Statistical Methods for Cancer Reporting in Switzerland

Version 1.0

Authors (alphabetically)

Karim Abawi, FSO
Anita Feller, NACR
Matthias Lorez, NACR
Eleftheria Michalopoulou, ChCR
Elodie Roy, FSO
Ben Spycher, ChCR
Rolf Weitkunat, FSO

https://www.cancer.bfs.admin.ch
https://www.nacr.ch
https://www.childhoodcancerregistry.ch

Documentation of statistical methods used by the National Agency for Cancer Registration (NACR), the Childhood Cancer Registry (ChCR), and the Federal Statistical Office (FSO)

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

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

Kinderkrebsregister
Registre du cancer de l'enfant
Registro dei tumori pediatrici
Childhood Cancer Registry

Zürich, Bern, Neuchâtel 2022

Table of Contents

Purpose of the present document

The Federal Statistical Office (FSO), the National Agency for Cancer Registration (NACR), and the Childhood Cancer Registry (ChCR) play complementary roles for cancer monitoring and reporting in Switzerland. The FSO is responsible for the annual publication of routine monitoring data, as well as the report "Cancer in Switzerland" every 5 years. The NACR and the ChCR are - in addition to contributing to the publications of the FSO - responsible for the publication of cancer data within a more general framework of health reporting. They publish relevant statistical results for the research community, the government, and the general public, and make unpublished results available to third parties upon request.

To ensure consistent and reliable results, create and preserve trust in official cancer statistics, and facilitate their correct interpretation, the FSO, NACR, and ChCR formed an agreement on the statistical methods used for calculating the main epidemiological endpoints. The present document provides a detailed description of these methods. As statistical methods for cancer reporting continue to evolve, the present document is intended to be dynamic in nature and shall be updated as needed.

Summary

In Switzerland, registration for adults, children, and adolescents diagnosed with cancer became mandatory on January 1, 2020 when the Cancer Registration Act (CRA) and the corresponding Cancer Registration Ordinance (CRO) went into effect. The new legislation standardized national cancer reporting in Switzerland.

New cases of cancer from all cantons are reported to a cantonal cancer registry (CCR). There are thirteen CCRs, located in either German-speaking parts of Switzerland or French/Italian-speaking parts, which is determined based on the language spoken by the majority of the population within a canton. CCRs submit pseudonymized individual cancer diagnoses to the National Cancer Dataset managed by the National Agency for Cancer Registration (NACR). Cases of cancer diagnosed among children and adolescents are reported to the Childhood Cancer Registry (ChCR).

The statistical analyses of cancer registry data by NACR, ChCR, and the FSO are described in this report.

Topics in this report include

- describing sources of data, such as determining cancer cases among children, adolescents, and adults, and passive and active vital status follow up
- reporting inclusion and exclusion criteria, point estimates and confidence intervals, and possible sources of error for
  - incidence and mortality
  - survival, including estimation of absolute survival proportion and relative survival
  - prevalence, including estimation and projection of limited-duration prevalence.

Abbreviations

Abbreviation	Term or Phrase
CRA	Cancer Registration Act
CRO	Cancer Registration Ordinance
CCR	Cantonal Cancer Registry
CCO	Central Compensation Office
ChCR	Childhood Cancer Registry
CI	Confidence Interval
DCO	Death Certificate Only
DASIR	Directly Age-Standardized Incidence Rates
DASMR	Directly Age-Standardized Mortality Rates
ENCR	European Network of Cancer Registries
ESPOP	Statistik des jährlichen Bevölkerungsstandes (1981—2010)
FSO	Federal Statistics Office
IR	Incidence Rate
IARC	International Agency for Research on Cancer
IACR	International Association of Cancer Registries
ICSS	International Cancer Survival Standards
ICC3	International Classification of Childhood Cancer
ICD	International Classification of Diseases and Related Health Problems
LDP	Limited-Duration Prevalence
LDPC	Limited-Duration Prevalence Count
LDPP	Limited-Duration Prevalance Proportion
MR	Mortality Rate
NACR	National Agency for Cancer Registration
OASI	Old-Age and Survivors' Insurance
SCCR	Swiss Childhood Cancer Registry
STATPOP	Statistik der Bevölkerung und der Haushalte
WHO	World Health Organization
YPLL	Years of Potential Life Lost

Definitions

Term	Definitions
Absolute or observed survival	The proportion of patients alive at a specified time after cancer diagnosis.
Actuarial assumption	Subtracting half of patients censored during an interval from the number still alive at the interval start and assuming censoring occurs uniformly throughout the interval. Individuals with a censored survival time are at risk—on average—for half of the interval.
Age-standardized survival	The weighted average of age-specific survival.
Cancer prevalence	The number of people alive at a given reference date who were previously diagnosed with cancer.
Cancer prevalence proportion	The number of prevalent people divided by the population at risk at the index date (usually December 31st of given year), usually expressed as fraction of 100'000 or as percentage.
Case ascertainment	The extent all diagnosed neoplasms in the resident population are included in a registry database.
Cause-specific survival	An approach that removes some heterogeneity at patient survival by only considering death from other causes and diseases in question as right-censored events.
Cohort method	Identifies a cohort of patients whose vital status was followed up for at least as long as the period of interest.
Complete method	Similar to the cohort approach, it includes all patients with diagnosis dates within the reporting period.
Complete or total prevalence	Estimates the number of prevalent persons at the index date diagnosed with a disease at any time in their life (irrespective of time since diagnosis).
Directly observed prevalent persons	Eligible patients diagnosed during the limited-duration period known to be alive at the index date.
Flexible Parametric Models	Modified Weibull models where the time after diagnosis is incorporated as a restricted cubic spline function.
Hazard function	The instantaneous death rate of patients at a given time after cancer diagnosis.
Incidence rate	A measure of disease occurrence in a population per time unit.
Limited-duration prevalence	Represents the number of prevalent persons at the index date who were diagnosed with cancer within a given time period preceding the index date.
Mortality rate	A measure of mortality occurrence in a population per time unit.
Period method	Estimates survival rates from patients selected based on their time of death or last known vital status, which should fall in the reporting period.
Reference or index date	Date for which a statistic (e.g., prevalence) is reported—usually December 31st of given year.
Relative survival	Equal to the ratio of survival among patients with a disease and expected survival among a comparable group of people who do not have the disease; does not rely on causes of death.
Scaled event counts	Counts multiplied by a factor.
Survival function	The probability that of survival as a function of time after diagnosis.
Survival monitoring schemes	Alternative schemes for selecting participants to estimate survival.
Survival time	The time interval between diagnosis until death or until the last known date when the patient was alive.

Notation

Variable or index	Description
i	Language region: German, French/Italian*
k	Age (single years or groups of years)
g	Sex: male, female
j	Calendar year (e.g., 2011)
q	Calendar period (e.g., 2011—2015)
N_{ikg}^{j} (N_{ikg}^{q})	Mid-year Swiss population during calendar year j (sum of mid-year populations over all years in period q) for language region i, age k, and sex g (combination $i \times k \times g$)
$\widetilde{N}_{ikg}^{j}$ $(\widetilde{N}_{ikg}^{q})$	Swiss population covered by CCR during calendar year j (sum of mid-year populations over all years in period q) for the combination $i \times k \times g$
E_{ikg}^{j} (E_{ikg}^{q})	End-year Swiss population during calendar year j (sum of mid-year populations over all years in period q) for the combination $i \times k \times g$
$N_{k}^{\star}$	European standard population for the age group k
n_{ikg}^{j} (n_{ikg}^{q})	Observed number of incident cases during calendar year j (or in period q) for the combination $i \times k \times g$
d_{ikg}^{j} (d_{ikg}^{q})	Observed number of cancer deaths during calendar year j (or in period q) for the combination $i \times k \times g$
$\hat{n}_{ikg}^{j}$ $(\hat{n}_{ikg}^{q})$	Estimated number of incident cases during calendar year j (or in period q) for the combination $i \times k \times g$
w_{ikg}^{j} (w_{ikg}^{q})	Extrapolation weights during calendar year j (or in period q) for the combination $i \times k \times g$
$\hat{p}_{ikg}^{j}$ $(\hat{p}_{ikg}^{q})$	Estimated raw incidence rate during calendar year j (or in period q) for the combination $i \times k \times g$
$\hat{m}_{ikg}^{j}$ $(\hat{m}_{ikg}^{q})$	Estimated raw mortality rate in calendar year j (or in period q) for the combination $i \times k \times g$
$\hat{p}_{ikg}^{j\star}$ $(\hat{p}_{ikg}^{q\star})$	Estimated age-standardized incidence rate during calendar year j (or in period q) for the combination $i \times k \times g$
$\hat{m}_{ikg}^{j\star}$ $(\hat{m}_{ikg}^{q\star})$	Estimated age-standardized mortality rate during calendar year j (or in period q) for the combination $i \times k \times g$
$\chi_{f,l}^{2}$	lth quantile of the Chi-Square distribution with f degrees of freedom
$(\cdots)_{L}$	Lower confidence interval (CI)
$(\cdots)_{U}$	Upper confidence interval (CI)
T	Random variable for survival time
t	Time in years
Δt	Time difference [e.g., time from last available follow-up to the index date $(31.12.j)$]
$Pr(\cdots)$	Probability
$S(t)$	Survival probability at time t
$h(t)$	Hazard rate at time t
$R(t)$	Relative survival at time t
$S^{\star}(t)$	Expected survival proportion at time t
x	Time interval in years for limited-duration prevalence
$n_{prev,ikg}^{31.12.j}$	Observed number of prevalent persons at the index date $(31.12.j)$ for the combination $i \times k \times g$
$\hat{n}_{prev,ikg}^{31.12.j}$	Expected number of prevalent persons for the combination $i \times k \times g$ among those lost to follow-up or with missing active follow-up at the index date $(31.12.j)$
$\left(\widehat{LDPC}_{ikg}^{x}\right)_{31.12.j}$	Estimated x-year limited-duration prevalence counts for the combination $i \times k \times g$ at the index date $(31.12.j)$
$\left(\widehat{LDPP}_{ikg}^{x}\right)_{31.12.j}$	Estimated x-year limited-duration prevalence proportions for the combination $i \times k \times g$ at the index date $31.12.j$

Abbreviations: CCR, Cantonal Cancer Registry. CI, confidence interval
*Cantons included: Geneva (GE), Vaud (VD), Valais (VS), Fribourg (FR), Neuchâtel (NE), Jura (JU), Ticino (TI).

Tables and Figures

Table or Figure	Caption
Table 1.	European standard population (1976) by age group.
Figure 1.	Case selection of patients diagnosed 2003—2007 (cohort approach), and selected based on follow-up dates 2013—2017 (period approach). The numbers in the cells indicate the minimum number of complete years of follow-up available for patients diagnosed between 2003—2017 (vertical axis) and who survived until the end of a given year up to 2017 (horizontal axis).

1. Data sources

1.1 Cancer cases

1.1.1 Adult cancer registration

Registration for adults, children, and adolescents diagnosed with cancer became mandatory in Switzerland on January 1, 2020 when the Cancer Registration Act (CRA) went into effect. Although some cantons began registering patients at all ages diagnosed with cancer in the 1970s–1980s, other cantons began much later (see https://www.bfs.admin.ch/bfs/en/home/statistics/health/surveys/ke.assetdetail.21184747.html). Yet despite noncompulsory cancer registration, levels of completeness of case ascertainment proved acceptable.[1]

New cancer cases from all cantons are reported to a cantonal cancer registry (CCR). For the purpose of national cancer monitoring and reporting, CCRs submit pseudonymized individual cancer diagnoses for the National Cancer Dataset managed by the National Agency for Cancer Registration (NACR). There are thirteen CCRs, located in either German-speaking parts of Switzerland or French/Italian-speaking parts, which is determined based on the language spoken by the majority of the population within a canton. Cantons register adults diagnosed with cancer who have completed their 20th year of life.[a]

1.1.2 Child and adolescent cancer registration

Before the CRA, the Swiss Childhood Cancer Registry (SCCR)[2] recorded childhood cancer diagnoses starting in 1976, yet high coverage was only reached toward the end of the 1980s. Since 1995, the SCCR is estimated to include over 95% of all childhood cancers diagnosed in Switzerland.[3] The SCCR only obtained nationwide coverage for ages 0–14. It collected data on cancer diagnoses among adolescents ages 16–21 from existing CCRs, which was restricted to cantons with a CCR thus duplicating NACR data.

Now, the Childhood Cancer Registry (ChCR) records all cancers diagnosed among children and adolescents who have not yet completed their 20th year of life. Core data from SCCR about childhood cancer diagnoses are an integral part of ChCR. Since 2020, ChCR forwards abstracted information for new cases to CCRs. Thus, CCRs databases are complete for all ages.

1.1.3 Death certificate only (DCO)

Death certificate only (DCO) cases are cancers solely known because they are indicated on death certificates. After active trace-back, cancer cases first notified to a registry via a death certificate are classified as DCO if no documents can be retrieved to confirm or chronologically classify the diagnosis from other sources, such as hospitals, reporting physicians, nursing homes, and other medical personnel or facilities. Trace-back was not possible for the SCCR prior to 2018; thus, the percentage of DCO cases is high among children and adolescents.

1.2 Vital status follow-up

A registered patient's vital status is systematically ascertained using two forms of follow-up: passive vital status follow-up and active vital status follow-up.

[a] Age is defined by the number of completed life years.

1.2.1 Passive vital status follow-up

Using old-age and survivors' insurance (OASI) numbers, deaths among people with a registered cancer diagnosis are identified by data linkage with the nationwide Central Compensation Office (CCO). Registered patients with cancer diagnoses who are deceased are linked by the Federal Statistical Office (FSO) with their mortality monitoring and cause of death data using the OASI number. The causes of death are reported back to CCRs. In addition, the FSO reports death by cancer, which had not been previously registered by CCRs.

In accordance with the CRA, the FSO reports the following variables to CCRs:

- OASI number,
- date of birth,
- date of death,
- sex,
- canton of residence,
- city/municipality of residence,
- underlying cause of death,
- immediate cause of death,
- concomitant diseases, and
- contact details of certifying physician and/or institution for0 trace-back.

The quality of passive follow-up may be compromised by either incomplete CCO data, vital statistics, or cancer registry records; false negative or false positive linkage errors; or emigration of patients diagnosed with cancer outside of Switzerland.

With passive follow-up alone, “dead” is the only vital status known positively. If the only means of follow-up is passive, the status “alive” is assumed in absence of reported death. Failure to identify matching death records for deceased patients (the problem of so-called “immortals”) may overestimate survival if the only means of follow-up is passive.

1.2.2 Active vital status follow-up

CCRs actively collect available information on the vital status of registered patients from cantonal and municipal population registers in their cantons. Patients whose vital status cannot be determined at the time of follow-up, for reasons such as moving to a different canton, are assigned to the category “lost to follow-up” and the last known date alive is retained.

All CCRs perform passive and active follow-up at least annually. However, due to cantonal differences in record linkage with population registries, the completeness of active follow-up was different between registries before the Cancer Registration Ordinance (CRO) was enacted in 2020. The CRO now provides equal opportunities for all CCRs and ChCR through automated vital status updates.

1.3 Swiss population data

FSO provides mid- and end-year resident populations at completed years of life (i.e., age at last birthday) as well as predictions for future mid- and end-year populations stratified by canton, age, and sex at STAT-TAB: “Die interaktive Statistikdatenbank” (ESPOP, STATPOP). The first available year in ESPOP/STATPOP is 1981. Population numbers in earlier years differ by definition of resident population.

Since official cancer statistics start in 1980, FSO estimated the 1980s population after applying ESPOP/STATPOP definitions. In one instance [canton Basel Landschaft (BL) 1994—2012], a district’s

population data (Laufenthal) are subtracted from the cantonal population because the district was excluded from cancer registration during that time interval.

1.4 Cause of death statistics

The FSO collects and reports mortality monitoring data (rapid process) based on the civil status register database and cause of death statistics (less rapid process) once certifying physicians provide coded medical death certificates. The FSO determines the principle cause of death for all residents of Switzerland.

Until 1994, causes of death in Switzerland were coded according to the World Health Organization (WHO) 8th revision of the International Classification of Diseases and Related Health Problems (ICD).[4] Since 1995, the 10th ICD revision coding system has been used[5] to code principle causes of death.[5]

Prior to 1995, coding principle causes of death followed national regulations, which deviated from international standards. For instance, "cancer" was coded as cause of death if "tumor" was noted as either the primary (underlying) or associated cause of death and neither accident, poisoning, trauma, nor influenza was noted as the cause. For a comprehensive assessment over both periods, correction factors depending on cancer types, age groups, and sex are available,[6] yet not applied routinely for cancer mortality monitoring.

2. Incidence and Mortality

Since our methods for incidence and mortality rates are largely identical, we describe them in parallel.

2.1 Introduction

The *incidence rate* (IR)—a measure of disease occurrence in a population per time unit—is quantified as

$$IR = \frac{number\ of\ cancers\ diagnosed\ during\ the\ person\ years\ at\ risk\ observed}{number\ of\ person\ years\ at\ risk}\ . \quad (1)$$

Similarly, the cancer specific *mortality rate* (MR)—a measure of mortality occurrence in a population per time unit—is quantified as

$$MR = \frac{number\ of\ cancer\ deaths\ during\ the\ person\ years\ at\ risk\ observed}{number\ of\ person\ years\ at\ risk}\ . \quad (2)$$

In principle, individuals can be at risk and under observation for different periods of time. Adding up these periods over all individuals yields the number of person years at risk. For our purposes, IRs always relate to a specific calendar year j or calendar period q; thus, we consider the approximate number of person years estimated by the mid-year population in that year or the sum of mid-year populations over all the years of the calendar period.

Data on the number of diagnosed cases—numerator in (1)—are taken from NACR and/or ChCR. Data on deaths due to cancer—numerator in (2)—are obtained from the FSO.

2.2 Inclusion and exclusion criteria

Cancer cases

- With the exception of non-melanotic skin cancer (ICD-10 code C44), all malignant primary diagnoses are included in the calculations for incidence.
- We define primary tumors according to International Association of Cancer Registries (IACR) and International Agency for Research on Cancer (IARC) (Multiple Primary Rules) criteria.
- For childhood and adolescent cancer, we include all diagnoses of the 12 main diagnostic cancer groups according to the 3rd version of the International Classification of Childhood Cancer (ICCC3).[7]
- Depending on the publication, ChCR may also report incidence for Langerhans cell histiocytosis.
- We include all multiple primary tumors diagnosed for the same individual in the denominator of the incidence rate.

Cancer deaths

- We include all deaths caused by malignant cancer in mortality calculations based on the FSO's coding of principle causes of death (section 1.4) (C00-C97, ICD-10).

2.3 Raw rates

We stratify rates by language region i, as well as by age group k and sex g, in single calendar year or multiple calendar year intervals.

IRs are often calculated for different strata of variables known to be associated with cancer risk, such as age and sex. In Switzerland, the prevalence of putative risk factors rates, such as smoking or dietary exposures, vary between language regions.

For a given population stratum $i \times k \times g$, we represent the number of person years at risk in year j by the mid-year Swiss population N_{ikg}^{j}, which we calculate as the mean of subsequent end-year (31.12) populations,

$$N_{ikg}^{j} = \frac{E_{ikg}^{j-1} + E_{ikg}^{j}}{2} \,. \tag{3}$$

The person years at risk during the period q is accordingly the sum of the mid-year populations for the years included in q, i.e., $N_{ikg}^{q} = \sum_{\forall j \in q} N_{ikg}^{j}$ (by slight abuse of notation, q here simultaneously represents the index values of the period in question and the set of years contained in that period).

The mid-year population we use is a linear interpolation of populations at the end of the year. When we assume the population followed a linear trend throughout the year, it precisely represents person years at risk.

We denote the observed annual number of cancer diagnoses for a given population stratum $i \times k \times g$ with n_{ikg}^{j}. Consequently, $n_{ikg}^{q} = \sum_{\forall j \in q} n_{ikg}^{j}$.

For ages 0—15, we directly obtain the number of cancer diagnoses from ChCR. However, cancer incidence among adults and before 2020 for ages 16—19 require special attention.

Because data for these years of age rely on existing CCRs, we reweight the observed number of cases, n_{ikg}^{q}, to obtain the estimated number of cases $\hat{n}_{ikg}^{q}$ with weights equal to $w_{ikg}^{q} = N_{ikg}^{q} / \widetilde{N}_{ikg}^{q}$, where $\widetilde{N}_{ikg}^{q} = \sum_{\forall j \in q} \widetilde{N}_{ikg}^{j}$ and with $\widetilde{N}_{ikg}^{j}$ being the population covered by CCRs in a given year j.

We present IR formulas with these weights, though they are required only for the years of age $k \geq 16$. For the younger age groups ($k < 16$), we set weights to 1.

We estimate the IR per 100′000 person years as

$$\hat{p}_{ikg}^{q} = \frac{100'000}{\widetilde{N}_{ikg}^{q}} \, n_{ikg}^{q} = \frac{100'000}{N_{ikg}^{q}} \, w_{ikg}^{q} n_{ikg}^{q} = \frac{100'000}{N_{ikg}^{q}} \hat{n}_{ikg}^{q} \tag{4}$$

We describe the so-called raw, crude, or non-standardized, national IRs in equation (5). Obviously, the second term is preferred for computation; however, we explicitly show the formula in terms of reweighted counts since these are relevant for estimating national IRs.

We estimate the national IR for age and sex stratum $k \times g$ as

$$\hat{p}_{kg}^{q} = \frac{100'000}{N_{kg}^{q}} \sum_{i} w_{ikg}^{q} n_{ikg}^{q} = \frac{100'000}{N_{kg}^{q}} \sum_{i} \hat{n}_{ikg}^{q} \tag{5}$$

where $N_{kg}^{q} = \sum_{i} N_{ikg}^{q}$ represents the person years at risk for stratum $k \times g$ at the national level.

We similarly obtain raw MRs as

$$\widehat{m}_{ikg}^{q} = \frac{100'000}{N_{ikg}^{q}}\, d_{ikg}^{q} \qquad (6)$$

where d_{ikg}^{q} is the observed number of deaths due to cancer during the period q.

Since the data used for the computation of (6) cover the whole of Switzerland, extrapolating the observed number of cancer deaths is not required here.

Accordingly, we obtain national mortality rate estimates as

$$\widehat{m}_{kg}^{q} = \frac{100'000}{N_{kg}^{q}}\, d_{kg}^{q}\ . \qquad (7)$$

Confidence intervals (CIs)

We use the method shown here when computed rates are simply *scaled event counts*—counts multiplied by a factor—as in (6). The method also applies to rates calculated according to equation (5) for $k < 16$ because weights w_{ikg}^{q} for this age group are always 1.

However, when the estimated rate is a weighted sum of counts with differing weights, as in equation (5) for $k \geq 16$, the method for computing CIs corresponds with direct standardized rates (section 2.4).

We assume the number of events (i.e., incident cases n or deaths due to cancer d) is a Poisson distributed random variable, which underlies our CI estimates. We use the following CIs for single-scaled Poisson parameters, while we use CIs for directly standardized rates (section 2.4) for linear combinations of Poisson parameters.

We first compute the CIs for the expected number of events λ based on an observed number of events n using a relationship between the Poisson and the Chi-squared distribution:

$$\tfrac{1}{2}\chi^{2}_{2n,\,a/2} \leq \lambda \leq \tfrac{1}{2}\chi^{2}_{2n+2,\,1-a/2} \qquad (8)$$

where $\chi^{2}_{f,l}$ is the lth quantile of the Chi-Square distribution with f degrees of freedom.[8]

We obtain the lower and upper limits of the 95% CI for deaths in the population stratum $i \times k \times g$ during period q as $\left(d_{ikg}^{q}\right)_{L} = \frac{1}{2}\chi^{2}_{2d_{ikg}^{q},0.025}$ and $\left(d_{ikg}^{q}\right)_{U} = \frac{1}{2}\chi^{2}_{2d_{ikg}^{q},0.975}$. We determine the lower and upper limits for the 95% CI of raw MRs by substituting $\left(d_{ikg}^{q}\right)_{L}$ and $\left(d_{ikg}^{q}\right)_{U}$ for d_{ikg}^{q} in equation (6), respectively.

When computing raw IRs according to equation (5) for age $k < 16$ years, for which $w_{ikg}^{q} = 1$, we analogously obtain CIs by substituting $\left(n_{ikg}^{q}\right)_{L}$ and $\left(n_{ikg}^{q}\right)_{U}$ for n_{ikg}^{q} (note n_{ikg}^{q} observed counts for the entire language region i, not estimated counts). We do not treat mid-year populations as a random variable for these computations.

2.4 Directly standardized rates

Cancer incidence varies greatly by age. When comparing data from Switzerland with other countries, it is important to correct for differences in population age structures. We report directly age-standardized

incidence rates (DASIRs) for each sex (g) over a period q, using the European standard population (Table 1) as reference, i.e.,

$$\hat{p}_g^{q\star} = \frac{100'000}{\sum_k N_k^\star} \sum_k \frac{N_k^\star}{100'000} \hat{p}_{kg}^q = \sum_k \frac{N_k^\star}{N_{kg}^q} \sum_i w_{ikg}^q n_{ikg}^q = \sum_k \frac{N_k^\star}{N_{kg}^q} \sum_i \hat{n}_{ikg}^q = \sum_k \frac{N_k^\star}{N_{kg}^q} \hat{n}_{kg}^q \quad (9)$$

where $N_k^\star$ is the European standard population (Table 1, note $\sum_k N_k^\star = 100'000$).

We calculate the directly age-standardized mortality rates (DASMRs) similarly; however, since mortality data are available for the entire country, we calculate DASMRs without the intermediate step of extrapolating to the whole of Switzerland:

$$\hat{m}_g^{q\star} = \frac{100'000}{\sum_k N_k^\star} \sum_k \frac{N_k^\star}{100'000} \hat{m}_{kg}^q = \sum_k \frac{N_k^\star}{N_{kg}^q} \sum_i d_{ikg}^q = \sum_k \frac{N_k^\star}{N_{kg}^q} d_{kg}^q \ . \quad (10)$$

Table 1. European standard population (1976) by age group.[9]

Age group	European standard population
0—4	8′000
5—9	7′000
10—14	7′000
15—19	7′000
20—24	7′000
25—29	7′000
30—34	7′000
35—39	7′000
40—44	7′000
45—49	7′000
50—54	7′000
55—59	6′000
60—64	5′000
65—69	4′000
70—74	3′000
75—79	2′000
80—84	1′000
85+	1′000
Total	100′000

Confidence intervals (CIs)

Following recommendations from Ng et al.'s[10] simulation study results, we compute CIs for directly age-standardized rates according to Fay and Feuer's[11] method (method G1 in Ng et al.). As an example, we present calculations for directly age-standardized sex-specific incidence rate, $\hat{p}_g^{q*}$ with variance estimated as

$$\widehat{Var}\left(\hat{p}_g^{q*}\right) = \widehat{Var}\left(\sum_k \frac{N_k^\star}{N_{kg}^q}\,\hat{n}_{kg}^q\right) = \sum_k \left(\frac{N_k^\star}{N_{kg}^q}\right)^2 \widehat{Var}\left(\hat{n}_{kg}^q\right) = \sum_k \left(\frac{N_k^\star}{N_{kg}^q}\right)^2 \widehat{Var}\left(\sum_i w_{ikg}^q n_{ikg}^q\right) = \sum_k \left(\frac{N_k^\star}{N_{kg}^q}\right)^2 \sum_i \left(w_{ikg}^q\right)^2 n_{ikg}^q = \sum_k \sum_i \left(w_{ikg}^{q\star}\right)^2 n_{ikg}^q \quad (11)$$

where $w_{ikg}^{q\star} = \frac{N_k^\star}{N_{kg}^q} w_{ikg}^q$.

The lower bound of the 95% CI for $\hat{p}_g^{q\star}$ is computed as

$$\left(\hat{p}_g^{q\star}\right)_L = \frac{\widehat{Var}\left(\hat{p}_g^{q\star}\right)}{2\hat{p}_g^{q\star}} \chi^2_{\left(2\left(\hat{p}_g^{q\star}\right)^2 / \widehat{Var}\left(\hat{p}_g^{q\star}\right)\right),0.025} \quad (12)$$

and the upper bound of the 95% CI as

$$\left(\hat{p}_g^{q\star}\right)_U = \frac{\widehat{Var}\left(\hat{p}_g^{q\star}\right)+\tilde{w}_g^2}{2\left(\hat{p}_g^{q\star}+\tilde{w}_g\right)} \chi^2_{\left(2\left(\hat{p}_g^{q\star}+\tilde{w}_g\right)^2 / \left(\widehat{Var}\left(\hat{p}_g^{q\star}\right)+\tilde{w}_g^2\right)\right),0.975} \quad (13)$$

where $\tilde{w}_g = max_{k,i} w_{ikg}^{q\star}$.

Analogous calculations yield 95% CIs for estimated crude national IRs, which also represent weighted sums of observed counts.

2.5 Possible sources of error

Incidence

The quality of incidence statistics depends on completeness of *case ascertainment*—the extent all diagnosed neoplasms in the resident population are included in the registry database. Underestimated rates may occur for diagnoses in outpatient settings (e.g., nursing homes or private practices), which potentially evaded capture and registration when compared with hospital-based diagnoses. Another possible cause of error are long delays between diagnosis date and date of reporting the case to the registry (>1 year). Overestimated rates may occur if secondary diagnoses (metastases), recurrences, or transformations are incorrectly categorized as primary diagnoses.

Mortality

Underlying causes of death serve as classification characteristics for estimating MRs. Due to deviations from international standards determining underlying causes of death before 1995 (section 1.4), rates prior to 1995 are slightly higher because the international standard was not used.[12,13] Deviations depend on sex, age, and cancer type, with a median of 97% (interquartile range 93—100%).

3. Survival

3.1 Introduction

Measuring the survival experiences of patients diagnosed with cancer is helpful for assessing the overall effectiveness of health care systems. Of particular interest are changes in survival over time and comparisons of survival between sub-populations within one country or between populations of different countries.

Survival time is the time interval between diagnosis until death or until the last known date when the patient was alive. It is expected that successes in the fight against cancer lead to prolonged survival times due to postponing—or even preventing—death from cancer.

However, longer survival times cannot always be equated with improved survival. Survival time may also be prolonged by earlier diagnosis. In many cases, earlier diagnosis effectively postpones or prevents death due to cancer, yet it is difficult to correctly attribute surplus survival time to either earlier diagnosis or postponed death. Therefore, survival estimates should always be interpreted in concert with incidence and mortality data and contextual health system information.

Absolute or *observed survival* is the proportion of patients alive at a specified time t after diagnosis with cancer. It serves as an estimate of the survival function $S(t)$. *Survival function* is the probability that survival time T exceeds t [i.e., $S(t) = Pr(T > t)$]. The survival function is linked to the hazard function $h(t)$. The *hazard function* is the instantaneous death rate at time t after patients are diagnosed with cancer through

$$S(t) = e^{-\int h(t)dt}. \quad (14)$$

Giving rise to more steeply decaying survival curves early after diagnosis, the hazard is usually high shortly after diagnosis and smaller at later intervals.

Absolute or observed survival estimates are not well suited for regional or national comparisons or comparisons over time because they depend heavily on patient characteristics, such as comorbidities and other risk factors, which are often unknown at the individual-level.

Analysis of *cause-specific survival* is an approach that removes some heterogeneity at patient survival by only considering death from other causes, such as other types of cancer, and from cancers in question as right-censored events (i.e., only part of the true cancer survival time is known).

However, some deaths are only indirectly related to cancer, such as treatment toxicity, suicides, and late effects like cardiovascular disease, second malignancies, or infection. They are unlikely to be classified as deaths from cancer on death certificates, which means they may not be captured in cause-specific analysis. Furthermore, it can be difficult to define a single main cause of death among multimorbid patients.

Following the customary approach in population-based cancer reporting, NACR estimates relative survival. *Relative survival* is equal to the ratio of survival among patients with cancer and expected survival among a comparable group of people free of the cancer in question, i.e.,

$$R(t) = \frac{S(t)}{S^{\star}(t)} \quad (15)$$

with $S^{\star}(t)$ being the expected survival at time t since diagnosis; it does not rely on causes of death.

Typically, people from the general population are used as a comparison group, even though they may not all be cancer free. For example, 50% relative survival shows that patients diagnosed with cancer survived

half as well as their counterparts in the general population. Surplus mortality is assigned to the cancer in question. A relative survival of 100% shows that deaths in the patient group were as frequent as deaths in the general population. In other words, there is no excess mortality associated with cancer.

The group of people from the general population must be comparable with the group of patients diagnosed with cancer with respect to age, sex, calendar year, and region of residence (cantons or canton groups). The life tables are stratified for these matching variables. We also assume any potential risk factors not specifically controlled for are distributed similarly among patients diagnosed with cancer and people from the general population, which is arguably a strong assumption. However, our sensitivity analyses showed that even for patients diagnosed with lung cancer who often smoke and carry a higher risk for other diseases than the general population, it did not have a concerning impact.[14]

3.2 Inclusion and exclusion criteria

Cancer cases

- With the exception of non-melanotic skin cancer (ICD-10 code C44), we include all malignant primary diagnoses.
- For childhood and adolescent cancers, we include all diagnoses of the 12 main ICCC3 groups.[7]
- Depending on the publication, ChCR may also report survival for Langerhans cell histiocytosis.
- We include people with multiple primary malignant cancer diagnoses in different cancer reporting groups separately in each cancer group.

  For example, if a person has melanoma as first primary cancer and lung tumor as second primary cancer, we include the melanoma occurrence in calculations for melanoma and the lung tumor in calculations for lung cancer.
- For people with multiple primary malignant diagnoses in a single cancer group, we include only the first diagnosis.
- For survival estimates of all cancers combined, we only count the first primary malignant diagnosis of any type.

  Hence, we ignore the lung tumor of the example patient above when we calculated all cancers combined. We applied this rule to survival estimates for children and adolescents.
- We exclude DCO cases.
- We do not exclude people who might be considered cured from cancer.

3.3 Survival estimation

For national survival statistics in Switzerland, we censor all survival times to the last known follow-up date. Our approach is more conservative than applying the assumption used in some countries that all people without known dates of death are still alive (excluding lost-to-follow up cases). In contrast to cancer incidence statistics, we do not estimate survival at the national level by weighting region-specific language estimates. We prefer our simple approach because there is little evidence for differences in cancer survival by language region,[15] while differences in the distribution of cancer risk factors are well known.[16–18]

3.3.1 Estimation of absolute survival proportion

We estimate the absolute survival proportion at time t_c after diagnosis $\hat{S}(t_c)$ non-parametrically with the actuarial or lifetable approach.[19] For this purpose, we divide time up to 10 years after diagnosis into consecutive intervals of increasing length, assuming constant hazard within intervals. We index intervals with c ($c = 1, 2, \ldots, 13$) and refer to the upper bounds of the intervals with t_c. The upper bounds (or cutpoints) are set to $0.1, 0.2, 0.6, 1, 2, 3, \cdots, 10$ years after diagnosis.

We estimate the survival proportion at time t_c as the product of all interval-specific survival proportions until interval c:

$$\hat{S}(t_c) = \prod_{j:\, t_j \leq t_c} \left(1 - \frac{d_c}{r_c'}\right) \quad (16)$$

where d_c is the number of observed deaths in the interval c, and r_c' is the adjusted number at risk during the interval c.

For the adjustment, we use the *actuarial assumption*. We subtract half of the patients who are censored during the interval from the number still alive at the interval start. We assume censoring occurs uniformly throughout the interval so individuals with a censored survival time are at risk—on average—for half of the interval.

ChCR reports only absolute survival and uses the Kaplan-Meier method for estimates.

3.3.2 Estimation of relative survival

Relative survival is the ratio of absolute survival and expected survival. Relative survival adjusts for the non-cancer background mortality.

Expected survival

NACR derives expected survival from population lifetables according to the Ederer II method.[20] The basic Ederer II method assumes that matched individuals from the population are at risk until the corresponding patient diagnosed with cancer dies or is censored. Individuals from the population are matched to patients for sex, age at death, year of death, and canton at death or canton group at death.

For example, the 5-year expected survival algorithm:

(1) averages lifetable survival probabilities for the first time-interval among people matched to all patients in the cancer group and provides first interval-specific expected survival;
(2) averages lifetable survival probabilities for the second time-interval among people matched to those patients who survived the first interval and provides second interval-specific expected survival;
(3) repeats the algorithm for every time-interval until 5-year survival time is reached;
(4) multiplies all interval-specific expected survival estimates to calculate the cumulative 5-year expected survival.

Construction of lifetables

NACR processes official vital statistics (all-cause deaths) and official population statistics (ESPOP and STATPOP) into age-abridged lifetables for individual cancer registries, stratified by sex, 5-year age groups, and 1-calendar year time periods.

To obtain complete life tables by single calendar years j and single years of age k, we smooth these with the following algorithm by

STEP 1. Estimating the mortality rate m_k^j for each age k and calendar year j as the mortality in year j of the corresponding 5-year age-group encompassing k.

STEP 2. Calculating running averages for mortality rates by replacing m_k^j by the mean of m_k^{j-1}, m_k^j, and m_k^{j+1}.

> In the following steps, we apply calculations to all k and j; we omit the subscripts j for ease of notation.

STEP 3. Deriving the probability for dying at age k as $m_k' = 1 - exp(-m_k)$ assuming constant m_k within the age-intervals.

STEP 4. Deriving the number of people surviving until age k, s_k, using the lifetable approach assuming a radix of 100'000 people with age $k = 0$.

STEP 5. Applying the Elandt-Jonhson method[21] to interpolate age-abridged s_k values that are distinct for individual years of age.

> In essence, the Elandt-Johnson method provides smoothing formulae and empirical coefficients for three interpolation schemes depending on age.

STEP 6. Estimating the probability of dying at age k as $m_k' = d_k/s_k$, with d_k denoting the number of deaths at age k and s_k, the estimated (interpolated) number of people from the radix surviving to age k.

STEP 7. Obtaining estimated age-specific mortality rates $m_k = -ln(1 - m_k')$.

Relative survival

We use estimated age-specific mortality rates m_k^j from the age-abridged lifetables to calculate the expected survival $S^*(t)$ (see Ederer II method), which forms the denominator in equation (15) for the relative survival estimate.

3.3.3 Age-standardization of survival estimates

To compare the survival estimates between sub-groups of different age-structures within Switzerland with other countries—or over time—, we directly age-standardize survival estimates using cancer-specific weights suggested by International Cancer Survival Standards (ICSS).[22] *Age-standardized survival* is the weighted average of age-specific survival. The age-groups are those defined by the ICSS weights.

Alternatively, if sparseness of data in certain age groups is a problem and age-specific survival estimates are unreliable or even unavailable,[23] Brenner et al. proposed another method. In their approach, the ICSS weights are individually assigned to all patients and a weighted lifetable is constructed. Thus, there is no need to calculate age-specific survival.

ChCR does not apply age-standardization to survival estimates.

3.4 Survival monitoring schemes

Figure 1 introduces *survival monitoring schemes* or alternative schemes for selecting patients to estimate survival. For our discussion, we use 10-year survival for the 2013—2017 reporting period as the estimate of interest.

Typically, we compare survival for a reporting period with earlier periods. In these situations, we use the cohort approach to obtain survival for earlier periods and the period approach for reporting periods. In other instances, the complete approach is more appropriate because it maximizes the number of included cases.

3.4.1 Cohort method for survival monitoring

The *cohort method*[24] identifies a cohort of patients whose vital status has been followed up for at least as long as the period of interest (e.g., 10 years). Figure 1 displays all patients in the diagnosis period 2003—2007 (rows marked yellow) who have at least 10 years of follow-up by the end of 2017 (green and grey area).

This method is attractive because of its conceptual simplicity. However, the requirement of a 10-year follow-up period means resulting survival must estimate the survival experience of patients diagnosed and treated during the 10 years preceding the reporting period. These estimates thus represent historic information and disregard more recent survival experiences of patients diagnosed and treated during the reporting period.

		Calendar year of death or last known date alive														
		2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	**2013**	**2014**	**2015**	**2016**	**2017**
Calendar year of diagnosis	**2003**	0	**1**	2	3	4	**5**	6	7	8	9	**10**				
	2004		0	**1**	2	3	4	**5**	6	7	8	9	**10**			
	2005			0	**1**	2	3	4	**5**	6	7	8	9	**10**		
	2006				0	**1**	2	3	4	**5**	6	7	8	9	**10**	
	2007					0	**1**	2	3	4	**5**	6	7	8	9	**10**
	2008											**5**	6	7	8	9
	2009											4	**5**	6	7	8
	2010											3	4	**5**	6	7
	2011											2	3	4	**5**	6
	2012											**1**	2	3	4	**5**
	2013											0	**1**	2	3	4
	2014												0	**1**	2	3
	2015													0	**1**	2
	2016														0	**1**
	2017															0

Figure 1. Case selection of patients diagnosed 2003—2007 (cohort approach), and selected based on follow-up dates 2013—2017 (period approach). The numbers in the cells indicate the minimum number of complete years of follow-up available for patients diagnosed between 2003—2017 (vertical axis) and who survived until the end of a given year up to 2017 (horizontal axis).

3.4.2 Period method for survival monitoring

Period method[24] estimates survival rates from patients selected based on their time of death or last known vital status, which should fall into the reporting period. For our example, 2013—2017 (columns marked yellow in Figure 1) for patients who were diagnosed between 2003—2017 (blue and grey area).

Patients only contribute person years at risk to the analysis during the reporting period. Thus, information on surviving the first years after diagnosis is estimated from recently diagnosed patients. For example, we estimate 1-year survival from patients diagnosed between 2012—2016.

Conversely, the probabilities for longer periods of time are increasingly informed by patient groups diagnosed in earlier periods who survived at least to the beginning of the reporting period. For instance, the probability of dying in the 10th year conditional on surviving to the first 9 years is estimated from patients diagnosed between 2003—2007.

Thus, period survival estimates are always based on the most recent survival information available. Figure 1's grey area indicates combinations of years of diagnosis and years after diagnosis when the conditional probabilities of dying in both the period and cohort analysis are informed by the same data.

Based on the assumption that survival rates remained stable between 2013—2017 when all patients have at least 10 years of follow-up, a 10-year period estimate can be interpreted as a prediction of 10-year cohort survival estimates for patients diagnosed between 2013—2017. The period survival estimate is conceptually more difficult to understand than cohort survival, yet it is more up-to-date.

3.4.3 Complete method for survival monitoring

Complete method is another monitoring method. Complete method is very similar to the cohort approach; however, it includes patients with diagnosis dates after 2007 (Figure 1), and it uses their survival experience even if they cannot be followed up for at least 10 years.

3.5 Possible sources of error

Survival time analysis is demanding. It requires complete registration of diagnoses and continuous effort to keep vital statuses of registered patients diagnosed with cancer updated until the case is closed by death. Complete and recent vital status follow-up information is a prerequisite for valid survival statistics.

Another requirement for valid survival statistics are low proportions of registered DCO cases. Since DCO case diagnosis dates are unknown, diagnosis dates are set to dates of death. By excluding DCO cases that represent patients with shorter survival times, valid survival statistic survival times are potentially overestimated.[25,26]

4. Prevalence

4.1 Introduction

Cancer prevalence is the number of people alive at a given reference date who were previously diagnosed with cancer. The *reference date* is usually the last day of a given year (j) and referred to as *index date* (denoted by $31.12.j$).

Cancer prevalence proportion is the number of prevalent people divided by the population at risk at the index date, usually expressed as fraction of 100′000 or as percentage.

Limited-duration prevalence (LDP) represents the number of prevalent persons at the index date who were diagnosed with cancer within the past x years (e.g., $x = 1, 2, 5,$ or 10 years). For example, the 10-year LDP on 31.12.2005 ($j = 2005$) comprises all people alive at that date diagnosed between 01.01.1996 and 31.12.2005. We describe LDP estimates later (section 4.3).

Complete or total prevalence estimates the number of prevalent persons at the index date diagnosed with cancer at any time in their life (irrespective of time since diagnosis). Complete prevalence typically requires statistical modelling to estimate the number of survivors diagnosed before the starting date of cancer registration. Complete prevalence estimates are not part of this report.

We consider measures of LDP more informative than complete prevalence. The 1- and 2-year LDPs estimate the number of patients with the strongest demands on public health services, including efforts for cancer staging, primary treatment, and supportive care. The 2- to 5-year LDP includes patients who are still likely to be under close clinical assessment for recurrence. Yet the 5- to 10-year LDP, or higher, includes many people who no longer receive cancer-related treatments; however, they might use services for late or long-term effects of their cancer diagnosis and treatment.[37]

4.2 Inclusion and exclusion criteria

Cancer cases

- We only include CCRs with at least 10 years of data before the index date—a criterion not applied by ChCR.
- With the exception of non-melanotic skin cancer (ICD-10 code C44), we include all malignant primary diagnoses.
- For childhood and adolescent cancers, we include all diagnoses of the 12 main ICCC3 groups.[7]
- Since a patient with multiple primary cancers in different cancer groups contributes to more than one cancer-specific prevalence measure, we include people with multiple primary malignant cancer diagnoses in different cancer reporting groups separately in each cancer group.
- For people with multiple primary malignant diagnoses in a single cancer group, we include only the first diagnosis.
- A designation as “first cancer” takes into account all cancers diagnosed during the entire time period of registration, not just during the LDP period.
- We consider prevalence measures for the group of all cancer types combined (excluding diagnoses for non-melanotic skin cancer) based only on the first primary malignant diagnosis irrespective of cancer group.

For example, if melanoma is the first primary cancer and a lung tumor is second primary cancer, we include the melanoma occurrence in calculations for melanoma and the lung tumor in the calculations for lung cancer. However, we ignore the lung tumor in calculations for all cancers combined.

- We include all cases irrespective of whether their vital status is known at the index date. Those with unknown vital status are patients who either cannot be traced any more (i.e., lost to follow-up) or who were never actively followed-up (i.e., missing active follow-up).
- We exclude DCO cases.
- Cases are ineligible if diagnoses were made at death, so we exclude diagnoses made at death.

Swiss population data

Since prevalence is calculated at end-year index dates (31.12.j), end-year populations are used as denominators in the calculations of prevalence proportions. For predictions of future incidence rates, which are needed for projecting prevalence (section 4.4), we use mid-year populations. We base predictions on the "reference" (same as "middle," or scenario A) for future growth as reported by the FSO (see Bevölkerungs-Entwicklung-Schweiz).

4.3 Estimation of limited-duration prevalence (LDP)

4.3.1 Counting prevalent persons

Number of directly observed prevalent persons, $n_{prev}^{31.12.j}$

Directly observed prevalent persons are eligible patients (section 4.2) diagnosed during the LDP known to be alive at the index date. We only count a patient with several primary cancer diagnoses within a cancer group as a prevalent case if the first of these is made during LDP.

Number of expected prevalent persons, $\hat{n}_{prev}^{31.12.j}$

For patients with unknown vital status at the index date, we impute their vital status in the LDP calculation. We estimate the probability of each patient with unknown vital status at the index date as still being alive at the index date, conditional on the length of observed survival (i.e., the conditional probability),

$$\widehat{Pr}(T > t + \Delta t \mid T > t) = \frac{\widehat{Pr}(T>t+\Delta t)}{\widehat{Pr}(T>t)} = \frac{\hat{S}(t+\Delta t)}{\hat{S}(t)} \quad (17)$$

with Δt being the time from last available follow-up t to the index date.[27]

To estimate the conditional survival probabilities in (17) continuously (and not only at certain cut points as in non-parametric survival analysis), we fit flexible parametric survival models to the observed survival data.[28-30]

In brief, *flexible parametric models* are modified Weibull models where the time after diagnosis is incorporated as a restricted cubic spline function. Complete type survival analysis (section 3.4.3) is performed by selecting a cohort of patients with diagnosis dates sufficiently long before the index date. For instance, within 10 years before the index date to estimate a maximum 10-year LDP. For simplicity, proportional hazards are assumed (i.e., the relative effects of covariates are constant in time after diagnosis). Non-parametric Kaplan-Meier estimation is used to visually validate the survival functions derived with flexible parametric modelling.

Adjusting for the variables age at diagnosis (k), sex (g), cancer registry, and cancer type, NACR fits flexible parametric survival models of observed survival. ChCR fits flexible parametric survival models separately for the two most frequent cancer types among children and adolescents, which are leukemias and CNS tumors.

If non-parametric Kaplan-Meier estimation of the observed survival differed by age and period of diagnosis within these two cancer types, we adjust the flexible parametric models for the age group at diagnosis (k = <1, 1—4, 5—9, 10—14, 15—19 years) and 5-year periods of diagnosis. For other cancers, we group cancer types with similar survival curves as assessed by non-parametric Kaplan-Meier estimation. We fit a single flexible parametric survival model adjusted for cancer type, age group at diagnosis (k = <1, 1—4, 5—9, 10—14, 15—19 years), and 5-year period of diagnosis.

By postestimation, the model derived predicted values for $\hat{S}(t + \Delta t)$ and $\hat{S}(t)$ for each patient, specific for age at diagnosis (k), sex (g), cancer registry, and cancer type; we apply equation (17). Then, if the estimated conditional survival probability is larger than a randomly generated uniform number between 0 and 1, we flag each patient who was either lost to follow-up or did not have active follow-up as prevalent.

The number of patients we flag as survivors are the expected prevalent persons ($\hat{n}_{prev}^{31.12.j}$), which we add to those directly observed alive at the index date ($n_{prev}^{31.12.j}$). ChCR calculates the expected number of prevalent persons $\hat{n}_{prev}^{31.12.j}$ as the sum of the estimated conditional survival probabilities over all individuals lost to follow-up or without active follow-up before the index date.[30]

4.3.2 Limited-duration prevalence counts (LDPC)

We calculate the x-year limited-duration prevalence count (LDPC) at the index date for each language region (i), attained age group (k), and sex (g) by summing the observed $n_{prev}^{31.12.j}$and expected $\hat{n}_{prev}^{31.12.j}$ number of prevalent cases and extrapolating them to the entire population in the language region using the ratio of N_{ikg}^{q} to $\widetilde{N}_{ikg}^{q}$as weights

$$\left(\widehat{LDPC}_{ikg}^{x}\right)_{31.12.j} = \frac{N_{ikg}^{q}}{\widetilde{N}_{ikg}^{q}} \left(n_{prev,ikg}^{31.12.j} + \hat{n}_{prev,ikg}^{31.12.j}\right), \quad (18)$$

where q is the x-year calendar period starting on 01.01. of the year $j - x + 1$ and ending at the index date, $31.12.j$.

ChCR does not flag patients as survivors or non-survivors as described above (section 4.3.1; last paragraph).

We estimate LDPC for whole Switzerland as the sum of extrapolated language-region-specific LDPCs

$$\left(\widehat{LDPC}_{kg}^{x}\right)_{31.12.j} = \sum_i \left(\widehat{LDPC}_{ikg}^{x}\right)_{31.12.j} = \sum_i \frac{N_{ikg}^{q}}{\widetilde{N}_{ikg}^{q}} \left(n_{prev,ikg}^{31.12.j} + \hat{n}_{prev,ikg}^{31.12.j}\right). \quad (19)$$

For cancers diagnosed among adults, NACR reports $\left(\widehat{LDPC}_{kg}^{x}\right)_{31.12.j}$ for each cancer type separately.

Variances and confidence intervals

Clegg et al. and Gigli et al. approximated the sum of prevalent persons, directly observed prevalent cases ($n_{prev}^{31.12.j}$), and expected prevalent cases ($\hat{n}_{prev}^{31.12.j}$) as Poisson distributed for rare diseases such as cancer.[27,31] Simulation showed that confidence intervals (CIs) based on this assumption have nominal coverage for high and medium levels of prevalence and supranominal (i.e., conservative) coverage at low prevalence.

Under this assumption, we estimate the variance of x-year LDPC at the index date for each language region (i), attained age-group (k), and sex (g) as

$$Var\left(\left(\widehat{LDPC}_{ikg}^{x}\right)_{31.12.j}\right) = Var\left(\frac{N_{ikg}^{q}}{\tilde{N}_{ikg}^{q}}\left(n_{prev,ikg}^{31.12.j} + \hat{n}_{prev,ikg}^{31.12.j}\right)\right) = = \left(n_{prev,ikg}^{31.12.j} + \hat{n}_{prev,ikg}^{31.12.j}\right)\left(\frac{N_{ikg}^{q}}{\tilde{N}_{ikg}^{q}}\right)^{2}$$

(20)

For whole Switzerland, the estimated variance is

$$Var\left(\left(\widehat{LDPC}_{kg}^{x}\right)_{31.12.j}\right) = \sum_i Var\left(\left(\widehat{LDPC}_{ikg}^{x}\right)_{31.12.j}\right).$$

(21)

We calculate 95% CIs based on the normal approximation and asymptotic variance of log-transformed upscaled counts in language regions and whole Switzerland. Using the delta method, the asymptotic variance of a prevalence count measure of interest (Y) is $Var[ln(Y)] = Var(Y)/Y^2$. A 95% CI is given by $Y \cdot e^{\pm z_{0.975}\sqrt{Var[ln(Y)]}}$, where z_q is the q-quantile of the standard normal distribution.

4.3.3 Limited-duration prevalence proportions (LDPP)

We calculate x-year limited-duration prevalence proportion (LDPP) per 100′000 people at the index date $31.12.j$ for each language region (i), attained 10-year age-group (k), and sex (g) by the sum of observed and expected number of prevalent cases divided by the population covered by CCRs ($\tilde{N}_{ikg}^{q}$)

$$\left(\widehat{LDPP}_{ikg}^{x}\right)_{31.12.j} = \frac{100'000}{\tilde{N}_{ikg}^{q}}\left(n_{prev,ikg}^{31.12.j} + \hat{n}_{prev,ikg}^{31.12.j}\right)$$

(22)

where q is the x-year calendar period starting on 01.01. of the year $j - x + 1$ and ending at the index date $31.12.j$.

We base the LDPP on CCR-covered parts of a language region and assume these representative for the whole language region. We estimate the x-year LDPP per 100′000 persons at the index date for whole Switzerland as the weighted average of language-region-specific LDPPs; we also use population ratios of the corresponding end-year populations as weights

$$\left(\widehat{LDPP}_{kg}^{x}\right)_{31.12.j} = \frac{1}{E_{kg}^{q}}\sum_i E_{ikg}^{q}\left(\widehat{LDPP}_{ikg}^{x}\right)_{31.12.j}$$

(23)

where q is the x-year calendar period starting on 01.01. of the year $j - x + 1$ and ending at the index date $31.12.j$.

Variances and confidence intervals

We estimate variance of $\left(\widehat{LDPP}_{ikg}^{x}\right)_{31.12.j}$ by

$$Var\left(\left(\widehat{LDPP}_{ikg}^{x}\right)_{31.12.j}\right) = \frac{100'000^2}{\left(\tilde{N}_{ikg}^{q}\right)^2}\left(n_{prev,ikg}^{31.12.j} + \hat{n}_{prev,ikg}^{31.12.j}\right).$$

(24)

Following Clegg et al.,[31] we derive CIs based on the relationship between Poisson and Chi-Square and derive lower $\left(\left(\widehat{LDPP}_{ikg}^{x}\right)_{31.12.j}\right)_L$ and upper $\left(\left(\widehat{LDPP}_{ikg}^{x}\right)_{31.12.j}\right)_U$ limits at confidence level $1 - a$ as[8]

$$\left(\left(\widehat{LDPP}_{ikg}^{x}\right)_{31.12.j}\right)_{L} = \frac{100'000}{2\widetilde{N}_{ikg}^{q}}\chi^{2}_{2\cdot\left(n_{prev}^{31.12.j}+\hat{n}_{prev}^{31.12.j}\right),\,a/2}$$

$$\left(\left(\widehat{LDPP}_{ikg}^{x}\right)_{31.12.j}\right)_{U} = \frac{100'000}{2\widetilde{N}_{ikg}^{q}}\chi^{2}_{2\cdot\left(n_{prev}^{31.12.j}+\hat{n}_{prev}^{31.12.j}+1\right),\,1-a/2} \quad (25)$$

where $\chi^{2}_{f,l}$ is the lth quantile of the Chi-Square distribution with f degrees of freedom.

The variance of $\left(\widehat{LDPP}_{kg}^{x}\right)_{31.12.j}$ is given by

$$Var\left(\left(\widehat{LDPP}_{kg}^{x}\right)_{31.12.j}\right) = \frac{1}{\left(E_{kg}^{q}\right)^{2}}\sum_{i}\left(E_{ikg}^{q}\right)^{2}Var\left(\left(\widehat{LDPP}_{ikg}^{x}\right)_{31.12.j}\right). \quad (26)$$

We calculate 95% CIs based on the normal approximation for the log-transformed LDPP for the whole of Switzerland, $\left(ln\left(\left(\widehat{LDPP}_{kg}^{x}\right)_{31.12.j}\right)\right)$, with variance

$$Var\left[ln\left(\left(\widehat{LDPP}_{kg}^{x}\right)_{31.12.j}\right)\right] = Var\left(\left(\widehat{LDPP}_{kg}^{x}\right)_{31.12.j}\right)/\left(\left(\widehat{LDPP}_{kg}^{x}\right)_{31.12.j}\right)^{2}. \quad (27)$$

We then back-transform the Wald limits.

4.4 Prevalence projection

LDPC and LDPP projections are helpful for planning public health resources. A duration suggested by Pisani et al.,[32] NACR carry out medium-term projections for periods up to 5 years into the future.

Pisani et al.'s method predicts future incidence and future survival, then combines both predictions to derive the predicted prevalence—a process done for single years of age and each calendar year as follows

$$\left(\widehat{LDPC}_{k}^{x}\right)_{31.12.j} = \sum_{z=1}^{x}\hat{p}_{k+1-z}^{j+1-z}\cdot\hat{S}_{k+1-z}(z-0.5) \quad (28)$$

where $\hat{p}_{k+1-z}^{j+1-z}$ is the age-specific incidence in year $j+1-z$ and $\hat{S}_{k+1-z}(z-0.5)$ the age-specific survival at time $t=z-0.5$ years after cancer diagnosis.[32] For example, the 2-year LDPC for age-class k at the end of 2015, $\left(\widehat{LDPC}_{k}^{2}\right)_{31.12.2015}$, is estimated as expected incidence in age-class k in 2015 times the survival probability for age-class k at time $t=0.5$ plus the expected incidence for age-class $k-1$ in 2014 times the survival probability for age-class $k-1$ at time $t=1.5$

$$\begin{aligned}\left(\widehat{LDPC}_{k}^{2}\right)_{31.12.2015} &= \sum_{z=1}^{2}\hat{p}_{k+1-z}^{2015+1-z}\cdot\hat{S}_{k+1-z}(z-0.5) = \hat{p}_{k+1-1}^{2015+1-1}\cdot\hat{S}_{k+1-1}(1-0.5) + \hat{p}_{k+1-2}^{2015+1-2}\cdot\hat{S}_{k+1-2}(2-0.5)\\ &= \hat{p}_{k}^{2015}\cdot\hat{S}_{k}(0.5) + \hat{p}_{k-1}^{2014}\cdot\hat{S}_{k-1}(1.5)\end{aligned}$$

Our simplifying assumption is that all cancer diagnoses occur at mid-year.

Incidence projection

To predict future incidence rates based on age at diagnosis and calendar period of diagnosis, NACR uses relatively short periods of projection, such as 5 years, and employs conventional age-period modeling.

NACR's approach assumes that cohort effects have minimal influence on rates over 5-year time spans. NACR employs a Stata macro provided by the European Network of Cancer Registries (ENCR) based on the method by Hakulinen and Dyda.[33,34]

In short, age-period modelling assumes that incident cases are Poisson distributed and rates are extrapolated as simple log-linear or even linear trends (the latter is recommended for increasing trends in order to avoid an explosion of predicted incidence produced by exponential models).

For stable or decreasing incidence trends, we model projections as

$$ln\left(n_k^j/N_k^j\right) = \alpha_k + \beta_k j \quad (29)$$

For increasing incidence trends, we model projections as

$$n_k^j/N_k^j = \alpha_k(1 + \beta \cdot j) \quad (30)$$

where α_k is the age effect, β_k the calendar-period effect for age-group k, and β the overall calendar-period effect.

We base predictions on trends within a chosen observation period ending in year $j-1$, and we use the official projections of population growth published by the FSO for the calendar years $j+1, \cdots, j+5$.

We apply locally weighted regression to stabilize the observed trends in the incidence rates we used as the basis for prevalence projections.[35]

After we fit the models in (29) and (30) using 5-year age groups, we apply restricted cubic spline regression to age-group specific observed and projected incidences to derive incidence predictions for single years of age—a prerequisite for Pisani et al's method.[32] We place knots empirically; for example, at ages $k = 5, 15, 35, 45, 65, 75, 85, 95$.

Survival projection

We estimate future survival by applying the period method (section 3.4.2) to patients with very recent dates of follow-up.[24,36] And also by fitting flexible parametric survival models[28-30] for single years of age (k), separately for cancer sites, and each combination of sex (g) and language region (i).

We include models with covariate age at diagnosis as a linear term. We always assume proportional hazards. Survival probabilities at $t = 0.5, 1.5, \cdots, 9.5$ years after diagnosis for single years of age from age $k = 0$ to age $k = 95$ are used for the Pisani method.

Variance and confidence intervals

We derive the variance of the observed incidence counts based on the Poisson assumption and the variance for the forecasted incidence counts based on the prediction model for incidence. We derive the variance of survival probabilities from the flexible parametric survival model.

We approximate the variance of the projected LDPC from Pisani et al.'s method[32] by applying error propagation rules for sums and products of random, uncorrelated, and correlated variables as

$$Var\left(\left(\widehat{LDPC}_k^x\right)_{31.12.j}\right) = \sum_{z=1}^{x}\Big[Var\left(\hat{p}_{k+1-z}^{j+1-z}\right)\cdot\left[\hat{S}_{k+1-z}(z-0.5)\right]^2 + Var\left[\hat{S}_{k+1-z}(z-0.5)\right]\cdot\left(\hat{p}_{k+1-z}^{j+1-z}\right)^2 + Var\left(\hat{p}_{k+1-z}^{j+1-z}\right) \cdot Var\left[\hat{S}_{k+1-z}(z-0.5)\right]\Big] + 2\cdot\sum_{z<w} Cov\left[\hat{p}_{k+1-z}^{j+1-z}\cdot\hat{S}_{k+1-z}(z-0.5), \hat{p}_{k+1-w}^{j+1-w}\cdot\hat{S}_{k+1-w}(w-0.5)\right].$$

(31)

4.5 Possible sources of error

We counsel caution when using prevalence estimates for health service planning purposes at national levels and international comparisons. Prevalence estimates are susceptible to the same biases that affect incidence and survival estimates, particularly for those cancers that are not uniformly and rapidly fatal.

The observed trends in prevalence estimates can be influenced by changes in data quality and coding conventions, such as

- differences in levels of vital-status loss to follow-up,
- the proportion of cases known only from death certificates,
- the migration of patients with cancer in and out of the registered population,
- the diagnosis of multiple cancers in the same person, and
- most importantly, the achieved completeness of case ascertainment by cancer registration.

Completeness of case ascertainment

If diagnoses are underregistered, the prevalence will be underestimated. NACR previously assessed completeness of case ascertainment using the flow method[37] and the method of comparing mortality to incidence ratio with survival proportions[38,39] without detecting signs of overt underregistration.[1]

Quality of passive and active vital status follow-up

All registries perform passive follow-up via linkage with the nationwide CCO, as well as with official vital statistics (section 1.2). Active follow-up encompasses regular assessment of the vital status of each registered person. Completeness of active follow-up differs between registries.

If deaths are underregistered, the prevalence will be overestimated. Depending on cancer type and attained age, cases without known vital status at the index date may be substantial. However, we account for this issue (section 4.3.1; Number of expected prevalent persons $\hat{n}_{prev}^{31.12.j}$).

5. Other statistical endpoints

5.1 Years of potential life lost

The number of years of potential life lost (YPLL) is an indicator for premature mortality. It is calculated as the sum of differences between age at death and a theoretically defined age limit (usually set to 70 years) reflecting life expectancy in the population. YPLL can also be presented as a rate. Reporting YPLL draws attention to certain cancers that contribute much to society's cancer burden—not because they are common, but because they occur early in life.

6. Bibliography

1. Lorez, M. *et al*. Evaluation of completeness of case ascertainment in Swiss cancer registration. *Eur. J. Cancer Prev.* **26**, S139–S146 (2017).
2. Michel, G. *et al*. The Swiss Childhood Cancer Registry: rationale, organisation and results for the years 2001-2005. *Swiss Med. Wkly.* **137**, 502–509 (2007).
3. Schindler, M. *et al*. Death certificate notifications in the Swiss Childhood Cancer Registry: assessing completeness and registration procedures. *Swiss Med. Wkly.* **145**, w14225 (2015).
4. World Health Assembly, 19. Eighth revision of the International Classification of Diseases. (Fourteenth plenary meeting, 20 May 1966, Committee on Programme and Budget, sixth report) (1966).
5. *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines.* (World Health Organization, 1992).
6. Lutz, J.-M., Pury, P., Fioretta, G. & Raymond, L. The impact of coding process on observed cancer mortality trends in Switzerland. *Eur. J. Cancer Prev. Off. J. Eur. Cancer Prev. Organ. ECP* **13**, 77–81 (2004).
7. Steliarova-Foucher, E., Stiller, C., Lacour, B. & Kaatsch, P. International Classification of Childhood Cancer, third edition. *Cancer* **103**, 1457–1467 (2005).
8. Sahai, H. & Khurshid, A. Confidence Intervals for the Mean of a Poisson Distribution: A Review. *Biom. J.* **35**, 857–867 (1993).
9. Waterhouse, J., Muir, C. S. & Correa, P. *Cancer incidence in five continents*. (IARC, 1976).
10. Ng, H. K. T., Filardo, G. & Zheng, G. Confidence interval estimating procedures for standardized incidence rates. *Comput. Stat. Data Anal.* **52**, 3501–3516 (2008).
11. Fay, M. P. & Feuer, E. J. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat. Med.* **16**, 791–801 (1997).
12. Lutz, J.-M., Pury, P., Fioretta, G. & Raymond, L. The impact of coding process on observed cancer mortality trends in Switzerland. *Eur. J. Cancer Prev. Off. J. Eur. Cancer Prev. Organ. ECP* **13**, 77–81 (2004).
13. Schmidlin, K. *et al*. Impact of unlinked deaths and coding changes on mortality trends in the Swiss National Cohort. *BMC Med. Inform. Decis. Mak.* **13**, 1 (2013).
14. Hinchliffe, S. R., Rutherford, M. J., Crowther, M. J., Nelson, C. P. & Lambert, P. C. Should relative survival be used with lung cancer data? *Br. J. Cancer* **106**, 1854–1859 (2012).
15. Fisch, T. *et al*. Variation in survival after diagnosis of breast cancer in Switzerland. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **16**, 1882–1888 (2005).

16. Aebi, N. J. *et al.* Patterns of cross-sectional and predictive physical activity in Swiss adults aged 52+: results from the SAPALDIA cohort. *Swiss Med. Wkly.* (2020) doi:10.4414/smw.2020.20266.
17. Faeh, D., Bopp, M., & Swiss National Cohort Study Group. Educational inequalities in mortality and associated risk factors: German- versus French-speaking Switzerland. *BMC Public Health* **10**, 567 (2010).
18. Krieger, J.-P. *et al.* Dietary Patterns and Their Sociodemographic and Lifestyle Determinants in Switzerland: Results from the National Nutrition Survey menuCH. *Nutrients* **11**, E62 (2018).
19. Cronin, K. A. & Feuer, E. J. Cumulative cause-specific mortality for cancer patients in the presence of other causes: a crude analogue of relative survival. *Stat. Med.* **19**, 1729–1740 (2000).
20. Ederer, F. & Hernan, H. Instructions to IBM 650 programmers in processing survival computations. No. 10. *Methodological note* (1959).
21. Baili, P., Micheli, A., Montanari, A. & Capocaccia, R. Comparison of Four Methods for Estimating Complete Life Tables from Abridged Life Tables Using Mortality Data Supplied to EUROCARE-3. *Math. Popul. Stud.* **12**, 183–198 (2005).
22. Corazziari, I., Quinn, M. & Capocaccia, R. Standard cancer patient population for age standardising survival ratios. *Eur. J. Cancer Oxf. Engl. 1990* **40**, 2307–2316 (2004).
23. Brenner, H., Arndt, V., Gefeller, O. & Hakulinen, T. An alternative approach to age adjustment of cancer survival rates. *Eur. J. Cancer Oxf. Engl. 1990* **40**, 2317–2322 (2004).
24. Brenner, H. & Gefeller, O. An alternative approach to monitoring cancer patient survival. *Cancer* **78**, 2004–2010 (1996).
25. Brenner, H. & Holleczek, B. Deriving valid population-based cancer survival estimates in the presence of nonnegligible proportions of cancers notified by death certificates only. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **20**, 2480–2486 (2011).
26. Silcocks, P. & Thomson, C. S. Correcting population-based survival for DCOs - why a simple method works and when to avoid it. *Eur. J. Cancer Oxf. Engl. 1990* **45**, 3298–3302 (2009).
27. Gigli, A. *et al.* Estimating the variance of cancer prevalence from population-based registries. *Stat. Methods Med. Res.* **15**, 235–253 (2006).
28. Lambert, P. C. & Royston, P. Further Development of Flexible Parametric Models for Survival Analysis. *Stata J. Promot. Commun. Stat. Stata* **9**, 265–290 (2009).
29. Royston, P. Flexible Parametric Alternatives to the Cox Model, and more. *Stata J. Promot. Commun. Stat. Stata* **1**, 1–28 (2001).

30. Royston, P. & Parmar, M. K. B. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat. Med.* **21**, 2175–2197 (2002).

31. Clegg, L. X., Gail, M. H. & Feuer, E. J. Estimating the variance of disease-prevalence estimates from population-based registries. *Biometrics* **58**, 684–688 (2002).

32. Pisani, P., Bray, F. & Parkin, D. M. Estimates of the world-wide prevalence of cancer for 25 sites in the adult population. *Int. J. Cancer* **97**, 72–81 (2002).

33. Hakulinen, T. & Dyba, T. Precision of incidence predictions based on Poisson distributed observations. *Stat. Med.* **13**, 1513–1523 (1994).

34. Dyba, T., Hakulinen, T. & Päivärinta, L. A simple non-linear model in incidence prediction. *Stat. Med.* **16**, 2297–2309 (1997).

35. Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* **74**, 829–836 (1979).

36. Talbäck, M., Rosén, M., Stenbeck, M. & Dickman, P. W. Cancer patient survival in Sweden at the beginning of the third millennium--predictions using period analysis. *Cancer Causes Control CCC* **15**, 967–976 (2004).

37. Bullard, J. *et al.* Completeness of cancer registration: a new method for routine use. *Br. J. Cancer* **82**, 1111–1116 (2000).

38. Asadzadeh Vostakolaei, F. *et al.* The validity of the mortality to incidence ratio as a proxy for site-specific cancer survival. *Eur. J. Public Health* **21**, 573–577 (2011).

39. Parkin, D. M. & Bray, F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. *Eur. J. Cancer Oxf. Engl. 1990* **45**, 756–764 (2009).

37. Lorez M, Galli F, Arndt V, and the NICER Working Group. Swiss Cancer Prevalence and Language region. Swiss Cancer Bulletin, **38**(1), 86-93 (2018).