Foundation
National Institute
for Cancer Epidemiology
and Registration

Methods for the estimation and projection of cancer prevalence

Contents

A. Introduction

The cancer prevalence is defined as the number of persons alive at the index date and who have previously been diagnosed with cancer. We estimated prevalence for index dates at yearly intervals from 31.12.2000 to 31.12.2015. The cancer prevalence proportion is the number of prevalent persons divided by the population at risk at index-date, usually expressed as fraction of 100'000 or as percentage. Persons are stratified by sex and age attained at the index-date.

Limited-duration prevalence represents the number of persons alive on the index-date who had a diagnosis of cancer within the past x years (e.g. x = 2, 5 or 10 years). For example, the ten-year limited-duration prevalence at the index-date 31.12.2010 comprises all persons alive at the index-date with a diagnosis within 1.1.2001 to 31.12.2010. Limited-duration prevalence is accessible by a simple counting method and is described below (chapter B: "Counting prevalent persons for index date ≤2010").

Complete or total prevalence on the other hand, is an estimate of the number of persons alive on the index-date who had been diagnosed with cancer, no matter how long ago that diagnosis was. Complete prevalence estimation typically requires statistical modelling to estimate the number of survivors diagnosed before the starting date of cancer registration and is described in chapter D: "Modelling of complete prevalence using the PIAMOD method".

The two-year limited-duration prevalence is relevant for assessing current demands on public health services regarding cancer treatment. They likely include primary treatment and supportive care for recovery from its effects. Prevalence at two to five years after diagnosis likely includes patients under close clinical assessment for recurrence. On the other hand, the five to ten-year limited duration prevalence, or higher, includes many persons who no longer receive cancer related treatment but may be utilizing services for late and long-term effects of their cancer diagnosis and treatment.

B. Counting prevalent persons for index date ≤2010.

B.1. Data sources.

Anonymized individual cancer diagnoses were taken from the National Cancer Dataset managed by the Foundation National Institute for Cancer Epidemiology and Registration (NICER) for the purpose of national cancer monitoring in Switzerland. Ten of the 16 Swiss cantons registering cancer up to diagnosis year 2010 were eligible to be included in this study.

The predominately German-speaking part of Switzerland was represented by eight cantons: Zurich (ZH), Glarus (GL), Basel City (BS), Basel Land (BL), Grison (GR), Appenzell Ausserrhoden (AR), Appenzell Innerrhoden (AI) and St. Gallen (SG). The predominately French/Italian-speaking part of Switzerland was represented by two cantons: Valais (VS) and Geneva (GE). The cancer registries of Neuenburg, Jura and Vaud did not provide survival information required to assess prevalence. Other cantons were not included because they did not have enough years of registration to

derive the ten-year prevalence at the first index date (Ticino, Fribourg, Lucerne). Table 1 lists the contributing Swiss cantons and the calendar years for which diagnoses were available.

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
ZH																				
GL																				
BS																				
BL																				
AR																				
AI																				
SG																				
GR																				
VS																				
GE																				

Table 1: Cancer registration data used for prevalence estimation (counting method). Predominately German-speaking cantons are shown in dark green and predominately French/Italian-speaking cantons in light green.

In seven cantons (ZH, AR, AI, SG, GR, VS, GE) registration covered the full time period 1991 through 2010 required for the ten-year prevalence at the first index date. In three cantons, individual registration years were lacking. Cancer registration started in 1992 in GL, and BS and BL did not provide diagnoses for 2009 and 2010. The missing prevalence estimates for GL, BS, and BL were carried forward from the nearest year with registration data.

B.2. Completeness of case ascertainment.

Completeness of case ascertainment has been assessed by the proportion of microscopically/histologically verified cases and the mortality to incidence ratio without detecting signs of overt under-registration. For the registry of canton Zürich, a certain amount of under-registration is possible because of partially restricted access to case data since 2007 (Cantonal report 2012 available online:
http://www.krebsregister.usz.ch/Publikationen/Seiten/Berichte.aspx).

B.3. Extrapolation to whole Switzerland.

Counted prevalence was extrapolated specific for cancer site, attained age (six classes), sex and language-region to all existing nineteen predominately German-speaking cantons, all existing seven predominately French/Italian-speaking cantons and to all Swiss cantons combined. Since we cannot exclude regional differences in cancer risk and survival, the derived prevalence trends are a combination of true trends and a reflection of increases in cantonal coverage.

B.4. Case inclusion/exclusion criteria.

Diagnoses were coded according to the "International Classification of Disease for Oncology" (ICD-O Third Edition), checked for validity and consistency using the IARCcrgTools package version 2.05 http://www.iacr.com.fr/iarccrgtools.htm). Whether a diagnosis was primary or secondary was decided according to international rules issued by IARC, IACR and ENCR (http://www.iacr.com.fr/MPrules_july2004.pdf). Conversion from ICD-O-3 into ICD-10 ("International Statistical Classification of Diseases and Related Health Problems", 10th Revision) was done using IARCcrgTools version 2.05.

Persons with multiple primary malignant cancer diagnoses in different sites were included separately for each site-specific prevalence. For persons with multiple primary malignant diagnoses in a single site, we included only the first primary diagnosis. The prevalence estimate for all cancer sites combined excluded diagnoses for non-melanotic skin cancer (ICD-O3 topography code C44.0-C44.9 with any morphology code except 8720/3) and counted the first primary malignant diagnosis in a person's lifetime.

For DCO (death-certificate-only) cases, the diagnosis date is unknown. Such cases are infrequent in Swiss cancer registration (<5%) for the majority of sites, with the exception of hepatic or pancreatic cancer. Because there was no available diagnosis date we excluded DCO cases for prevalence estimation. We assume that this exclusion results in a slight underestimation of the prevalence. Cases discovered at death were also excluded.

B.5. Correction for lost to follow-up.

For person's lost to follow-up, the vital status at the latest index date is unknown. The probability P_Δ of each patient lost to follow-up still being alive at the index date, conditional on the length of observed survival $P(fu)$, was estimated as

$$P_\Delta = P(fu + \Delta)/P(fu) \qquad (1)$$

with Δ being the time from last follow-up to index-date. These survival probabilities were estimated by flexible parametric models of observed survival using the *stpm2* command written for the Stata® language (Royston and Lambert 2009). In brief, flexible parametric models are modified Weibull models where the time after diagnosis is incorporated as restricted cubic spline function. Cancer registry, sex and age at diagnosis were used as covariates. We assumed proportional hazards (i.e. the relative effects of covariates were constant in time after diagnosis). Non-parametric Kaplan-Meier estimation was used to validate the survival functions derived with flexible parametric modelling.

Complete type survival analysis was performed for each index-date by selecting a cohort of patient with diagnosis within ten years before the index date.
Cases lost to follow-up were then selected according to the probability to be alive at index date P_Δ. The expected survivors were added to those cases observed alive at index-date.

B.6. Quality of passive and active follow-up (FU).

All registries performed passive FU via annual linkage to the official vital statistics. In ZH, passive FU was incomplete because linkage has been attempted only if cancer was mentioned in the death certificate. Active follow-up encompasses regular assessment of the vital status of each registered person. Completeness of active follow-up as of 31.12.2010 was different between cantons. Complete active follow-up was provided by SG, AR, AI, and GE. In ZH, GL, BS/BL, GR and VS, the follow-up date was somewhere between diagnosis date and 31.12.2010 in 13%, 17%, 20%, 13% and 18% of the cases, respectively. Such cases without recently updated vital status were treated as lost to follow-up. In ZH, there were additional 16% of cases without active follow-up. To cases without any active follow-up information, a survival time of 1-day was assigned after which the case was considered lost to follow-up.

B.7. Estimating confidence intervals for observed (counted) prevalences ≤2010.

Clegg et al. (2002) and Gigli et al. (2006) proposed to approximate the sum of patients observed alive and those expected alive among those lost to follow-up (see B.5.) as Poisson distributed. The variance of the corresponding prevalence proportion is thus estimated by

$$Var(P(x)) = \frac{C_1(x) + C_2(x)}{N(x)^2} \quad (2)$$

where $P(x)$ is the prevalence proportion in age-class x, $C_1(x)$ are the number of cases in age class x observed alive, $C_2(x)$ are the number of cases in age class x expected alive and $N(x)$ is the general population in age class x. Confidence intervals were derived by relying on the relationship between Poisson and Chi-Square distribution (Johnson and Kotz, 1969). Lower (P_L) and upper (P_U) confidence limits at levels 1-α are derived as

$$P_L = \chi^{-2}_{2(c_1+c_2),\alpha/2} \big/ 2N \quad (3)$$

$$P_U = \chi^{-2}_{2(c_1+c_2+1),1-\alpha/2} \big/ 2N \quad (4)$$

where c_1+c_2 is the observed value of C_1+C_2 and $\chi^{-2}_{f,p}$ is the pth quantile of the Chi-Square distribution with f d.f.

Exact Poisson 95% confidence limits are given for the sum of observed and expected prevalence counts.

B.8. Population at risk.

End- and mid-year populations at completed age (i.e. age at last birthday) for 1981-2010 as well as predictions for future end-year populations 2011-2020 are available online and stratified by canton, age and sex provided by the Swiss Federal Statistical Office (SFSO) at STAT-TAB: “Die interaktive Statistikdatenbank”. http://www.bfs.admin.ch/bfs/portal/de/index/infothek/onlinedb/stattab.html . For predictions, we employed the recently updated “middle” scenario for future growth (downloaded document “px-d-01-4C01”; accessed 24.6.2013). Since we presented prevalence proportions for end-year index-dates, end-year populations were used as denominators. For predictions of event rates 2011 to 2015, mid-year populations were employed, derived as the mean of successive end-year populations (see C.1.).

Sex	Area*	Year	Age_0_19		Age_20_49		Age_50_59		Age_60_69		Age_70_79		Age_80_up	
			Count	%#	Count	%#	Count	%#	Count	%#	Count	%#	Count	%#
Men	SA	2000	609'603	100	1'122'620	100	332'074	100	223'828	100	148'593	100	66'666	100
Men	SA	2005	588'461	97	1'150'998	103	348'662	105	254'145	114	161'542	109	79'656	119
Men	SA	2010	580'225	95	1'193'671	106	376'583	113	296'416	132	176'567	119	92'841	139
Men	SA	2015	580'256	95	1'201'130	107	430'196	130	318'698	142	207'862	140	108'262	162
Men	SR	2000	246'300	100	451'169	100	132'753	100	90'148	100	59'739	100	26'710	100
Men	SR	2005	251'898	102	468'445	104	137'756	104	102'898	114	64'285	108	31'854	119
Men	SR	2010	261'650	106	496'275	110	148'341	112	119'176	132	71'841	120	37'442	140
Men	SR	2015	270'568	110	525'765	117	174'660	132	126'788	141	84'837	142	43'482	163
Men	CH	2000	855'903	100	1'573'789	100	464'827	100	313'976	100	208'332	100	93'376	100
Men	CH	2005	840'359	98	1'619'443	103	486'418	105	357'043	114	225'827	108	111'510	119
Men	CH	2010	841'875	98	1'689'946	107	524'924	113	415'592	132	248'408	119	130'283	140
Men	CH	2015	850'823	99	1'726'895	110	604'856	130	445'485	142	292'698	140	151'744	163
Women	SA	2000	573'511	100	1'108'615	100	326'568	100	248'549	100	207'131	100	136'827	100
Women	SA	2005	555'726	97	1'136'037	102	343'425	105	269'995	109	212'864	103	155'233	113
Women	SA	2010	545'189	95	1'170'253	106	369'061	113	304'517	123	219'048	106	172'767	126
Women	SA	2015	547'937	96	1'178'105	106	417'606	128	324'169	130	241'690	117	185'922	136
Women	SR	2000	234'737	100	457'790	100	137'498	100	100'963	100	85'091	100	56'767	100
Women	SR	2005	240'491	102	475'552	104	142'725	104	112'781	112	86'634	102	65'053	115
Women	SR	2010	249'231	106	501'668	110	149'328	109	130'038	129	89'817	106	72'965	129
Women	SR	2015	258'947	110	523'352	114	172'898	126	135'606	134	101'169	119	77'231	136
Women	CH	2000	808'248	100	1'566'405	100	464'066	100	349'512	100	292'222	100	193'594	100
Women	CH	2005	796'217	99	1'611'589	103	486'150	105	382'776	110	299'498	102	220'286	114
Women	CH	2010	794'420	98	1'671'921	107	518'389	112	434'555	124	308'865	106	245'732	127
Women	CH	2015	806'883	100	1'701'457	109	590'503	127	459'775	132	342'859	117	263'153	136

* SA: predominately German speaking cantons; SR: predominately French/Italien speaking cantons; CH: all Swiss cantons.

# of count in 2000

Datasources: BFS Alters-Kantonsstatistik (2000 to 2010); BFS Kantonale Bevölkerungsszenarien (2015).

Table 2: Swiss Population dynamics (mid-year population at completed age).
Note that prevalence proportions were expressed as proportion of end-year populations.

C. Medium-term projection of prevalence index dates 2011-2015.

We projected observed data until 2010 for 5 years to 2015. We estimated the future expected incidence and future expected survival and combined both estimates to derive the expected prevalence as suggested in Pisani, Bray and Parkin (2002).

C.1. Incidence projection.

We employed the conventional simplified approach to predict future event rates as depending predominately on age_at_diagnosis and calendar-period_of_diagnosis, assuming that rates are similar if different birth cohorts were compared. The relatively short period of projection (5 years) justified the simplicity of the approach. Stata® commands written by Hakulinen and Dyba (1994) were used for age-period modelling. In short, the model assumed that incident cases are Poisson distributed and that rates can be extrapolated as simple log-linear or even linear trends (the latter is recommended for increasing trends in order to avoid an explosion of prediction produced by exponential models). For stable or decreasing incidence trends, projections were derived with

$$\ln(c_{it} / n_{it}) = \alpha_i + \beta_i * t \qquad (5)$$

and for increasing incidence trends with

$$c_{it} / n_{it} = \alpha_i * (1 + \beta * t) \qquad (6)$$

where c_{it} is the number of cases in age group i = 1, 2,…, 18 for age-groups (0-4), (5-9),…, (80-84), (85+) and time t for the calendar years, n_{it} is the number of person-years for age-group i and year t, α_i is the age effect, and β_i is the calendar-period effect for age-group i. β is an overall calendar-period effect. The projections take the official predicted population growth in 2011 to 2015 into account (see B.8.). The projection is based on observed trends within the period 2002-2010. To stabilize the observed trends in incidence rates as basis for prevalence projections, locally weighted regression was applied.

The prediction model for increasing trends (equation 6) does not assume the same constant proportional change in incidence over time in each age group, which is unrealistic. It is allows steeper changes at higher baseline rates, creating a special case of age-specific trend (Dyba, Hakulinen and Päivärinta (1997)).
Ninety-five percent confidence limits of the prediction interval, accounting for uncertainty of the slope as well as the random error in future observations, assumed asymptotic normality.

Five-year age-groups with fewer than 0.5 observations per year on average during the observation period 2000-2010 were not projected with the method of Hakulinen and Dyba (1994), but using the simplifying assumption of a constant incidence rate for the years 2011-2015. This rate was set to the average rate during 2008-2010.

Restricted cubic spline regression was applied to observed and projected incidences in 5-year age-groups for calendar years 2010 to 2015 in order to interpolate them for single years of age, a prerequisite for the method of

Pisani et al. (2002) (see C.3.). Knots were placed at the middle point of 5-year age-groups. For the age-group 85+ years, a maximum age of 95 years was applied with the upper boundary knot placed at 90 years of age.

C.2. Survival projection.

The future survival may be estimated by applying the period approach to patients with very recent dates of follow-up (Brenner and Gefeller 1996; Talbäck et al. 2004). In short, period analysis selects cases by follow-up dates. This is achieved by left truncation of person-times at risk at the beginning of the specified follow-up period in addition to right censoring at its end. The specified follow-up period encompassed the latest five available years in the National Cancer Dataset (2006-2010). The survival functions were estimated with flexible parametric models (Royston and Lambert, 2009). Separate models were fitted for cancer sites and each combination of sex and language area. Models included the covariate age at diagnosis as linear term. Proportional hazard was always assumed. Survival probabilities at 0.5, 1.5, …, 9.5 years after diagnosis for single years of age from age 0 to age 95 were predicted for use in the Pisani et al. (2002) method (see C.3.).

C.3. Combining incidence and survival.

If future expected incidence counts and survival probabilities have been estimated for single years of age and for each calendar year, the future prevalence counts may be derived as proposed by Pisani, Bray and Parkin (2002) as

$$P_k(n-year) = \sum_{i=1}^{n} I_{k+1-i} * S_{k+1-i}(i-0.5) \qquad (7)$$

where k is the age-class, n is either the 2-, 5- or 10-years limited duration period, I_k is the annual incidence for age-class k and S_k is the survival for age-class k. The simplifying assumption is that all incidences occur at mid-year. For example, the 2-year prevalence for age-class k at the end of 2015 would be estimated as expected incidence in age-class k for 2015 times the survival probability *S(0.5 years)* for age-class k plus the expected incidence in age-class k-1 for 2014 times the survival probability *S(1.5 years)* for age-class k-1.

C.4. Estimating the variance and confidence intervals for projected prevalences 2011-2015.

The variance of the incidence counts in 2000 to 2010 was derived from the Poisson assumption, while the variance for the predicted incidence counts in 2011 to 2015 was derived from the Stata® command written by Hakulinen and Dyba (1994). The variance of survival probabilities was derived from the flexible parametric models of Royston and Lambert (2009). The variance of the prevalence derived with the method of Pisani et al. (2002) was approximated by applying error transformation rules for sums and products of random, uncorrelated variables as

$$Var[P_k(n-year)] = \sum_{i=1}^{n} (Var[I_{k+1-i}] * S_{k+1-i}(i-0.5)^2 + Var[S_{k+1-i}(i-0.5)] * I^2{}_{k+1-i} + Var[S_{k+1-i}(i-0.5)] * Var[I_{k+1-i}]) \qquad (8)$$

using the same notation as in equation (7).

D. Modelling of complete prevalence using the PIAMOD method.

D.1. Introduction.

The PIAMOD method is implemented as part of the statistical software MIAMOD (De Angelis et al. 1994). It allows estimation and projection of cancer prevalence by using yearly (a) cancer incidence, (b) relative survival data, (c) all-cause mortality data, and (d) population data.

In short, an Age-Period-Cohort (APC) model is constructed for the observed incidence, assumed to be a Poisson random variable. The Age-Period-Cohort model describes the natural logarithm of the incidence rates as a sum of non-linear age_at_diagnosis-, (calendar)period_at_diagnosis- and birth_cohort-effects. It also projects incidence rates out of the observation period into the past and future and assumes that both age and cohort effects persist over the whole projection periods, whereas the period effect is simply the linear extrapolation of the trend within a user selected observation period.

D.2. Data inputs.

Incidence: Persons with multiple primary malignant cancer diagnoses in different sites were included separately for each site-specific incidence. For persons with multiple primary malignant diagnoses in a single site, we included only the first primary diagnosis. The estimate for all cancer sites combined excluded diagnoses for non-melanotic skin cancer (ICD-O3 topography code C44.0-C44.9 with any morphology code except 8720/3) and counted the first primary malignant diagnosis in a person's lifetime. Since the MIAMOD software accepts only a single population input file, which serves as denominator for both, the cancer-specific incidence and the general mortality data, incidences had to be up-scaled to simulate complete registration coverage from 1981 to 2010 for the pool of 12 cantons used (ZH, GL, FR, BS, BL, AR, AI, SG, GR, TI, VS, GE). Up-scaling was done by 1-year age-groups, sex and diagnosis year.

Survival: The appropriate σ_{ij} in equation 10 (below) is the probability to survive the extra death hazard due to cancer, i.e. the relative survival. Relative survival was calculated as the ratio of the observed probability of survival of cancer cases and the expected survival of persons in the general population matching in age, sex and calendar year of death (Ederer et al. 1961). Expected cancer survival proportions were estimated using the Ederer II method applied to combined all-cause mortality tables for the cantons included in the present work (Ederer and Heise, 1959). Relative survival ratios were estimated using the *strs* command (version 1.3.7) written by Dickman et al. (in press) for the Stata® Statistical Software.

The relative survival data was provided in tabular form. Survival trend data for periods 1981-85, 1986-1990, 1991-1995, 1996-2000 and 2001-2005 were generated with complete type survival analysis. For diagnosis period 2006-2010 the period type survival analysis was used. For modelling, the simplifying assumption was used that future relative survival does not improve but remains the same as the latest available estimate (2010) and past survival was not worse but the same as the first available estimate (1981).

All-cause mortality: The official vital statistics (SFSO, STAT-TAB; Alters-Kantonsstatistik, 1981-2010) provided number of deaths stratified by completed age (0,1,2,3, ... ,99+), canton, year and sex. The numbers for the 12 cantons in the study were pooled.
http://www.bfs.admin.ch/bfs/portal/de/index/infothek/onlinedb/stattab/01.topic.9.html (last access 30.11.2012).

Population: see B.8.

D.3. Theory.

In more detail (Verdecchia et al., 2002), incidence rate μ at age i and year t is modelled as a polynomial function of age, period of diagnosis and birth cohort using a log link function ϕ

$$\phi(\mu_{it}) = const + \sum_{k=1}^{k_1} a_k i^k + \sum_{k=1}^{k_2} b_k t^k + \sum_{k=1}^{k_3} c_k (t-i)^k \qquad (9)$$

The degree of the polynomials k_1, k_2 and k_3 have been chosen to give the best fit as indicated by a likelihood ratio statistics. The parameters to be estimated are a, b and c for age, period and cohort effects. Since the cohort term is a linear combination of age and year (cohort = year - age), the linear term (k=1) of period of diagnosis is excluded in order to avoid colinearity problems when estimating parameters.

The prevalence is estimated for a specific birth cohort based on incidence and relative survival as

$$\nu_i = \sum_{j=0}^{i-1} (1-\nu_j)\mu_j \sigma_{ij} \qquad (10)$$

with ν_i is the cohort-specific prevalence at attained age i, expressed as the summation over all ages up to i of the rate μ_j of becoming ill between age j and j+1, j less than i, times the probability to survive up to age i, σ_{ij}. The term (1-ν_j) is the birth cohort proportion still at risk at attained age j.

A system of equations like (10) for each birth cohort involved in the observation period allows estimation of all cancer patients with a past history of cancer, irrespective of date of diagnosis, that is, complete prevalence.

D.4. Model selection and validation.

D.4.1. Model selection

The optimal degrees of the polynomials for age and cohort were chosen based on the Likelihood ratio statistics. The period effect of the incidence trend was always fitted using a restricted cubic spline function with knots at 1981, 1991, 2001 and 2010.

D.4.2. Validation 1 based on limited duration prevalence

We compared observed (counted) 5-year limited duration prevalence with modelled 5-year limited duration prevalence counts for 2000-2010. We accepted the modelled complete prevalence estimate if absolute deviations for 5-year limited duration prevalence were ≤ 10% on average.
[Note: This criterion flagged potentially underestimated prevalence for the localisations lung (both sexes), prostate, liver (both sexes) and pancreas (both sexes).]

D.4.3. Validation 2 based on cause-specific mortality

We compared observed cause-specific mortality counts for the cancer in question during 1995-2010 with modelled cause-specific mortality counts. The period start at 1995 was chosen because of a change in the coding rules in the official mortality statistics. We accepted the modelled complete prevalence estimate if absolute deviations for mortality counts were ≤ 15% on average.
[Note: This criterion flagged the localisations all sites combined (women), corpus uteri, female breast, urinary bladder (both sexes), liver and pancreas in women with potentially overestimated site-specific mortality.]

Applying both aforementioned validation rules, we restrained from reporting estimates of complete prevalence for cancer of the liver and pancreas. The two sites yielded, especially in women, substantial deviation between observed and modelled prevalence (liver: 23%, pancreas: 25%) and cause-specific mortality (liver: 27%, pancreas: 15%). Placing higher weight on validation 1, the complete prevalence for lung and prostate cancer is reported with a warning about possible underestimation.

D.5. Extrapolation to whole Switzerland

APC modelling was based on the combined incidence, all-cause mortality and survival trends of twelve Swiss cantons (ZH, GL, FR, BS, BL, AR, AI, SG, GR, TI, VS, GE). Modelled prevalence counts by sex and six age-groups were up-scaled to whole Switzerland using the population ratios. For years 2011-2015, the MIAMOD software projected the future risk population to be slightly smaller compared with the official population projection provided by the SFSO (see B.8.). The up-scaling of prevalence counts for 2011-2015 took this difference into account.

E. Possible sources of errors.

The observed trends in prevalence estimates can be influenced by changes in data quality and coding conventions, e.g. differences in levels of loss to follow-up, the proportion of cases known only from death certificates, the migration of cancer patients in and out of the registered population, the diagnosis of multiple cancers in the same person and, perhaps most importantly, the completeness of prevalence estimates prior to cancer registration.

These sources of possible error are directly related to the quality of cancer registry functions and to the length of time that a registry has been in operation. The proportion of death-certificate-only diagnoses (reported as an index of probable underestimation) ranged from 0% to maximally 13% (liver cancer in a single cancer registry), but remained normally <5% which we regarded as negligible. Loss to follow-up was corrected for. Caution should be applied when using these data for health service planning purposes at a national level or for international

comparisons. Prevalence estimates are susceptible to the forces that drive the incidence of, and survival from, specific cancer types, particularly with respect to those cancers that are not uniformly and rapidly fatal.

F. References.

Brenner H and Gefeller O (1996). An alternative approach to monitoring cancer patient survival. *Cancer* **78**(9), 2004-2010.

Clegg L, Gail M and Feuer E (2002). Estimating the Variance of Disease-Prevalence Estimates from Population-Based Registries. *Biometrics* **58**, 684-688.

De Angelis G, De Angelis R, Frova L and Verdecchia A. (1994). MIAMOD: a computer package to estimate chronic disease morbidity using mortality and survival data. *Comput Programs Biomed* **44**, 99-107. Software download available here: http://www.eurocare.it/MiamodPiamod/tabid/60/Default.aspx .

Dickman PW, Coviello E and Hills M. Estimating and modelling relative survival. *The Stata Journal* (in press).

Dyba T, Hakulinen T and Päivärinta L (1997). A simple non-linear model in incidence prediction. *Statistics in Medicine* **16**, 2297-309.

Ederer F, Axtell LM and Cutler SJ (1961). The relative survival rate: a statistical methodology. Natl Cancer Inst Monogr. **6**: p. 101-21.

Ederer F and Heise H (1959). Instructions to IBM 650 programmers in processing survival computations. Methodological note no. 10. End Results Evaluation Section: National Cancer Institute. Bethesda, MD.

Hakulinen T and Dyba T (1994). Precision of incidence predictions based on Poisson distributed observations. *Stat Med* **13**, 1513-23.
Software download available here: http://www.cancer.fi/syoparekisteri/en/general/links/.

IARC (2004). International Rules for Multiple Primaries (ICD-O Third Edition). IARC, IACR & ENCR. Internal Report No. 2004/02. IARC, Lyon.

Johnson NL and Kotz S (1969). Discrete distributions. Houghton-Mifflin, Boston/ J Wiley & Sons, New York.

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

Royston P and Lambert P (2009). Further development of flexible parametic models for survival analysis. *The Stata Journal*, **9**(2), 265-290.

Royston P and Lambert P (2011). Flexible Parametric Survival Analysis using STATA: Beyond the Cox Model. Stata Press, ISBN-10: 1-59718-079-3.

Talbäck M, Rosen M, Stenbeck M and Dickman PW (2004). Cancer patient survival in Sweden at the beginning of the third millennium – predictions using period analysis. *Cancer Causes and Control* **15**, 967-976.

Verdecchia A, De Angelis G and Capocaccia R (2002). Estimation and projections of cancer prevalence from cancer registry data. *Statist. Med*. **21**, 3511-3526.

F. Contact

Matthias Lorez, PhD MAS
National Institute for Cancer Epidemiology and Registration (NICER)
Seilergraben 49
CH-8001 Zürich
ml@nicer.org

Zürich, 2014