Nationale Krebsregistrierungsstelle
Organe national d'enregistrement du cancer
Servizio nazionale di registrazione dei tumori
National Agency for Cancer Registration

New Data Quality Concept

within the framework of the Federal Law on the Registration of Cancer (Cancer Registration Act, CRA; SR 813.33)

Version 1.0

Authors

Matthias Lorez, NACR and Ulrich Wagner, NACR
Contributions by Anita Feller, NACR und Anne Schmidt, NACR

12.1.2022

CONTENTS

1. ABREVIATIONS

aDQR	(annual) Data Quality Report
AHVN13	AHV-Number with thirteen digits
FOPH	Federal Office of Public Health
FSO	Federal Statistical Office
CHOP	Schweizerische Operationsklassifikation
CI-V	Cancer Incidence in Five Continents
DCN	Death Certificate Notified
DCO	Death Certificate Only
ICD	International Classification of Diseases
ICD-O	International Classification of Diseases for Oncology
JRC/ENCR	Joint Research Centre/European Network of Cancer Registries
ChCR	Childhood Cancer Registry
CCR	Cantonal Cancer Registry
CRA	Cancer Registration Act
CRO	Cancer Registration Ordinance
NACR	National Agency for Cancer Registration
NCD	National Cancer Dataset
NCD-D	National Cancer Data Dictionary
NCD-S	Database of the NACR
NICER	National Institut for Cancer Epidemiology and Registration
PSU	Primary Site uncertain/unknown
QCS	JRC/ENCR Quality check software
QI	Quality Indicator
RSW	Cancer Registration Software
SCHB	Swiss Coding Handbook
SOP	Standard Operating Procedure
UICC	Union for International Cancer Control
UPI	Unique Person Identification Database of ZAS
WHO	World Health Organisation
ZAS	Zentrale Ausgleichsstelle der AHV; Centrale de compensation CdC; Central Compensation Office CCO

2. INTRODUCTION

On January 1st, 2020, the Cancer Registration Act (CRA; SR 813.33) and the Cancer Registration Ordinance (CRO; SR 818.331) came into force[1]. Uniform and nationwide cancer registration has been made mandatory in Switzerland. All healthcare providers are now legally obliged to report cancer-relevant information specified in the ordinance.

The cantonal data and the National Cancer Dataset (NCD) serve to inform the public on defined aspects of the cancer burden, support research on cancer, as well as guide health policy. This entails the comprehensive monitoring of cancer trends, the evaluation of prevention and early detection measures, the evaluation of the quality of diagnosis, care, and treatment, for the optimization and management of the health care system at the cantonal and national level. For

[1] Articles 36, 37, 38, and 40 CRO came into force already on the 1st June 2018.

the cancer datasets to be meaningful, data must be complete, accurate, comparable and available in a timely manner.

Article 14 paragraph 1[2] and Article 18 paragraph 2[3] CRA, as well as Article 27 b[4] CRO assign the task of evaluating the quality of the cancer registry data, of reporting deficiencies, and of supporting quality improvements to the National Agency of Cancer Registration (NACR).

3. OVERVIEW

The present concept paper is addressed to the population-based Cantonal Cancer Registries (CCR's), and the population-based Childhood Cancer Registry (ChCR).
It specifies instruments and measures that are already available and tested for the evaluation and further development of data quality, and announces new instruments and measures that still need to be implemented. The concept explains how they will be applied, e.g. how instruments and measures are prioritized, what e.g. determines the focus of round robin tests, and how to move from quality assessment to concrete improvements of data quality at cantonal and national level.
The concept indicates the chronological order in which the instruments will be introduced (a graphical overview is provided in chapter 10) and, in the case of quality indicators yet to be developed, the planned implementation dates.

3.1. The classical four dimensions of data quality, and timeliness

Any data quality concept must consider the four classical dimensions. Evaluation of the quality of cancer registry data within these dimensions is primarily performed with the help of selected Quality Indicators (QI's).

Comparability

There must be temporal and geographical comparability of cancer statistics generated for different groups of analysis (analysis groups may be several cantons, Switzerland vs. other countries, genders, age groups, different calendar years, etc.). Comparability is achieved by adherence to national and international guidelines for cancer registration and the standardization of practices. This leads to comparable data along the calendar time axis within each analysis group, but also between different analysis groups at defined points in time. This poses a formidable task in Switzerland with fourteen independent cancer registries[5].

[2] Art. 14 KRG: Überprüfung, Erfassung und Aufbereitung der Daten.
1. Die nationale Krebsregistrierungsstelle überprüft die Daten, welche die kantonalen Krebsregister ihr weitergeleitet haben, und informiert die betreffenden Register über allfällige Mängel.

[3] Art. 18 KRG: Sicherstellung der Datenqualität
2. Sie überprüft regelmässig die Qualität der Datenregistrierung der kantonalen Krebsregister und des Kinderkrebsregisters. Sie kann zu diesem Zweck die registrierten Daten mit Ausnahme der personenidentifizierenden Daten bei den kantonalen Krebsregistern und dem Kinderkrebsregister stichprobenweise einsehen.

[4] Art. 27 KRV: Weitere Aufgaben
b. Sie trifft die erforderlichen Massnahmen zur Sicherstellung der Datenqualität. Sie kann insbesondere Ringversuche unter den kantonalen Krebsregistern durchführen. Die Ergebnisse werden den beteiligten Stellen bekanntgegeben.

[5] There are just 7 countries worldwide with a larger number of regional cancer registries: USA, China, India, France, Germany, Italy, and Spain. CI-V XI, IARC Scientific Publications No. 166 (2017).

Completeness of case ascertainment
For completeness of case ascertainment (or case finding), the question is whether all legally reportable cancer diagnoses made in a defined population have actually been recorded in the cancer registries databases. The observed number of registered diagnoses is compared to the unobservable expected number of cancer diagnoses made in the respective population.

Case completeness
Case completeness is sufficient when all mandatorily reportable information for a cancer case documented in the medical reports has been transferred to the registry database. Missing information is not relevant for the quality of cancer registration data if the data item in question has not been generated or investigated during the medical procedures. The data collection is still considered complete in that case. If, on the other hand, mandatory information is missing because of failure to report it to the registry, or from failure of the registry to record it, the case is considered incomplete.

Accuracy
The accuracy (or validity) of the registered cancer data refers to the correspondence between the registered information and the information documented in the medical reports. Note that cancer registries are expected to ask the reporting physician for confirmation or correction of highly suspicious or conflicting reported information. The accuracy depends on the precision of the source documents and the level of expertise in abstracting, coding and recording, both in the clinic and the registry.

Timeliness
Timeliness as a further aspect of data quality is novel to the quality assessment of Swiss cancer registry data, and is explicitly targeted by the CRO[6].
In order for the cancer registry to contribute to development and optimization of the health care system, the registered information must be sufficiently recent. The timeliness corresponds to the time between diagnosis and the date when the case is included in cancer statistics. This time interval breaks down into different parts: (1) when the information is known until it is reported to the cancer registry, (2) from recording it in the registry database until the finalization of quality checks on the registry level, (3) from submission to the NACR until the finalization of the NCD and the publication date of the first statistical report. It is noteworthy that there are conflicts between data timeliness and other aspects of data quality, in particular completeness.

3.2. Instruments for evaluation and development of data quality

Quality Checks as part of the annual data submission
CCR's submit annually all cases since the beginning of registration to NACR. This rule allows for additions or corrections of cases irrespective of the year of diagnosis. Before the data are

[6] «Hat sich das neue System der Krebsregistrierung einmal etabliert, werden diese Fristen um ein Jahr gekürzt.» (S.31, Art. 39, Erläuterungen zur KRV).

submitted, the NACR requests the correction of errors, or verification of unusual findings, indentified by a predefined release of the quality check software (QCS) of the Joint Research Centre/European Network of Cancer Registries (JRC/ENCR). The submitted data is re-checked with the QCS and in addition, separate NACR-derived checks are performed. The findings are reported back to the CCR for correction/verification (see Appendix for NACR-derived checks). Only if all findings are resolved or commented, the data is integrated into the NCD.

The annual Data Quality Report (aDQR)

The NACR combines basic and supplementary cancer data from all CCR's into a single NCD. This opens the possibility for systematic comparisons between CCR's. To identify also implausible deviations in all CCR's, comparisons with acclaimed reference values will be made (e.g. WHO's Cancer in Five Continents). Outlying values of QI's for individual CCR's will be defined with high specificity in mind, rather than high sensitivity, in order to minimize false positive findings.

The aDQR typically compares the most recent year submitted to the NACR with previous years (unless otherwise specified). The evaluation is per cancer registry and per canton, in the case of noticeable differences between canton and registry. The QI's are described and their selection is justified. For comparison of QI's between registries or with reference values from international sources, the test methods are described.

The aDQR will be issued to all CCR's at first in draft form. Each registry will be able to compare itself with other registries and track changes in its own data quality over time. Registries with statistically outlying QI's will be invited for comment. The finalized aDQR with consolidated statements from the CCR's and the NACR on certain findings, will be made available at the NACR website. It serves as accompanying quality documentation for the cancer statistics that is provided with the website, as well as for third parties using the NCD.

Round Robin Tests

In round robin tests, preselected cancer cases or specific coding problems are delivered to the cancer registries[7]. The individual registrations are compared with ideal solutions prepared in advance by selected experts. The selection of experts depends on the type of cancer or coding problem and is done in accordance with the cancer registry and the NACR. The results are described in a final report prepared jointly by the NACR and the registries. Measures are jointly decided, if necessary, to remedy any shortcomings.

The focus of round robin tests is usually determined by indications of problematic coding practices, either from the aDQR, or from issues raised by individual CCR's, or other organisations. This usually involves clarifications or changes to the Swiss Coding Handbook (SCHB)[8], as well as cancer registry workshops dedicated to the problem. Round Robin tests are able to quantify the initial problem, but also to assess the successful implementation of new practices.

Round robin tests may vary considerably in scale (large/small). In 2022, the NACR will focus on surveying current coding practices in narrowly defined areas.

[7] In Switzerland, a large scale round robin test was conducted by NICER for the first time at the end of 2014 under the project title "VARICO". The identified heterogeneities in registration practices led to the revision of specifications for the National Cancer Dictionary from version 3.0 (2013) to 4.0 (2016).

[8] Details are given in the NACR document «Definite Procedure for Cooperation within the Framework of Optimizing the SCHB»; 1.3.2021.

Random Sampling of Registered Cases

The NACR may randomly select a number of cases from the NCD and ask the registries to submit the anonymized original case reports. The NACR compares the submitted case report data with the existing coding in the NCD. The results are reported back to each registry and measures to remedy any potential deficiencies are jointly decided. Random sampling of registered cases is planned for the first time in 2023. For this purpose, a procedure jointly agreed between NACR, CCR's, and ChCR will be prepared for the beginning of 2023.

Registration Workshops ("CoreDays") and Working Groups

For identified data quality problems, focused training events will be held in which improved registration can be explained and practiced. Other possible topics for workshops are changes in national or international guidelines for cancer registration, or in classification systems (currently these are the CHOP classification of cancer treatment[9] and the ICD-O-3.2 classification of cancer). A new concept paper for organization of workshops by the NACR is under development and will be presented in 2023.

Coding questions on the registry level are collected and discussed regularly in working groups. Working groups constitute of experienced cancer registration persons (at least six years coding experience) and at least one coding experienced person of NACR. Working groups also discuss different coding practices, exchange recent knowledge, evaluate ideal solution and develop drafts for auxiliary material or constructive adaptation request for ENCR/IACR.

For the CHOP area a working group is planned to improve collecting, abstracting and reporting treatment data, based on the intended use of the data.

Special Reports

Special reports are dedicated to important topics and are published on an irregular basis. The first special report, published in 2017, was dedicated to completeness of case ascertainment[10]. The second special report is scheduled for 2023 and describes the quality, completeness and usability of the UICC Stage data available in the NCD for routine national cancer reporting.

Uniform Registration Software

The use of the same registration software (RSW) for collecting and processing cancer registry data in all CCR's and the ChCR supports standardized practices. Interfaces between reporting entities or institutions and the registration software have been created, and will be further improved. Furthermore, the NCD Dictionary[11] (NCD-D) is designed to allow classification systems which are used routinely by reporting parties be directly exported to the cancer registries (e.g. CHOP codes for treatments, or cancer-specific grading systems).

3.3. Childhood Cancer Registry data

[9] NACR-Workshops about CHOP were held 3., 16., 17.12.2021.

[10] Lorez et al. Evaluation of completeness of case ascertainment in Swiss cancer registration. *EJCP* 2017; **26**, 139-146.

[11] The National Cancer Data Dictionary consists of three parts: (1) the variables of the basic data for adults, adolescents, and children, (2) the variables of the supplementary data for adults, and (3) the variables of the supplementary data for children and adolescents.

The data quality of cancer cases in children and adolescents (< 20 years) is in the responsibility of the ChCR. The 'Concept for Publication of Cancer Data' states that the NACR publishes indicators to assess data quality including all age groups, while the ChCR publishes quality indicators for data on children and adolescents[12].
The ChCR submits annually abstracted data for new cases in children and adolescents to the CCR's. The quality of the submitted data is reported to the CCR's and the NACR, starting with diagnoses 2020.

4. TIMELINE for data quality concept development

The data quality concept was developed by the NACR in agreement with the CCR's, and the ChCR. The agreed upon processes enable regular exchange between the NACR and the registries on the topic of data quality and improvement measures. The opinion of the cancer registries on the present concept is obtained (in English) in writing and documented.

Datum	**Prozessschritte**
10.03.2021	Q-Concept first draft finished in DE
1.04.2021	FOPH feedback (formal legal aspects) received
25.09.2021	Second draft finished
30.09.2021	Translation DE to EN
15.11.2021	CCR feedback received
31.12.2021	Q-Concept version 1.0 in EN
Jan – April 2022	Practical application of QI's as basis for QI's refinement
May/June 2022	Workshop for QI's refinement with CCR's/ChCR
June 2022	Latest date for data submissions to NACR (incl. diagnoses 2019)
August 2022	Q-Concept version 1.1 available in EN, DE, FR
September 2022	Reporting findings of aDQR (incl. diagnoses 2019) to CCR's/ChCR for comment
October 2022	The first aDQR (incl. diagnoses 2019) based on Q-Concept version 1.1 available online (EN)

5. COMPARABILITY

Registry data are only comparable if uniform registration standards and definitions are available and adhered to by the registries. The possibility that individual errors may occur in the coding of cancer information despite known rules is not dealt with here, but in the accuracy dimension.

5.1. Prerequisites

The publication of the NCD-D and the SCHB form the basis for comparable data. Both documents are regularly revised. NCD-D and SCHB also specify which classification systems are to be used for registration, and when to switch from older to newer versions. Because the NCD-D and the

[12] "Konzept für die Auswertung und Veröffentlichung von Krebsdaten» version 1.0 (Dez 2020). Chapter 6.3.2 (Informationen zur Datenqualität).

SCHB are based on international guidelines whenever possible, international comparability of the registered cancer data has also been assured.

5.2. Evaluation of comparability

QI's in the annual Data Quality Report

Contingency tables

The relative distributions of codings for the main variables in the basic data (see below list) are determined and compared between CCR's using contingency tables. The evaluation is limited to the five most common cancers (female breast cancer, colorectal cancer, lung cancer, prostate cancer, melanoma) and for total cancer (C00-C97 excluding C44). Stratification is by age group (20-59, 60-74, 75+) and year of diagnosis. Testing for categorical variables is performed using the Chi-square test for independence of the distribution in the variable from the cancer registry (unadjusted, but corrected for multiple testing). Findings will be published in the aDQR.
Prioritized basic data variables include the following, in particular those that have generated issues in the past or have not previously been quality-checked:

>Person information

Sex; Civil status; Vital status

>Diagnosis information

Method of 1st detection; Most valid basis of diagnosis; ICD-O Topography; ICD-O Behaviour; Associated in situ tumour; ICD-O Histological grade; Laterality; y-Prefix of cTNM; cT*; cN*; cM*; a-Prefix of pTNM; y-Prefix of pTNM; pT*; m-Suffix of pT; pN*; pM*; TNM stage group; Lymphatic invasion; Venous invasion; Perineural invasion; Topography of metastasis at diagnosis.
(*only the 1st digit after T, N, M)

>Breast cancer only (women): prognostic factor information

Elston/Ellis grading system; Her2 receptor status.

>Prostate cancer only: prognostic factor information

WHO Grade group.

>Colon cancer only: prognostic factor information

Circumferential resection margins; Microsatellite instability.

>Treatment prognostic factor information

Residual invasive; Residual in-situ tumour; Sentinel lymph node assessment.

>1st Treatment complex information

Basis of 1st treatment complex decision; 1st treatment complex goal(s); Type of recurrence(s)/transformation(s); Topography of post-diagnosis metastases.

>Additional data in adults: colon, breast, or prostate cancer only

Inherited Predispositions; Charlson Index.

>Calendar date information

Accuracy for date of birth; - vital status; - incidence; - 1st treatment complex decision; - 1st treatment complex start; - date of event(s).

Implementation plan for comparability QI's

The details of the contingency tables analysis as parts of the aDQR have not yet been developed. This is planned for 2022 (based diagnoses up to and including 2018), and in 2023 (based on diagnoses including 2019). Inclusion in the aDQR is planned for 2023 (based on diagnoses including 2020). The change in context between diagnoses registered before or after 1.1.2020 (CRO) is taken into account. An overview of implementation plans is given in chapter 10.

5.3. Further development of comparability

Processing coding questions

Uncertainties about the correct application of the NCD-D and the rules set down in the SCHB in daily registration practice are reported to NACR as an ongoing activity. Ambiguities, misunderstandings and errors, or omissions in the NCD-D or SCHB are resolved and the solutions communicated to all registries. An online help desk for posting request and receiving guidance has been created by the NACR[13].

Registration Workshops and Working Groups

Registration workshops bring registrars from different cancer registries and the NACR together in order to practice common understanding and application of the coding rules. The workshop agendas will be determined, among other things, by the most frequent or important request posted by registries at the online help desk, or by the observations made in the aDQR. Workshops are planned and described in a separate concept paper.

Uniform Registration Software

The use of the same registration software for collecting and processing cancer registry data in all CCR's and the ChCR supports standardized and comparable data on the national level. An implementation plan to achieve this goal has been prepared by the FOPH[14].

6. COMPLETENESS OF CASE ASCERTAINMENT

The exact number of all new cancer cases diagnosed in a calendar year is an unknown quantity. In order to evaluate the completeness of case ascertainment in the registry empirically, different methods for the estimation of the true number of diagnoses are available[15], but none is universally accepted.

Apart from "undercounting", the opposite problem of "overcounting" can arise when registries make errors in combining information belonging to the same case and instead assume independent cases, or when registries incorrectly assign different diagnoses belonging to the

[13] https://nicerswiss.sharepoint.com/sites/NACR-CCR/Lists/HelpDesk/AllItems.aspx

[14] "Vorschlag zum Vorgehen der Kantone und der kantonalen Krebsregister bei der Umstellung auf die nationale Registrierungssoftware» (4.4.2019). BAG.

[15] Bullard et al. 2000. *British J Cancer* **82**, 1111-16; Silcocks and Robinson 2007. *J Public Health* **29**, 455-462; Schmidtmann 2008. *Biometr J* **50**, 1077-92; Parkin and Bray 2009. *Eur J Cancer* **45**, 747-755.

same person to two or more persons. Another variation of the "overcount" problem occurs when the same case is recorded in multiple registries, so that two or more independent diagnoses for the same person are incorrectly assumed after merging of the data at the national level (NCD).

6.1. Prerequisites

In principle, the mandatory reporting put into force since 1.1.2020 should automatically ensure that every cancer diagnosis made reaches the cancer registry. However, possibilities to assess the compliance with mandatory reporting are limited. Differences in reporting compliance as well as differences in activities of registries to detect lack of reporting compliance represent an uncertainty factor in the interpretation of regional data.

6.2. Tasks of the cancer registries

Identity verification

Reported information is allocated to an individual person by using the unique personal identification number (UPI) of the social security system (AHVN13) whenever possible. Registries are required to ask for this information if it is missing. Each registry must verify the unique personal identification number with the UPI database of the central compensation office (ZAS) before it is used for the first time (Art. 18 para. 1 let. a CRO).

Clarification of case responsibility

Case responsibility by main residence and age must be determined by comparing the personal identification number with the cantonal or communal population registry (Art. 18 para. 2 CRO). This results in the transfer of all case data to another cancer registry if the diagnosis was made while the person resided outside the catchment area of the registry, or if the age at diagnosis was < 20. It must be prohibited that several cancer registries are processing data on the same diagnosis because they are unaware of conflicts of case responsibility. Duplicate registrations are avoided by querying the information system of the NCDS maintained by the NACR (Art. 18 para. 1 let. b CRO). If the same person is listed in two or more registries, the registries must negotiate that they have not processed identical diagnoses.

Checking reporting activity

The cancer registries are in direct contact with the reporting persons and institutions and should know the expected number of new cancer cases by hospital, department or group practice based on various criteria, such as treatment volume and focus, e.g. from the values of the past three years. It is in the responsibility of the CCR/ChCR to define reporting entities (hospital, department or group practice). If there are obviously fewer (>20% less) new cancer cases reported than expected, inquiries must be made at the institution in question.

Implementation plan for registry tasks

The tasks of identity verification as well as clarification of case responsibility has started 1.1.2020 for diagnoses 2020, in accordance with the CRO.

The earliest date for the task of CCR/ChCR to evaluate reporting activity is 2023 after having received diagnoses of 2020 and 2021.

6.3. Evaluation of completeness of case ascertainment

QI's in the annual Data Quality Report

A single QI alone cannot inform on completeness of case ascertainment because general agreement on a gold standard method is lacking. Therefore, the joint consideration of several QI's is required.

Historical Trend

Semi-quantitative methods assess completeness of case ascertainment indirectly, without attempting to quantify the number of missing cases. These include methods that examine the stability of incidence numbers or rates over time ("historical methods") or make comparisons with standard values, if such standards are available[16].

For the aDQR, a simple count of the number of cases in the most recent diagnosis year is applied in order to identify potential problems of finding cases. The method cannot identify systematic under-, or over-counts. Prominent incidence trends and rare cancer types may be difficult to interpret. For a number of non-malignant diagnoses, mandatory reporting and registration does not begin until diagnosis year 2020. Such diagnoses cannot be assessed before diagnosis year 2023 at the earliest.

QI Name (abbreviation)	Minuend	Subtrahend	Time	Remarks
Difference between submitted and expected case number (**diffSE**)	Case number in one diagnosis year submitted to NACR	Average case number of the 3 adjacent diagnosis years submitted to NACR	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by cancer type (see list in Appendix) and malignancy. Flagging of unusually high or low values if |diffSE| > 15% of the subtrahend, and numerical |diffSE| > 15.

High Proportion of Morphologically Verified Cases

Another indirect approach relates to the proportion of diagnoses with the highest diagnostic validity, which are those using morphological methods (MV%). An unusually high proportion of diagnoses based on histology or cytology/haematology could indicate an excessive reporting contribution of pathology laboratories and thus potential under-registration of diagnoses from other sources[17].

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Proportion of morphologically verified cases (**MVhi**)	Case number with codes 5 (cytology), 6 (histology of metastasis)[18], and 7 (histology of primary tumour) in variable	All cases with valid code in variable «Most valid basis of diagnosis» of the NCD-D	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by cancer type (see list in Appendix) and malignancy.

[16] Curado et al. 2007. Cancer Incidence in five continents, vol IX, IARC Scien Public **160**.; Hackl et al. 2011. *Statistische Nachrichten*, **9**, 848-859.

[17] Bray and Parkin 2009. *Eur J Cancer* **45**, 747-755.

[18] Inclusion of histology of metastasis is according to: Standards and Guidelines for Cancer Registration in Europe (2003). IARC Technical Publication No.40.

	«Most valid basis of diagnosis» of the NCD-D			Flagging of unusually high values with statistical test after Parkin and Plumer in CI-V vol. VIII[19]

Death Certificate Notifications

Another indirect measure of completeness of case ascertainment is the proportion of cases in which registration was triggered by a death certificate (DCN). When registrations are often triggered by death certificates, missed diagnoses are likely due to the known imprecision and lack of specificity in the certified causes of death[20]. The likelihood of a cancer diagnosis appearing on the death certificate at all decreases with time after diagnosis[21]. Note that the proportion of cases which are registered solely based on a death certificate (DCO) are not a good estimator of completeness of case ascertainment because this proportion is usually only a fraction of the DCN due to follow-back enquiries which often successfully identify relevant medical information.

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Proportion of cases with a death certificate as earliest notification (**DCN**)	Case number with code 1 in variable «DCN flag» of the NCD-D	All cases with valid code in variable «DCN flag» of the NCD-D	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by cancer type (see list in Appendix) and malignancy. Flagging of unusually high values if DCN ≥ 10%

Mortality-to-Incidence Rate Ratio

The completeness of case ascertainment of the cancer registry can also be assessed by comparing the mortality-to-incidence (MI) rate ratio with reference registries that are considered complete and have the same expected ratios[22]. Reference MI rate ratios are not required when comparing MI rate ratios to estimates of relative survival from the same registry, as both are determined by the case-mortality rates prevailing in the population[23].

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Ratio of the crude mortality to the crude incidence rate (**MIRR**)	crude mortality rate in the defined time period	crude incidence rate in the defined time period	Interval of 3 consequtive years	Analysis is stratified by cancer type (see list in Appendix) and malignancy. Flagging of conspicuous values with statistical test after Parkin and Bray (2009) between a registry and the pool of all registries[24]

[19] The NACR has gained experience with the applicability of this test in the special report: "Evaluation of Completeness of Case Ascertainment in Swiss Cancer Registration." (2017) by Lorez M, Bordoni A, Bouchardy C, Bulliard JL, Camey B, Dehler S, Frick H, Konzelmann I, Maspoli M, Mousavi SM, Rohrmann S and Arndt V., published in *EJCP* **26**, 139-146.

[20] Mathers C, Fat D, Inoue M, Rao C, Lopez A (2005). Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bulletin of the World Health Organization* **83**(3), 171-77.

[21] Bullard J, Coleman M, Robinson D, Lutz J, Bell J, Peto J (2000). Completeness of cancer registration: a new method for routine use. *British J Cancer* **82**(5), 1111-16.

[22] Haberland J, Bertz J, Görsch B, Schön D (2001). Krebsinzidenzschätzungen für Deutschland mittels log-linearer Modelle. *Gesundheitswesen* **63**, 556-60. // Hofferkamp, J (Ed). Standards for Cancer Registries Volume III: Standards for Completeness, Quality, Analysis, Management, Security and Confidentiality of Data. Springfield (IL): North American Association of Central Cancer Registries, August 2008.

[23] Parkin DM, Bray F (2009). Evaluation of data quality in the cancer registry: Principles and methods Part II. *Eur J Cancer* **45**, 756-64. // Vostakolaei FA, Karim-Kos HE, Janssen-Heijnen MLG, Visser O, Verbeek ALM, Kiemeney L (2010). The validity of the mortality to incidence ratio as a proxy for site-specific cancer survival. *Eur J Pub Health* **21**(5), 573-577.

[24] The NACR has gained experience with the applicability of this test in the special report: "Evaluation of Completeness of Case Ascertainment in Swiss Cancer Registration." (2017) by Lorez M, Bordoni A, Bouchardy C, Bulliard JL, Camey B, Dehler S, Frick H, Konzelmann I, Maspoli M, Mousavi SM, Rohrmann S and Arndt V., published in *EJCP* **26**, 139-146.

Implementation plan for case finding QI's

Refinements to QI's in a workshop together with CCR's/ChCR are planned for 2022 (based diagnoses up to and including 2018). Inclusion in the aDQR is planned for 2022 (based diagnoses up to and including 2019). The change in context between diagnoses registered before or after 1.1.2020 (CRO) is taken into account. An overview of implementation plans is given in chapter 10.

6.4. Further development of completeness of case ascertainment

Clearingprozess ZAS/UPI

Despite the use of AHVN13, there may still be problems with the identification of persons due to the possibility of incorrect AHV numbers in the population registers. This can occur if a person does not register with the same name as it is stored in ZAS/UPI, or because the AHV number in ZAS/UPI is an incorrect/outdated. Of course, this can lead to problems when querying address information (e.g. for case responsibility) in population registers. ZAS/UPI has worked out a procedure to be followed in order to correct errors (clearing process). Cancer registries are required to forward any person identification problems into the ZAS/UPI clearing process to avoid future conflicts[25].

Date of informing the patient

Without documented date of information about cancer registration and the patient's right to object to registration, registration cannot be performed. Failure to provide information, or failure to report the date of information to registries, may thus result in under-reporting of cancer cases. The NACR supports the CCR's and the ChCR in information campaigns to draw the attention of physicians to this problem.

Due to this problem, the CRO has been recently changed: registration of case data is possible if no veto has been issued until 3 months after the notification date. The new rule will be in force after 1.1.2022[26].

7. ACCURACY

On the one hand, the quality dimension of accuracy deals with the frequency of registration errors. On the other hand, the accuracy assesses the extent to which the data collection as a whole is based on high-quality information provided by the reporting parties, or on documents with known quality-related deficiencies, such as death certificates (see dimension Completeness of case ascertainment: QI's of MVhi and DCN).

The accuracy does not address the issue of systematic error due to misinterpretation of the registration rules laid down in the Swiss Coding Handbook (see dimension Comparability).

[25] https://www.zas.admin.ch/zas/de/home/partenaires-et-institutions-/unique-person-identification--upi-/rectification-des-donnees.html

[26] file:///C:/Users/lom/AppData/Local/Temp/%C3%84nderungserlass_DE-3.pdf

7.1. Prerequisites

[to be added].

7.2. Tasks of the cancer registries

The responsibility for error-free registration lies with the cancer registries. It is also the responsibility of the registries to trace-back when sources with known deficiencies in the content of diagnostic or treatment information (e.g. death certificates) are notified.

7.3. Evaluation of accuracy

QI in the annual Data Quality Report

Formal validity of the statement

The proportion of invalid codes in a variable.

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Invalid, or undefined code (**INVAL**)	Number of cases with codes not in compliance with the NCD-D	All cases, excluded those with codes for missing information	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by cancer type (see list in Appendix) and malignancy. Flagging of unusually high values (> 15%).

Version- and site-specificity of UICC T, N, M

All UICC T, N, M codes and stage groups correspond to the assigned UICC versions and associated cancer localizations. Note that these tests are, in principle, implemented in the JRC/ENCR Quality Check Software (QCS). They are carried out only if the UICC TNM stage group is not missing. As long as the variable "TNM stage group" of the NCD-D is incompletely recorded, these tests are performed with NACR-generated checking code. Findings are reported as warnings. Tests performed by the NACR may be requested in the form of STATA program code.

Logical contradictions

Contradictions of codes in one variable with those in another variable.

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Contradictory data (**CONTRA**)	Number of comparisons with contradictions between variables A and B	All comparisons between variables A and B (cases with missing codes in A or B excluded)	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by cancer type (see list in Appendix) and malignancy. Flagging of unusually high values (> 5%).

Findings are documented as errors and must always be corrected. The information may be incorrect in one or the other variable, or both.

The list of all tests includes (1) all tests implemented in the JRC/ENCR QCS, and (2) all NACR-defined tests. All tests performed in (1) and (2) are documented in the Appendix.

Implausible Combinations

The combination of codes in different variables is not logically impossible, but occurs rarely and a verification of the coding is justified. Verified codings require the setting of a "checked" variable so that future checking rounds do not ask for repeated verifications[27].

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Implausible data (**IMPL**)	Number of comparisons with unlikely combinations of codes in variables A and B	All comparisons between variables A and B (cases with missing codes in A or B excluded)	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by cancer type (see list in Appendix) and malignancy. Flagging of unusually high values (> 5%).

Findings are documented as warnings. The list includes (1) all tests implemented in the JRC/ENCR QCS, and (2) all NACR-defined tests. All tests performed in (1) and (2) are documented in the appendix.

Low Proportion of Morphologically Verified Cases

The proportion of 'morphologically' (used synonymously with 'microscopically') verified cases. The test checks whether the proportion of cases with the highest validity (microscopically verified) is unusually low. (Note the different interpretation of this proportion in completeness of case ascertainment!).

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Proportion of morphologically verified cases (**MVlo**)	Case number with codes 5 (cytology), 6 (histology of metastasis)[28], and 7 (histology of primary tumour) in variable «Most valid basis of diagnosis» of the NCD-D	All cases with valid code in variable «Most valid basis of diagnosis» of the NCD-D	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by cancer type (see list in Appendix) and malignancy. Flagging of unusually low values with statistical test after Parkin and Plumer in CI-V vol. VIII[29]

Death Certificate Only

Proportion of cases which are registered only with data in the death certificate. It is tested whether this proportion is unusually high in a certain registry.

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Proportion of cases with a death certificate as only notification (**DCO**)	Case number with code 0 in variable «Most valid basis of diagnosis» of the NCD-D	All cases with valid code in variable «Most valid basis of diagnosis» of the NCD-D	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by cancer type (see list in Appendix) and malignancy. Flagging of unusually high values ≥ 10%.

[27] At present, only a single checkbox variable is available. It serves to indicate that the entirety of JRC/ENCR QCS findings have been cleared. If the case is updated as a later stage, the checked status has to be removed. The NACR plans a solution which enables the flagging of individual findings having been cleared. Then, only the checked status for the updated information of a case has to be removed, and the checked status for unchanged information ca be retained.

[28] Inclusion of histology of metastasis is according to: Standards and Guidelines for Cancer Registration in Europe (2003). IARC Technical Publication No.40.

[29] The NACR has gained experience with the applicability of this test in the special report: "Evaluation of Completeness of Case Ascertainment in Swiss Cancer Registration." (2017) by Lorez M, Bordoni A, Bouchardy C, Bulliard JL, Camey B, Dehler S, Frick H, Konzelmann I, Maspoli M, Mousavi SM, Rohrmann S and Arndt V., published in *EJCP* **26**, 139-146.

Cases with Primary Site Uncertain

Proportion of cases registered with unknown/nonspecific primary tumour (PSU, primary site uncertain/unknown). These are ICD-10: C26, C39, C48, C75, C76, C80[30]. It is tested if this proportion is unusually high. Since this QI assesses diagnostic quality (i.e., precision, or failure to detect the primary site when sampling from a metastasis), a high value indicates problematic data quality but says nothing about the quality of the registration.

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Proportion uncertain/unspecific malignant primary diagnoses (**PSU**)	Number of primary diagnoses ICD-10: C26, C39, C76, C80	All malignant primary diagnoses	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by malignancy. Flagging of unusually high values with statistical test after Parkin and Plumer in CI-V vol. VIII

Implementation plan for accuracy QI's

Refinements to QI's in a workshop together with CCR's/ChCR are planned for 2022 (based diagnoses up to and including 2018). Inclusion in the aDQR is planned for 2022 (based diagnoses up to and including 2019), with the exception of PSU. The change in context between diagnoses registered before or after 1.1.2020 (CRO) is taken into account. The details of the evaluation of the QI PSU as part of the aDQR have not yet been developed. Implementation of this QI is planned for the publication date 2023 (based on diagnoses including 2020). An overview of implementation plans is given in chapter 10.

7.4. Further development of accuracy

Additional Tests

The list of tests for logical inconsistencies, or implausible coding will be evaluated annually and expanded if needed.

Uniform Registration Software

The use of the same software program for collecting and processing cancer registry data in all CCR's and the ChCR helps to minimize errors in data entry, in single data fields and with respect to the relation between fields[31].

Use of Artificial Intelligence

Support from artificial intelligence can free employees from tedious routine work. In times of scarce personnel resources, employees can concentrate on the essentials. NACR has made initial contacts with companies that have experience with machine reading of pathology reports in cancer registration. Computer programs may support the coding process, but also perform plausibility checks at a later stage.

Artificial intelligence can favourably influence not only the accuracy, but also the comparability of the data: the quality of human coding is dependent on the inhomogeneous competence and

[30] Bray and Parkin 2009. *Eur J Cancer* **45**, 747-755.

[31] This includes adapting the drop-down selection options to the tumour type at hand, or implementing data checks directly during data entry.

experience of the individual coding professional. Artificial intelligence, on the other hand, provides more consistent output. NACR is monitoring these developing technical possibilities. There are various providers in the field of medical coding that offer solutions for the automatic recording of diagnoses and treatments on the basis of unstructured and structured reports. A deeper examination of existing solutions, an exchange of experience with users and an analysis of transferability would be advisable. As a first step, a survey to assess needs of users (cancer registries) would be necessary to find out in which points of coding/registration they can benefit most in terms of time or quality from a (partially) automated solution in order to set priorities for (a) pilot project(s).

Random Case Sampling
The gold standard method for checking the coding accuracy is to take a random sample of already registered cases and re-register them independently by highly qualified personnel using the original reports. From the extent of agreement between the previously registered data and the separately derived codes, the frequency of errors in the overall data collection can be estimated.

Implementation plan for random case sampling
Evaluating of a random sample of cases from CCR's/CHCR databases has not yet been attempted. Negotiation of the details with the registries is planned for 2022, so that this test can be carried out for the first time in 2023. An overview of implementation plans is given in chapter 10.

8. CASE COMPLETENESS

This dimension of data quality is concerned about failure to process reported information, or the registration of code "unknown" in spite of existing information in the reported data. Note that data providers are responsible to delete non-mandatory cancer information from routine medical reports before sending them to the cancer registry. Data providers are not required to generate information solely for the purpose of cancer registration. No notification to the registry is required that certain data items of the National Cancer Data Dictionary are not available. Only if relevant data items are assessed as part of diagnosis or treatment, the findings have to be reported to the responsible cancer registry.

8.1. Prerequisites

In principle, the mandatory reporting in force since 1.1.2020 should ensure the completeness of the information on cancer diagnoses. This information is defined in the NCD-D. However, control possibilities regarding compliance with the reporting obligation are limited.

8.2. Tasks of the cancer registries

The cancer registries are required to prevent incomplete registration by making enquiries with the data providers about expected data items which are missing from the report[32]. What data

[32] Citation from: «Erläuterungen zur KRV» to Art. 8: «Um den für die Krebsregistrierung entstehenden Aufwand für die Meldepflichtigen in überschaubaren Grenzen zu halten, dürfen die Meldepflichtigen gemäss Absatz 2 dem Krebsregister Berichte weiterleiten, die sie im Rahmen

items can be expected by experienced registration personal depends on the type of cancer and the data provider.

8.3. Evaluation of completeness

QI in the annual Data Quality Report

Missing information

The proportion of cases with missing information in a variable.
All CCR's and the ChCR have migrated their "historical" data (i.e. information on diagnoses before 2018) into the format required by the NCD-D. It was unavoidable in this process that information was missing for novel variables of the NCD-D, or for novel categories of pre-existing variables[33]. This will be accounted for when evaluating the completeness of "historical" data. Also, data registered since 1.1.2020 for diagnoses in 2018 and 2019 are expected to have missing information because the reporting is only mandatory for diagnoses after 1.1.2020.

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Missing code (**MISS**)	Number of cases with missing code*	All cases*	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by cancer type (see list in Appendix) and malignancy. Flagging of unusually high values (> 20%).

*Excluded are cases, where no code is expected (e.g. resection margin without resection performed). Expectation criteria are defined for each variable evaluated regarding missingness.

Code «Unknown»

The proportion of cases with code "unknown" in a variable.

QI Name (abbreviation)	Nominator	Denominator	Time	Remarks
Unspecific code (**UNSPEC**)	Number of cases with code «Unknown»	All cases with code given (missing excluded)	Most recent diagnosis year submitted to NACR (if not otherwise specified)	Analysis is stratified by cancer type (see list in Appendix) and malignancy. Flagging of unusually high values (> 20%).

Implementation plan for case completeness QI's

Refinements to QI's in a workshop together with CCR's/ChCR are planned for 2022 (based diagnoses up to and including 2018). Inclusion in the aDQR is planned for 2022 (based diagnoses up to and including 2019). The change in context between diagnoses registered before or after 1.1.2020 (CRO) is taken into account. An overview of implementation plans is given in chapter 10.

8.4. Further development of completeness

ihrer beruflichen Tätigkeit zu Dokumentationszwecken ohnehin erstellen. Darunter fallen beispielsweise Tumourboard-, Operations-, Pathologie-, Histologie-, Zytologie- oder Spitalaustrittsberichte, Arztbriefe oder Auszüge aus der Krankengeschichte.».

[33] New categories were defined, for example, with variable 2.7 - Highest achieved diagnostic certainty: the "Imaging" category.

Random Case Sampling
The same applies here as for quality dimension Accuracy. In addition, in the cases tested, if the information is missing, inquiries are made with the data providers to determine whether the information was not assessed during diagnosis or treatments, or has not been reported to the registry.

9. TIMELINESS

With current procedures, the National Cancer Dataset (NCD) is ready to be used in monitoring, reporting, or supporting research at three years after the most recently included diagnosis year. For example, national statistical reports including the new incident cases and cancer deaths in 2020 can only be produced in the course of 2023. It should be kept in mind that a certain time lag between diagnosis and data reporting is unavoidable, especially with regard to complete ascertainment of all diagnoses.

9.1. Prerequisites

The CRA and the CRO provide a framework that should lead to a noticeable acceleration of the processes of reporting, registration, and data submission to the NACR. The goal is to improve timeliness by one year until 2023 (Art. 39 in the explanatory notes to the CRO).
In order to enable the registration of data as promptly as possible, Article 6 in paragraph 1 of the CRO demands that mandatory cancer data must be reported to the responsible cancer registry within 4 weeks of generating the information. It is assumed that data providers are increasingly managing medical histories of their cancer patients electronically, and the information will be reported to the cancer registry automatically in a structured data export.
The Federal Statistical Office (FSO) will deliver the vital statistics to the cancer registries 1 year earlier as presently. Because the trace-back of death certificates is very time-consuming for the registries, this will enable the submission of cancer data to the NACR, and from the NACR to the FSO, also one year earlier as at present.
The time period between diagnosis and inclusion in statistical reports breaks down into three logical segments: (1) from the date of the data item to be known until notification of the registry, (2) processing time at the registry level until submission to the NACR, and (3) subsequent processing time at the NACR until the data is reported.

9.2. Evaluation of timeliness

QI in the annual Data Quality Report

Time to 90% registered

An important factor of timeliness is the lag between date of diagnosis and date of first notification of the case to the cancer registry. Experience has shown that this is highly dependent on the type of cancer[34].

QI Name (abbreviation)	Calculation	Time	Remarks
Time until 90% of the finally registered cases are notified to the cancer registry (**TT90**)	TT90 is derived from the empirical cumulative distribution function of the time intervals between «Date of incidence» and «Date of Notification» of the NCD-D	Interval of 3 consecutive years	Analysis is stratified by cancer type (see list in Appendix).

Implementation plan for timeliness QI's

Refinements to QI's in a workshop together with CCR's/ChCR are planned for 2022 (based diagnoses up to and including 2018) and for 2023 (based diagnoses up to and including 2019). Inclusion in the aDQR is planned for 2023 (based on diagnoses including 2020). The change in context between diagnoses registered before or after 1.1.2020 (CRO) is taken into account.

9.3. Further development of timeliness

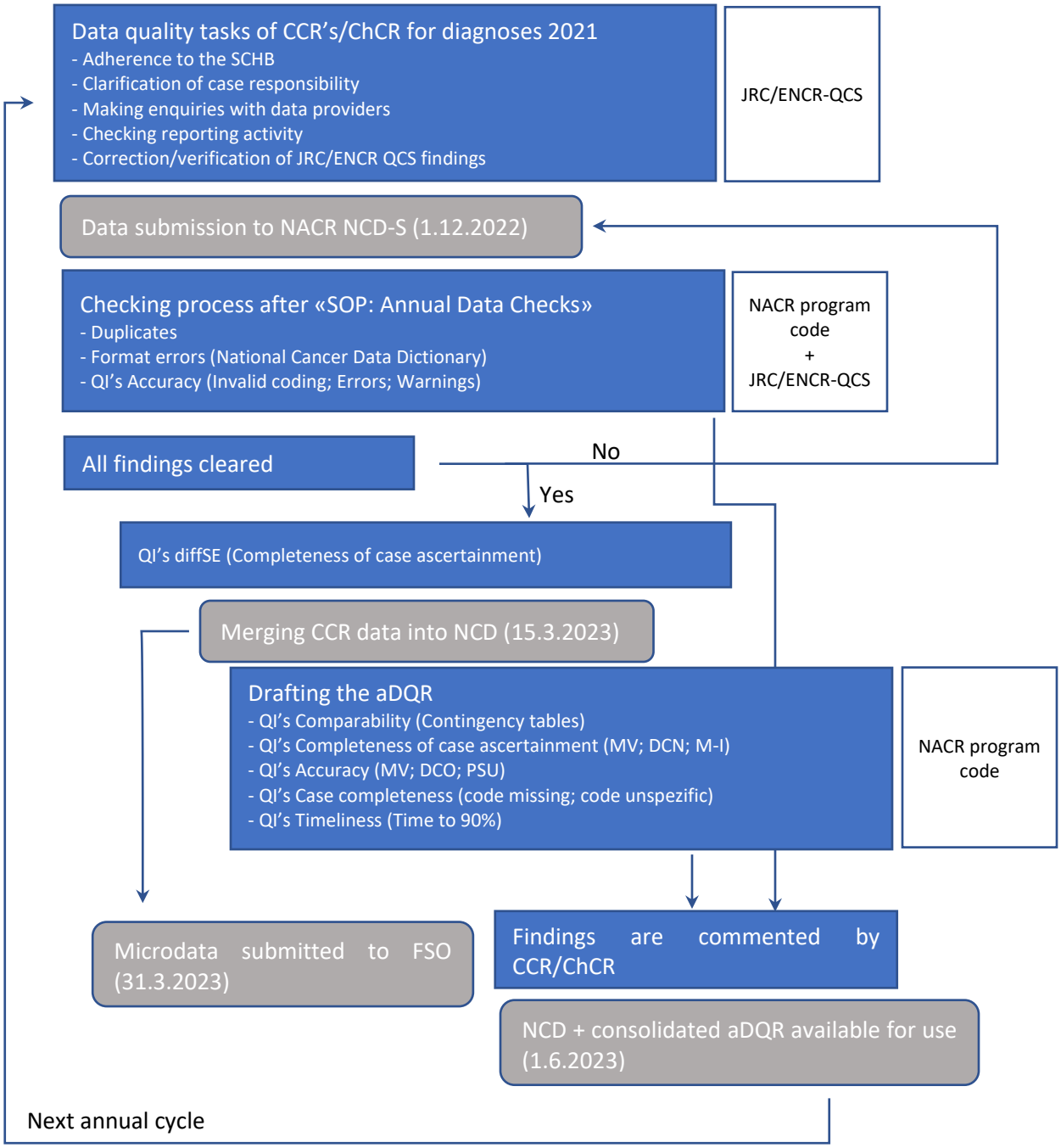
[to be added.]

[34] Lorez et al. Evaluation of completeness of case ascertainment in Swiss cancer registration. EJCP 2017; 26, 139-146. Supplementary digital content No 3.

10. GRAPHICAL OVERVIEWS

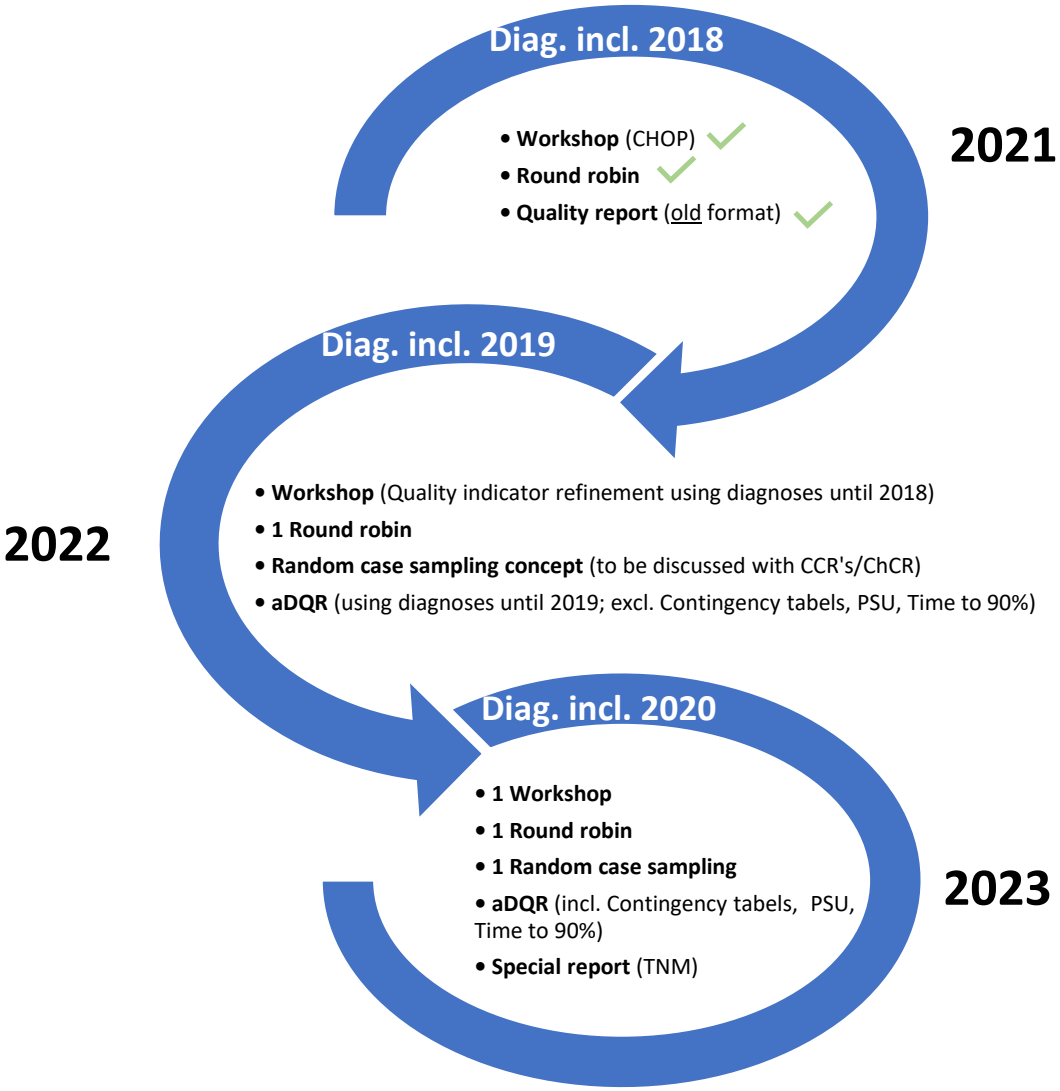
10.1. Annual Quality Assessment and Development Cycle

The graphics shows how, starting with the data quality tasks of the CCR's and ChCR, the different instruments for evaluation and development of data quality are working together. The cycle is repeated each calendar year. As an example, the cycle of calendar year 2023 is shown.



10.2. Implementation Plan Summary

Graphical overview of implementation plans for several quality improvement measures: workshops, round robin tests, annual Data Quality Report (aDQR) und random case sampling.



11. APPENDIX

11.1. Cancer types for separate stratified analyses

Lip, Oral Cavity, Pharynx
Oesophagus
Stomach
Colon, rectum, anus
Liver & Intrahepatic Bile Ducts
Pancreas
Lung, Bronchus, Trachea
Skin Melanoma
Breast (female)
Cervix
Corpus & uterus NOS
Ovary
Prostate
Testis
Kidney
Bladder
Eye, Brain, Central Nerves
Thyroid
Hodgkin Lymphoma
Non-Hodgkin Lymphoma
Multiple Myeloma
Lymphoid Leukaemia
Myeloid Leukaemia
All sites (except non-melanotic skin cancer)

11.2. NACR-defined checks

Table 1. Single variable checks and multivariable checks defined by the NACR. They are applied in addition to the checks of the JRC/ENCR QCS. Variables are identified by the abbreviations used for annual data submission to the NACR.

code	type	message	comment
18	error	variable "ncid" (Case number) > 10 digits	
19	error	variable "ncid" (Case number) with point	
20	error	variable "ncid" (Case number) < 0	
39	error	variable "sex" (Sex) with undefined code	
49	error	variable "d_birth" (Date of birth) with day not 15	
50	error	variable "d_birth" (Date of birth) with month not 1-12	
51	error	variable "d_birth" (Date of birth) with year<1880 or >yyi	
52	warning	variable "d_birth" (Date of birth) missing	Not even the year imputable?
55	error	variable "dacc_birth" (Accuracy for date of birth) with undefined code	
56	warning	variable "dacc_birth" (Accuracy for date of birth) missing	Not even the year imputable?
69	error	variable "nat" (Nationality) with undefined code	
70	warning	variable "nat" (Nationality) missing	Code “unknown” exists

79	error	variable "bd" (Most valid basis of diagnosis) with undefined code	
80	error	variable "bd" (Most valid basis of diagnosis) missing	
89	error	variable "dcn" (DCN flag) with undefined code	
90	warning	variable "dcn" (DCN flag) missing	Truly unknown?
98	error	variable "d_i" (Date of incidence) with day not 15	
99	error	variable "d_i" (Date of incidence) with month not 1-12	
100	error	variable "d_i" (Date of incidence) with year<1970 or >recent	
101	warning	variable "d_i" (Date of incidence) missing	Not even the year imputable?
103	error	variable "age_i" (Age at incidence) missing	
104	error	variable "age_i" (Age at incidence) > 5 digits	
105	error	variable "age_i" (Age at incidence) with point	
106	error	variable "age_i" (Age at incidence) < 0	
107	warning	variable "age_i" (Age at incidence) > 110 years	
109	error	variable "dacc_i" (Accuracy for date of inc) with undefined code	
110	warning	variable "dacc_i" (Accuracy for date of inc) missing	Not even the year imputable?
130	error	variable "lat" (laterality) with undefined code	
140	error	variable "topo" (topography) with point	
141	error	variable "topo" (topography) missing	Code “unknown” exists
145	error	variable "mph" (morphology) missing	Code “unknown” exists
150	error	variable "beh" (behaviour) with undefined code	
151	error	variable "beh" (behaviour) missing	Code “unknown” exists
170	error	variable "icdo_v" (ICD-O-Version) with undefined code	
171	error	variable "icdo_v" (ICD-O version) missing	Not imputable?
180	error	variable "d_notif" (Date of notification) with day not 15	
181	error	variable "d_notif" (Date of notification) with month not 1-12	
182	error	variable "d_notif" (Date of notification) with year<1970 or >recent	
183	warning	variable "d_notif" (Date of notification) missing	Not even the year imputable?
190	error	variable "age_notif" (Age at notification) missing, with known d_birth/d_notif	
191	error	variable "age_notif" (Age at notification) > 5 digits	
192	error	variable "age_notif" (Age at notification) with point	
193	error	variable "age_notif" (Age at notification) < 0	
194	warning	variable "age_notif" (Age at notification) > 110 years	
240	error	variable "checked" (exception_verified) with undefined code	
250	error	variable "civ" (civil status) with undefined code	
260	error	variable "detec1" (Method of first detection) with undefined code	
261	warning	variable "detec1" (Method of first detection) missing	Code “unknown” exists
264	error	variable "detec2" (Diagnostic method(s) used) with blank(s)	
265	error	variable "detec2" (Diagnostic method(s) used) with undefined code(s)	
266	warning	variable "detec2" (Diagnostic method(s) used) missing	Code “unknown” exists
270	error	variable "clinsize" (clinical tumour size) with undefined values	
271	error	variable "pathsize" (pathological tumour size) with undefined values	
280	error	variable "tnm_v" (UICC TNM version) with undefined values	
281	error	variable "tnm_v" (UICC TNM version) missing, but TNM coded	
288	error	invalid "y_ctnm", independent on cancer type	
290	error	invalid "cT", independent on cancer type	
300	error	invalid "cN", independent on cancer type	
310	error	invalid "cM", independent on cancer type	
320	error	invalid "y_ptnm", independent on cancer type	
322	error	invalid "a_ptnm", independent on cancer type	
330	error	invalid "pT", independent on cancer type	
335	error	invalid "m_pt" (m suffix pT), independent on cancer type	
340	error	invalid "pN", independent on cancer type	
360	error	invalid "pM", independent on cancer type	
361	warning	variable "lymph_inv" (lymphatic invasion) missing, but pT info exist	
362	warning	variable "ven_inv" (veneous invasion) missing, but pT info exist	

363	warning	variable "pn_inv" (perineural invasion) missing, but pT info exist	
364	error	variable "lymph_inv" (lymphatic invasion) with undefined code	
365	error	variable "ven_inv" (veneous invasion) with undefined code	
366	error	variable "pn_inv" (perineural invasion) with undefined code	
368	error	variable "st_tnm" (TNM Stage Group) with undefined code	
370	error	invalid entry in variable "grade of differentiation" (item 1.37; NCDv3)	
400	error	variable "rln_exam" (Number of examined regional lymph nodes) with undefined code	
410	error	variable "rln_inv" (Number of involved regional lymph nodes) with undefined code	
500	error	variable "sfu" (vital status) with undefined code	
501	error	variable "sfu" (vital status" missing	Code “unknown” exists
510	error	variable "d_fu" (Date of vital-status) with day not 15	
511	error	variable "d_fu" (Date of vital-status) with month not 1-12	
512	error	variable "d_fu" (Date of vital-status) with year<1970 or >recent	
513	warning	variable "d_fu" (Date of vital-status) missing	Not even the year imputable?
514	error	variable "dacc_fu" (Accuracy for date of vitalstatus) with undefined code	
515	warning	variable "dacc_fu" (Accuracy for date of vitalstatus) missing	Not even the year imputable?
517	error	variable "age_fu" (Age at vitalstatus) missing, with known d_birth/d_fu	
518	error	variable "age_fu" (Age at vitalstatus) > 5 digits	
519	error	variable "age_fu" (Age at vitalstatus) with point	
520	error	variable "age_fu" (Age at vitalstatus) < 0	
521	warning	variable "age_fu" (Age at vitalstatus) > 110 years	
530	error	variable "cd_princ" (principle cause of death) with wrong number of digits	
531	error	variable "cd_princ" (principle cause of death) with point	
533	error	variable "cd_cod1" (primary cause of death) with wrong number of digits	
534	error	variable "cd_cod1" (primary cause of death) with point	
536	error	variable "cd_cod2" (secondary cause of death) with wrong number of digits	
537	error	variable "cd_cod2" (secondary cause of death) with point	
539	error	variable "cd_cod3" (first tertiary cause of death) with wrong number of digits	
540	error	variable "cd_cod3" (first tertiary cause of death) with point	
542	error	variable "cd_cod4" (second tertiary cause of death) with wrong number of digits	
543	error	variable "cd_cod4" (second tertiary cause of death) with point	
545	error	variable "cd_v" (ICD-version for causes of death) with undefined code	
546	warning	variable "cd_v" (ICD-version for causes of death) missing	Not imputable?
610	error	variable "bc_erec" (estrogen receptor status C50) with undefined code	
620	error	variable "bc_prec" (progesteron receptor status C50) with undefined code	
630	error	variable "bc_her2rec" (HER2 receptor status C50) with undefined code	
640	error	variable "bc_tpl" (tumour proliferation labeling C50) with undefined code	
650	error	invalid entry in variable "method of detection" (item 1.26, NCDv4)	
680	error	variable "gr_icdo" (ICD-O Histological grade) with undefined code	
800	error	duplicate "ncid" (national case identifier)	
801	error	variable "ncid" (national case identifier) not coding for canton at diagnosis	
805	error	unequal sexes within one patient	
809	error	unequal month of birth within one patient	
810	error	unequal year of birth within one patient	

815	error	unequal month of follow-up within one patient	
816	error	unequal year of follow-up within one patient	
820	error	unequal principle cause of death within one patient	
821	error	unequal primary cause of death within one patient	
822	error	unequal secondary cause of death within one patient	
823	error	unequal first tertiary cause of death within one patient	
824	error	unequal second tertiary cause of death within one patient	
825	error	unequal ICD-version (cause of death information BFS) within one patient	
830	error	impossible dates (birth, diagnosis) vs age at diagnosis	
833	error	date incidence < date birth	
840	error	impossible dates (birth, vitalstatus) vs age at vitalstatus	
843	error	age follow-up < age incidence	
845	error	DCO case, but date_incidence not equal date_follow-up	
846	error	DCO case, but age_incidence not equal age_follow-up	
850	error	basis of diagnosis = DCO, but DCN = no	
851	error	DCO case, but vital status not dead	
852	error	DCN-case, but vital status not dead	
860	warning	known vital status at follow-up, but year of follow-up uncertain	
861	warning	known vital status at follow-up, but unknown age at follow-up	
862	warning	unknown vital status at follow-up, but known date of follow-up	
863	error	unknown age at vital status, but known date of follow-up	
870	error	implausible combination of dates of diagnosis and vitalstatus and survival duration	
875	error	registration year < incidence year	
2290	warning	beh=2, but pT1-T4 or pN1-N3 or c/pM1 codes	
2300	warning	unexpected value in cT (site-specific, version-specific check)	
2310	warning	unexpected value in pT (site-specific, version-specific check)	
2320	warning	unexpected value in cN (site-specific, version-specific check)	
2330	warning	unexpected value in pN (site-specific, version-specific check)	
2340	warning	unexpected value in cM (site-specific, version-specific check)	
2350	warning	unexpected value in pM (site-specific, version-specific check)	

11.3. JRC/ENCR QCS checks

[See the JRC documents:

(1) “Carmen Martos, Emanuele Crocetti (Coordinator), Otto Visser, Brian Rous, Francesco Giusti 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, EUR 29089 EN, Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79-77889-6, doi:10.2760/429053, JRC105078”

(2) Francesco Giusti, Carmen Martos, Stefano Adriani, Manuela Flego, Antonino Brunetto, Tadeusz Dyba, Lena Voith von Voithenberg, Luciana Neamtiu, Raquel N. Carvalho, Giorgia Randi, Nadya Dimitrova, Nicholas Nicholson, Revveka Trigka, Emanuele Crocetti, Manola Bettio, Enrico Ben, The JRC-ENCR Quality Check Software (QCS) for the validation of cancer registry data: user compendium – version 2.0, European Commission, Ispra 2021, JRC127031]

Documents are available on request.