in the protocol. Due to the large number of data submissions to JRC-ENCR deviating from the format requested in the data call protocol, either because of time or technical constraints from the PBCR side, the CSV Data Layout Converter has been developed as an auxiliary tool to facilitate the preparation of the incidence file by the PBCR before running the software.

In the second step, the PBCR should run the software on the four different files to be submitted (incidence, population, mortality, life table). All records having some issues will produce errors and/or

warnings in the output files, allowing the PBCR to verify and possibly correct the issues before data submission.

When the files are ready, the PBCR submits them through the secure JRC Data Submission Portal.

After data is submitted by the PBCR, JRC re-runs the JRC CSV Data layout converter on incidence data to check the adherence of the incidence file to the format required by the ENCR protocol and verifies that the file can be validated by the JRC-ENCR QCS. The preliminary format check is carried out by the JRC for all data files submitted by the PBCR. The results of this first check are communicated to the PBCR via the JRC Data Submission Portal with a preliminary format check report.

In case of critical format errors, it might not be possible for the JRC to correct them and run the software. In this case the PBCR will be asked to correct the format and submit again the dataset to JRC.

If the files are received in the correct format, the JRC will check them running the software and will prepare an internal consistency report. A summary of the submitted data, with the number of records for each type of error and warning, is included in this report. Issues raised by the JRC-ENCR QCS for which there is a clear solution (e.g. a prostate cancer case with topography C61 instead of C61.9) are included in the internal consistency report only to inform the PBCR on the proposed solution (e.g. topography is changed to C61.9).

If needed, the PBCR will fix the issues reported by the JRC and will re-submit the updated file through the JRC portal.

In order to apply the international rules for multiple primary tumours (18, 19), the JRC is defining and developing a new MPT algorithm for automatic selection of multiple primary tumours according to current international rules for these tumours. The MPT algorithm will be applied on incidence data only after the cleaning process.

The new advanced features that will be available with the JRC-ENCR QCS version 2.1, will offer several improvements:

- Full implementation of all the 10 protocols
- Better usage of the computer's memory during validation
- Better definition of multiple primary criteria (updated to the latest guidelines)
- More precise identification and classification of MPTs
- More precise filtering criteria for MPT with behaviour < 3 (non-malignant behaviours)
- A richer validation feedback to the user (both in the GUI and in the output reports)
- Updates of all univariate and multivariate checks according to the latest guidelines

Some more advanced improvements are planned for version 2.2, such as:

- Possibility to exclude duplicate records from the dataset
- Possibility to merge duplicate records into a unique one

The final steps performed at JRC are data aggregation, calculation of quality indicators such as incidence rates, and publication of the results in the ECIS web application.

Figure 2 illustrates the validation workflow of an incidence file. Parallel considerations could be done for the other types of files (mortality, population, and life table).

3.2 History and roadmap of the JRC-ENCR QCS development

The development of the software occurred in subsequent versions, following incorporation of feedback from several PBCRs. In 2015 the first version of JRC-ENCR QCS (1.4) was released for testing by European PBCRs. This version contained all basic checks (univariate and multivariate), but not the MPT checks. Some versions (1.5 and 1.6) were released in 2016 informally (at the JRC level, and to few pilot PBCRs) to test and verify the accuracy of the MPT check.

An updated version 1.7 of the software was officially released in November 2016: this was the first public available version which included the MPT check, and the possibility to run also on MacOS and Linux operating systems.

This release was followed by version 1.8 in December 2018, which included new features and many corrections, thanks to the feedback received by PBCRs. The main new features of JRC-ENCR QCS 1.8 were the creation of a separate file reporting on MPTs and the update of morphology families used by the MPT checks according to the 2011 update of ICD-O-3.

The latest release of the JRC-ENCR QCS is version 2.0, based on the guidelines of the ENCR data call protocol 2022 and on the experience gathered in validating more than 35 million cases submitted in the first JRC-ENCR 2015 data call from around 150 PBCRs based in 35 European countries.

Version 2.0 introduced TNM (Tumour/Nodes/Metastasis) (16) consistency checks and further updates of ICD-O-3 morphology codes. Moreover, the new 2.0 version introduced a completely new architecture, which moved all protocol data from the source code to some configuration files. This new approach made it possible to update the logic of the majority of checks without the need to release a new version of the software, but simply updating the configuration files.

PBCRs can download the latest version of the JRC-ENCR QCS toolkit, which include the *CSV Data Layout Converter* tool and the latest version of the JRC-ENCR QCS, from the ENCR website (20).

The JRC in collaboration with the ENCR has been organising several training sessions to familiarise PBCRs staff with the software (21).

3.3 Using the JRC-ENCR QCS

The software use is rather straightforward; after launching it the user interface window is opened (Figure 3).

The user can choose the type of file and the data call protocol from the drop-down menu. By selecting "*Incidence 2014 (56 variables)*", the validation checks are performed according to the 2015 data call protocol, whereas "*Incidence 2020 (39 variables)*" corresponds to the rules of the 2022 data call.

The "*Select File*" button allows browsing and selecting the file to be checked, and the "*Start checks*" button starts the validation

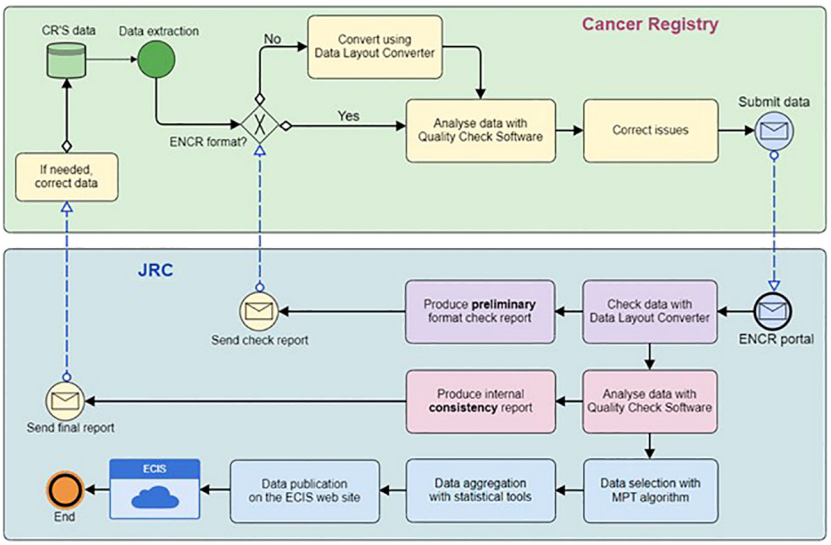


FIGURE 2
ECIS cancer burden data flow from submission to publication.

process. While the software is running, the number of the checked records appears in the display text box. Once the validation is ended, the output window displays a short report about the completed process, while the "*Open*" button allows accessing the output folder containing all the output report files.

Similarly, mortality, population and life table files can be checked by selecting the corresponding type of file from the drop-down menu.

Figures 4–7 show some examples of errors and warnings given by the JRC-ENCR QCS. Error codes start with an E and warning codes start with a W. The following are examples of a univariate check and of some multivariate checks within the same record. The last example regards a check between different records of the same patient.

For each error and warning the Patient (*Pat*) and Tumour (*Tum*) identifiers are reported. Some essential information on the cancer case is also included: *BoD* (Basis of Diagnosis), *Topo* (Topography), *Morpho* (Morphology), *Beh* (Behaviour), *Sex*, *DoI* (Date of Incidence) and *DoB* (Date of Birth). Finally, the variable(s) that triggered the warning/error are reported (*Var_Name*) with the respective value(s) (*Var_Value*) and resulting code (*Error_Code*).

In addition, the JRC-ENCR QCS is reporting a summary table with type of warnings and errors, and the number of records for each one, to give a general overview of data quality to the user, and help in setting the priorities for reviewing the data (Figure 8).

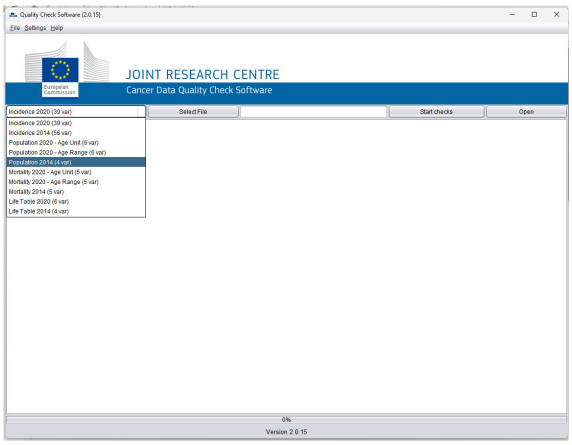


FIGURE 3
The user interface of the JRC-ENCR QCS (Version 2.0.15 example).

3.4 Use of the JRC-ENCR QCS: downloads and trainings

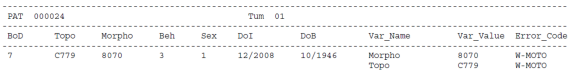
During the period August 2022 - January 2023 there were overall 139 unique downloads of the software. Thirty-two users were from Spain, 28 from Italy, 11 from Germany, 10 from France, 9 from Poland, while the remaining were from about 20 additional countries.

The JRC used the JRC-ENCR QCS for validating the data submitted by the European PBCRs in the 2015 data call for the calculation of incidence and mortality indicators in ECIS. A total of more than 35 million cases from around 150 PBCRs based in 35 countries were processed.

Several training sessions were organised by JRC in collaboration with the ENCR to familiarise PBCRs staff with the software, with around 300 participants trained European-wide.

PAT	000001				Tum	02			
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C427	9800	3	2	9/2010	2/1924	Topo	C427	E-OUTR

FIGURE 4
Error for out of range value for topography (E-OUTR). Topography C427 does not exist in the ICD-O-3 classification.



PAT	000024							Tum	01
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C779	8070	3	1	12/2008	10/1946	Morpho Topo	8070 C779	W-MOTO W-MOTO

FIGURE 5
Warning for morphology and topography combinations (W-MOTO). The combination of topography=C779 (Lymph node, NOS) and morphology=8070 (squamous cell carcinoma, NOS) is probably a metastasis and topography should be coded as C809 for the unknown primary site.

4 Discussion

The JRC-ENCR QCS is a Java standalone desktop application, under development since 2015, created to support PBCRs in their data validation processes.

The software is freely downloadable from the ENCR website (19), allowing the user to work locally and to share data at a later stage in an anonymised/pseudonymised format for European projects. This feature is particularly relevant with respect to compliance to the data protection rules detailed in the GDPR (13).

Since the first release in 2015, the software has been incorporating the updated European and International recommendations and classifications, such as the ICD-O-3.2 (22), the morphology grouping table for the purpose of defining multiple tumours based on the ICD-O-3.2 (19), TNM 8th revision (16) and the ENCR recommendations (23, 24).

Checks related to stage at diagnosis have been improved in the software since its first release. The latest JRC-ENCR QCS version introduced the consistency check of TNM and stage values.

Until the development of the JRC-ENCR QCS, the majority of European PBCRs used *IARCcrgTools* for data validation, a tool developed by the International Agency for Research on Cancer (IARC) (25). Some of the checks performed by the JRC-ENCR QCS related to the core variables are similar to the ones implemented by the *IARCcrgTools*. Therefore, it is possible for non-European PBCRs to use the JRC-ENCR QCS in their data validation process.

The main differences between the two software are the checks related to the TNM staging system, which are implemented in the JRC-ENCR QCS (26). This is a major strength of the software, given the increased number of PBCRs collecting stage, and therefore the necessity to validate this key information. Additionally, the JRC-ENCR QCS also differs from IARC's software as for checking topography and morphology combinations. While *IARCcrgTools* follows the groups of families included in Appendix 1 of the "Check and Conversion Programs for Cancer Registries" manual (25), the JRC-ENCR QCS is developed around table 8 of the JRC Technical Report (11) which considers each ICD-O-3 topography and morphology code.

In addition, the checks implemented in the JRC-ENCR QCS follow the Call for Data Protocols for European PBCRs variables, formats and allowed values (5, 13). Checks within records (within a single variable or between variables) and between records (duplicate and multiple primary tumours) are performed according to the JRC Technical reports on data quality checks (9–11). The fact that the JRC-ENCR QCS follows the specifications of the Call for Data Protocols for European PBCRs contributes to the improvement of the harmonisation and comparability of data across Europe.

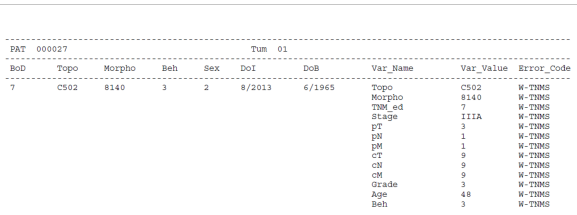
The file format, the number and type variables and the specific ENCR Recommendations should be taken into account by non-European PBCRs using the JRC-ENCR QCS.

To facilitate the submission to the JRC and improve the quality of PBCRs data, the ENCR and the JRC recommend checking beforehand the format of required files and data internal consistency with the JRC-ENCR QCS.

To support the use of the software, the JRC organised in collaboration with the ENCR and other stakeholders several training sessions, with an overall participation of around 300 PBCRs staff so far. In addition, the JRC developed the JRC CSV Data layout converter, a tool to further facilitate the use of the software by PBCRs.

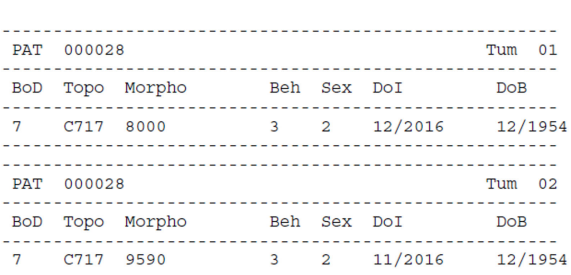
The JRC-ENCR QCS is used by the 15 Swiss PBCRs for the annual submission to the National Agency for Cancer Registration (NACR) (27). The findings are reported back to Swiss PBCRs for correction/verification, and once resolved the data is integrated into the Swiss National Cancer Dataset. Moreover, the software has been used to check data in several studies from European PBCRs and other research institutes (28–35).

The feedback from the JRC-ENCR QCS users has been essential for improving and adapting the software to European PBCRs needs. This feedback allowed to refine the software algorithms in each



PAT	000027							Tum	01
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C502	8140	3	2	8/2013	6/1965	Topo Morpho TNM_ed Stage pT pN pM cT cN cM Grade Age Beh	C502 8140 7 IIIA 3 1 1 9 9 9 3 48 3	W-TNMS W-TNMS W-TNMS W-TNMS W-TNMS W-TNMS W-TNMS W-TNMS W-TNMS W-TNMS W-TNMS W-TNMS W-TNMS

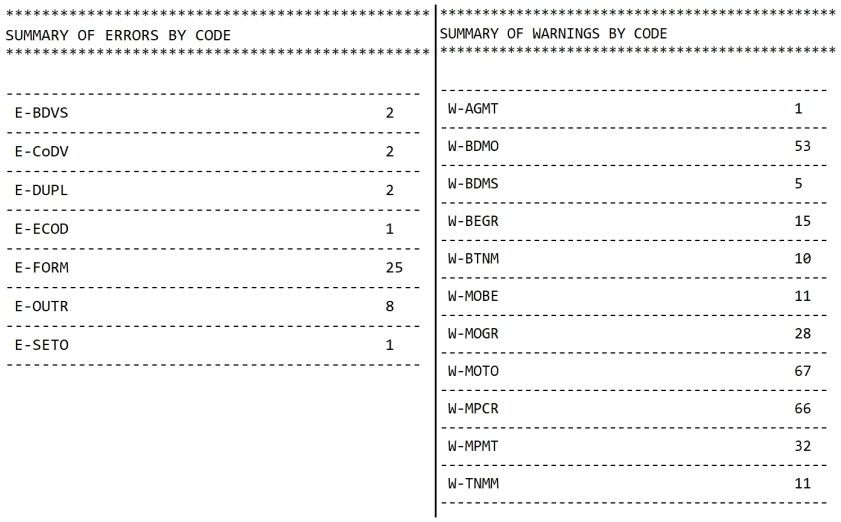
FIGURE 6
Warning for inconsistency between TNM and stage (W-TNMS). This case is a breast carcinoma with pT=3, pN=1, pM=1 and Stage=IIIA. This combination is not consistent with the TNM classification (7th edition); this means that either pM is actually 0, or stage is equal to IV.



PAT	000028					Tum	01
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	
7	C717	8000	3	2	12/2016	12/1954	

PAT	000028					Tum	02
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	
7	C717	9590	3	2	11/2016	12/1954	

FIGURE 7
Warnings for multiple primary tumours. In this example of multivariate check, the software gives warning for multiple primary tumours because the two records are reporting the same tumour, and only one should be considered for the calculation of cancer burden indicators.



SUMMARY OF ERRORS BY CODE	
E-BDVS	2
E-CoDV	2
E-DUPL	2
E-ECOD	1
E-FORM	25
E-OUTR	8
E-SETO	1

SUMMARY OF WARNINGS BY CODE	
W-AGMT	1
W-BDMO	53
W-BDMS	5
W-BEGR	15
W-BTNM	10
W-MOBE	11
W-MOGR	28
W-MOTO	67
W-MPCR	66
W-MPMT	32
W-TNMM	11

FIGURE 8
Example of a summary table of errors and warnings provided by the JRC-ENCR QCS.

release and to increase the precision, the completeness and clarity of the software reports.

The flexibility and the performance of the JRC-ENCR QCS has been enhanced over time. In the latest release an innovative approach was introduced, moving all protocol data from the source code to configuration files. This new approach has made it possible to update the logic of majority of checks without the need to release a new software version, but simply updating the configuration files. If needed, user can customise different tables, modifying or introducing specific values.

Data availability statement

The original contributions presented in the study are included in the article. The JRC-ENCR QCS is accessible at: https://encr.eu/sites/default/files/QCS/jrc-qcs-2.0.zip. The user compendium of the JRC-ENCR QCS is accessible at: https://encr.eu/sites/default/files/QCS/JRC127031_jrc127031_jrc-encr_qcs_user_compendium_v2_0_20220929.pdf. The executables of the JRC-ENCR QCS will be published at: https://code.europa.eu/ecis/jrc-encr-qcs-binaries. Further inquiries can be directed to the corresponding authors.

Author contributions

The first draft of the manuscript was written by FG and CM. EB supervised the IT process. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors acknowledge Carine Dochez (Belgian Cancer Registry) for her editorial assistance in the preparation of this manuscript. The authors acknowledge Antonino Brunetto (JRC) for the technical expertise provided in the preparation of this manuscript. The authors acknowledge the following institutions and entities for the invaluable feedback provided running the present and previous versions of the JRC-ENCR QCS: • East Switzerland Cancer Registry • OmniSoftory Engineering S.A., Switzerland • National Institute for Cancer Epidemiology and Registration (NICER), Switzerland • National Agency for Cancer Registration (NACR), Switzerland • Swiss Childhood Cancer Registry • Basel Cancer Registry, Switzerland • Bern Cancer Registry, Switzerland • St. Gallen Appenzell Cancer Registry, Switzerland • Samara Cancer Registry, Russian Federation • Ragusa and Caltanisetta Cancer Registry, Italy Trapani Cancer Registry, Italy • Cancer Registry of Puglia – Section of Health Local Unit Barletta – Andria – Trani, Italy • Veneto Cancer Registry, Italy • Varese Cancer Registry, Italy • Castellon Cancer Registry, Spain • Basque Country Cancer Registry, Spain • Girona Cancer Registry, Spain • Tarragona Cancer Registry, Spain • Greater Poland Cancer Registry • Polish National Cancer Registry • Cancer Registry of Republic of Slovenia • Morphological Tumour Registry, Luxembourg • Estonian Cancer Registry • Finnish Cancer Registry • Cancer Registry of Norway • Isère Cancer Registry, France • Loire-Atlantique/Vendée Cancer Registry, France • Association of the French Cancer Registries (FRANCIM) • Association of Population-based Cancer Registries in Germany (GEKID) • Cancer Registry of Baden-Württemberg, Germany • National Pediatric Cancer Registry of Hungary • National Cancer Registry Ireland.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Parkin DM. The role of cancer registries in cancer control. *Rev Int J Clin Oncol* (2008) 13(2):102–11. doi: 10.1007/s10147-008-0762-6

2. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Action Against Cancer*. European Partnership. /* COM/2009/0291 final */.

3. ECIS - European Cancer Information System. (2023). Available at: https://ecis.jrc.ec.europa.eu (Accessed 12 March 2023). © European Union.

4. *JRC Data Submission Portal*. Available at: https://portal-encr.jrc.ec.europa.eu/ (Accessed 31 March 2023).

5. *Call for data protocol for European population-based cancer registries*. Available at: https://www.encr.eu/sites/default/files/Data_call/ECIS%20call%20for%20data%20protocol_20221124.pdf (Accessed 31 March 2023).

6. Bray F and Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: Comparability, validity and timeliness. *EJC* (2009) 45:747–55. doi: 10.1016/j.ejca.2008.11.032

7. Parkin DM and Bray F. Evaluation of data quality in the cancer registry: principles and methods. Part II: Completeness. *EJC* (2009) 45:756–64. doi: 10.1016/j.ejca.2008.11.033

8. Bray F, Colombet M, Mery L, Piñeros M, Znaor A, Zanetti R and Ferlay J eds. *Cancer incidence in five continents, vol. XI*. Lyon: International Agency for Research on Cancer (2021). Available at: https://publications.iarc.fr/597.

9. Martos C, Crocetti E, Visser O, Rous B and the Data Quality Check Working Group. *A proposal on cancer data quality checks: one common procedure for European cancer registries*. Luxembourg (Luxembourg: Publications Office of the European Union (2014). JRC93456. EUR 27008.

10. Martos C, Crocetti E, Visser O, Rous B, Giusti F and the Cancer Data Quality Checks. *A proposal on cancer data quality checks: one common procedure for European cancer registries – version 1.1*. Luxembourg: Publications Office of the European Union (2018). JRC105078. doi: 10.2760/429053

11. Martos C, Giusti F, Van Eycken E and Visser O. A common data quality check procedure for European cancer registries. *Technical Report* Ispra, Italy: European Commission (2023). JRC132486

12. Giusti F, Martos C, Adriani S, Flego M, Brunetto A, Dyba T, et al. *The JRC-ENCR Quality Check Software (QCS) for the validation of cancer registry data: user compendium – version 2.0*. Ispra: European Commission (2022). JRC127031.

13. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General Data Protection Regulation).

14. European Network of Cancer Registries. *ENCR-JRC call for data* (2015). Available at: https://encr.eu/sites/default/files/Data_call/2015_ENCR_JRC_Call_for_Data_Version_1_1.pdf (Accessed 12 March 2023).

15. ENCR Recommendations. Available at: https://encr.eu/ENCR-Recommendations (Accessed 7 August 2023).

16. Brierley JD, Gospodarowicz MK and Wittekind C eds. *The TNM classification of Malignant tumours. 8th edition*. Oxford: Wiley Blackwell (2017).

17. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al. *International classification of diseases for oncology. 3rd ed*. World Health Organization, Geneva, Switzerland (2013). First Revision (ICD-O-3.1).

18. *International rules for multiple primary cancers (ICD-O third edition)* (2004). Lyon: IARC. Available at: http://www.encr.eu/images/docs /recommendations/MPrules_july2004.pdf (Accessed 31 March 2023).

19. *Groups of Malignant neoplasms considered to be histologically 'different' for the purpose of defining multiple tumours*. Available at: http://www.iacr.com.fr/index.php?Itemid=577 (Accessed 31 March 2023). ICD-O-3.2.

20. ENCR. *Tools for registries*. Available at: https://encr.eu/sites/default/files/QCS/JRC-QCS-KIT-V1.0.zip (Accessed 31 March 2023).

21. ENCR. *Materials from trainings and workshops*. Available at: https://encr.eu/node/535 (Accessed 7 August 2023).

22. *International classification of diseases for oncology*. World Health Organization. Available at: http://www.iacr.com.fr/index.php?Itemid=577 (Accessed 31 March 2023). Second Revision (ICD-O-3.2).

23. ENCR Recommendations. *Recording and reporting of urothelial tumours of the urinary tract* (2022). Available at: https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation_UT_Jun2022_EN.pdf (Accessed 31 March 2023).

24. ENCR Recommendations. *Basis of diagnosis* (2022). Available at: https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation%20BoD_Oct2022_EN_241022.pdf (Accessed 31 March 2023).

25. Ferlay J, Burkhard C, Whelan S and Parkin DM. *Check and Conversion programs for cancer registries* (IARC/IACR Tools for cancer Registries). Lyon (2005). IARC technical Report No. 42.

26. Tagliabue G, Perotti V, Fabiano S, Tittarelli A, Barigelletti G, Contiero P, et al. Comparison between two cancer registry quality check systems: functional features and differences in an Italian network of cancer registries dataset. *Front Oncol* (2023) 13:1197942. doi: 10.3389/fonc.2023.1197942

27. Matthias Lorez, NACR and Ulrich Wagner, NACR. *New data quality concept within the framework of the federal law on the registration of cancer (Cancer Registration Act, CRA; SR 813.33)* (2022). Available at: https://www.nkrs.ch/assets/files/uploads/nacr-data-quality-concept-v1.pdf (Accessed 20 January 2023).

28. Spitale A, Ortelli L, Mazzucchelli L and Bordoni A. Quality indicators of colorectal cancer care in Southern Switzerland: results from a population-based study. *Swiss Med Wkly* (2017) 147:w14530. doi: 10.4414/smw.2017.14530

29. Ortelli L, Spitale A, Mazzucchelli L and Bordoni A. Quality indicators of clinical cancer care for prostate cancer: a population-based study in Southern Switzerland. *BMC Cancer* (2018) 18:733. doi: 10.1186/s12885-018-4604-2

30. Blum M and Mousavi M. Die häufigsten Krebserkrankungen nach Geschlecht in den Kantonen Graubünden und Glarus in den Jahren 2016 bis 2020. (2022).

31. Sanatatii M. *Sistem informatic pentru registrele de sănătate – "RegInterMed". Caiet de sarcini pentru servicii de dezvoltare și implementare sistem informatic, inclusiv servicii de instruire*. Available at: https://management-documente.ms.ro/media/documents/Caiet-de-sarcini-RegInterMed.pdf (Accessed 20 January 2023).

32. Stegmaier C, Hentschel S, Hofstädter F, Katalinic A, Tillack A and Klinkhammer-Schalke M. *Das Manual der Krebsregistrierung*. München: W. Zuckschwerdt Verlag (2019).

33. Meléndez PG, Sanz RÁ, Iglesias LE and López SG. *Registro poblacional de tumores infantiles de Castilla y León. Incidencia y supervivencia de los tumores infantiles en Castilla y León 2010 – 2020*. Valladolid: Junta de Castilla y León. Consejería de Sanidad (2022). Dirección General de Salud Pública.

34. *Malignant neoplasms in Montenegro 2013*. Podgorica: Institute of Public Health of Montenegro, Center for Control and Prevention of Non-communicable Diseases, Registry of Malignant Neoplasms of Montenegro (2018).

35. Guevara M, Molinuevo A, Salmerón D, Marcos-Gragera R, Carulla M, Chirlaque MD, et al. Cancer survival in adults in Spain: a population-based study of the spanish network of cancer registries (REDECAN). *Cancers* (2022) 14(10):2441. doi: 10.3390/cancers14102441

frontiers | Frontiers in Oncology

TYPE Methods
PUBLISHED 12 December 2023
DOI 10.3389/fonc.2023.1250549

OPEN ACCESS

EDITED BY
Hajo Zeeb,
Leibniz Institute for Prevention Research and Epidemiology (LG), Germany

REVIEWED BY
Paulo S. Pinheiro,
University of Miami, United States
Juan Carlos Núñez-Enríquez,
Instituto Mexicano del Seguro Social, Mexico
Lois Dickie,
National Cancer Institute (NIH), United States

*CORRESPONDENCE
Otto Visser
o.visser@iknl.nl

RECEIVED 30 June 2023
ACCEPTED 22 November 2023
PUBLISHED 12 December 2023

CITATION
Visser O, Kościańska B, Caetano dos Santos FL, Cuccaro F, Forjaz G, Michalek IM, Mousavi M, Sulkowska U, Martos C and Giusti F (2023) 2022 revised European recommendations for the coding of the basis of diagnosis of cancer cases in population-based cancer registries. *Front. Oncol.* 13:1250549. doi: 10.3389/fonc.2023.1250549

2022 revised European recommendations for the coding of the basis of diagnosis of cancer cases in population-based cancer registries

Otto Visser[1]*, Beata Kościańska[2], Florentino Luciano Caetano dos Santos[3], Francesco Cuccaro[4], Gonçalo Forjaz[5], Irmina Maria Michalek[3], Mohsen Mousavi[6], Urszula Sulkowska[7], Carmen Martos[8] and Francesco Giusti[8]

[1]Department of Registration, Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, Netherlands, [2]Lublin Cancer Registry, Lublin, Poland, [3]Polish National Cancer Registry, Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland, [4]Cancer Registry of Puglia, Section of Health Local Unit Barletta–Andria–Trani, Barletta, Italy, [5]Westat, Rockville, MD, United States, [6]Graubünden and Glarus Cancer Registry, Chur, Switzerland, [7]Mazowiecki Cancer Registry, Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland, [8]European Commission, Joint Research Centre, Ispra, Italy

The basis of diagnosis recommendations for population-based cancer registries aim to provide a standardized coding tool that reflects the certainty of cancer diagnosis, especially when pathological confirmation is lacking. The proportion of clinical diagnoses serves as an indicator of data quality. Given the evolving nature of diagnostic techniques, regular revision of the basis of diagnosis rules is crucial. To address this, a working group comprising representatives from the steering committee and member registries of the European Network of Cancer Registries was established. The original 1999 recommendations were comprehensively reviewed, resulting in the publication of an updated version. These new recommendations came into effect for incident cancer cases starting from January 1, 2023. The updated recommendations comprise an adapted code list for the basis of diagnosis, optional codes for histology cases, revisions related to flow cytometry, liquid biopsy, and cytogenetic/molecular testing, consolidation of histology codes 6 and 7, introduction of a new code 8 for cytogenetic/molecular confirmation, and establishment of new criteria for registering specific morphology codes in cancers lacking pathological confirmation.

KEYWORDS

basis of diagnosis, recommendations, cancer, cancer registration, Europe

Introduction

The methods used to diagnose cancer have greatly improved over time. While pathological diagnosis is still the gold standard, an increasing number of cancers, such as hepatocellular carcinoma, can be diagnosed using modern imaging techniques with acceptable certainty without pathological confirmation (1), Imaging techniques are especially relevant for cancer cases which require invasive (and potentially harmful) techniques to obtain a sample for pathological examination, such as tumors of the pancreas, liver, and central nervous system.

The most valid basis of diagnosis is one of the key variables in population-based cancer registries (2). International studies show that there is a large variation in the distribution of the basis of diagnosis of registered cancer cases. For example, in a study of Berrino et al. the proportion of microscopically verified cases ranged from 79% in Poland to 98% in Sweden (3). This may be due to real variation but may also be caused by differences in registration practices or in interpretation of the coding. Clear guidelines for the coding of the basis of diagnosis should reduce these differences in interpretation and contribute to the comparability of the data.

The aim of the Basis of Diagnosis Recommendations by the European Network of Cancer Registries (ENCR) is to provide guidelines to European cancer registries for defining the level of certainty of the diagnosis of cancer (4). This is particularly relevant in the absence of a pathological confirmation of cancer. The proportion of clinical diagnoses (basis of diagnosis codes 1, 2, or 4) is an indicator of the quality of the data of a cancer registry. While a high proportion of clinical diagnoses in a cancer registry may well reflect the extent of the clinical and pathological investigations in the registry area, especially in developing countries, it may also indicate an overestimation of the cancer incidence. For example if non-malignant lesions without pathology are erroneously included in a cancer registry, cancer incidence will be inflated. Besides, cancer survival will be overestimated, as the risk of dying from a non-malignant disease will generally be much lower than from a malignancy.

In registries with a (very) low proportion of clinical diagnoses, there may be an underestimation of cancer incidence due to incomplete notification of clinically and/or radiologically confirmed cancer cases. In many cancer registries, notification of pathologically confirmed cases is better organized than notification of cancer cases with a clinical diagnosis only. Consequently, these cancer registries run the risk of incompleteness of cancers, such as lung cancer, pancreatic cancer, and several hematological malignancies which are not confirmed pathologically in a considerable proportion of cases.

Traditionally, cancer cases without pathological confirmation were coded by cancer registries with an unspecified morphology code according to ICD-O, i.e. 9990/3 in the first edition (5) and 8000/3 as of the second edition (6). For several cancer entities, exceptions were made to this rule, as indicated by the 1999 ENCR recommendations (7). Since then, imaging techniques have improved, and additional techniques have become available, such as molecular diagnostics, which has increased the number of cancer entities which may be diagnosed with reasonable certainty in the absence of pathology. Therefore, these recommendations required further revision.

Methods

During the summer of 2021, the ENCR initiated a call for expressions of interest from member registries to form a working group (WG) with the purpose of updating the ENCR recommendations on the basis of diagnosis, originally published in 1999. The primary objective of this WG with expertise in cancer registration, epidemiology, pathology and radiotherapy, was to enhance the comparability of incidence and survival data between different European registries and countries. Following the establishment of the WG, a proposal was formulated by one of its members (OV). Subsequently, an online meeting took place on October 27, 2021, during which the proposal was deliberated upon. An amended proposal, agreed upon by all members, was circulated. The draft recommendations were then scrutinized and endorsed by the ENCR Steering Committee (SC) on November 9, 2021. Following the SC's approval, the recommendations were disseminated to all ENCR members for consultation. Fourteen cancer registries provided feedback, which was subsequently discussed among the WG members. Based on this discussion several modifications were incorporated. On June 8, 2022, the SC granted final approval to the revised recommendations, which were subsequently published on the ENCR website on October 20, 2022. Lastly, on November 30, 2022, a webinar was organized, specifically for registry staff from ENCR institutions, to provide a detailed explanation of the new recommendations.

Results and discussion

The recommendations (8) include an adapted code list for the basis of diagnosis, as presented in Table 1. Additionally, Table 2 provides optional codes for cases with histology as the basis of diagnosis. The revisions made to the previous version of the recommendations pertain primarily to flow cytometry, liquid biopsy, and cytogenetic and/or molecular testing. Furthermore, the original code 6 (histology of metastasis) has been merged with code 7 (histology of primary tumor) into consolidated code 7, which now includes histology of primary tumor, histology of metastasis, and histology at autopsy. As a result, code 6 is no longer used in the updated recommendations. Additionally, a new code 8 has been introduced for cancer cases with cytogenetic or molecular confirmation of the diagnosis, which was not present in the original recommendations.

Furthermore, a compilation of cancers has been created, which may be registered with a specific morphology based on clinical information or clinical investigations when pathology results are unavailable. The list is presented in Table 3. In exceptional cases, other specific cancers may be diagnosed through clinical investigations; however, assignment of a specific morphology code should only be performed after careful evaluation by a coding expert from the cancer registry.

TABLE 1 Basis of diagnosis codes.

Code	Description	Criteria
0	Death certificate only (DCO)	Information provided is from a death certificate.
1	Clinical	Diagnosis made before death, but without any of the following (codes 2-8).
2	Clinical investigation	All diagnostic techniques, including X-ray, endoscopy, imaging, ultrasound, exploratory surgery (such as laparotomy), and autopsy, without a tissue diagnosis.
4	Specific tumor markers	Including biochemical and/or immunologic markers that are specific for a tumor site.
5	Cytology	Examination of cells from a primary or secondary site, including fluids aspirated by endoscopy or needle; also includes the microscopic examination of peripheral blood and bone marrow aspirates, immunophenotyping by flow cytometry and a liquid biopsy[#] in the absence of pathology.
7	Histology	Histologic examination of tissue from the tumor (primary or metastatic), however obtained, including all cutting techniques and bone marrow biopsies; also includes autopsy specimens of the tumor.
8	Cytogenetic and/or molecular testing	Detection of tumor-specific genetic abnormalities or genetic changes in the tumor, including techniques such as karyotyping, FISH (fluorescent *in situ* hybridization), PCR (polymerase chain reaction), DNA sequencing
9	Unknown	

[#] a liquid biopsy is a sample of blood or another body fluid (liquor, etc.) for the detection of cancer cells or DNA-fragments of these tumor cells.

Table 4 presents a roster of cancers that can be diagnosed using elevated tumor markers in conjunction with clinical investigations and when pathology is not available.

When utilizing these tables in registration practice, the following rules should be observed.

1) Use the highest code from the range 1-8 (Table 1), unless it is a 'death certificate only' (DCO) case (basis of diagnosis 0) or if the basis of diagnosis cannot be determined (basis of diagnosis 9).

The order of the codes for the basis of diagnosis (from 1 to 8) represents an increasing reliability of the cancer diagnosis. The highest code within the range should be assigned to represent the most reliable basis of diagnosis.

TABLE 2 Optional codes for cases with histology basis of diagnosis.

Code	Description	Criteria
7.1	Histology of the primary tumor	Histologic examination of tissue from the primary tumor, however obtained, including all cutting techniques and bone marrow biopsies.
7.2	Histology of a metastasis	No histology of the primary tumor
7.3	Histology at autopsy	No histology before autopsy

2) Use code 0 when trace back from the death certificate is not possible. DCO cases should be registered with morphology code 8000, unless the morphology code can be derived from the ICD code (C43 [8720/3], C45 [9050/3], C46 [9140/3], and C81-C96/D45-D47 [9590/3-9989/3]) or from the text on the death certificate (e.g., 'adenocarcinoma of the stomach' or 'rhabdomyosarcoma').

Limited information is generally available for DCO cases, but even with only a coded cause of death, the morphology can be deducted in several instances. Some registries have access to detailed information on the death certificate, which should be used for morphology coding if available.

3) Code 1 should only be used for cancers that are detected by physical examination only. This includes cancers of head and neck, eye, breast, skin and superficial soft tissues, external genitals, vagina, cervix, anus, rectum, and prostate. It is almost impossible to diagnose a cancer in most inner organs (such as the lung, stomach, colon, or kidney) with physical examination only, but rare exceptions are possible.

Only a few cancers may be diagnosed with physical examination alone. As physical examination is typically followed by a biopsy and/or imaging in most cases, the number of cases with physical examination as the basis of diagnosis is very small.

4) Codes 1 and 2 may be used when a diagnosis of cancer is at least **likely** ('probably cancer'). If clinical investigations reveal that a cancer diagnosis is possible, the case should not be registered in the absence of pathological confirmation (basis of diagnosis 5-8).

To avoid overestimating the number of cancers, cases should only be registered when the symptoms or appearances are most likely caused by cancer. If multiple disorders, including cancer, could explain the symptoms or appearances, the case should not be registered. For example, if the diagnosis includes 'large lesion in the left cerebellum, differential diagnosis arterial malformation, low grade neuronal tumor' the case should not be registered, as a non-malignant disorder could also explain the symptoms or appearances.

5) Cancers registered with basis of diagnosis 1 or 2 should be assigned morphology code 8000/3 (8000/0 or 8000/1 are also allowed for benign and borderline malignant tumors of the central nervous system). Exceptions to this rule are listed in Table 3. These exceptions apply only to cases where a specific diagnosis is at least **likely**. If the diagnosis is only **possible** or multiple diagnoses are mentioned in the clinical file or report, the case should be registered with morphology code 8000/3 (8000/0 or 8000/1 are also allowed for benign and borderline malignant tumors of the central nervous system).

Table 3 provides an overview of specific tumor entities that may be diagnosed using imaging or physical examinations. If the diagnosis in the clinical report is relatively certain, that specific diagnosis should be coded. For example, if the report states 'lesion in the frontal lobe, typical for glioblastoma', the morphology code of glioblastoma (9440/3) should be used in combination with basis of diagnosis 2 (clinical investigation). However, if the report states 'low-grade lesion in the temporal lobe; differential diagnosis DNET, ganglioglioma, low-grade astrocytoma' morphology code 8000/1 should be used.

TABLE 3 Cancers that may be registered with a specific morphology based on clinical information (basis of diagnosis code 1) or clinical investigations (basis of diagnosis code 2).

Cancer type	Basis of diagnosis code	ICD-O topography code	ICD-O morphology code
Melanoma			
- Melanoma of the skin	1	C44	8720/3
- Melanoma of the eye	1 or 2	C69.0, C69.3, C69.4	8720/3
Solid childhood cancers (age <15 years)			
- Nephroblastoma	2	C64	8960/3
- Hepatoblastoma	2	C22	8970/3
- Retinoblastoma	1 or 2	C69.2	9510/3
Hepatocellular carcinoma	2	C22.0	8170/3
Cholangiocarcinoma	2	C22.1, C24.0, C24.9	8160/3
Non-functioning neuroendocrine tumors (NETs)			
- Non-functioning NET of the pancreas	2	C25.4	8150/3
- Non-functioning NET of the small intestine	2	C17	8240/3
Intraductal papillary mucinous neoplasm (IPMN)	2	C25	8453/2, 8453/3
Sarcoma			
- Sarcoma, NOS	2	*	8800/3
- Liposarcoma	2	*	8850/3
- Leiomyosarcoma	2	*	8890/3
- Angiosarcoma	1** or 2	*	9120/3
- Kaposi sarcoma of the skin	1	C44	9140/3
- Osteosarcoma - Chondrosarcoma	2 2	C40, C41 C40, C41	9180/3 9220/3
- Chordoma	2	C41.0	9370/3
Central nervous system (CNS) tumors			
- Mature teratoma, cystic teratoma	2	C71, C75.1, C75.3	9080/0
- Teratoma, NOS	2	C71, C75.1, C75.3	9080/1
- Immature teratoma, malignant teratoma	2	C71, C75.1, C75.3	9080/3
- Hemangioblastoma	2	C71, C72.0	9161/1
- Craniopharyngioma	2	C75.2	9350/1
- Pinealoma	2	C75.3	9360/1
- Pineocytoma	2	C75.3	9361/1
- Pineoblastoma	2	C75.3	9362/3
- Glioma, NOS	2	C71, C72.0	9380/39
- Low grade glioma	2	C71, C72.0	9380/32
- High grade glioma	2	C71, C72.0	9380/33
- Subependymoma	2	C71.5, C71.7	9383/1
- Subependymal giant cell astrocytoma	2	C71.5, C71.7	9384/1
- Choroid plexus papilloma	2	C71.5, C71.7	9390/0
- Atypical choroid plexus papilloma	2	C71.5, C71.7	9390/1

(Continued)

TABLE 3 Continued

Cancer type	Basis of diagnosis code	ICD-O topography code	ICD-O morphology code
- Choroid plexus carcinoma	2	C71.5, C71.7	9390/3
- Ependymoma	2	C71.5, C71.7, C72.0	9391/3
- Anaplastic ependymoma	2	C71.5, C71.7, C72.0	9392/3
- Myxopapillary ependymoma	2	C72.0, C72.1	9394/1
- Papillary tumor of the pineal region	2	C75.3	9395/3
- Astrocytoma, NOS	2	C71, C72.0	9400/39
- Low grade astrocytoma	2	C71, C72.0	9400/32
- High grade/anaplastic astrocytoma	2	C71, C72.0	9401/33
- Desmoplastic infantile astrocytoma/desmoplastic infantile ganglioglioma	2	C71	9412/1
- Dysembryoplastic neuroepithelial tumor	2	C71	9413/0
- Pilocytic astrocytoma	2	C71, C72.0	9421/1
- Optic nerve glioma, optic chiasm glioma in children	2	C72.3	9421/1
- Glioblastoma	2	C71, C72.0	9440/3
- Oligodendroglioma, NOS	2	C71	9450/39
- Low grade oligodendroglioma	2	C71	9450/32
- High grade/anaplastic oligodendroglioma	2	C71	9451/33
- Medulloblastoma, NOS	2	C71.6	9470/3
- Embryonal tumor of the CNS, NOS	2	C71, C72.0	9473/3
- Gangliocytoma	2	C71, C72.0, 75.1	9492/0
- Dysplastic gangliocytoma of the cerebellum	2	C71.6	9493/0
- Ganglioglioma	2	C71, C72.0	9505/1
- Neurocytoma	2	C71	9506/1
- Multinodular and vacuolating neuronal tumor	2	C71	9509/0
- Glioneural tumor	2	C71, C72.0	9509/1
- Meningioma, NOS	2	C70	9530/0
- Atypical meningioma	2	C70	9539/1
- Anaplastic (malignant) meningioma	2	C70	9530/3
- Schwannoma	2	C72.4, C72.5	9560/0
Hematological malignancies			
- Primary lymphoma of the central nervous system	2	C71	9590/3
- Langerhans cell histiocytosis	2	C34, C41, C71***	9751/3

Other specific cancers not listed here may be diagnosed through clinical investigations. A specific morphology code should only be applied after evaluation by a coding expert of the cancer registry.
NOS, not otherwise specified.
* Sarcomas can be localized at any site, but mostly occur in the soft tissues, including the retroperitoneum and the mediastinum.
** Angiosarcoma of the (skin of the) breast following radiotherapy of the breast.
*** Other sites are possible.

6) Code 4 (specific tumor markers) should always be used in combination with a clinical diagnosis of cancer and/or a clinical investigation showing cancer since many tumor markers, such as prostate-specific antigen (PSA), may also be increased in the absence of cancer. The cancers that may be registered with basis of diagnosis 4 are listed in Table 4.

Although tumor markers may be increased in many cancers, they can also be increased in the absence of cancer. Therefore, when

TABLE 4 Cancers that can be diagnosed based on an elevated tumor markers in combination with clinical investigations.

Cancer type	Tumor marker	ICD-O Morphology code
Colorectal cancer	Carcinoembryonic antigen (CEA)	8000/3
Hepatocellular carcinoma	Alfa-fetoprotein (AFP)	8170/3
Pancreatic cancer, cancer of the gallbladder/bile ducts	Cancer antigen 19-9 (CA 19-9)	8000/3
Ovarian cancer	Cancer antigen 125 (CA-125)	8000/3
Prostate cancer	Prostate-specific antigen (PSA)	8000/3
Choriocarcinoma of the placenta	Human chorionic gonadotropin (HCG)	9100/3
Germ cell tumor	Human chorionic gonadotropin (HCG)	9064/3
	Alfa-fetoprotein (AFP) (+/- HCG)	9065/3
Neuroendocrine tumor	Chromogranin A	8240/3
Functioning neuroendocrine tumors (excluding pituitary gland tumors)	Insulin Glucagon Gastrin Vasoactive intestinal peptide (VIP) Somatostatin Serotonin Adrenocorticotropic hormone (ACTH) and other hormones	8151/3 8152/3 8153/3 8155/3 8156/3 8241/3 8158/3
Medullary thyroid carcinoma	Calcitonin	8345/3
Neuroblastoma	Catecholamine degradation products (Homovanilic acid [HVA], Vanillylmandelic acid [VMA])	9500/3
Prolactinoma	Prolactin	8271/0
Other functioning pituitary gland tumors	Growth hormone, Follicle-stimulating hormone (FSH), Luteinizing hormone (LH), Adrenocorticotropic hormone (ACTH), thyroid stimulating hormone (TSH)	8272/0
Phaeochromocytoma	Catecholamines, Chromogranin A	8700/3
Multiple myeloma	M-protein (IgG, IgM, IgA) >30g/L	9732/3
Waldenström's macroglobulinemia	IgM	9761/3

coding basis of diagnosis 4, it should always be accompanied by a clinical diagnosis (for example increased PSA in combination with a malignant appearance of the prostate at rectal examination) or a clinical investigation (for example increased alfa-fetoprotein in combination with LiRADS 6).

7) Flow cytometry is often used for the diagnosis of leukemia and lymphoma, such as chronic lymphocytic leukemia.

Flow cytometry is classified with the same code as cytology since it utilizes cell suspension.

8) If a genetic abnormality specific to cancer is found through 'liquid' biopsy (in combination with a clinical diagnosis of cancer, but in the absence of pathological confirmation), basis of diagnosis 5 should be applied.

A liquid biopsy involves detection of cancer cells or fragments of DNA from these cancer cells in blood or other body fluids. In cases where no pathological information is available, a liquid biopsy is classified with the same code as cytology.

9) Codes 7.1-7.3 are optional for cases with histology.

Although not necessary for international comparison, several registries may choose to distinguish various categories of histology: histology of the primary tumor, histology of a metastatic site, and histology at autopsy. While these categories are equal in terms of the certainty of the diagnosis, the different categories may be useful for other purposes, such as staging or cross-checks (e.g., coding histology from a metastatic site (7.2) means that the patient has metastatic disease).

10) Many tumors have genetic abnormalities, but only a few are specific to the diagnosis of a certain cancer. Basis of diagnosis 8 should be used only when the genetic abnormality is specific for that cancer. In most cases, the abnormality should be present (e.g., CML, BCR-ABL1+ is 9875/3), but there are also cancer diagnoses characterized by the absence of a genetic abnormality (e.g., glioblastoma IDH wild type is 9445/3). Basis of diagnosis 8 applies to both examples.

Our understanding of cancer cells and their genetic properties has improved significantly in recent decades. Specific genetic abnormalities have been identified in an increasing proportion of cancers, leading to the classification of cancer entities based on these abnormalities. While some cancers already have separate morphology codes for cases with and without cytogenetic/molecular confirmation, others do not.

Hence, basis of diagnosis 8 was introduced to distinguish cases with and without cytogenetic/molecular confirmation until specific morphology codes are available for cancer entities defined by genetic abnormalities. Basis of diagnosis 7 should be used for cases in which cytogenetic/molecular diagnostics were not performed, but a pathological diagnosis was available. This code may become obsolete in the future if specific morphology codes will become available for cancer entities that are defined by genetic abnormalities.

Conclusion

The updated recommendations introduce an adapted code list for the basis of diagnosis and incorporate new techniques, while maintaining consistency with the original version. Consolidating histological codes and introducing a new code for cytogenetic/molecular confirmation enhances the accuracy and specificity of cancer diagnoses. Additionally, the inclusion of specific morphological codes based on clinical information or

investigations improves the classification of cancers without pathological confirmation. These updates will have a minimal impact on cancer registry operations and contribute to reducing the number of cases with unspecified morphological codes, particularly for central nervous system tumors, leading to enhanced international comparability of cancer registry data.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

All authors contributed to the development of the ENCR Recommendations. The first draft of the manuscript was written by OV. All authors critically reviewed and approved the final manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Chernyak V, Fowle KJ, Kamaya A, Kielar AZ, Elsayes KM, Bashir MR, et al. Liver imaging reporting and data system (LI-RADS) version 2018: imaging of hepatocellular carcinoma in at risk patients. *Radiology* (2018) 289(3):816–30. doi: 10.1148/radiol.2018181494

2. Jensen O, Parkin D, MacLennan R, Muir C, Skeet R. Cancer registration: principles and methods. (1991).

3. Berrino F, De Angelis R, Sant M, Rosso S, Bielska-Lasot M, Coebergh JW, et al. Survival for eight major cancers and all cancers combined fro European adults diagnosed in 1995-99: results of the EUROCARE-4 study. *Lancet Oncol* (2007) 8 (9):773–83. doi: 10.1016/S1470-2045(07)70245-0

4. Tyczynski JE, Démaret E, Parkin DM. *Standards and Guidelines for Cancer Registration in Europe* Vol. 40. . Lyon: IARC Technical Publication No (2003).

5. World Health Organization. *International classification of diseases for oncology. 1st ed*. Geneva: World Health Organization (1976).

6. Percy C, Van Holten V, Muir C. *International classification of diseases for oncology. 2nd ed*. Geneva: World Health Organization (1990).

7. Smith J, Frost R, Teppo L, Visser O. *Recommendations for coding of basis of diagnosis.* (1999). European Network of Cancer Registries. Available at: https://encr.eu/sites/default/files/pdf/basisd.pdf (Accessed June 30, 2023).

8. Visser O, Kościańsk B, Caetano dos Santos FL, Cuccaro F, Forjaz G, Michalek IM, et al. *ENCR Recommendations Basis of diagnosis.* (2022). European Network of Cancer Registries. Available at: https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation%20BoD_Oct2022_EN_241022.pdf (Accessed June 30, 2023).

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 February 2024
DOI 10.3389/fonc.2024.1250107

OPEN ACCESS

EDITED BY
Maria Paula Curado,
A.C.Camargo Cancer Center, Brazil

REVIEWED BY
Chenghao Zhanghuang,
Kunming Children's Hospital, China
Renjun Peng,
Central South University, China

*CORRESPONDENCE
Francesco Giusti
francescogiusti@hotmail.com
Manola Bettio
JRC-ENCR@ec.europa.eu

†PRESENT ADDRESS
Carmen Martos,
Foundation for the Promotion of Health and Biomedical Research in the Valencian Region (FISABIO), Valencia, Spain

RECEIVED 29 June 2023
ACCEPTED 22 January 2024
PUBLISHED 20 February 2024

CITATION
Giusti F, Martos C, Bettio M, Negrão Carvalho R, Zorzi M, Guzzinati S and Rugge M (2024) Geographical and temporal differences in gastric and oesophageal cancer registration by subsite and morphology in Europe.
Front. Oncol. 14:1250107.
doi: 10.3389/fonc.2024.1250107

Geographical and temporal differences in gastric and oesophageal cancer registration by subsite and morphology in Europe

Francesco Giusti[1,2*], Carmen Martos[1†], Manola Bettio[1*], Raquel Negrão Carvalho[1], Manuel Zorzi[3,4], Stefano Guzzinati[4] and Massimo Rugge[4,5]

[1]European Commission, Joint Research Centre (JRC), Ispra, Italy, [2]Belgian Cancer Registry, Brussels, Belgium, [3]Regional Epidemiological Service Unit, Azienda Zero, Padova, Italy, [4]Veneto Tumor Registry, Azienda Zero, Padova, Italy, [5]Department of Medicine - DIMED, Surgical Pathology and Cytopathology Unit, Università degli Studi di Padova, Padova, Italy

Background: Gastric and oesophageal cancers pose a serious public health concern. In 2020 a total of 189,031 incident cases (136,038 stomach, 52,993 oesophagus) and 142,508 deaths (96,997 stomach, 45,511 oesophagus) were estimated in Europe. Oesophago-gastric cancers are a heterogeneous disease, with different aetiology and epidemiology for the various topographic subsites and main histopathological types. Topography subsite and morphology is key information to allow differentiating oesophago-gastric cancers. Correct registration and coding of such variables are fundamental in allowing proper description of the epidemiology of different subsites and histopathological types of oesophago-gastric cancers. The aim of this article is to highlight geographical and temporal variability in topography and morphology of oesophago-gastric cancers observed in Europe in the considered period.

Methods: Data collected in the framework of the ENCR-JRC (European Commission's Joint Research Centre) data call and feeding the European Cancer Information System (ECIS) were used to assess the variability of topography and morphology registration of gastric and oesophageal cancer in Europe in the period 1995-2014. Malignant cancers of the stomach and the oesophagus were selected following, respectively, topography codes C16 and C15 of the International Classification of Diseases for Oncology, third edition (ICD-O-3). Analyses were performed by subsite, morphology group, year, sex, and European region.

Results: A total of 840,464 incident cases occurring in the period 1995-2014 – 579,264 gastric (67.2%) and 276,260 (32.8%) oesophageal carcinomas – was selected for the analysis. Data was recorded by 53 PBCRs (9 based in Northern Europe, 14 in Western Europe, 3 in Eastern Europe and 27 in Southern Europe) from 19 countries.

10.3389/fonc.2024.1250107

Conclusion: A wide variability in oesophago-gastric cancers topographic subsites and histopathological types patterns was observed, with a corresponding improvement in accuracy of registration in the analysis period. PBCRs are ideally placed to guide the epidemiological evaluations of such a complex group of diseases, in collaboration with clinicians, patients and other public health stakeholders.

KEYWORDS

cancer registry, gastric cancer, Oesophageal cancer, data quality, Stomach/Oesophagus data quality

1 Introduction

Gastric and oesophageal cancers pose a serious public health concern. In 2020 a total of 189,031 incident cases (136,038 stomach, 52,993 oesophagus) and 142,508 deaths (96,997 stomach, 45,511 oesophagus) were estimated in Europe. Although incidence rates of gastric cancer have been decreasing in Europe since decades (1–4), due to ageing the burden of oesophago-gastric cancer is expected to further rise in absolute terms the in next decades, with an estimated 25% increase for gastric cancer (170,027 cases) and a 22% increase in oesophageal cancer (64,720 cases) by 2040 (5).

Oesophago-gastric cancers are a heterogeneous disease, with different aetiology and epidemiology for the various topographic subsites and main histopathological types. Stomach cancers can be classified in two anatomic sites, cardia gastric cancers (CGCs) (the upper part, next to the gastro-oesophageal junction) and non-cardia gastric cancers (NCGCs) (the lower part of the organ), and two main histological groups, diffuse and intestinal type. As for the stomach, oesophageal cancer are mainly adenocarcinomas; the two main histopathological subtypes are oesophageal squamous cell carcinoma (OSCC) and oesophageal adenocarcinoma (OAC) (6–8).

Helicobacter pylori infection is the most common cause of NCGCs, whereas it is not a known risk factor in CGC, and likely has a protective effect in OACs (9, 10). Tobacco smoking and alcohol consumption are the main risk factors for OSCCs, whereas Gastroesophageal reflux disease (GORD) and obesity are associated with OACs and CGCs (11, 12).

Population-based cancer registries (PBCRs) collect incidence data on all reportable neoplasms within a defined area. PBCRs have been in operation since the 1940s in a growing number of European countries, adhering to the international standards of the International Association of Cancer Registries (IACR) and the European Network of Cancer Registries (ENCR). PBCRs are currently evolving beyond their traditional role, adding critical information such as stage, treatment and biomarkers to their records. Among other purposes, such information can be also used to evaluate public health interventions, or inequalities in access to care (13–17).

Topography subsite and morphology is key information to allow differentiating oesophago-gastric cancers. Correct registration and coding of such variables is fundamental in allowing proper description of the epidemiology of different subsites and histopathological types of oesophago-gastric cancers.

This study aims to highlight geographical and temporal variability in topography and morphology of oesophago-gastric cancers observed in Europe in the period 1995-2013.

2 Methodology

Data collected in the framework of the ENCR-JRC (European Commission's Joint Research Centre) data call (18) and feeding the European Cancer Information System (ECIS) (1) were used to assess the variability of topography and morphology registration of gastric and oesophageal cancer in Europe in the period 1995-2014, corresponding to the latest 20 years currently available in ECIS. In order to reduce fluctuations in data patterns, population-based PBCRs which submitted at least 17 out of 20 incidence years in the considered time interval were included in the study. Since the number of cases was smaller, figures do not show data for the year 2014.

Malignant cancers of the stomach and the oesophagus were selected following, respectively, topography codes C16 and C15 of the International Classification of Diseases for Oncology, third edition (ICD-O-3) (19). Carcinomas and unspecified types of cancer were considered for the study, whereas haematological malignancies (particularly primary lymphomas), gastrointestinal stromal tumours, sarcomas and neuroendocrine tumours were excluded. Carcinomas and unspecified types of cancer were 98.3% of all solid oesophago-gastric cancers. Tumours with a death certificate as only basis of diagnosis (representing 4.1% of solid tumours) were also excluded.

For the analyses by subsite, ICD-O-3 topography codes were grouped as follows:

- C15.0-C15.3, **Upper one-third of the oesophagus**.
- C15.4, **Middle one-third of the oesophagus**.
- C15.5, **Lower one-third of the oesophagus**.

- C15.8, **Overlapping lesion of the oesophagus**.
- C15.9, **Oesophagus, Not Otherwise Specified (NOS)**.
- C16.0, **CGC**.
- C16.1-C16.6 **NCGC**.
- C16.8, **Overlapping lesion of stomach**.
- C16.9, **Stomach, NOS**.

Morphology groups with corresponding ICD-O-3 codes for cancers of the oesophagus were:

- **NOS neoplasm or carcinoma** (8000-8005, 8010-8015, 8020-8022, 8050).
- **OAC** (8140-8149, 8160-8162, 8190-8221, 8260-8337, 8350-8551, 8570-8576, 8940-8941).
- **OSCC** (8052-8078, 8083-8084).

For stomach cancers, histological categories, based on the Laurén (7) classification, were:

- **NOS neoplasm or carcinoma** (8000-8005, 8010-8015, 8020-8022, 8050).
- Adenocarcinoma, **intestinal type** (8144), including mucinous adenocarcinoma (8480-8481) and tubular adenocarcinoma (8211).
- Carcinoma, **diffuse type** (8145), including signet-ring cell carcinoma (8490).
- **Adenocarcinoma, NOS** (8140).
- **Other specified carcinomas**.

For geographical comparison, Europe was considered as divided in regions (Northern Europe, Western Europe, Eastern Europe and Southern Europe) following the United Nations Statistics Division scheme (20) (see also Figure 1).

Age-standardised incidence rates were calculated by year, sex, topography, morphology and region considering the 2013 European Standard Population (21). Two-sided 95% confidence intervals were calculated using the normal approximation method and reported next to incidence rates.

The analysis was performed using the statistical software SAS Version 9.3 (SAS Institute Inc., Cary, NC, USA).

3 Results

A total of 840,464 incident cases occurring in the period 1995-2014 – 579,264 gastric (67.2%) and 276,260 (32.8%) oesophageal carcinomas – was selected for the analysis. Data was recorded by 53 PBCRs (9 based in Northern Europe, 14 in Western Europe, 3 in Eastern Europe and 27 in Southern Europe) from 19 countries (Figure 1).

Overall, cancer cases in males were 539,830 (64.2%) and 300,634 in females (35.8%). Carcinomas occurred in males and females respectively for 61.1% and 38.9% of gastric cases, and 70.7% and 29.3% of oesophageal cases. The median age at diagnosis was 73 [interquartile range (IQR) 64-78] for gastric cancer (72 [IQR 63-77] for males and 74 [IQR 67-80] for females), and 71 [IQR 62-77] for oesophageal cancer (69 [IQR 60-75] for males and 74 [IQR 66-82] for females).

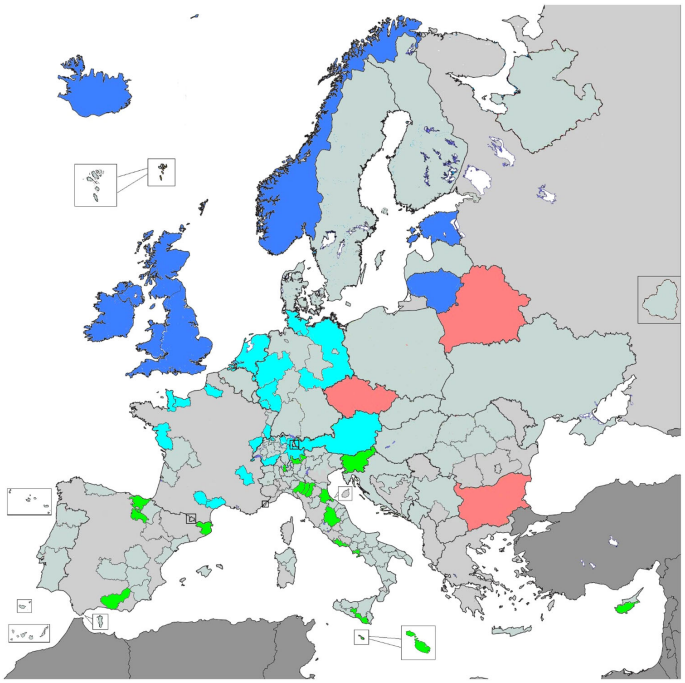


FIGURE 1
Selected PBCRs for the analysis, by European subregion. Blue corresponds to Northern Europe, cyan to Western Europe, red to Eastern Europe and green to Southern Europe (plus Cyprus).

3.1 Topography

3.1.1 Overall temporal analysis

Oesophageal cancer incidence grew overall from 9.2 [9.0-9.4] cases per 100,000 in 1995 to 10.5 [10.3-10.7] cases in 2013 (Supplementary Figure 1). An increase in incidence was observed in Northern Europe (from 12.6 [12.3-13.0] to 14.5 [14.1-14.9] cases per 100,000), Western Europe (from 6.9 [6.4-7.4] to 9.5 [9.1-9.8]) and Eastern Europe (from 3.5 [3.1-3.9] to 5.6 [5.2-6.0]), whereas oesophageal cancer incidence decreased in Southern Europe (from 7.1 [6.5-7.8] to 5.9 [5.2-6.5]) between 1995 and 2013 respectively (Supplementary Figure 2).

Analysis by anatomical subsite showed that incidence of carcinomas in the upper third of the oesophagus remained stable in the study period (0.9 [0.8-0.9] cases per 100,000 in 1995 – 1.0 [1.0-1.1] cases in 2013), whereas incidence in the middle third increased from 1.2 [1.1-1.2] to 1.8 [1.7-1.9] in the same period (Figure 2). Incidence of carcinomas to the lower third of the oesophagus rose from 2.6 [2.5-2.7] cases per 100,000 in 1995 to 5.2 [5.1-5.4] in 2013, showing opposite trends as compared to NOS cancers (likely to arise for the majority from the lower third), which decreased from 4.4 [4.3-4.6] to 2.2 [2.1-2.3] cases per 100,000 in the same period. Overall, the sum of lower third and NOS oesophageal cancer rates grew from 7.0 [6.9-7.1] cases per 100,000 in 1995 to 7.4 [7.3-7.5] in 2013 (Figure 2). Overlapping lesion was the subsite with the lowest incidence, with rates between 0.2 [0.2-0.2] and 0.3 [0.3-0.4] cases per 100,000 in 1995-2013.

An overall decline of gastric cancer incidence was observed in the period 1995-2013, with rates decreasing from 25.0 [24.6-25.3] to 15.0 [14.8-15.2] cases per 100,000 (Supplementary Figure 3). The decrease in incidence was observed in all European regions: Northern Europe (from 22.5 [22.0-23.0] to 11.4 [11.1-11.7] cases per 100,000), Western Europe (from 21.9 [21.2-22.6] to 14.4 [14.0-14.8]), Eastern Europe (from 31.2 [30.3-32.1] to 22.9 [22.1-23.7]) and Southern Europe (from 28.7 [27.5-29.8] to 15.7 [14.9-16.6]) between 1995 and 2013 respectively (Supplementary Figure 4).

The decreasing trend was due to the NOS subsite of stomach decline, which fell from 10,6 [10.3-10.8] to 3,4 [3.3-3.6] cases per 100,000, as well as to NCGCs, decreasing from 9.1 [8.9-9.3] to 6.9 [6.8-7.1] cases per 100,000. Rates for CGCs and overlapping lesions of the stomach were more stable, with respectively a mean rate of 3.7 [3.5-3.8] and 1.4 [1.4-1.5] cases per 100,000 in the analysis period (Figure 3).

3.1.2 Temporal analysis by geographical area

High variability was observed in the analysis by region, subsite and sex in oesophago-gastric carcinomas.

NCGC declined in all European regions, both for men and women between 1995 and 2013. For males, NCGCs had higher incidence rates than oesophageal carcinomas throughout all the analysis period in Southern and Eastern Europe, whereas oesophageal carcinomas took over NCGCs as the most incident oesophago-gastric cancer in 2002 in Northern Europe and in 2013 in Western Europe. Notably, NCGC rates remained high in Eastern Europe for both sexes, decreasing from 39.2 [37.4-40.9] to 29.2 [27.8-30.6] cases per 100,000 and from 20.0 [19.1-20.9] cases to 14.0 [13.2-14.8] cases per 100,000 for men and women respectively.

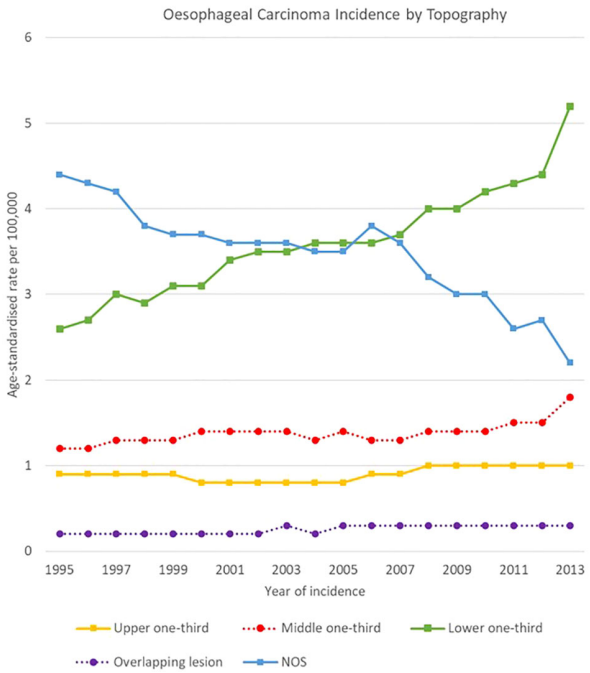


FIGURE 2
Time trends in oesophageal carcinoma incidence by topography, 1995-2013.

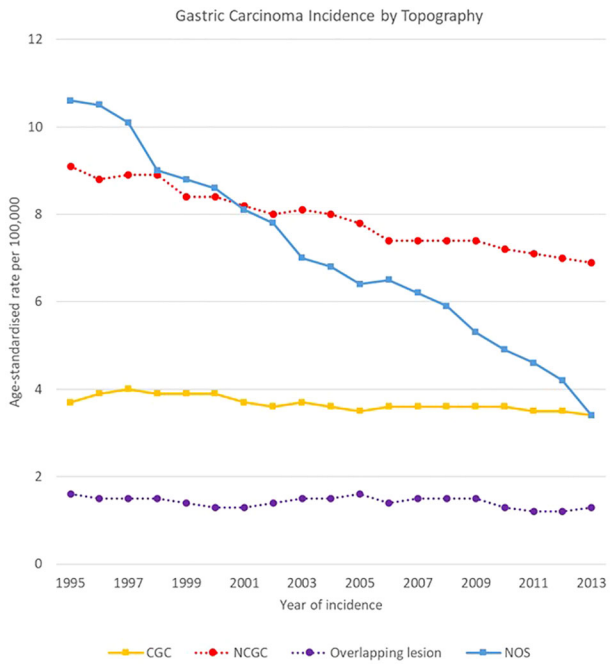


FIGURE 3
Time trends in gastric carcinoma incidence by topography, 1995-2013.

CGC rates were stable between 1995 and 2013 in all regions and for both sexes, with around 5-7 cases per 100,000 for men and around 1-2 cases per 100,000 for women (Figures 4–7).

3.2 Morphology

3.2.1 Oesophageal carcinomas

The geographical comparison of oesophageal carcinomas morphology groups showed rather different trends and ratios for males across the four European regions, while less variability occurred in females.

In Northern Europe, OSCC rates were stable (from 4.9 [4.4-5.3] to 4.8 [4.3-5.2] cases per 100,000), while OAC rates increased from 9.0 [8.5-9.6] to 15.0 [14.4-15.7] cases per 100,000. The ratio between OSCCs and OACs rose from 1 to 1.8 in 1995 to 1 to 3.1 in 2013. In Western Europe, OSCCs rates were also stable (from 5.8 [5.2-6.5] to 5.9 [5.4-6.3] cases per 100,000), and OACs increased (from 4.6 [4.0-5.2] to 8.8 [8.2-9.4] cases per 100,000). In Eastern Europe, incidence rates for both OACs and OSCCs increased between 1995 and 2013 from 1.1 [0.7-1.6] to 3.0 [2.4-3.5] cases and from 3.7 [3.0-4.4] to 7.1 [6.3-8.0] cases per 100,000 respectively; OACs were less common than OSCCs, with a ratio of 1 to 3.4 in 1995 and 1 to 2.4 in 2013. In Southern Europe, a sharp decline in OSCCs incidence was observed from 10.5 [9.2-11.8] to 5.8 [4.9-6.7] cases per 100,000, while OACs rates went from 1.9 [1.2-2.5] to 4.1 [3.3-4.8] cases per 100,000.

For females, incidence of OSCCs was higher than for OACs consistently in the four European subregions and for all the period 1995-2013. Among morphology groups, in 2013 the higher incidence in OSCC occurred in Northern Europe with 4.4 [4.0-4.8] cases per 100,000, and the lowest in Eastern Europe with 0.9 [0.7-1.2] cases per 100,000 (Figures 8–11).

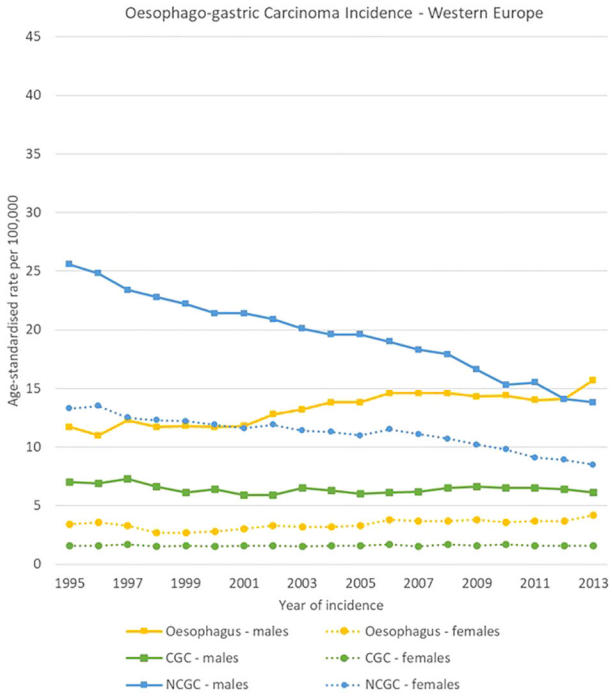


FIGURE 5
Time trends for oesophago-gastric carcinomas by topography and sex in Western Europe, 1995-2013.

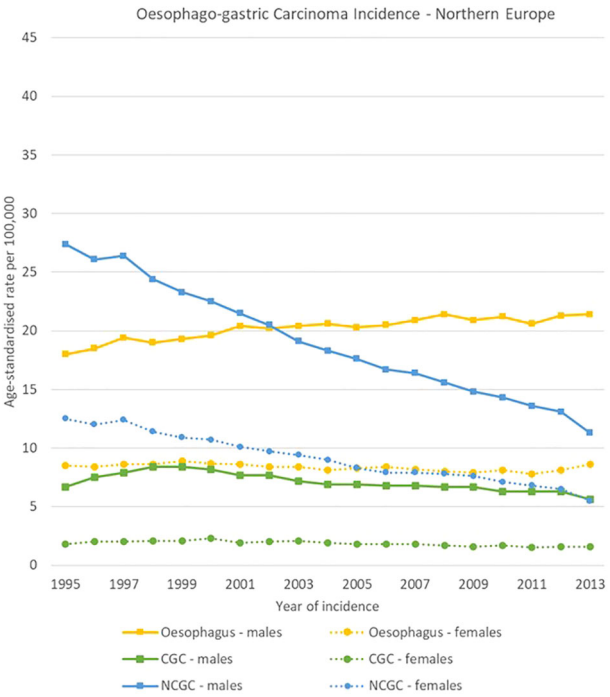


FIGURE 4
Time trends for oesophago-gastric carcinomas by topography and sex in Northern Europe, 1995-2013.

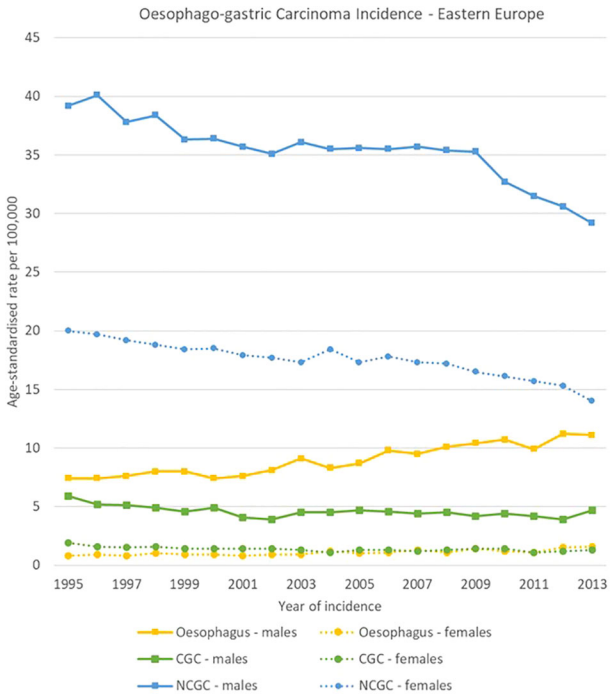


FIGURE 6
Time trends in oesophago-gastric carcinomas by topography and sex in Eastern Europe, 1995-2013.

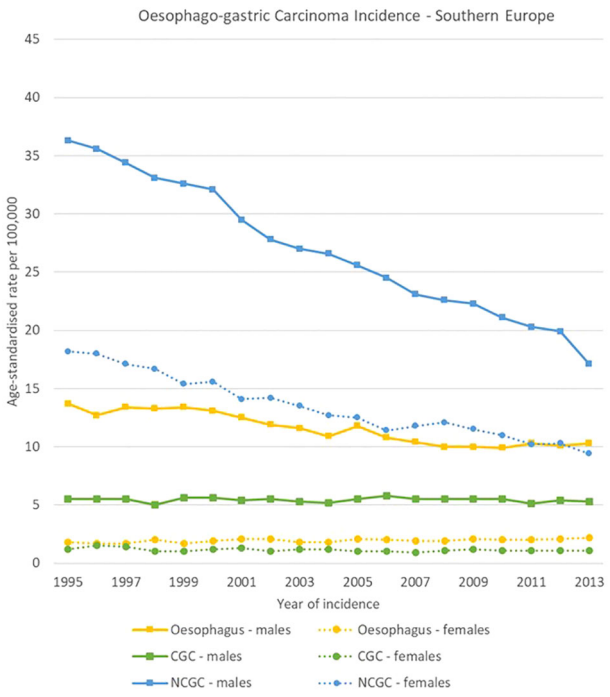


FIGURE 7
Time trends in oesophago-gastric carcinomas by topography and sex in Southern Europe, 1995-2013.

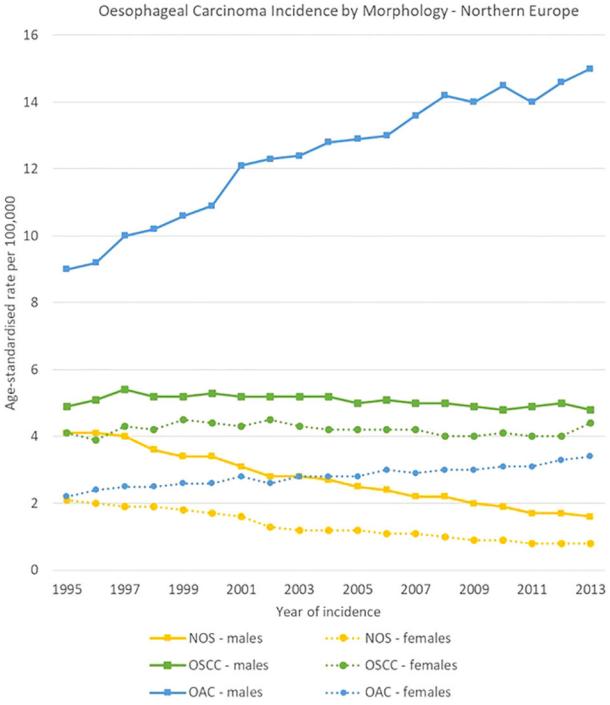


FIGURE 8
Time trends in oesophageal carcinomas by morphology and sex in Northern Europe, 1995-2013.

3.2.2 Gastric carcinomas

Morphology groups of stomach carcinomas also showed different trends in the analysis period. Incidence rates generally decreased in the two less specific groups (adenocarcinomas NOS and NOS neoplasms or carcinomas) and were stable instead for most of the other groups. In particular, NOS neoplasms or carcinomas decreased sharply in Eastern Europe, from 12.5 [11.9-13.0] cases per 100,000 in 1995 to 2.5 [2.3-2.8] cases per 100,000 in 2013, and adenocarcinomas NOS incidence fell from 11.8 [11.0-12.6] to 4.9 [4.4-5.4] cases per 100,000 in Southern Europe.

Specified morphology groups also showed variability among European regions. For instance, intestinal-type adenocarcinomas incidence changed from 5.4 [4.9-6.0] to 3.0 [2.5-3.4] cases per 100,000 in Southern Europe, and from 1.3 [1.2-1.4] to 1.0 [0.9-1.2] cases per 100,000 in Northern Europe. Higher incidence occurred for diffuse type carcinomas in Southern Europe, where it decreased from 4.1 [3.7-4.6] to 3.1 [2.7-3.5] cases per 100,000 in 1995-2013 respectively. Incidence rates in Northern Europe instead decreased from 1.4 [1.3-1.6] to 1.0 [0.9-1.1] cases per 100,000 in the same period (Figures 12–15).

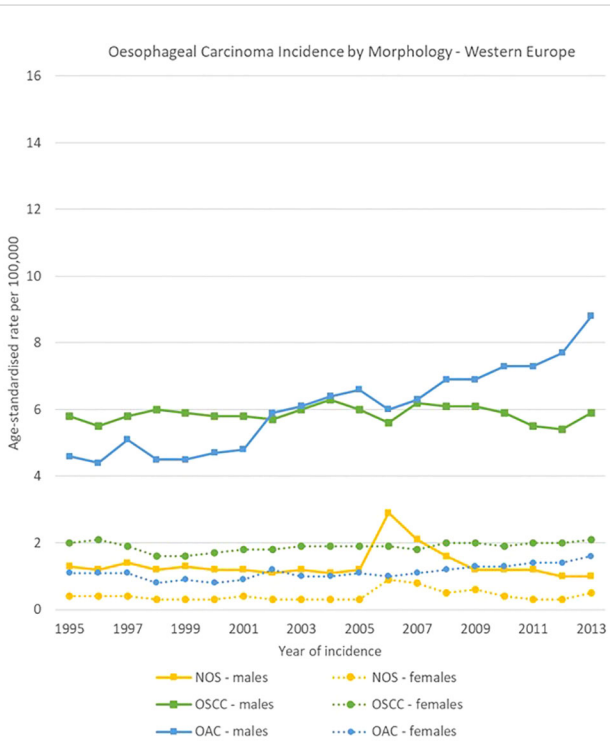


FIGURE 9
Time trends in oesophageal carcinomas by morphology and sex in Western Europe, 1995-2013.

4 Discussion

To our best knowledge, this is the first analysis addressing long-term changes in topography and morphology registration of both gastric and oesophageal cancers in Europe by area and sex.

Within upper gastrointestinal tract, cancer topography, combined with histological subtyping, have crucial impact on cancer prevention strategies, diagnostic procedures, the best choices of treatments and,

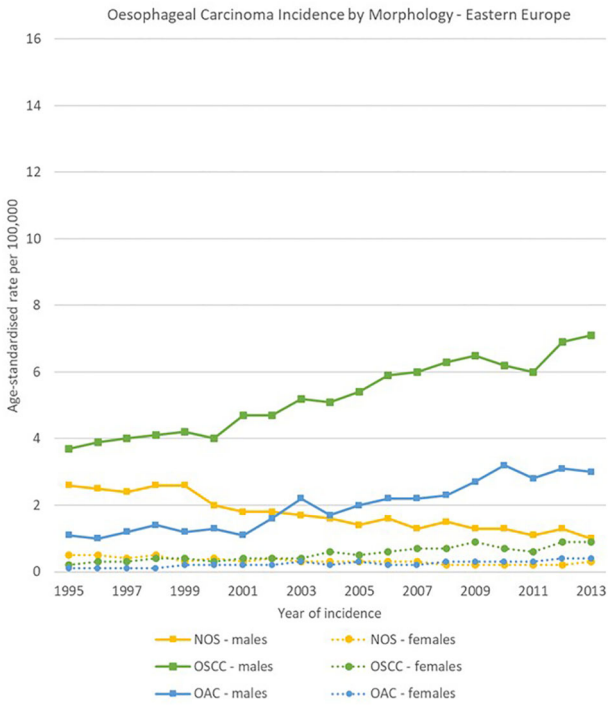


FIGURE 10
Time trends in oesophageal carcinomas by morphology and sex in Eastern Europe, 1995-2013.

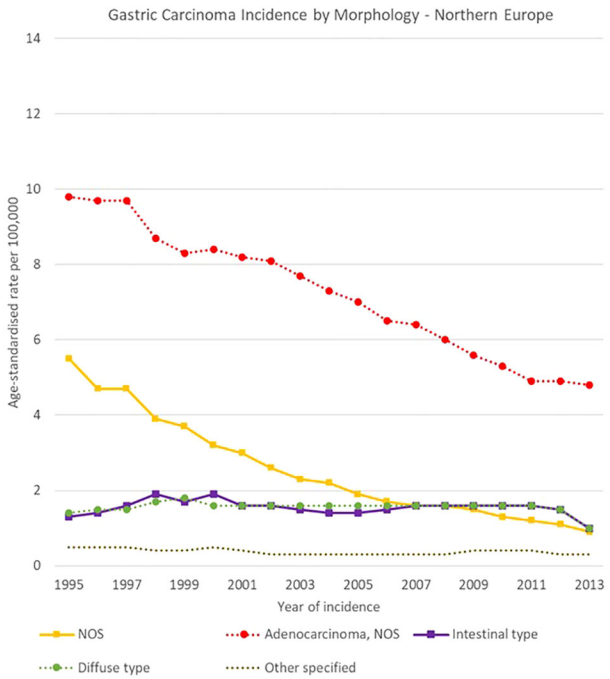


FIGURE 12
Time trends in gastric carcinomas by morphology in Northern Europe, 1995-2013.

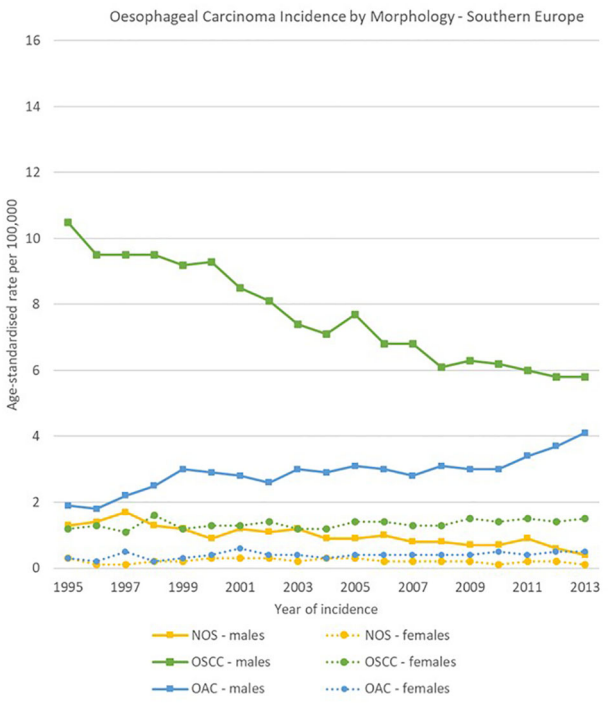


FIGURE 11
Time trends in oesophageal carcinomas by morphology and sex in Southern Europe, 1995-2013.

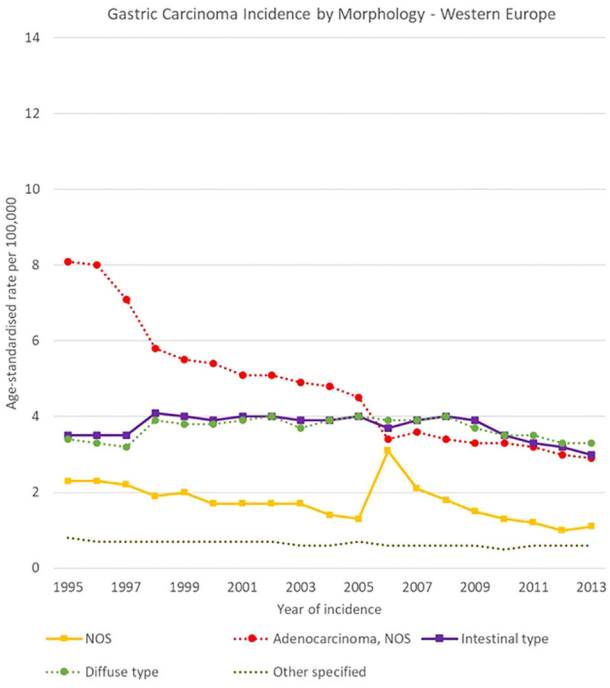


FIGURE 13
Time trends in gastric carcinomas by morphology in Western Europe, 1995-2013.

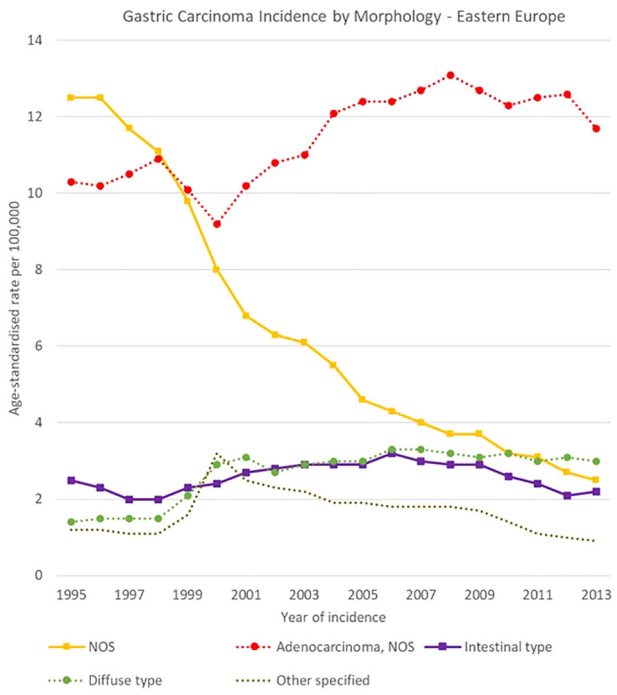


FIGURE 14
Time trends in gastric carcinomas by morphology in Eastern Europe, 1995-2013.

ultimately, in the estimates of prognosis. Unfortunately, these important data are unequally provided by non-high-resolution cancer registration, which results in fragmentary, and sometimes divergent, epidemiological information.

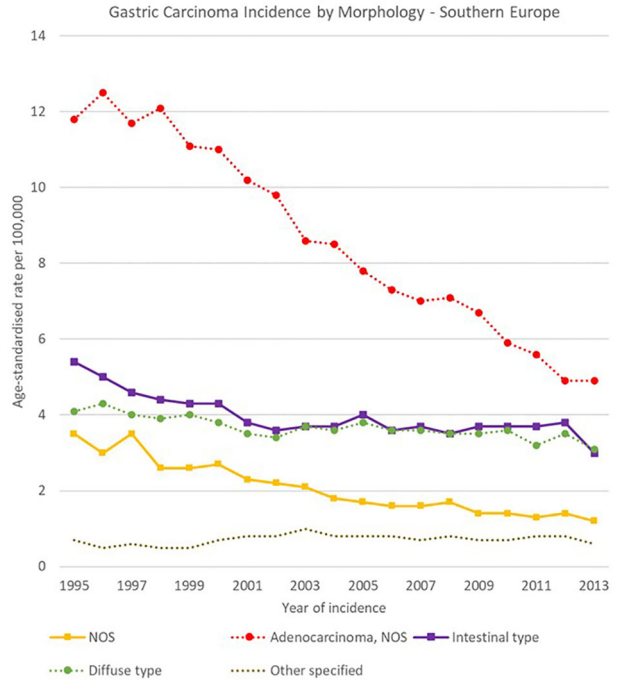


FIGURE 15
Time trends in gastric carcinomas by morphology in Southern Europe, 1995-2013.

Among the main strengths of the study are the large number of the selected cases (840,464 oesophago-gastric cancers), the length of the analysis period (1995-2014) and the geographical representation (53 PBCRs from Northern, Western, Eastern and Southern Europe).

In order to assess the representativity of data by European region, checks on incidence rates for gastric and oesophageal cancer in areas for which registration could not cover at least 17 years in the analysis period were performed. The incidence of gastric and oesophageal cancers was checked for all PBCRs with data available in 1995 and in 2013, and the rates were close overall to those reported in the study.

As a possible limitation, variations in data collection and coding practices could partly limit the generalisability of the study, even though a constant improvement in data quality has been reported for European PBCRs in the same period (22).

4.1 Different epidemiological patterns of cancer incidence across Europe

The present study revealed notable differences in the reporting and registration of upper gastrointestinal cancers across Europe, over time. Overall, it was found that there was an improvement in identifying specific subsites of oesophageal and gastric cancers between 1995 and 2013.

As for the oesophagus, there was a rising incidence of cancer in the lower one-third of the oesophageal channel, while the registration of malignancies with unspecified histology showed a decrease. Other subsites, however, did not exhibit significant changes.

Between 1995 and 2014, OAC increased consistently across European regions, in line with what already reported (6, 23–25). However, the extent of this increase varied by region and gender. Females, in particular, had higher incidence rates in Northern Europe than in Southern and Eastern Europe. These differences could be attributed to the varying effects of cancer-promoting factors such as gastro-oesophageal reflux, obesity, and low *Helicobacter pylori* prevalence in Northern European countries (12, 26–28) versus a cancer-protective effect observed in Southern and Eastern European populations due to higher rates of *Helicobacter pylori* infection (9, 10, 29).

The occurrence of OSCC varied among European regions. OSCC decreased only in males of Southern Europe, remained steady in Northern and Western Europe, and increased in both sexes in Eastern Europe. These differences could be attributed to the greater effects of tobacco smoking and alcohol consumption among Eastern European populations (11, 30).

Consistently with the current literature, gastric cancer incidence decreased dramatically (10, 23, 31, 32). In particular, NCGC rates became lower than those of oesophageal cancer in 2003 in Northern Europe, and in 2013 in Western Europe. The variability in gastric/oesophageal cancers ratio between European areas likely results from the differences in the prevalence of the different risk factors, particularly *Helicobacter pylori* infection (29).

Such evidence should prompt the implementation of internationally validated strategies for gastric cancer prevention (primary and secondary) (33, 34).

The decrease in NOS subsite proportion of gastric cancer coupled with stable rates of cardia-gastric adenocarcinoma. This was evidenced by the CGC : NCGC ratio, which remained steady from 1995 (1:2.5) to 2013 (1:2.1).

4.2 Data accuracy of cancer registration

4.2.1 Oesophageal and gastroesophageal-junction malignancies

Between 1995 and 2014, the quality of registration of oesophageal cancer improved significantly across Europe despite some variations among the regions. The incidence of malignancies with either unspecified subsite and/or morphology reduced overall across Europe.

Notably, in Northern European countries, the incidence of malignancies with unspecified morphology dropped from 4.1 to 1.6 cases per 100,000 in males, and specific cancer morphological groups such as oesophageal adenocarcinoma became more frequently identified, with an increase in incidence from 9.0 to 15.1 cases per 100,000. However, the incidence of OSCC did not show any significant change during this period for both sexes.

These findings underscore the importance of reducing the proportion of oesophageal cancers with unspecified morphology, given the different treatment strategies for OAC and OSCC (6, 35–39).

The improvement in diagnostic tools, such as high magnification and ultrasound endoscopy, played a significant role in promoting cancer histological subtyping (40). Along with that, the advent of digital pathology databases has helped to ensure consistent delivery of histological information from local pathology archives to centralised cancer registrations. Lastly, the HER-2 revolution has played a crucial role in the development of cancer-personalised therapies, making molecular profiling of oesophageal adenocarcinoma mandatory (38, 41).

4.2.2 Gastric malignancies

A strong decrease in the proportion of stomach NOS subsite was observed, with a decline in incidence from 10,6 to 3,4 cases per 100,000 between 1995 and 2013. The increase in the accuracy of registration of gastric cancers either as CGCs or NCGCs is crucial, given the different clinical characteristics and outcomes between the two subsites (6, 35, 37, 42, 43).

In addition, due to the differences (such as clinical features, genetics, surgery) between intestinal and diffuse type gastric cancers it is important for PBCRs to record the morphology of incident cancers with the best possible accuracy (44, 45).

With the only notable exception of Eastern Europe, the quality improvement in gastric cancer registration showed a trend even more favourable than that of oesophageal malignancies. Moreover, the drop in the incidence of NOS adenocarcinoma histotypes and NOS malignancies mirrors the lowering incidence of primary gastric epithelial malignancies as consistently documented (with the abovementioned exception) all over the considered European regions. This situation is consistent with the regional prevalence of the leading risk factor of gastric adenocarcinoma (*Helicobacter pylori*, with exceedingly higher prevalence in Eastern Europe) (29).

5 Conclusion and way forward

A wide variability in oesophago-gastric cancers topographic subsites and histopathological types patterns was observed, with a corresponding improvement in accuracy of registration in the analysis period. PBCRs are ideally placed to guide the epidemiological evaluations of such a complex group of diseases, in collaboration with clinicians, patients and other public health stakeholders.

JRC and ENCR have been supporting high level of quality and the harmonisation of European population-level cancer incidence data, though several initiatives such as drafting of recommendations for PBCR data coding, the organisation of trainings for PBCR personnel, and the developments of IT tools for data quality checks (46–50).

In the context of such activities, the present analysis on geographical and temporal differences in gastric and oesophageal cancer registration is a first step towards a more in-depth evaluation of the burden of these diseases in Europe.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Author contributions

The first draft of the manuscript was written by FG, CM and MR. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors acknowledge Carine Dochez (Belgian Cancer Registry) for her editorial assistance in the preparation of this manuscript. Grateful acknowledgement is given to the European cancer registries for feeding the ECIS with data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1250107/full#supplementary-material

References

1. ECIS - European Cancer Information System. (2023). European Union. Available at: https://ecis.jrc.ec.europa.eu (Accessed 12 March 2023).

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492

3. Brenner H, Rothenbacher D, Arndt V. Epidemiology of stomach cancer. *Methods Mol Biol* (2009) 472:467–77. doi: 10.1007/978-1-60327-492-0_23

4. Dyba T, Randi G, Bray F, Martos C, Giusti F, Nicholson N, et al. The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers. *Eur J Cancer* (2021) 157:308–47. doi: 10.1016/j.ejca.2021.07.039

5. *Cancer tomorrow* (2023). International Agency for Research on Cancer. Available at: https://gco.iarc.fr/tomorrow/en (Accessed 3 May 2023).

6. Arnold M, Ferlay J, van Berge Henegouwen MI, Soerjomataram I. Global burden of oesophageal and gastric cancer by histology and subsite in 2018. *Gut* (2020) 69 (9):1564–71. doi: 10.1136/gutjnl-2020-321600

7. Laurén P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand* (1965) 64:31–49. doi: 10.1111/apm.1965.64.1.31

8. Enzinger PC, Mayer RJ. Esophageal cancer. *N Engl J Med* (2003) 349(23):2241–52. doi: 10.1056/NEJMra035010

9. de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health* (2020) 8(2):e180–90. doi: 10.1016/S2214-109X(19)30488-7

10. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol Biomarkers Prev* (2014) 23(5):700–13. doi: 10.1158/1055-9965.EPI-13-1057

11. Pandeya N, Olsen CM, Whiteman DC. Sex differences in the proportion of esophageal squamous cell carcinoma cases attributable to tobacco smoking and alcohol consumption. *Cancer Epidemiol* (2013) 37(5):579–84. doi: 10.1016/j.canep.2013.05.011

12. Lindblad M, Rodríguez LA, Lagergren J. Body mass, tobacco and alcohol and risk of esophageal, gastric cardia, and gastric non-cardia adenocarcinoma among men and women in a nested case-control study. *Cancer Causes Control* (2005) 16(3):285–94. doi: 10.1007/s10552-004-3485-7

13. International Association of Cancer Registries. (2022). Available at: https://www.iacr.com.fr/ (Accessed 2 May 2023).

14. International Agency for Research on Cancer. (2022). Available at: https://www.iarc.who.int/events/ (Accessed 2 May 2023).

15. Siesling S, Louwman WJ, Kwast A, van den Hurk C, O'Callaghan M, Rosso S, et al. Uses of cancer registries for public health and clinical research in Europe: Results of the European Network of Cancer Registries survey among 161 population-based cancer registries during 2010-2012. *Eur J Cancer* (2015) 51(9):1039–49. doi: 10.1016/j.ejca.2014.07.016

16. Giusti F, Martos C, Trama A, Bettio M, Sanvisens A, Audisio R, et al. Cancer treatment data available in European cancer registries: Where are we and where are we going? *Front Oncol* (2023) 13:1109978. doi: 10.3389/fonc.2023.1109978

17. Voith von Voithenberg L, Crocetti E, Martos C, Dimitrova N, Giusti F, Randi G, et al. Cancer registries - guardians of breast cancer biomarker information: A systematic review. *Int J Biol Markers* (2019) 34(2):194–9. doi: 10.1177/1724600819836097

18. *ENCR-JRC Project 2015 - Updating population-based cancer incidence and mortality in Europe* (2015). Available at: https://www.encr.eu/sites/default/files/ENCR-JRC2015_Protocol.pdf (Accessed 12 March 2023). Call.

19. World Health Organization (WHO). *International classification of diseases for oncology 3rd ed*. Geneva: World Health Organization (2013). (ICD-O-3), 2nd revision.

20. *United Nations Statistics Division, Standard country or area codes for statistical use (M49)* . Available at: https://unstats.un.org/unsd/methodology/m49/ (Accessed 12 March 2023).

21. Eurostat. *Revision of the European Standard Population - Report of Eurostat's task force* (2013), edition. Luxembourg: Publications Office of the European Union, 2013.

22. Giusti F, Martos C, Negrão Carvalho R, Van Eycken L, Visser O, Bettio M. Quality indicators: completeness, validity and timeliness of cancer registry data contributing to the European Cancer Information System. *Front Oncol* (2023) 13:1219128. doi: 10.3389/fonc.2023.1219128

23. Feller A, Fehr M, Bordoni A, Bouchardy C, Frick H, Mousavi M, et al. Trends in incidence of oesophageal and gastric cancer according to morphology and anatomical location, in Switzerland 1982-2011. *Swiss Med Wkly* (2015) 145:w14245. doi: 10.4414/smw.2015.14245

24. Castro C, Bosetti C, Malvezzi M, Bertuccio P, Levi F, Negri E, et al. Patterns and trends in esophageal cancer mortality and incidence in Europe (1980-2011) and predictions to 2015. *Ann Oncol* (2014) 25(1):283–90. doi: 10.1093/annonc/mdt486

25. Machii R, Saika K. Morphological distribution of esophageal cancer from Cancer Incidence in Five Continents Vol. X. *Jpn J Clin Oncol* (2015) 45(5):506–7. doi: 10.1093/jjco/hyv061

26. McColl KEL. What is causing the rising incidence of esophageal adenocarcinoma in the West and will it also happen in the East? *J Gastroenterol* (2019) 54(8):669–73. doi: 10.1007/s00535-019-01593-7

27. El-Serag HB, Sweet S, Winchester CC, Dent. Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* (2014) 63(6):871–80. doi: 10.1136/gutjnl-2012-304269

28. OECD/European Union. *Health at a glance: Europe 2020: state of health in the EU cycle* Paris: OECD Publishing (2020). doi: 10.1787/82129230-en

29. Venneman K, Huybrechts I, Gunter MJ, Vandendaele L, Herrero R, Van Herck K. The epidemiology of Helicobacter pylori infection in Europe and the impact of lifestyle on its natural evolution toward stomach cancer after infection: A systematic review. *Helicobacter* (2018) 23(3):e12483. doi: 10.1111/hel.12483

30. GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet* (2020) 396(10258):1223–49. doi: 10.1016/S0140-6736(20)30752-2

31. Rawla P, Barsouk A. Epidemiology of gastric cancer: global trends, risk factors and prevention. *Prz Gastroenterol* (2019) 14(1):26–38. doi: 10.5114/pg.2018.80001

32. Thrift AP, El-Serag HB. Burden of gastric cancer. *Clin Gastroenterol Hepatol* (2020) 18(3):534–42. doi: 10.1016/j.cgh.2019.07.045

33. Tepes B, Pérez-Aisa Á, Vaira D, Caldas M, Bujanda L, Castro-Fernandez M, et al. European Registry on Helicobacter pylori management (Hp-EuReg): patterns and trends in first-line empirical eradication prescription and outcomes of 5 years and 21 533 patients. *Gut* (2021) 70(1):40–54. doi: 10.1136/gutjnl-2020-321372

34. Rugge M, Genta RM, Graham DY, Di Mario F, Vaz Coelho LG, Kim N, et al. Chronicles of a cancer foretold: 35 years of gastric cancer risk assessment. *Gut* (2016) 65 (5):721–5. doi: 10.1136/gutjnl-2015-310846

35. Aragonés N, Izarzugaza MI, Ramos M, Chirlaque MD, Almar E, Martínez C. Oesophago-gastric Cancer Working Group. Trends in oesophago-gastric cancer incidence in Spain: analysis by subsite and histology. *Ann Oncol* (2010) 21 Suppl 3: iii69–75. doi: 10.1093/annonc/mdq083

36. Pohl H, Welch HG. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. *J Natl Cancer Inst* (2005) 97(2):142–6. doi: 10.1093/jnci/dji024

37. Lindblad M, Ye W, Lindgren A, Lagergren J. Disparities in the classification of esophageal and cardia adenocarcinomas and their influence on reported incidence rates. *Ann Surg* (2006) 243(4):479–85. doi: 10.1097/01.sla.0000205825.34452.43

38. Sugano K, Spechler SJ, El-Omar EM, McColl KEL, Takubo K, Gotoda T, et al. Kyoto international consensus report on anatomy, pathophysiology and clinical significance of the gastro-oesophageal junction. *Gut* (2022) 71(8):1488–514. doi: 10.1136/gutjnl-2022-327281

39. Obermannová R, Alsina M, Cervantes A, Leong T, Lordick F, Nilsson M, et al. Oesophageal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol* (2022) 33(10):992–1004. doi: 10.1016/j.annonc.2022.07.003

40. Hammoud GM, Hammad H, Ibdah JA. Endoscopic assessment and management of early esophageal adenocarcinoma. *World J Gastrointest Oncol* (2014) 6(8):275–88. doi: 10.4251/wjgo.v6.i8.275

41. Fassan M, Ludwig K, Pizzi M, Castoro C, Guzzardo V, Balistreri M, et al. Human epithelial growth factor receptor 2 (HER2) status in primary and metastatic esophagogastric junction adenocarcinomas. *Hum Pathol* (2012) 43(8):1206–12. doi: 10.1016/j.humpath.2011.09.004

42. Ekström AM, Signorello LB, Hansson LE, Bergström R, Lindgren A, Nyrén O. Evaluating gastric cancer misclassification: a potential explanation for the rise in cardia cancer incidence. *J Natl Cancer Inst* (1999) 91(9):786–90. doi: 10.1093/jnci/91.9.786

43. Lv L, Liang X, Wu D, Wang F, Zhang Y, Cang H, et al. Is cardia cancer a special type of gastric cancer? A differential analysis of early cardia cancer and non-cardia cancer. *J Cancer* (2021) 12(8):2385–94. doi: 10.7150/jca.51433

44. Pernot S, Voron T, Perkins G, Lagorce-Pages C, Berger A, Taieb J. Signet-ring cell carcinoma of the stomach: Impact on prognosis and specific therapeutic challenge. *World J Gastroenterol* (2015) 21(40):11428–38. doi: 10.3748/wjg.v21.i40.11428

45. Machlowska J, Baj J, Sitarz M, Maciejewski R, Sitarz R. Gastric cancer: epidemiology, risk factors, classification, genomic characteristics and treatment strategies. *Int J Mol Sci* (2020) 21(11):4012. doi: 10.3390/ijms21114012

46. Martos C, Giusti F, Van Eycken E, Visser O. *A common data quality check procedure for European cancer registries, JRC132486* (2023). Ispra, Italy: European Commission. Available at: https://encr.eu/sites/default/files/Recommendations/JRC132486_cancer_data_quality_checks_procedure_report_2.0.pdf (Accessed 7 August 2023).

47. ENCR. *Materials from trainings and workshops* Available at: https://encr.eu/node/535 (Accessed 7 August 2023).

48. *ENCR e-learning modules on cancer coding* Available at: https://encr.eu/news/e-learning (Accessed 7 August 2023).

49. *ENCR registries reviews* Available at: https://encr.eu/registries-reviews (Accessed 7 August 2023).

50. Giusti F, Martos C, Adriani S, Flego M, Negrão Carvalho R, Bettio M, et al. The Joint Research Centre-European Network of Cancer Registries Quality Check Software (JRC-ENCR QCS). *Front Oncol* (2023) 13:1250195. doi: 10.3389/fonc.2023.1250195

frontiers | Frontiers in Oncology

TYPE Perspective
PUBLISHED 25 July 2024
DOI 10.3389/fonc.2024.1438805

OPEN ACCESS

EDITED BY
Stefano Guzzinati,
Azienda Zero, Italy

REVIEWED BY
Matthias Lorez,
National Institute for Cancer Epidemiology and Registration (NICER), Switzerland
Francesco Cuccaro,
Azienda Sanitaria Locale della Provincia di Barletta Andria Trani (ASL BAT), Italy

*CORRESPONDENCE
Francesco Giusti
francescogiusti@hotmail.com
Manola Bettio
jrc-encr@ec.europa.eu

RECEIVED 26 May 2024
ACCEPTED 01 July 2024
PUBLISHED 25 July 2024

CITATION
Giusti F, Martos C, Carvalho RN, Zadnik V, Visser O, Bettio M and Van Eycken L (2024) Facing further challenges in cancer data quality and harmonisation.
Front. Oncol. 14:1438805.
doi: 10.3389/fonc.2024.1438805

Facing further challenges in cancer data quality and harmonisation

Francesco Giusti[1,2]*, Carmen Martos[1,3], Raquel N. Carvalho[1], Vesna Zadnik[4], Otto Visser[5], Manola Bettio[1]* and Liesbet Van Eycken[2]

[1]European Commission, Directorate General Joint Research Centre, Ispra, VA, Italy, [2]Belgian Cancer Registry, Brussels, Belgium, [3]Foundation for the Promotion of Health and Biomedical Research in the Valencian Region (FISABIO), Valencia, Spain, [4]Slovenian Cancer Registry, Institute of Oncology Ljubljana, Ljubljana, Slovenia, [5]Department of Registration, Netherlands Comprehensive Cancer Organisation (IKNL), Utrecht, Netherlands

This article highlights the recent and ongoing activities of European population-based cancer registries (PBCRs) in data quality and harmonisation in the framework of the collaboration between the European Network of Cancer Registries (ENCR) and the Directorate-General Joint Research Centre (JRC), the science and knowledge centre of the European Commission. The article concludes the Frontiers in Oncology's Research Topic "Joining Efforts to Improve Data Quality and Harmonization Among European Population-Based Cancer Registries", which has been an opportunity for several European researchers to share their experience on cancer data quality and harmonisation. Such experience will be helpful for PBCRs in view of future challenges and opportunities in cancer epidemiology, with a few examples discussed in the present article.

KEYWORDS

cancer registry, data quality, data harmonisation, challenges, Europe

1 Introduction

During recent decades, the role of population-based cancer registries (PBCRs) has advanced beyond their traditional focus on providing cancer incidence and survival data, enlarging it to data providers for health-service management (1–4). In this respect, PBCRs face further challenges of data quality and harmonisation issues.

Since 1990, the European Network of Cancer Registries (ENCR) has been operational with the aim to connect PBCRs in Europe.

The ENCR plays a crucial role in supporting PBCRs to improve the quality (including comparability) and availability of cancer incidence data and paves the way for the use of data collected by PBCRs in cancer control, health-care planning and research. Cancer data comparability between countries and regions is particularly important for the European policy makers, who rely on the European Cancer Information System for accurate and up-to-date cancer burden statistics computed with data from the almost 200 PBCRs currently

active in Europe. ENCR activities have a global impact, also due to its collaborations with the International Association of Cancer Registries (IACR) and the International Agency for Research on Cancer (IARC) and the fact that ENCR recommendations and guidelines regularly serve as models endorsed within the IACR. An example of collaboration between ENCR and IACR was the joint ENCR-IACR 2023 Scientific Conference, which took place in Granada, Spain, in November 2023 and was attended by more than 350 participants (5).

The *Frontiers in Oncology* Research Topic "*Joining Efforts to Improve Data Quality and Harmonization Among European Population-Based Cancer Registries*" has been an opportunity for European researchers to share their experience on cancer data quality and harmonisation (6).

In this light, this article refers to all the contributions to the Research Topic and summarises the present situation in European PBCRs related to data quality and harmonisation, as well as the currently implemented activities carried out by ENCR and JRC to improve them. Of particular note, the activity of several ENCR working groups and the update of ENCR recommendations will be described. Moreover, the European Cancer Information System (ECIS) (7) as the ultimate outcome of data quality and harmonisation efforts will be presented.

2 Current advances of cancer registration in Europe

Since 2012, the ENCR Secretariat has been hosted at the Joint Research Centre (JRC), the science and knowledge centre of the European Commission. In this scenario, several initiatives were carried out in the last decade (8) aimed at improving cancer data quality and harmonisation of European PBCRs: the JRC and ENCR coordinated thematic expert working groups to draft guidelines and recommendations on data collection, coding, and reporting, organised trainings, including on the revised recommendations, and developed common rules and related validation software to check data compliance to agreed European standards (9).

European PBCRs are very heterogeneous in terms of geographical coverage, either national or regional, and can cover very different population sizes, translating in datasets ranging from around 125,000 to over 50 million cancer records. Additionally, they differ regarding registration practices, for example in relation to data sources, definitions and procedures. Therefore, common rules and definitions are necessary in order to harmonise data from different PBCRs and ensure their comparability at European level.

To this purpose, the following recommendations, reports and documents were published during the period 2022–2024 on the ENCR website (10).

2.1 ENCR recommendations

2.1.1 Data quality checks for European cancer registries

Recognising the pivotal importance of comparability, completeness, validity, and timeliness in ensuring the reliability and utility of PBCR data, in 2013 the ENCR and JRC launched the Data Quality Checks Working Group to address the fragmented landscape of data validation methods across European PBCRs.

To achieve this objective, a series of workshops were convened in 2013 and 2014. These meetings served as forums for stakeholders from diverse backgrounds, including PBCR experts, epidemiologists, and data analysts, to collaboratively deliberate on the establishment of a harmonised framework for data quality assessment.

Following the work of the Cancer Data Quality Checks Working Group (11) the first agreed quality control checks among European PBCR's were proposed, aimed at validating the internal consistency of cancer incidence variables. The report, and later update (12) formed the basis for the JRC-ENCR Quality Check Software (QCS), described in one contribution of the current Research Topic (13).

2.1.2 Standard dataset for the European network of cancer registries (2023)

This recommendation updates a previous document released in 2005 (14), to provide the minimum dataset to be collected by European PBCRs. Given the great expansion of PBCRs role in cancer control, quality assessment of cancer care, clinical and epidemiological research in the latest years, additional standardised data items were deemed necessary for registration. Thanks to the rapid growth of electronic records in the health care sector, many items may now be collected by linkage to existing data sources, as part of routine operations or on an *ad hoc* basis. However, the abundance of available data may be at the expense of standardisation and comparability. While the level of automation may increase access to growing amounts of data, the legal basis for access to and linkage with health data, varying greatly across Europe, may jeopardize the capacity to check the quality of such data.

The 2023 revision of the standard dataset recommendation (15, 16) was drafted to preserve the possibilities for comparisons on cancer incidence between European and non-European PBCRs, to share data definitions for in-depth and wide-scale collaborative efforts and identify variables that may support an expanded role of PBCRs in cancer control.

2.1.3 Basis of diagnosis (2022)

The 2022 recommendations updated the previous ones from 1999 (17–19).

Basis of diagnosis is a key variable, including information both on the way in which the tumour is diagnosed and the level of likeness of the diagnosis itself. It is also influenced by the ability of individual PBCRs to intercept the different (pathological, cytological, molecular…) reports.

Guidance in the latest recommendations is particularly relevant in the absence of pathological confirmation of the tumour. The proportion of clinical diagnoses (basis of diagnosis values 1, 2 and 4) is a data quality indicator. While a high proportion of clinical diagnoses in a PBCR may reflect the situation with regard to clinical and pathological investigations in the area covered by the PBCR, it may also indicate overdiagnosis and overestimation of cancer incidence, possibly taking into account tumours that would never

have caused symptoms or death. On the other hand, PBCRs with a very low proportion of clinical diagnoses might underestimate incidence rates, potentially missing cancer cases that should be counted.

Among the modifications introduced, the new value 8 (Cytogenetic and/or molecular testing) for coding the basis of diagnosis is particularly relevant in view of the fast evolution of diagnostic techniques, such as karyotyping, FISH (fluorescent *in situ* hybridization), PCR (polymerase chain reaction) and DNA sequencing.

2.1.4 Cancer cases in migrant population (2022)

In the wake of the increase in the number of migrants (including refugees) in European countries, and with a particular consideration of the millions of refugees from Ukraine to Europe, in 2022, a new ENCR recommendation was released to clarify and harmonise whether to register migrant individuals without a legal residency at the date of incidence (20).

2.1.5 Recording and reporting of urothelial tumours of the urinary tract (2022)

Following the previous publication in 1995 of "*Recommendations for coding bladder cancers*" (21) and IARC's 2003 book on "*Standards and guidelines for cancer registration in Europe*" (22), knowledge about the biology and pathology of urinary tract tumours and their classification has increased considerably (23). Great variability has been observed among European PBCRs in the recording (i.e. registration) and the reporting (i.e. in presenting cancer burden statistics) of these tumours (24).

The 2022 ENCR recommendation aimed at improving comparability of data on urothelial tumours of the urinary tract in Europe by defining criteria mainly for registration, taking into account multiple aspects of these tumours such as primary topography, histological type, grade, extent of invasion, multi-centricity, progressions and time interval between tumors (25, 26). An example of the rules that should lead to greater data harmonisation and comparability is the suggestion not to record the "Urothelial proliferation of uncertain malignant potential", which in any case are not reportable.

2.1.6 Coding incidence date (2023)

The previous recommendation on the coding of incidence date was released in 1995 and revised in 1997 (27). The detection of inconsistencies in its application among European PBCR's led to the creation of a working group which re-prioritized events considered for the registration of incident date considering modern methods of diagnosis such as flow cytometry, molecular testing, screening tests and more recent radiological and imaging techniques (28). An increased standardisation of incidence date, in addition to allowing more accurate cancer incidence statistics, also improves the consistency of survival estimates.

2.1.7 ENCR endorsement of the Toronto childhood cancer stage guidelines (2016)

In 2016 the ENCR Steering Committee endorsed and encouraged the active use of the Toronto Childhood Cancer Stage Guidelines by European PBCRs, in order to promote the consistency of stage data for childhood malignancies (29–31). Moreover, the Toronto childhood cancer stage has been included in the latest 2022 ECIS data call protocol to European PBCRs.

One article of the current Research Topic shared the experience of the International Benchmarking of Childhood Cancer Survival by Stage (BENCHISTA) project in encouraging the implementation of the Toronto Childhood Cancer Stage Guidelines (32). The extensive application of the Toronto staging allows for instance to study whether the differences in survival of patients with childhood cancers between countries are due to a different diagnostic timing or to differences in access to care and treatment protocols, which is the main objective of the BENCHISTA project.

2.2 The European cancer information system

The JRC has been developing since 2012 ECIS as a comprehensive infrastructure, consisting of several components to manage a central data repository and to coordinate in an efficient and sustainable way the activities of data validation, analysis, and dissemination. A key component of the ECIS is a web-based tool launched in February 2018 (33) to report and disseminate cancer burden indicators such as incidence, mortality, survival and prevalence. Indicators in ECIS are derived from European PBCRs data. The ECIS web application (34) allows the visualisation of such indicators across European areas and time dimension.

The first data call to feed ECIS was launched in 2015. The database feeding ECIS is dynamic and is updated as new data becomes available.

The ECIS web-application is modular and currently, its data explorer section consists of the following modules:

- *Incidence and mortality estimates*– latest release year is 2022 as the outcome of a collaborative project between JRC and the IARC, in collaboration with the ENCR;
- *Long-term incidence and mortality estimates up to 2040*, evaluating the impact of different demographic scenarios by 2040 on the cancer burden;
- *Survival estimates*, reporting on the results of the latest published EUROCARE-5 study (35);
- *Incidence and mortality historical data*, including indicators computed from PBCRs observed data;
- *Childhood cancer incidence historical data*, reported according to the International Classification of Childhood Cancer (ICCC), third edition.
- Prevalence estimates in 2020, reporting on the results of the EUROCARE-6 study (36).

2.2.1 The 2015 Call for Data protocol

The 2015 Call for data protocol required the submission from European PBCRs of a cancer case file, a population file, a mortality file, life tables and a data submission questionnaire (37). Data were harmonized at central level, incidence and mortality indicators were

computed by the JRC and disseminated through the ECIS web application, released in 2018.

The variables required by the protocol for the incidence file included demographic and tumour characteristics like sex, age, topography, morphology, considered as core variables for reporting incidence indicators and which were the focus of data quality evaluations. Additionally, the protocol included variables for survival analysis, as well as stage and treatment information.

2.2.2 The 2015 European dataset

ENCR-affiliated PBCRs contributed data to the 2015 Call for Data (Figure 1). Over 34.5 million incident cases were collected from general PBCRs (all ages and all cancer sites) and specialised (childhood or site specific) PBCRs. Data harmonisation procedures, such as correction of errors detected by the JRC-ENCR Quality Check Software and the implementation of multiple primary tumours rules were performed centrally at JRC and by the submitting PBCRs. Following data harmonisation, around 30 million cases from 145 PBCRs (with incidence years between 1953 and 2014) were validated for the ECIS web application.

2.2.3 The 2022 data call protocol

A second ECIS call was launched in 2022 to the ENCR PBCRs (38). While the core variables from the 2015 protocol were retained, the experience gained from the previous call led to a few changes in the 2022 protocol, namely:

- the case definition was changed: in situ/non-invasive tumours requested only for breast, urothelial tumours, ovary and skin melanoma, whereas, according to the ICD-O-3.2 the only benign tumours should be those of the central nervous system and gastrointestinal stromal tumours (GIST);
- A finer geographical detail was requested, specifying the geographical area of residence at diagnosis for incident cases according to the NUTS classification level 2 (NUTS2) (39);
- Toronto childhood cancer stage was introduced;
- Better specification of treatment (e.g. for different systemic therapies) and related timing (e.g. neo-adjuvant vs adjuvant) was added.

Novelties in the 2022 ECIS protocol implied the definition and implementation of additional validation rules, and related work for the update of IT tools.

2.3 Data quality aspects addressed in the current research topic

As shown in the present Research Topic, the quality of incidence data reported by European PBCRs improved between 1995 and 2014 (40). The analysis of 28,776,562 cases from 130 PBCRs in 30 European countries reported worse data quality for the oldest age groups and for cancer sites with poor survival. No differences were found between males and females, whereas high variability in data quality was detected across European PBCRs. The use of electronic health records, steadily increasing over the years, might be one of the contributing factors for a more accurate and timely registration of data.

A second contribution of the Research Topic focused on geographical variability and data quality in gastric and oesophageal cancer. A wide variability in oesophago-gastric cancers topographic subsites and histopathological types was observed, with a corresponding improvement in accuracy of registration in the study period (1995–2014) (41).

One article of the Research Topic focuses on the JRC-ENCR Quality Check Software (QCS) (13), as the IT tool developed by the JRC to check the internal consistency of PBCRs data.

Another valuable article of the present Research Topic thoroughly compared the functional characteristics of the JRC-ENCR QCS with the check tool developed by the IARC and the IACR (42). The paper concluded that it would be advisable to use both systems for data quality control, since they provide checks on different groups of variables (stage, follow-up) or on the same variables but with different modalities.

Finally, one important aspect of the improvement in data quality in European PBCRs is related to the enhanced possibility to analyse long-term cancer incidence trends. One example of such investigation is the article focusing on the incidence pattern of haematological neoplasms in Spanish children between 1983 and 2018, and its comparison with other southern European countries (43).

2.4 Current focus of JRC- ENCR activities

Harmonisation activities continue to be one major focus of the collaboration between the ENCR and the JRC. More specifically, the following topics are the subject of active ENCR Working Groups (9):

2.4.1 Working group on treatment data harmonisation

As reported in the present Research Topic, a growing number of European PBCRs are collecting treatment data (44). This overview, which combined data from a literature review and conference proceedings, together with data from 125 European PBCRs, has led to the creation of a working group which provided the first recommendations for treatment data collection and coding, and the invitation to PBCRs to improve data harmonisation and comparability in Europe.

2.4.2 Working group on cancer recurrences

The aim of the working group is to define a protocol for the standardised collection of cancer recurrence, progression and transformation data by PBCRs.

2.4.3 Working group on central nervous system tumours

Aimed at updating the previous ENCR recommendation, dated 1998.

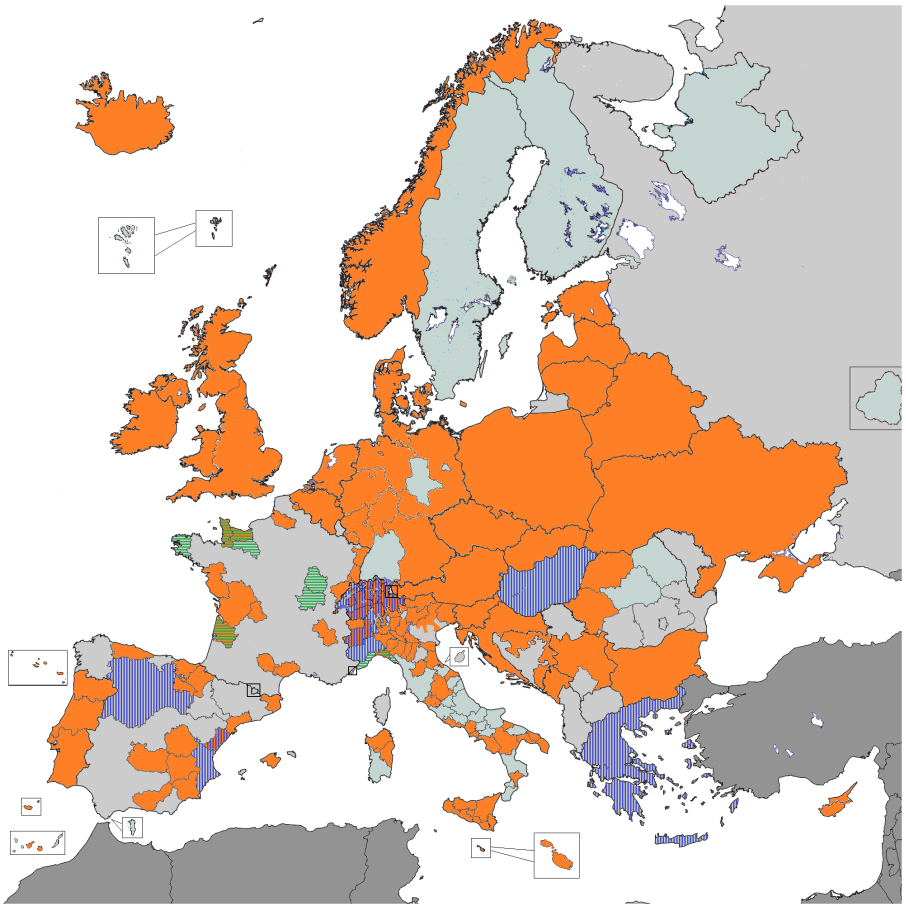


FIGURE 1
PBCRs contributing data 2015 ENCR-JRC Call for Data. Orange: all ages and all cancer sites PBCRs; Vertical stripes: childhood PBCRs; Horizontal stripes: site-specific PBCRs.

2.4.4 Working group on haematological malignancies

Aimed at updating the previous ENCR recommendation, dated 2014.

2.4.5 Working group on survival in ECIS

Aimed at defining the data standards and quality checks to be applied for publication of survival indicators in ECIS.

2.4.6 Working group on multiple primaries registration

Aimed at updating the previous ENCR recommendation, dated 2004.

2.5 ECIS in the context of the European commission's Europe's Beating Cancer Plan

The European Commission's Europe's Beating Cancer Plan (EBCP) (45), released in February 2021, is structured around four key action areas (Prevention, Early detection, Diagnosis and treatment, Improvement of quality of life) and is supported by 10 flagship initiatives, underscoring the European Union's commitment to support cancer prevention, treatment, and care across the continent. In this context, a Knowledge Centre on Cancer (46) and the European Cancer Inequalities Registry (47) were established in the framework of the EBCP.

Several activities and collaborations are ongoing to expand the information provided by the ECIS in line with commitments of the EBCP and demand for good quality data at population level will continue. The following is a list of developments in line with such commitments, including:

- Providing cancer incidence data at regional level, following the NUTS (Nomenclature of Territorial Units for Statistics) classification level. The availability of more granular data will facilitate ecological comparisons (for instance, with environmental and socio-economic data) and align with the overarching EBCP actions aiming to address inequalities between and within EU Member States. On this point, it will be important to monitor possible issues of reidentification of patients that might arise with more granular data;

- Displaying of cancer prevalence data, necessary for proper quantification to support EBCP objectives of reducing the burden of cancer, improving cancer outcomes, and enhancing the quality of life for all cancer survivors across Europe;
- Reporting on cancer stage data, which guide evidence-based decision-making tracking advancements towards cancer control goals and promoting quality improvement in cancer care;
- Exploring the expansion to cancer screening data monitoring, in line with the fourth EBCP flagship initiative, which aims to put forward a new EU-supported Cancer Screening Scheme to help Member States ensure that 90% of the EU population who qualify for breast, cervical and colorectal cancer screenings are offered screening by 2025. The CanScreen-ECIS project (48) paved the way towards this achievement.

3 Discussion

The improvement in quality and the harmonisation of PBCRs data will remain the focus of JRC and ENCR activities. As indicated by European PBCRs (49), a priority should be to develop a common mechanism for estimating the national cancer burden for countries with partial cancer registration, to enable direct and more accurate comparisons between countries. In addition, countries with absent or underdeveloped cancer registration should be assisted in establishing PBCRs. The quality indicators reported in the present Research Topic can be used as the baseline for monitoring PBCRs data quality indicators in Europe (40).

Reliable data from PBCRs are crucial for the effective implementation and evaluation of cancer control programmes. The standardisation of data and the harmonisation of procedures has led to an overall improvement in the description of neoplastic diseases and how incidence, survival, prevalence, mortality are all necessary (and somehow interlaced) indicators for understanding the epidemiology of tumors. The role of PBCRs has been expanding over the years; at the same time, thanks also to the essential action of ENCR and JRC, European PBCRs have made progress over the last decades with regard to data quality. This momentum should be sustained in order to further improve harmonisation and decrease resource disparities leading to quality disparities. Clear guidelines and policies offer the basis for this, with guiding principles for the equitable and effective operation of PBCRs providing a structured framework that enables registries to maximise their potential and contribute to cancer surveillance and research efforts, regardless of resource constraints.

Ongoing advances in technology can offer alternative models for data sharing and international comparisons, for instance a federated approach for data collection, as shown in the current Research Topic "*Joining Efforts to Improve Data Quality and Harmonization Among European Population-Based Cancer Registries*" with the description of the Nordcan.R tool. The article showed how the tool is used to compute statistics for the Nordic cancer statistics web platform NORDCAN, and demonstrated that it works also with non-Nordic data (50).

An innovative approach in view of federated data quality evaluations was also presented in the current Research Topic. The article presented an ontology created using a modular approach to handle specific checks for childhood cancers, leading to a simpler maintenance of data validation rules (51).

In this context, a key role is going to be played by the future European Health Data Space (EHDS), a European Commission initiative to build a common EU framework facilitating the use of health data for secondary purposes that could be beneficial to European PBCRs by facilitating cancer data sharing (52). This initiative aims to improve interoperability and accessibility of health data across Europe, fostering better research and improved public health outcomes. By creating a standardised environment for health data exchange, the EHDS will enable more efficient data sharing between PBCRs and researchers, helping to overcome current barriers related to data fragmentation and diverse national regulations. This will not only help streamlining the process of data harmonisation but will also promote innovation in cancer research, ultimately contributing to more effective cancer prevention and treatment strategies across Europe.

Three articles in the present Research Topic focus on methodologies for the computation of cancer prevalence. A first article showed two alternative approaches in the framework of the completeness index method, based on incidence and survival modelling, in order to provide comparable indicators on complete cancer prevalence (53). The second article described the procedures to derive complete prevalence and several indicators of cancer cure from PBCRs. Limited duration prevalence was calculated for 62 cancer types by sex and PBCR, presenting indicators which may be relevant for patients and clinical practice and reproducible in different European countries (54). Lastly, a new method to estimate short-time projections of cancer prevalence by phase-of-care was illustrated. Evidence from this method was addressed to policy makers for planning future cancer care, thus improving cancer survivorship experience for patients and care-givers (55).

Finally, in recent years, biomarkers have become more important in guiding diagnosis and treatment options as well as for the prognosis of several tumour types such as, for example, breast, oropharyngeal and lung cancer (56). The use of biomarkers is also important in predicting recurrences. For this reason, biomarkers should be taken into account in the future by the ENCR because it will be necessary to standardise data collection, coding and reporting of this key information.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

FG: Conceptualization, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. CM: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. RC: Conceptualization, Project administration, Writing – review & editing. VZ: Writing –

review & editing. OV: Writing – review & editing. MB: Conceptualization, Project administration, Supervision, Writing – review & editing. LVE: Conceptualization, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer FC declared a past collaboration with the authors OV, LVE, MB, RC, FG, and CM to the handling editor.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

1. Siesling S, Louwman WJ, Kwast A, van de Hurk C, O'Callaghan M, Rosso S, et al. Uses of cancer registries for public health and clinical research in Europe: Results of the European Network of Cancer Registries survey among 161 population-based cancer registries during 2010–2012. *Eur J Cancer*. (2015) 51:1039–49. doi: 10.1016/j.ejca.2014.07.016

2. Armstrong BK. The role of the cancer registry in cancer control. *Cancer Causes Control*. (1992) 3:569–79. doi: 10.1007/BF00052754

3. Bray F, Colombet M, Mery L, Piñeros M, Znaor A, Zanetti R, Ferlay J eds. Cancer Incidence in Five Continents, Vol. XI. IARC Scientific Publication No. 166. Lyon: International Agency for Research on Cancer (2021). Available at: https://publications.iarc.fr/597. Licence: CC BY-NC-ND 3.0 IGO.

4. Forsea A-M. Cancer registries in Europe–going forward is the only option. *Ecancer*. (2016) 10:641. doi: 10.3332/ecancer.2016.641

5. ENCR. IACR Scientific Conference (2023). Granada, Spain. Available online at: https://www.encr.eu/sites/default/files/inline-files/ENCR%20Scientific%20Meeting%202023_book%20abstracts (Accessed 15 June 2024). JRC135581.

6. Joining Efforts to Improve Data Quality and Harmonization Among European Population-Based Cancer Registries. *Front Oncol*. Available at: https://www.frontiersin.org/research-topics/33889/joining-efforts-to-improve-data-quality-and-harmonization-among-european-population-based-cancer-registries/magazine.

7. ECIS - European Cancer Information System. (2024). Available online at: https://ecis.jrc.ec.europa.eu (Accessed 12 March 2024).

8. European Network of Cancer Registries. (2024). Available online at: https://encr.eu/ (Accessed 12 March 2024).

9. ENCR. Working groups. Available online at: https://encr.eu/Activities/Working-groups (Accessed 18 January 2024).

10. ENCR. ENCR Recommendations. Available online at: https://encr.eu/ENCR-Recommendations (Accessed 18 January 2024).

11. Martos C, Crocetti E, Visser O, Rous B, Giusti Fthe Cancer Data Quality Checks Working Group. A proposal on cancer data quality checks: one common procedure for European cancer registries – version 1.1 (2018). Luxembourg: EUR 29089 EN, Publications Office of the European Union. Available online at: https://encr.eu/sites/default/files/inline-files/Cancer%20Data%20Quality%20Checks%20Procedure%20Report%20online.pdf (Accessed 18 January 2024).

12. Martos C, Giusti F, Van Eycken E, Visser O. A common data quality check procedure for European cancer registries (2023). Ispra, Italy: European Commission. Available online at: https://encr.eu/sites/default/files/Recommendations/Cancer%20Data%20Quality%20Checks%20Procedure%20Report%202.0_v20240506.pdf (Accessed 15 May 2024).

13. Giusti F, Martos C, Adriani S, Flego M, Negrão Carvalho R, Bettio M, et al. The joint research centre-european network of cancer registries quality check software (JRC-ENCR QCS). *Front Oncol*. (2023) 13:1250195. doi: 10.3389/fonc.2023.1250195

14. ENCR. Recommendations for a Standard Dataset for the European Network of Cancer Registries (2005). Available online at: https://encr.eu/sites/default/files/pdf/recommendations.pdf (Accessed 12 March 2024).

15. ENCR. ENCR recommendations, in: *Standard dataset for the European Network of Cancer Registries* (2023). Available online at: https://encr.eu/sites/default/files/Recommendations/ENCR-Recommendation-standard-dataset_Mar2023.pdf (Accessed 12 March 2024).

16. Michalek IM, Martos C, Caetano Dos Santos FL, Giusti F, Degerlund H, Neamtiu L, et al. Advancing data collection and analysis: 2023 revised European Network of Cancer Registries recommendations for standard dataset. *Eur J Cancer*. (2024) 199:113557. doi: 10.1016/j.ejca.2024.113557

17. ENCR. Recommendations for coding Basis of Diagnosis (1999). Available online at: https://encr.eu/sites/default/files/pdf/basisd.pdf (Accessed 2 March 2024).

18. ENCR. ENCR recommendations, in: *Basis of Diagnosis* (2022). Available online at: https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation%20BoD_Oct2022_EN_241022.pdf (Accessed 2 March 2024).

19. Visser O, Kościańska B, Caetano dos Santos FL, Cuccaro F, Michalek IM, Mousavi M, et al. revised European recommendations for the coding of the basis of diagnosis of cancer cases in population-based cancer registries. *Front Oncol Volume*. (2022) 13:1250549. doi: 10.3389/fonc.2023.1250549

20. ENCR. ENCR recommendations, in: *Cancer cases in migrant population* (2022). Available online at: https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation_Migrants_June2022_EN.pdf (Accessed 2 March 2024).

21. ENCR. Recommendations for coding bladder cancers, (1995). Available online at: https://encr.eu/sites/default/files/pdf/bladeng.pdf (Accessed 2 March 2024).

22. Standards and Guidelines for Cancer Registration in Europe (2003). Lyon: IARC Technical Publication No. 40. Available online at: https://publications.iarc.fr/Book-And-Report-Series/Iarc-Technical-Publications/Standards-And-Guidelines-For-Cancer-Registration-In-Europe-2003 (Accessed 2 March 2024).

23. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO classification of tumours of the urinary system and male genital organs—Part B: Prostate and bladder tumours. *Eur Urol*. (2016) 70:106–19. doi: 10.1016/j.eururo.2016.02.028

24. Daubisse-Marliac L, Grosclaude P, Carulla M, Parada D, Vilardell L, Ameijide A, et al. Registration of urothelial tumours in cancer registries: how to improve and make it more useful? *Int J Environ Res Public Health*. (2022) 19:2714. doi: 10.3390/ijerph19052714

25. ENCR. ENCR recommendations, in: *Recording and Reporting of Urothelial Tumours of the Urinary Tract* (2022). Available online at: https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation_UT_Jun2022_EN.pdf (Accessed 18 January 2024).

26. Galceran J, Parada D, Eden M, Tumino R, Warren AY, Martos C, et al. The 2022 ENCR Recommendations on recording and reporting of urothelial tumours of the urinary tract. *Front Oncol*. (2022) 12:1046239. doi: 10.3389/fonc.2022.1046239

27. ENCR. Recommendations for coding Incidence Date. Available online at: https://www.encr.eu/sites/default/files/pdf/incideng.pdf (Accessed 18 October 2023).

28. ENCR. ENCR recommendations, in: *Coding Incidence Date* (2022). Available online at: https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation%20DOI_Mar2022_0.pdf (Accessed 18 October 2023).

29. Gupta S, Aitken JF, Bartels U, Brierley J, Dolendo M, Friedrich P, et al. Paediatric cancer stage in population-based cancer registries: the Toronto consensus principles and guidelines. *Lancet Oncol*. (2016) 17:e163–72. doi: 10.1016/S1470-2045(15)00539-2

30. Gupta S, Aitken J, Bartels U, Bhakta N, Bucurenci M, Brierley JD, et al. Development of paediatric non-stage prognosticator guidelines for population-based cancer registries and updates to the 2014 Toronto Paediatric Cancer Stage Guidelines. *Lancet Oncol*. (2020) 21:e444–51. doi: 10.1016/S1470-2045(20)30320-X

31. ENCR. ENCR endorsement of the Toronto Childhood Cancer Stage Guidelines. Available online at: https://encr.eu/news/encr-endorsement-toronto-childhood-cancer-stage-guidelines (Accessed 12 March 2024).

32. Lopez-Cortes A, Didonè F, Botta L, Hjalgrim LL, Jakab Z, Canete Nieto A, et al. Cancer data quality and harmonization in Europe: the experience of the BENCHISTA Project - international benchmarking of childhood cancer survival by stage. *Front Oncol*. (2023) 13:1232451. doi: 10.3389/fonc.2023.1232451

33. Bettio M, Negrao Carvalho R, Dimitrova N, Dyba T, Flego M, Giusti F, et al. Measuring the cancer burden in Europe: The European cancer information system (ECIS). *EMJ Oncol* (2019) 7(1):48–9. doi: 10.1093/annonc/mdz263.014

34. Randi G, Zaro F N, Carvalho R, Martos C, Dimitrova N, Dyba T, et al. The European Cancer Information System (ECIS) web application. Available online at: https://ecis.jrc.ec.europa.eu/pdf/ECIS_Guide.pdf (Accessed 12 March 2024). JRC113106.

35. Minicozzi P, Otter R, Primic-Žakelj M, Francisci S. Survival of Cancer patients in europe, 1999 - 2007: the EUROCARE-5 study. *Eur J Cancer*. (2015) 51:2099–266. Available at: https://www.ejcancer.com/issue/S0959-8049(15)X0017-7.

36. De Angelis R, Demuru E, Baili P, Troussard X, Katalinic A, Chirlaque Lopez MD, et al. Complete cancer prevalence in Europe in 2020 by disease duration and country (EUROCARE-6): a population-based study. *Lancet Oncol*. (2024) 25:293–307. doi: 10.1016/S1470-2045(23)00646-0

37. ENCR-JRC Project 2015 - Updating population-based cancer incidence and mortality in Europe (2015). Call. Available online at: https://www.encr.eu/sites/default/files/ENCR-JRC2015_Protocol.pdf (Accessed 7 August 2023).

38. Call for Data Protocol for European Population-Based Cancer Registries (2022). (Luxembourg: Publications Office of the European Union). Available online at: https://www.encr.eu/sites/default/files/Data_call/ECIS%20call%20for%20data%20protocol_20221124.pdf (Accessed 12 March 2024).

39. European Commission. *Statistical regions in the European Union and partner countries NUTS and statistical regions 2021 - re-edition 2022*. Luxembourg: Publications Office of the European Union, (2022). doi: 10.2785/321792.

40. Giusti F, Martos C, Negrão Carvalho R, Van Eycken L, Visser O, Bettio M. Quality indicators: completeness, validity and timeliness of cancer registry data contributing to the European Cancer Information System. *Front Oncol*. (2023) 13:1219128. doi: 10.3389/fonc.2023.1219128

41. Giusti F, Martos C, Bettio M, Negrão Carvalho R, Zorzi M, Guzzinati S, et al. Geographical and temporal differences in gastric and oesophageal cancer registration by subsite and morphology in Europe. *Front Oncol*. (2024) 14:1250107. doi: 10.3389/fonc.2024.1250107

42. Tagliabue G, Perotti V, Fabiano S, Tittarelli A, Barigelletti G, Contiero P, et al. Comparison between two cancer registry quality check systems: functional features and differences in an Italian network of cancer registries dataset. *Front Oncol*. (2023) 13:1197942. doi: 10.3389/fonc.2023.1197942

43. Trallero J, Sanvisens A, Almela Vich F, Jeghalef El Karoni N, Saez Lloret I, Díaz-Del-Campo C, et al. Incidence and time trends of childhood hematological neoplasms: a 36-year population-based study in the southern European context, 1983-2018. *Front Oncol*. (2023) 13:1197850. doi: 10.3389/fonc.2023.1197850

44. Giusti F, Martos C, Trama A, Bettio M, Sanvisens A, Audisio R, et al. Cancer treatment data available in European cancer registries: Where are we and where are we going? *Front Oncol*. (2023) 13:1109978. doi: 10.3389/fonc.2023.1109978

45. Available online at: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_342 (Accessed 12 March 2024).

46. Available online at: https://knowledge4policy.ec.europa.eu/cancer/about_en (Accessed 12 March 2024).

47. Available online at: https://cancer-inequalities.jrc.ec.europa.eu/ (Accessed 12 March 2024).

48. The CanScreen-ECIS project. Available online at: https://canscreen-ecis.iarc.who.int/ (Accessed 18 January 2023).

49. Available online at: https://www.encr.eu/sites/default/files/inline-files/EUROPE_IACR%202023_report_final.pdf (Accessed 12 March 2024).

50. Larønningen S, Skog A, Engholm G, Ferlay J, Johannesen TB, Kristiansen MF, et al. Nordcan.R: a new tool for federated analysis and quality assurance of cancer registry data. *Front Oncol*. (2023) 13:1098342. doi: 10.3389/fonc.2023.1098342

51. Nicholson N, Giusti F, Martos C. An ontology design for validating childhood cancer registry data. *Front Oncol*. (2023) 13:1212434. doi: 10.3389/fonc.2023.1212434

52. European Commission. *Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space*. Strasbourg: COM (2022). 197 final 2022/0140(COD).

53. Demuru E, Rossi S, Ventura L, Dal Maso L, Guzzinati S, Katalinic A, et al. Estimating complete cancer prevalence in Europe: validity of alternative vs standard completeness indexes. *Front Oncol*. (2023) 13:1114701. doi: 10.3389/fonc.2023.1114701

54. Toffolutti F, Guzzinati S, De Paoli A, Francisci S, De Angelis R, Crocetti E, et al. Complete prevalence and indicators of cancer cure: enhanced methods and validation in Italian population-based cancer registries. *Front Oncol*. (2023) 13:1168325. doi: 10.3389/fonc.2023.1168325

55. Francisci S, Tursini F, Dal Maso L, Gigli A, Guzzinati S. Projecting cancer prevalence by phase of care: a methodological approach for health service planning. *Front Oncol*. (2023) 13:1201464. doi: 10.3389/fonc.2023.1201464

56. Voith von Voithenberg L, Crocetti E, Martos C, Dimitrova N, Giusti F, Randi G, et al. Cancer registries - guardians of breast cancer biomarker information: A systematic review. *Int J Biol Markers*. (2019) 34:194–9. doi: 10.1177/1724600819836097

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact