Original article

Annals of Oncology 13: 831–839, 2002
DOI: 10.1093/annonc/mdf152

Measuring cancer prevalence in Europe: the EUROPREVAL Project

R. Capocaccia[1]*, M. Colonna[2,4], I. Corazziari[1], R. De Angelis[1], S. Francisci[1], A. Micheli[3] & E. Mugno[3]
On behalf of the EUROPREVAL Working Group †

[1]*Istituto Superiore di Sanità, Rome, Italy;* [2]*Isère Cancer Registry, Grenoble, France;* [3]*Istituto Nazionale per lo Studio e la Cura dei Tumori, Milan, Italy;* [4]*FRANCIM, Toulose, France*

Received 4 May 2001; revised 12 December 2001; accepted 13 December 2001

Cancer prevalence is the proportion of individuals in a population who at some stage during their lifetime have been diagnosed with cancer, irrespective of the date of diagnosis. Cancer prevalence statistics have generally been provided by a limited number of well established cancer registries that have been in existence for several decades. The advent of systematic follow-up of life status of incident cases and the availability of new statistical methodologies, now makes it possible for registries established during the 1970s or 1980s to provide prevalence data. The main problems encountered in the estimation of prevalence are the inclusion of: (i) cases lost to follow-up; (ii) cases known only from their death certificate; (iii) cases diagnosed before the start of registration; and (iv) the treatment of multiple tumours and migrations. The main aim of this paper was to review these problems and discuss, through the experience gained with EUROPREVAL, how they can be overcome. A method is presented for the calculation of prevalence of all cancers combined in the populations covered by the 45 cancer registries participating in EUROPREVAL. Prevalence of cancer is estimated to be 2% on average, with the highest values (3%) in Sweden and the lowest in Eastern Europe, with a minimum of ~1% in Poland.
Key words: epidemiology, Europe, methods, prevalence, tumours

Introduction

Cancer prevalence is the proportion of individuals in a population who at some stage during their lifetime have been diagnosed with cancer, irrespective of the date of diagnosis. This definition assumes that cancer is an irreversible disease and diagnosed individuals remain cancer cases until death. Such people make greater demands on the health system than the general population. They require treatment, follow-up for cancer recurrence, screening for independent secondary cancers and may be permanently impaired or disabled as a result of their cancer. However, prevalent cancer cases are a highly heterogeneous group in terms of health status, as they include patients undergoing clinical treatment and those diagnosed many years previously who may be considered cured of their cancer and require few if any additional health care resources. Time from diagnosis is therefore an essential qualifier of cancer prevalence data.

Unlike cancer incidence or mortality, prevalence has not been a major focus of epidemiological statistics; nevertheless, several methods have been developed to provide estimates of prevalence mainly as a by-product of some other activity. For example, health surveys sampling the general population [1] can provide prevalence estimates from persons reporting they have been diagnosed with cancer. However, health surveys are expensive, require very large samples to obtain data on rarer cancers, and are prone to bias, as the compliance of seriously ill persons is expected to differ from that of healthy individuals. Direct methods [2–4] employ incidence and follow-up data collected by population-based cancer registries (CRs). In essence, they calculate prevalence by counting how many incident cases are still alive at a given index date; however, cancer registration must have been carried out for long enough so that essentially all surviving cancer cases are registered, otherwise the prevalence estimate will be low.

Indirect methods estimate prevalence by modelling the mathematical relationships between incidence, prevalence, survival and mortality. Depending on what data are available—incidence and survival [5], incidence and mortality [6] or mortality and survival [7]—data can be used to obtain prevalence functions. Mortality data are usually obtained from national statistics, while incidence and survival data are provided by CRs.

**Correspondence to:* Dr R. Capocaccia, Istituto Superiore di Sanità, Laboratorio di Epidemiologia e Biostatistica, Viale Regina Elena 299, 00161 Roma, Italy. Tel: +39-06-499-02541; Fax: +39-06-499-02230; E-mail: rcap@iss.it
†Members of the EUROPREVAL Working Group are listed in the Acknowledgements.

Cancer prevalence data are not systematically available at the level of national populations. Some prevalence data pertaining to a few US states and some European countries that have been covered by cancer registration for many decades have been published [2–4, 8–13].

EUROPREVAL is the first Europe-wide project to estimate the prevalence of the most important cancers in the participating European countries. The first objective of the project was to use direct methods to calculate prevalence data for European populations covered by cancer registration. Data were available from the EUROCARE-2 Study [14], which collected data on cancer patients diagnosed from 1978 to 1992 from 56 participating population-based CRs in 17 European countries. In order to ensure that the cancer prevalence estimates from the numerous participating CRs were comparable across the board, it was necessary to solve some important methodological problems pertaining to data registration, consistency and comparability. The aim of this paper was to review these problems and discuss, through the experience gained with EUROPREVAL, how they can be overcome. We then applied these solutions to the basic prevalence calculations for all cancers combined in the populations covered by the 45 cancer registries participating in EUROPREVAL. The results of the study are more extensively presented in a companion paper appearing in this issue [15].

Problems arising from direct methods of calculating prevalence

The direct method used to calculate cancer prevalence in a population covered by cancer registration was simply to count all incident cases of cancer that were still alive at a given date (the index date). The procedure basically consists of allocating all incident cases still alive at the index date to cells in a two-dimensional matrix according to their age at the index date and the number of years since the cancer was diagnosed. When divided by the corresponding population sizes, they provide prevalence figures as proportions that are specific for age and disease duration. However, the figures thus produced are likely to be incomplete or inaccurate due to loss of cases during follow-up, cases known only from their death certificate (DCO), problems arising from the treatment of migration and multiple cancers, and lack of completeness due to surviving cases diagnosed before the registry came into existence. These problems are discussed below.

Cases lost to follow-up

Some cases are inevitably lost to follow-up and their vital status at the index date is therefore uncertain. For most European registries, the percentage of cases lost is <1% [16]. Higher percentages are the norm in US registries [13], because follow-up procedures differ and there is a relatively high rate of state-to-state migration.

Cases lost to follow-up are considered in the analysis by attributing an estimated survival probability. The approach of Feldman et al. [4] was to derive this probability from a single life table calculated for the whole set of patients not lost to follow-up, irrespective of age and period of diagnosis. A more recently published approach [12] estimates the survival probability of each lost case from the subset of followed patients belonging to the same age and period of diagnosis. The approach of Feldman et al. [4] gives more stable estimates; however, it can produce erroneous results in patients considerably younger or older than the average. The second approach [12], in theory, provides more accurate probability estimates, but is subject to random variability when the total number of patients in the age group considered is small. Whatever method is used, the number of lost cases estimated alive at the index date is added to the number of prevalent cases determined by the direct method.

Cases known only from the death certificate (DCO)

Some cases are notified to registries only when they die and the death certificate reports cancer as the underlying cause. Whether DCO cases should be included in prevalence counts is controversial, as is the method of their inclusion. It can be argued that such cancers are diagnosed very close to death; that patients were not actually treated as cancer patients and therefore only contribute to the prevalent population for a negligible time. On the other hand, some patients are DCO not because they were first diagnosed at death, but because an earlier diagnosis failed to reach the CR. These patients should be included in the prevalence count; however, this is not a simple task. The problem is to estimate the number of cancer cases not observed by the registry who actually had a cancer diagnosis at the prevalence date; these cases will be registered as DCO *after this date*. Specific studies of trends and survival times of DCO cases will have to be carried out in order to estimate the number of such cases. In any event, if they are not included, the proportion of DCO to registered cases should be reported to provide an indication of the extent to which prevalence may have been underestimated. We have done this when presenting the EUROPREVAL data.

Migration

People moving away after cancer diagnosis and registration are usually not lost to follow-up even if they move to a different health area, and although they no longer make demands on the health area where they were diagnosed, are still included in the prevalence data. Conversely, patients who move into another health area after diagnosis are not counted in the prevalence of that area even though they are treated there. In such situations the prevalence of a region or health area is underestimated when the net flow is into that area, and is overestimated in the opposite case. However, the resulting error is small in most European countries, where the net migration rate is usually between –1 and +1% per annum. Prevalence estimates should not therefore be substantially affected by migration, unless there is a net migration of cancer patients

from one area to another. To our knowledge, this phenomenon has not been considered in any prevalence analysis, mainly because there is no systematic information on migration for health reasons. Migration was not considered when presenting the data from EUROPREVAL.

Multiple cancers

Prevalence can refer to the number of *people* with cancer or the number of *cancers* in the population. The difference lies in the way multiple primary malignant tumours are accounted for. *Person prevalence* considers only the first primary malignant cancer diagnosed in each person, and is a measure of the number of people actually making demands on health care resources for cancer. On the other hand, patients with two or more tumours are counted several times in *tumour prevalence*, which considers all primary malignant cancers in a person irrespective of whether they are the first or subsequent cancers. If the multiple cancers are treated independently, this second indicator is more pertinent to the demand for health care.

The difference between the two indicators may be substantial, particularly in older age groups. Inclusion of multiple tumours in comparative studies is complicated by a lack of uniformity among CRs in the application of cancer coding rules, particularly for tumours in paired organs [17]. Furthermore, the number of multiple tumours registered depends on the age of a CR: the older the registry, the greater the likelihood of registering previous diagnoses in multiple cancer patients. Recently established registries may know from clinical records that a given tumour is not the first primary, but may not have the resources or procedures necessary to access full information on previously diagnosed cancers. The EUROPREVAL project considered person prevalence only.

Completeness bias

Even when corrected to include DCO and lost-to-follow-up cases, prevalence measured on populations covered by CRs is still incomplete, as prevalent cases diagnosed before the registry began operating will not be recorded. Such *unobserved* cases are far from negligible, especially for recently established registries (<15 years) and for cancer sites with good prognoses [18]. It is vital, therefore, that these unobserved cases are estimated and included in the prevalence data. It is also essential, in a Europe-wide study involving numerous registries that have been operating for variable lengths of time (from only ~5 years to ≥40 years), that a uniform and unbiased way of dealing with completeness is used, so as to provide comparable prevalence estimates for populations covered by cancer registration for different lengths of time.

We defined the *observed prevalence* as that produced by the counting method described above, to which various (small) corrections have been applied to take account of the cases lost to follow-up. We then applied an appropriate correction factor to the observed prevalence, the *completeness index* [18], which is an estimate of non-registered cases still alive. The figure thus produced is defined as the *total prevalence*. The completeness index will vary according to the length of the registration period and the characteristics of the cancer being considered. Completeness indices were estimated for the Connecticut Cancer Registry and the total prevalence figures thus calculated were compared with observed prevalence, which, since the registry has been operating for >50 years, should have been almost complete; concordances were found to be satisfactory [19]. The same method was applied to Italian prevalence data [20] and to EUROPREVAL data to improve European estimates of prevalence.

Estimation of total prevalence

In a population covered by cancer registration for L years, the total prevalence (N_{tot}) is given by the sum of the observed prevalence, N_{obs} (the proportion of patients diagnosed after the start of registry activity), plus the unobserved prevalence (N_{unobs}) of patients diagnosed before that date, whose value is unknown. N_{tot} is not directly measurable but can be estimated indirectly by dividing N_{obs} by the completeness index, R [18], which depends on L, the length of time the registry has been operating

$$N_{tot} = N_{obs}/R$$

R is an estimate of the extent to which the observed prevalence represents the total prevalence, and is defined by

$$R = N_{obs}^{(m)}/N_{tot}^{(m)}$$

where $N_{obs}^{(m)}$ and $N_{tot}^{(m)}$ are *model-based estimates* of observed and total prevalence, respectively, and are derived from parametric models of age-specific cancer incidence probability and relative survival probability [18]. The completeness index, R, takes the value of 1 when all prevalent cases are observed, and approaches 0 as the proportion of prevalent cases that are observed decreases. The value of $N_{obs}^{(m)}/N_{tot}^{(m)}$ depends on the registration period (L), cancer site, sex and age class: all the factors that influence incidence and survival in the models.

Simple log-linear models can be used as *incidence functions* the major determinants of which are age at diagnosis and date of birth. These models are consistent, for a general class of cancers, with the multistage theory of carcinogenesis [21]. As shown by Capocaccia and De Angelis [18], R is not influenced by absolute incidence levels, but only by the age slope of incidence. R is larger for cancers the incidence of which rises steeply with age, e.g. prostate cancer, and is lower for cancers such as cervical cancer the incidence of which is largely independent of age.

Survival models with cure can be used as *relative survival functions*. This class of model assumes that only a proportion of patients, the so called *fatal* cases, have an excess death risk, while the remainder have the same mortality rate as the general population (not affected by the specific cancer) and

can thus be regarded as cured [22]. Cure models allow estimation of long-term survival, which must be estimated accurately as it plays a crucial role in estimating prevalence. Survival has a direct influence on R: cancers with poor survival are characterised by high R values, as only recently diagnosed patients are likely to be alive at the prevalence date. In contrast, a high proportion of cancers patients with good prognosis who were diagnosed before a young registry started operating will still be alive at the index date, so R will be low.

Similarly, R can be used to estimate the partial prevalence for a period longer than the observation period; for instance, the 15-year prevalence in a population observed only for 10 years. This method is useful to decompose estimated total prevalence by duration of disease.

Standard errors of prevalence estimates

In cases where all prevalent cases are observed and followed from diagnosis to the index date, a simple Poisson distribution can be used to derive the standard error (SE) of the number of cases [5]:

$$\mathrm{SE}\,(N_{\mathrm{tot}}) = \sqrt{N_{\mathrm{tot}}}$$

The standard error of the prevalence as a proportion is obtained by dividing $\sqrt{N_{\mathrm{tot}}}$ by the population count.

However, total prevalence is a composite estimator made up of: (i) a direct count, N_{obs}, of cases observed and followed; (ii) an estimated number (N_{lost}) of lost-to-follow-up cases surviving until the index date; and (iii) an estimated proportion $1/R$ of cases diagnosed before registration began. While the first term is Poisson distributed, the second term comes from an estimated life table—itself derived from the cases actually followed—and the last term is based on statistical models applied to the same or, in some cases, to an independent dataset. A formal theory of prevalence estimator sampling errors, that takes all these sources of variability and their interrelations into account, is being developed but is not yet available [23, 24].

A rough approximation to the SE of the total prevalence can be calculated assuming that the proportion of lost ($P_{\mathrm{lost}} = N_{\mathrm{lost}}/N_{\mathrm{obs}}$) and the completeness index, R, are without error:

$$\mathrm{SE}\,(N_{\mathrm{tot}}) = \sqrt{N_{\mathrm{obs}}}\,(1 + \mathrm{P}_{\mathrm{lost}})/R.$$

However, it is apparent that SEs calculated with this expression systematically underestimate the real variability of the prevalence figures, as potentially important sources of uncertainty are neglected and cannot be used for formal statistical testing.

Prevalence of all cancers combined from European registry data

In the EUROPREVAL Project, prevalence estimates for the most important cancer sites were produced for 17 European countries. The results of the study are extensively presented in a companion paper in this issue [15]. Here we report the methodological choices adopted in the study and present prevalence results in Europe for all cancers combined (Tables 1 and 2).

The EUROPREVAL project used incidence and follow-up data provided by the EUROCARE project [14], collecting data from 56 participating population-based CRs from 17 European countries. The EUROCARE-2 database contains data on cancer patients diagnosed from 1978 to 1992, the minimum information for each patient being sex, date of birth, date of diagnosis, date of end of follow-up, tumour site, morphology and life status.

To calculate consistent prevalence figures at a given date, incidence and follow-up data must be complete at that date. The most recent date for which these data are available for all registries participating in EUROCARE is 31 December 1992, and this was taken as the index date for prevalence computation.

A certain fraction of the records (1–2% in most European registries) had wrong or missing values for some patient variables. Often the month of birth or month of diagnosis was missing, but in a few cases both the month and year were missing; in other cases the sex was not specified. The exclusion of such cases from calculations leads to underestimation of the prevalence. Automatic procedures to correct data incompatibilities or to impute missing values were therefore used whenever possible.

We used the direct approach to estimate the prevalence of all cancers in European registries at the index date of 31 December 1992. Specifically developed software [25] was used for the calculation of prevalence. The basic data are shown in Table 1 along with the main steps in the calculation, so as to illustrate the problems discussed above. Data from most participating CRs were included; data from specialised registries (concerned with digestive system or haematological cancers, etc.) were not included. For each CR, the following are reported in Table 1: (a) number of years of cancer registration, prior to the index date, available from EUROCARE-2 database; (b) the number of cancer cases collected during the period and included in the analysis; (c) the number of cases alive at the index date; (d) the number of lost cases; (e) the number of lost cases estimated alive at the index date; (f) the observed prevalent cases [= (c) + (e)] up to (a) years after diagnosis; (g) the completeness index; (h) the total prevalent cases [= (f)/(g)]; (i) the population count per 100000; (j) the total prevalence per 100000 [= (h)/(i)]; and (k) the average yearly DCO cases as a percentage of the total prevalence.

The incidence period considered ranged from 5 years (1998–1992) in several southern European registries to 23 years (1970–1992) in Iceland, Saarland and Geneva. Some registries, mainly in northern Europe, where cancer registration started during the 1950s, also provided a complete set of data covering their entire registration period. These data were used to check the estimates of the completeness index, R.

Table 1. Calculation of prevalence for all cancers combined in European cancer registry areas (national registries are in capital letters)

Registry	Period length[a]	Cases observed[b]	Cases alive[c]	Cases lost[d]	Lost estimated alive[e]	Observed prevalent cases[f]	Completeness index[g]	Total prevalent cases[h]	Population (per 100 000)[i]	Total prevalence[j] (×100 000)	% DCO cases[k]
Tyrol (A)	5	11 027	6534	0	0	6534	0.42	15 550	6.41	2426	1.4
DENMARK	17	306 016	97 764	0	0	97 764	0.79	123 488	51.7	2389	0.0
Eindhoven (NL)	15	37 738	13 284	684	253	13 537	0.78	17 250	9.24	1867	0.0
East Anglia	14	95 451	27 325	4819	2596	29 921	0.73	40 840	20.89	1955	0.2
Mersey	8	71 560	27 433	0	0	27 433	0.56	49 186	24.12	2039	0.9
Oxford	14	119 304	38 637	1	0	38 637	0.73	52 620	25.82	2038	0.3
Thames	15	313 481	92 665	1483	701	93 366	0.75	124 198	67.56	1838	2.8
Wessex	14	163 327	55 287	0	0	55 287	0.73	75 699	29.93	2529	1.4
West Midlands	15	272 119	79 145	43	15	79 160	0.76	104 502	52.78	1980	0.5
Yorkshire	15	197 615	54 474	10	1	54 475	0.76	72 102	36.98	1950	0.9
ESTONIA	15	55 612	14 893	529	107	15 000	0.73	20 681	15.44	1339	0.0
FINLAND	15	205 377	73 179	74	31	73 210	0.78	94 132	50.42	1867	0.1
Somme (F)	11	20 360	7510	2040	1335	8845	0.68	12 941	5.49	2357	0.0
Saarland (G)	23	84 618	26 250	0	0	26 250	0.90	29 302	10.55	2777	1.4
ICELAND	23	13 668	4700	4	0	4700	0.89	5270	2.61	2019	0.0
Florence	8	44 373	19 619	327	150	19 769	0.59	33 748	11.82	2855	0.8
Genoa	7	24 550	10 714	46	19	10 733	0.54	19 818	6.79	2919	0.4
Latina	10	10 878	4523	57	12	4535	0.67	6804	4.79	1420	0.3
Modena	5	14 566	7525	155	0	7525	0.45	16 841	6.06	2779	0.0
Parma	15	27 262	8919	70	12	8931	0.79	11 364	3.92	2899	0.8
Ragusa	12	8338	2625	18	10	2635	0.72	3683	2.91	1266	0.1
Romagna	7	15 626	7860	17	10	7870	0.55	14 419	4.26	3385	0.3
Turin	8	31 888	13 094	1307	558	13 652	0.59	23 274	9.56	2435	0.9
Varese	15	45 803	16 363	415	130	16 493	0.79	20 944	7.99	2621	0.0
Cracow	17	24 784	5633	1156	212	5845	0.78	7464	7.13	1047	1.6
Warsaw	5	23 941	7649	2166	838	8487	0.43	19 862	16.25	1222	1.1
SCOTLAND	15	302 159	81 395	52	18	81 413	0.76	107 226	51.11	2098	0.8
SLOVAKIA	15	184 706	66 708	163	30	66 738	0.71	93 343	53.07	1759	2.0
SLOVENIA	10	51 487	17 848	273	77	17 925	0.62	28 783	19.96	1442	0.8
Basque Country	7	40 807	20 757	0	0	20 757	0.55	37 847	20.97	1805	1.6
Mallorca	5	10 426	5135	160	38	5173	0.44	11 689	5.86	1995	0.4
Navarra	8	12 522	6241	0	0	6241	0.58	10 671	5.22	2044	1.5
Tarragona	8	12 384	5658	73	33	5691	0.58	9786	5.53	1770	0.7
South Sweden	15	80 313	33 020	0	0	33 020	0.76	43 168	14.17	3046	0.0
Basel	12	17 127	7447	114	59	7506	0.69	10 860	4.33	2508	0.0
Geneva	23	29 784	8605	1375	486	9091	0.89	10 178	3.87	2630	0.2

[a]Number of years of cancer registration, before the index date of 31 December 1992.
[b]Number of cancer cases collected during the period and included in analysis.
[c]Number of cases alive at the index date.
[d]Number of cases lost.
[e]Number of cases lost estimated alive at the index date.
[f]Observed prevalent cases up to period length years after diagnosis [= cases alive + lost estimated alive].
[g]Proportion of the total prevalence observed by the registry.
[h]Total prevalent cases [= observed prevalent cases/completeness index].
[i]Population per 100 000.
[j]Total prevalence per 100 000 [= total prevalent cases/population (per 100 000)].
[k]Average number of cases known only from their death certificate (DCO) per year as a percentage of total prevalence.

Table 2. Total prevalence decomposition by duration of disease for all cancers combined in European countries at the common index date of 31 December 1992 and prevalence proportions per 100000 of population within 2, 5 and 10 years since diagnosis

Registry	Total prevalent cases	Two-year prevalence	Five-year prevalence	Ten-year prevalence	Total prevalence
Austrian registry	15550	514	1020	1565	2427
Denmark	123488	534	1028	1550	2389
Dutch registry	17250	431	813	1223	1867
English registries	519148	441	837	1252	2012
Estonia	20681	304	549	812	1339
Finland	94132	411	798	1211	1867
French registry	12941	598	1076	1541	2357
German registry	29302	557	1045	1656	2777
Iceland	5270	398	823	1271	2019
Italian registries	150895	588	1131	1718	2597
Polish registries	27326	283	506	738	1169
Scotland	107226	477	893	1322	2098
Slovakia	93343	327	609	930	1759
Slovenia	28783	343	625	898	1442
Spanish registries	69993	427	814	1244	1863
Swedish registry	43168	602	1218	1888	3046
Swiss registries	21038	530	1066	1632	2566
Total	1379533	448	852	1282	2042

Cases lost to follow-up were considered in the analysis by assuming they had the same survival probability as cases not lost to follow-up of the same age at diagnosis and number of years passed since diagnosis. For each calendar period considered, the number of lost cases estimated alive was added to the number of prevalent cases. The contribution of lost cases estimated alive to the observed prevalence was highest in Somme, where they represent ~15% of the observed prevalent cases, followed by Warsaw (10%) and East Anglia (9%). Percentages ranging from 2% to 6% were observed in three registries: Geneva (a registry with good follow-up procedures, but with problems due to patient migration), Torino and Cracow. In all the other registries, lost cases contributed <2% to the observed prevalence.

As noted, migration was not considered as its influence on prevalence can be assumed to be negligible. Only first primary tumours were included; patients with multiple tumours were considered as one prevalent case. DCO cases were not included, but formed a small proportion of the total prevalence (always <3%, and in most cases <1%). Lack of completeness generally has a major influence on the prevalence estimations. Thus, the completeness index was only ~90% in registries with >20 years of follow-up, and fell to <50% in registries with only 5 years of follow-up. In the latter cases, half or more of the total prevalence had to be estimated.

In Table 2, registries are grouped by country, while prevalence proportions have been decomposed by time from diagnosis.

The average estimated prevalence in European countries was $2042 \times 100\,000$, i.e. about two out of every 100 European citizens have had a previous cancer diagnosis. Sweden presents the highest estimated prevalence (3046). High levels were estimated also in Germany (2777), Italy and Switzerland (with >2500). Low prevalence was estimated in Eastern European countries, with the lowest levels in Poland (1169).

Approximately 20% (variable from 19% to 25% according to country) of total prevalent cases were diagnosed with cancer in the last 2 years, while ~40% (range 35–46%) and 60% (range 53–66%) were diagnosed in the last 5 and 10 years, respectively. The geographical variability of these proportions is lower than the absolute levels of prevalence, at least for all cancers combined. Indeed it is related to country-specific incidence distribution by cancer site and to survival levels, rather than to the length of follow-up.

Validation of total prevalence estimates

As statistical modelling is incorporated in total prevalence estimates, through the use of completeness index, validations against empirical-based estimates were carried out in EUROPREVAL study. To check the values of the completeness indices provided by the statistical models, we used data from the long-established CRs whose observation periods are long

enough to have registered virtually all prevalent cases. We calculated figures for the 15-year prevalence, ignoring cases registered prior to that, and then corrected these data using the corresponding completeness index, *R*. The result (the estimated total prevalence) was then compared with the observed total prevalence. This method was also used to check prevalences estimated in SEER registries up to 1993, using data collected from 1940 to 1993 by the Connecticut cancer registry [19].

The results of the validation analysis are shown in Table 3. Total prevalence estimates for the 10 cancer sites included in the study were analysed separately. For the Finnish and Danish registries, the estimated total prevalence was slightly below the observed total prevalence. This is probably because the observed prevalence figures supplied by these registries refer to cancers rather than persons with cancer. In contrast, for the Eindhoven, Estonia and Saarland registries (with observation periods of 20–25 years) estimated total prevalence figures were more frequently above the observed total prevalence. Differences between estimated and observed values were generally less than ±10%; the only exceptions were cervical cancer and Hodgkin's disease.

For Hodgkin's disease the estimated figures greatly exceeded the observed figures. For cervical cancer, the estimated total prevalence was lower than the observed prevalence in Finland and Denmark, but was >10% in all other registries except Iceland. These mismatches can be attributed to marked changes in the epidemiology of these two cancers. For Hodgkin's disease the use of new more effective therapies became widespread; for cervical cancer screening became widespread. Our modelling approach was based on EUROCARE-2 data for the incidence period from 1978 to 1989, and did not take into account these developments. Thus the excess of estimated prevalent cases of Hodgkin's disease can be explained by model-based overestimation of backward projections of survival; the lower prevalence estimates for cervical cancer in Finland and Denmark are the result of the high incidence levels in the pre-screening period that were not considered in the model.

Conclusions

As the population ages the proportion most at risk for developing cancer grows, while advances in cancer treatment are resulting in an increasing proportion of cancer patients living longer. Thus the demand for social services and especially health care by this sector of the population is growing, particularly in developed countries. Cancer prevalence is a vital indicator, as it is a measure of the number of cancer patients who require health and social services resources, and can be used to adequately plan future allocation of such resources. Providing and updating reliable and systematic prevalence

Table 3. Comparison of the estimated total prevalence (calculated using the completeness index at 15 years) with the observed prevalence for registries established for >20 years, prevalence values (number of prevalent cases) are referred to the index date of 31 December 1992

	Period (years)	Stomach		Colon		Rectum		Melanoma		Leukaemias	
		Estim.	Obs.	Estim.	Obs.	Estim.	Obs.	Estim.	Obs.	Estim.	Obs.
Denmark	50	1374	1430	9855	10968	6565	6141	7588	8104	2354	2375
Finland	40	3190	3487	5393	5233	3684	4078	5384	5128	1782	2003
Iceland	37	224	213	352	339	120	122	183	164	69	69
Estonia	23	1551	1498	1134	1051	1043	958	565	508	564	583
Saarland	23	1014	938	2512	2506	1656	1561	906	879	427	460
Eindhoven	22	493	461	2481	2566	–	–	706	675	269	275
Slovenia	43	1185	1209	1334	1279	1568	1435	959	819	619	641
Thames	37	1842	1784	9092	9758	5724	5580	4556	4784	2369	2613
	Period (years)	Breast		Cervix		Corpus		Prostate		Hodgkin's	
		Estim.	Obs.	Estim.	Obs.	Estim.	Obs.	Estim.	Obs.	Estim.	Obs.
Denmark	50	28526	28205	8896	11220	8064	8592	5681	5774	2777	1738
Finland	40	22644	23139	2855	3238	5974	6837	6476	7323	2331	1523
Iceland	37	1068	1076	269	249	242	256	500	499	89	79
Estonia	23	3286	3699	2846	2373	1792	1708	597	486	801	330
Saarland	23	5852	5911	2601	1724	1863	1631	1683	1759	479	311
Eindhoven	22	4113	4208	688	363	830	756	921	974	370	222
Slovenia	43	5107	5125	2632	1922	2589	2030	802	835	1221	440
Thames	37	31236	31754	6490	4338	5862	5770	6159	7203	3790	3115

Estim., estimated; Obs., observed.

statistics, obtained using uniform and validated methodologies such as we now have, on cancer incidence, survival and mortality is therefore important for all in European countries.

The breakdown of cancer prevalence figures according to time since diagnosis is an important first step towards the development of specific indicators of health care needs for specific sections of the population. A subsequent step will be to classify prevalent cases by disease stage at the index date. This will be even more informative for planning the allocation of health resources, as it will make it possible to identify four groups of patients: those recently diagnosed patients who are receiving primary treatment; those who can be considered cured of their cancer; those in the terminal phase of their illness; and the remainder with intermediate status, also referred to as 'continuing-phase' [26]. The groups thus identified are much more homogeneous in terms of predictable health needs than subgroups simply defined by time since diagnosis. Our approach to the estimation of the stage distribution of prevalence will be described in forthcoming papers from EUROPREVAL.

Acknowledgements

We thank Donald Ward for the English revision of the manuscript. This research has been supported by the EUROPREVAL Biomed-2 Programme, Contract No. BMH4 98 3899.

Members of the EUROPREVAL Working Group: Austria, W. Oberaigner (Cancer Registry of Tyrol); Denmark, H. Storm, G. Engholm (Danish Cancer Society 'Institute of Cancer Epidemiology'); Estonia, T. Aareleid (Estonian Cancer Registry); Finland, T. Hakulinen (Finnish Cancer Registry); France, G. Hédelin (Bas-Rhin Cancer Registry), H. Lefevre (Calvados Digestive Cancer Registry), J. Mace-Lesec'h (Calvados General Cancer Registry), J. Faivre (Côte d'Or Digestive Cancer Registry), G. Chaplain (Côte d'Or Gynaecologic Cancer Registry), P. M. Carli (Côte d'Or Malignant Haemopathies Registry), P. Arveux (Doubs Cancer Registry), J. Estève (University of Lyon), M. Colonna (Isère Cancer Registry), N. Raverdy, P. Jun (Somme Cancer Registry); Germany, J. Michaelis (German Registry of Childhood Malignancies), H. Ziegler, C. Stegmaier (Saarland Cancer Registry); Iceland, H. Tulinius (Icelandic Cancer Registry); Italy, R. Capocaccia (Project Leader), I. Corazziari, R. De Angelis, S. Francisci, S. Hartley, F. Valente, A. Verdecchia, A. Zappone (National Institute of Health, Rome), F. Berrino, G. Gatta, A. Micheli, E. Mugno, M. Sant (National Institute for the Study and Cure of Tumours, Milan), P. Crosignani (Lombardy Cancer Registry), E. Conti,V Ramazzotti (Latina Cancer Registry), M. Vercelli, C. Casella, A. Puppo (Liguria Cancer Registry, NCI, University of Genova, Genova), M. Federico (Modena Cancer Registry), M. Ponz De Leon (Modena Colorectal Cancer Registry), V. De Lisi (Parma Cancer Registry), R. Zanetti, S. Rosso (Piedmont Cancer Registry), C. Magnani (Piedmont Childhood Cancer Registry), L. Gafà, R. Tumino (Ragusa Cancer Registry), F. Falcini (Romagna Cancer Registry), E. Paci, E. Crocetti (Tuscany Cancer Registry), S. Guzzinati, P. Zambon (Venetian Cancer Registry); Poland, J. Rachtan (Cracow Cancer Registry), M. Bielska-Lasota (Warsaw Cancer Registry); Slovakia, I. Plesko (National Cancer Registry of Slovakia); Slovenia, V. Pompe-Kirn (Cancer Registry of Slovenia); Spain, I. Izarzugaza (Basque Country Cancer Registry), A. Izquierdo (Girona Cancer Registry), I. Garau (Mallorca Cancer Registry), E. Ardanaz, C. Moreno (Navarra Cancer Registry), J. Galceran (Tarragona Cancer Registry), V. Moreno (Catalan Institute of Oncologia); Sweden, T. Möller, H. Anderson (Southern Swedish Regional Tumour Registry); Switzerland, J. Torhorst (Basel Cancer Registry), C. Bouchardy, J. M. Lutz, M. Usel (Geneva Cancer Registry), J. E. Dowd (World Health Organisation, Geneva); The Netherlands, J. W. W. Coebergh, M. Janssen-Heijnen (Eindhoven Cancer Registry), R. A. M. Damuhis (Rotterdam Cancer Registry); Scotland, R. Black, V. Harris, D. Stockton (Scottish Cancer Intelligence Unit); UK, T. W. Davies (East Anglian Cancer Registry), M. P. Coleman, S. Harris (London School of Hygiene and Tropical Medicine), E. M. I. Williams (The Merseyside and Cheshire Cancer Registry), D. Forman, R. Iddenden (Northern and Yorkshire Cancer Registry and Information Service & Centre for Cancer Research), M. J. Quinn (Office for National Statistics), M. Roche (Oxford Cancer Intelligence Unit), J. Smith (South and West Cancer Intelligence Unit), H. Moller (Thames Cancer Registry), P. Silcocks (Trent Cancer Registry), G. Lawrence, K. Hemmings (West Midlands Cancer Intelligence Unit).

References

1. Byrne J, Kessler LG, Devesa SS. The prevalence of cancer among adults in the United States 1987. Cancer 1992; 69: 2154–2159.
2. Hakama M, Hakulinen T, Teppo L, Saxen E. Incidence, mortality and prevalence as indicators of the cancer problem. Cancer 1975; 36: 2227–2231.
3. Adami HO, Gunnarson T, Sparén P, Eklund G. The prevalence of cancer in Sweden 1984. Acta Oncol 1989; 28: 463–470.
4. Feldman AR, Kessler L, Myers MH, Naughton MD. The prevalence of cancer. Estimates based on the Connecticut tumor registry. N Engl J Med 1986; 315: 1394–1397.
5. Gail MH, Kessler D, Midthune D and Scoppa S. Two approaches for estimating disease prevalence from population-based registries of incidence and mortality. Biometrics 1999; 55: 1137–1144.
6. Colonna M, Hedelin G, Estève J et al. National cancer prevalence estimation in France. Int J Cancer 1987; 2000: 301–304.
7. Verdecchia A, Capocaccia R, Egidi V, Golini A. A method for the estimation of chronic disease morbidity and trends from mortality data. Stat Med 1989; 8: 201–216.
8. Polednak AP. Estimating the prevalence of cancer in the United States. Cancer 1997; 80: 136–141.
9. Coebergh JW, Janssen-Heijnen M, Louwman M, Voogd A. Cancer incidence, care and survival in the south of The Netherlands 1955–

1999. A report from the Eindhoven Cancer Registry (IKZ). Eindhoven, the Netherlands: Eindhoven Cancer Registry 2001.
10. Thomson H, Rahu M, Aaareleid T, Gornoi K. Cancer in Estonia 1968–1992. Incidence, Mortality, Prevalence, Survival. Tallinn, Estonia: Institute of Experimental and Clinical Medicine 1996.
11. Pompe-Kirn V, Golouh R, Lindtner J et al. Cancer Incidence in Slovenia 1998, Report 40. Ljubljana, Slovenia: Institute of Oncology 2001.
12. Micheli A, Francisci S, Krogh V et al. Cancer prevalence in Italian Cancer Registry areas: the ITAPREVAL study. The ITAPREVAL Working Group. Tumori 1999; 85: 309–369.
13. Ries LAG, Kosary CL, Hankey BF et al. (eds). SEER Cancer Statistics Review 1973–1996, NIH Pub. No. 99-2789. Bethesda, MD: National Cancer Institute 1999.
14. Berrino F, Sant M, Verdecchia A et al. Cancer patients survival in European countries. IARC Scientific Publication No. 132. Lyon, France: IARC 1995.
15. Micheli A, Mugno E, Krogh V et al. Cancer prevalence in European registries areas. EUROPREVAL Study. Ann Oncol 2002; 13: 840–865.
16. Capocaccia R, Gatta G, Chessa E, Valente F. The EUROCARE Working Group. The EUROCARE 2 Study. IARC Scientific Publication No. 151. Lyon, France: IARC 1999; 1–40.
17. Filali K, Hédelin G, Schaffer P et al. Multiple primary cancers and estimation of the incidence rates and trends. Eur J Cancer 1996; 32A: 683–690.
18. Capocaccia R, De Angelis R. Estimating the completeness of prevalence based on cancer registries data. Stat Med 1997; 16: 425–440.
19. Merrill RM, Capocaccia R, Feuer EJ, Mariotto A. Cancer prevalence estimates based on tumour registry data: the Surveillance, Epidemiology, and End Results (SEER) Program. Int J Epidemiol 2000; 29: 197–207.
20. Corazziari I, Mariotto A, Capocaccia R. Correcting the completeness bias of observed prevalence. Tumori 1999; 85: 370–381.
21. Cook PJ, Doll LR, Fellingham SA. A mathematical model for the age distribution of cancer in man. Int J Cancer 1969; 4: 93–112.
22. Verdecchia A, De Angelis R, Capocaccia R et al. The cure for colon cancer: result from the EUROCARE study. Int J Cancer 1998; 77: 322–329.
23. Gigli A, Mariotto A, Corazziari I, Capocaccia R. Estimating the variance of chronic disease prevalence. Quaderno (CNR) 2000; 1–24.
24. Clegg LX, Mitchell GH, Feuer EJ. Estimating the variance of disease prevalance estimates from registries. Biometrics 2002; In press.
25. Krogh V, Micheli A. Measure of cancer prevalence with a computerized program: an example of larynx cancer. Tumori 82; 1966: 1–4.
26. Brown M, Riley GF, Potosky AL, Ertzioni RD. Obtaining long-term disease specific cost of care. Application to Medicare enrollees diagnosed with colorectal cancer. Med Care 1999; 37: 1249–1259.