Cancer Epidemiology 37 (2013) 850–856

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Methodological aspects of estimating rare cancer prevalence in Europe: The experience of the RARECARE project

Sandra Mallone [a,*], Roberta De Angelis [a], Jan Maarten van der Zwan [b], Annalisa Trama [d], Sabine Siesling [b,c], Gemma Gatta [d], Riccardo Capocaccia [a] The RARECARE WG

[a] Department of Cancer Epidemiology, Istituto Superiore di Sanità, Rome, Italy
[b] Department of Research, Comprehensive Cancer Centre The Netherlands, Utrecht, The Netherlands
[c] Health Technology and Services Research (MIRA), University of Twente, Enschede, The Netherlands
[d] Evaluative Epidemiology Unit, Fondazione IRCSS, Istituto Nazionale dei Tumori, Milan, Italy

ARTICLE INFO

Article history:
Received 12 March 2013
Received in revised form 2 August 2013
Accepted 3 August 2013
Available online 16 September 2013

Keywords:
Rare cancer
Completeness index
Complete prevalence
Incidence
Survival
Statistical model quality assessment

ABSTRACT

This paper describes the usage and the performance evaluation of the completeness index method in the 'Surveillance of Rare Cancers in Europe project' (RARECARE) for estimating rare cancer prevalence in Europe. The 15-year prevalence at 1st January 2003 for 255 cancers is obtained from a pool of 22 RARECARE cancer registries (CRs). Incidence and survival models are applied to the RARECARE database to estimate the parameters from which the completeness indices are calculated. Complete prevalence is obtained adjusting the observed 15-year prevalence by the completeness index, to account for those cancer survivors diagnosed before the CR activity started. Main factors influencing the performance of the completeness index method for rare cancers are the same as for common cancers: age distribution of incidence and lethality of the cancer. For cancers occurring in the elderly, with low survival rates and consequently a restricted number of long-term survivors we obtained completeness indices higher than 0.9. Values lower than 0.7 correspond to those cancers with good prognosis and/or incidence more concentrated at the younger ages, indicating that 15 years of follow up are insufficient to detect all prevalent cases. Validation analysis shows that for a restricted subgroup of rare cancers with very low incidence and low survival, the completeness indices were not able to adequately correct the observed prevalence even considering a registration period of 20 years. On average, sensitivity analyses show a slight overestimation of complete prevalence for rare and common cancers whose increasing incidence is known in literature. RARECARE is the largest project on rare cancers conducted to date. Improving health care programs for cancer survivors is a public health priority and prevalence data which provides important information in this field should be regularly asked to Member States and included in the EU health statistics.

1. Introduction

Cancer prevalence is the count or proportion of cancer patients alive in a population at any given time, regardless of the date of diagnosis. Different methods are available to obtain cancer prevalence estimates. Direct methods employ population-based cancer registries (CRs) data to obtain the observed prevalence by enumerating how many incident cases are alive at a certain reference date [1–3]. The major problem arising when direct methods are used is the impossibility of accounting for those survivors diagnosed before the CR registration activity started, the so called 'unobserved' prevalence. The resulting underestimation in the prevalence depends on the length of the registration period and on the mean survival time of cancer patients varying by cancer site. Indirect methods estimate complete prevalence by applying statistical techniques to CRs data and/or health statistics [4–6]. The completeness index method is an alternative approach which allows an adjustment of the observed prevalence by a correction factor obtained modelling the prevalence as a combination of incidence and survival functions [7–10]. Aim of this paper is to describe how the completeness index method was applied to the 'Surveillance of Rare Cancers in Europe project' (RARECARE) database [11] for estimating the rare cancer prevalence in Europe

* Corresponding author at: Department of Cancer Epidemiology, National Centre for Epidemiology Surveillance and Health Promotion (CNESPS), Istituto Superiore di Sanità, Viale Regina Elena, 299, 00161 Rome, Italy. Tel.: +39 06 49904295; fax: +39 06 49904285.
E-mail address: sandra.mallone@iss.it (S. Mallone).

1877-7821/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.canep.2013.08.001

and how the method performance was evaluated. The focus is on methodological aspects, difficulties faced dealing with rare cancers and solutions adopted.

2. Materials and methods

2.1. Definition of terms

The observed prevalence – i.e. observed limited-duration prevalence (L-year prevalence) – is the proportion of cancer patients in a population diagnosed at age x within a given time period (L) and who are still alive at a certain reference date. It can be directly calculated ($N_O(x,L)$) in a CR active since L years before or estimated ($\hat{N}(x,L)$) on the basis of the mathematical relationship between prevalence, incidence and survival.

The 'unobserved' prevalence $N_U(x,L)$ is the proportion of cancer patients in a population aged x who are expected to be alive at a certain reference date, provided that they were diagnosed before the start of the CR registration activity. By definition, it is not observable and can only be estimated on the basis of the mathematical relationship between prevalence, incidence and survival.

Complete prevalence $N_C(x)$ is the proportion of cancer patients in a population aged x, who received a diagnosis of cancer at any age t ($t < x$) and are still alive at a certain reference date. $N_C(x)$ can be decomposed into the sum of $\hat{N}(x,L)$ and $N_U(x,L)$, and estimated on the basis of the mathematical relationship between prevalence, incidence and survival as follows:

$$N_C(x) = N_U(x,L) + \hat{N}(x,L) = \int_0^{x-L} I(t)\cdot S(t,x-t)\mathrm{d}t + \int_{x-L}^{x} I(t)\cdot S(t,x-t)\mathrm{d}t \quad (1)$$

where $I(t)$ is the incidence function for patients aged t, $S(t, x-t)$ is the relative survival function at age x for a person who has received a cancer diagnosis at age t ($t < x$). The incidence and survival functions in Eq. (1) can be expressed in a parametric form and parameters estimates can be obtained from the CRs data. For a more detailed evaluation of the above equation refer to Capocaccia & De Angelis [7].

2.2. The completeness index method

The completeness index [7–10] for cancer patients aged x and a length of registration activity equal L, is a measure of the estimated prevalence $\hat{N}(x,L)$ relative to the complete prevalence and is defined as:

$$R(x,L) = \frac{\hat{N}(x,L)}{N_C(x)} = \frac{\int_{x-L}^{x} I(t)\cdot S(t,x-t)\mathrm{d}t}{\int_0^x I(t)\cdot S(t,x-t)\mathrm{d}t} \quad (2)$$

A completeness index close to 0 indicates a large incompleteness of the prevalent cases directly observed by a CR. On the contrary, a completeness index close to 1 indicates a small incompleteness of the observed prevalence. This is the case for highly lethal cancer sites and long-established CRs which observe the great majority of all prevalent cases. From Eq. (2), complete prevalence can be obtained by dividing the observed limited-duration prevalence by the modelled completeness index:

$$N_C(x) = \frac{N_O(x,L)}{R(x,L)} \quad (3)$$

The completeness index method was applied to the RARECARE database considering the first primary cancer, for obtaining estimates of complete cancer prevalence in the following four steps:

(1) Two alternative parametric forms (polynomial and exponential) of the incidence data [12–14] and a set of cure-models [15] of the survival data are assumed. The model parameters are obtained by cancer site from the RARECARE database (Supplementary material) using the SAS software [16]. Incidence and survival data are calculated using SEER*STAT software [17] stratified by period of diagnosis (five-year calendar period from 1985 to 1999 for incidence, and three-year calendar period from 1988 to 1999 for survival), and age at diagnosis (0–4, ..., 85–99). Relative survival is calculated according to the Hakulinen method [18]. For the cohort of patients diagnosed in 1988–1999 and followed-up to 31st December 2003, cumulative relative survival curves are constrained to not increase over follow-up time.

(2) The 15-year prevalence at the reference date of 1st January 2003 by cancer site and age at prevalence date (0–4, ..., 75–99) is obtained from the RARECARE database with SEER*STAT software [17] using the counting method [1–3]. According to this approach, lost to follow-up cases taken into account by estimating their survival time from the life tables of the other patients successfully followed-up, and matched by age, sex and study period. Since only the first primary of each cancer is considered, we estimated the prevalence of *persons* with cancer and not the prevalence of *cancers*.

(3) The completeness indices $R(x, 15)$ by cancer site, age (0–4, ..., 75–99) and sex are calculated using the COMPREV software[19] on the basis of the incidence and survival parameter estimates (intercept, b or b_1, $\cdots$, b_6 for the exponential or polynomial model respectively; $1-P$, λ, γ, β_1, β_2), reference year and age for the survival models together with the elements of the parameters covariance matrix.

(4) The complete prevalence at 1st January 2003 for each of the cancers of the RARECARE list, is derived as the ratio between the observed 15-year prevalence and the corresponding completeness index as follows:

$$N_C = \sum_x N_C(x) = \sum_x \frac{N_O(x;15)}{R(x;15)} \quad x = 0\text{–}4, 5\text{–}9, \ldots, 75\text{–}99 \quad (4)$$

2.3. Cancer sites and CRs data

The RARECARE project provided a list of 260 cancers organized in 3 tiers: the tier-3 corresponds to the WHO names of a cancer, and their respective ICD-O-3 morphologies and topography codes. These cancers were then grouped into 203 tier-2 cancers which have to be considered as the reference entities for clinical management and research. The tier-2 cancers were finally grouped into tier-1 cancers, which are more general categories considered to require the same clinical expertise and patient referral structure. According to an annual crude incidence less than 6/100,000/year, 186 and 17 tier-2 cancers of the list turned out to be rare and common respectively [11]. In this work, we considered the whole RARECARE list of cancers, regardless their rarity. The definition of gastrointestinal stromal tumour and blastic plasmocytoid dendritic cell neoplasm was too recent for the prevalence estimation, while for pancreatoblastoma, myelodysplastic syndrome with 5q syndrome, carcinoma with osteoclast-like giant cells of pancreas, incidence and survival data were too sparse to be modelled. Thus, these 5 cancers were excluded by the analysis. The 15-year prevalence was calculated using those 22 CRs (Iceland, Norway, Sweden, Austria, Saarland, Amsterdam, Geneva, St. Gallen, Cracow, Slovakia, Firenze, Modena, Parma, Ragusa, Romagna, Slovenia, East Anglia, Northern and Yorkshire, Oxford, West Midlands, Scotland, Wales) having in the year 2003 at least 15 years of incidence and survival data. These CRs represent 12% of the total European

population [20]. In addition, 25 other CRs[1] with the last incidence year in 1999 or after are considered for incidence modelling, and 24 other CRs[2] with last incidence year between 1997 and 1999 and followed-up to 2002, are considered for survival modelling.

3. Evaluation analysis of complete prevalence

We performed two types of evaluation of the complete prevalence results using the RARECARE database.

3.1. Validation analysis

In order to validate the complete prevalence results we considered the observed prevalence calculated using the long-established CRs of Norway and Scotland (national wide, year of incidence in the period 1978–2002). Only cancers with at least 10 prevalent cases during the period 1978–2002 were analyzed. The limited-duration prevalence $\hat{N}(x, L)$ at a certain reference date can be estimated from Eq. (2) by truncating the complete prevalence to l years ($l < L$) as follows:

$$\hat{N}(x, L) = N_C R(x, L) = \left[\frac{N_O(x, l)}{R(x, l)} R(x, L)\right] \quad (5)$$

We calculated the observed 25-year prevalence $N_O(x, 25)$ at 1st January 2003 using Norway and Scotland data with the counting method [1–3]. We then obtained three different estimates of the 25-year prevalence $\hat{N}(x, 25)$ from Eq. (5) with $L = 25$ and $l = 10, 15, 20$ years. The $R(10)$, $R(20)$, $R(25)$ completeness indices are calculated on the basis of the same incidence and survival parameter estimates used for $R(15)$. Results are compared in terms of the absolute percentage differences between $N_O(x, 25)$ and $\hat{N}(x, 25)$:

$$\left|\frac{N_O(x, 25) - \hat{N}(x, 25)}{N_O(x, 25)}\right| \times 100 \quad (6)$$

3.2. Sensitivity analysis

We performed two sensitivity analyses of the complete prevalence estimates to changes in the registration period length considering the pool of the 22 RARECARE CRs. Only cancers with at least 10 observed prevalent cases were considered. We obtained the observed 15-year prevalence $N_O(x, 15)$ from the pool of 22 CRs with the counting method [1–3]. We then estimated the 15-year prevalence $\hat{N}(x, 15)$ at 1st January 2003 by truncating the complete prevalence to 15 years of estimated prevalence as in Eq. (5) using the completeness indices corresponding to $l = 5$ and $L = 15$ years of registration lengths. $R(5)$ is calculated on the basis of the same incidence and survival parameter estimates used for $R(15)$. Results are compared in terms of the absolute percentage differences between $N_O(x, 15)$ and $\hat{N}(x, 15)$:

$$\left|\frac{N_O(x, 15) - \hat{N}(x, 15)}{N_O(x, 15)}\right| \times 100 \quad (7)$$

Suppose that the complete prevalence estimates are independent of the CR registration period length, then we have:

$$[N'_C - N_O(x; L)][\hat{N}(x; L) - N_O(x; l)] = [N_C - N_O(x; L)] \quad [N_O(x; L) - N_O(x; l)] \qquad 1 < L \quad (8)$$

From Eq. (8), we obtained from the pool of 22 CRs the estimate of the complete prevalence for $l = 5$ and $L = 15$ as follows:

$$N'_C = N_O(x; 15) + [N_C - N_O(x; 15)]\left[\frac{N_O(x; 15) - N_O(x; 5)}{\hat{N}(x; 15) - N_O(x; 5)}\right] \quad (9)$$

where N_C is the complete prevalence calculated adjusting the 15-year observed prevalence by the corresponding completeness index $R(x, 15)$, $N_O(x, 5)$ and $N_O(x, 15)$ are the observed 5 and 15-year prevalence respectively, $\hat{N}(x, 15)$ is the estimated 15-year prevalence calculated by truncating the complete prevalence to 15 years of estimated prevalence as in Eq. (5).

Results are compared to the complete prevalence N_C in terms of the absolute percentage differences:

$$\left|\frac{N_C - N'_C}{N_C}\right| \times 100$$

4. Results

4.1. Completeness indices

Fig. 1 shows the $R(15)$ expressed as a percentage for the tier-2 rare and common cancers. The median values of completeness index are in the category 80–90 for common cancers (81%) and 70–80 for rare cancers (73%) sites, with interquartile range (IQRs) of 15% and 26% respectively. Fig. 2 shows the tier-1 cancers with $R(15)$ higher than 70%. The following cancers had an index higher than 80%: myelodisplastic syndrome, myeloproliferative neoplasm, malignant mesothelioma, and epithelial tumour of colon, trachea, stomach, rectum, oropharynx, lung, gallbladder and extrahepatic bile tract, oesophagus, pancreas, prostate and liver and intrahepatic bile tract. Fig. 3 shows cancers with $R(15)$ lower than 70%. The epithelial tumour of cervix uteri and corpus uteri, testicular and paratesticular cancers, malignant skin melanoma, showed completeness indices ranging from 47% to 70%.

4.2. Validation analysis

Fig. 4a–d report the results of the validation analysis for rare and common cancers separately. Each bar represents the absolute percentage differences between the observed and estimated 25-year prevalence based on different length of the registration period (10, 15, 20) for Norway and Scotland. In Norway the average of the percentage differences on rare cancers are 19, 11 and 3 for $R(10)$, $R(15)$, $R(20)$ respectively. In Scotland are 14, 11 and 4 for $R(10)$,

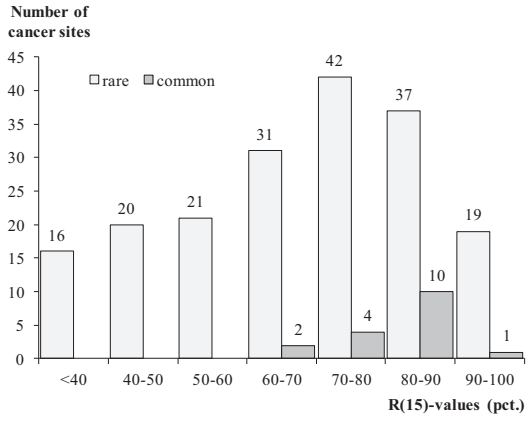


Fig. 1. $R(15)$-values for the tier-2 cancers of the RARECARE list by rarity.

[1] Flanders, Eindhoven, North Netherlands, Stedendriehoek-Twente, Basel, Ticino, Kielce, Warsaw, Alto Adige, Biella, Ferrara, Friuli V.G., Genova, Napoli, Palermo, Reggio Emilia, Salerno, Sassari, Trento, Umbria, Veneto, Albacete, Castellon, Girona, Northern Ireland.

[2] Bas Rhin, Calvados, Calvados digestive, Cote D'Or digestive, Cote D'Or hematologic, Doubs, Haut Rhin, Isére, Somme, Tarn, Eindhoven, Basel, Grisons, Valais, Zurich, Warsaw, Geneva, Varese, Veneto, Basque Country, Navarra, Tarragona, South Western, Trent.

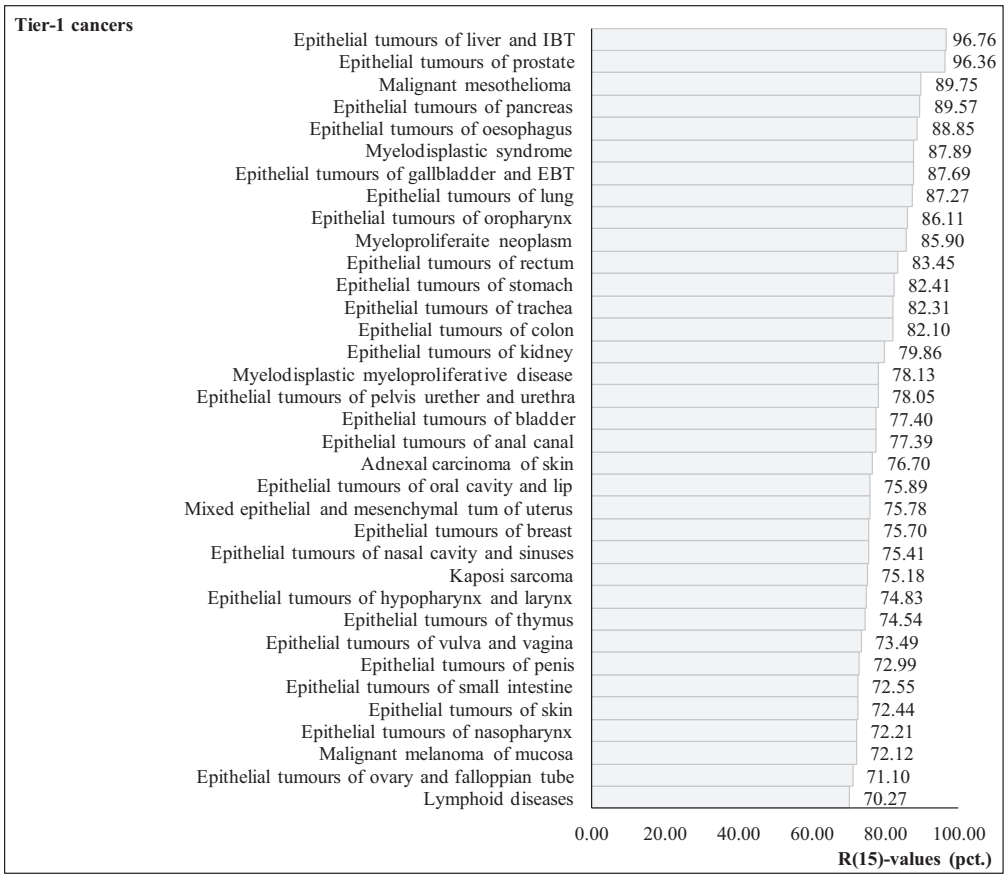


Fig. 2. *R*(15)-values higher than 70%. Tier-1 cancers of the RARECARE list.

R(15), *R*(20) respectively. Discrepancies higher than 15 (range = 16–100) are obtained in 8 and 21 rare cancers for the registry of Norway and Scotland respectively with incidence between 0.01 and 0.98/100,000/year and 5-year relative survival between 6.6% and 65.7%[11]: the undifferentiated carcinomas of stomach, oesophagus and lung, the soft tissue sarcoma of head and neck, paraorbital region and mediastinum, the squamous cell carcinomas and variant of gallbladder and extrahepatic bile tract, kidney, colon, corpus uteri, prostate. For common cancers, the average of the percentage differences considering *R*(10), *R*(15) and *R*(20) is 15 (range = 0–65), 8 (range = 0–27) and 3 (range = 0–9) for the registry of Norway, and for the registry of Scotland is 11 (range = 1–26), 5(range = 0–13) and 2 (range = 0–5).

4.3. Sensitivity analysis

Results of the sensitivity analysis are reported in Fig. 5a–d showing the observed and estimated 15-year prevalence for rare and common cancers separately, obtained on the base of the pool of 22 RARECARE CRs. On average, the differences are 11 percentage points for common and 20 for rare cancers respectively. The highest discrepancies (>20) are obtained for 52 rare cancers, most of them (69.0%) with less than 1.000 observed cases. Comparisons between the different estimates of complete prevalence are shown in Fig. 6. On average, the percentage discrepancies is 10 (SD 14) and the complete prevalence N'_C tends to be higher than N_C for the majority of the cancers (157 cancer entities). For 28 cancer sites (marked black in Fig. 5) we obtained an absolute difference greater than 20. Among these there are 4 tier-1 rare cancers (trophoblastic tumour of placenta, epithelial tumour of middle ear, glial tumour of the nervous autonomic nervous system and paraganglia, nonglial tumour of the central nervous system and pineal gland) with 4 of the corresponding tier-2 cancers. The remaining cancers are mainly soft tissue sarcomas of different sites (mediastinum, heart and pelvis, paraorbital region), squamous cell carcinomas (of pancreas, middle ear), the mast cell tumours, retinoblastoma and epatoblastoma.

5. Discussion

The completeness index method has been previously applied to CR data in Europe and USA to estimate complete prevalence for common cancers [10,21–23]. However, dealing with rare cancers implies a great difficulty in statistical modelling of CRs data and in computation of completeness index because of the low number of patients. This work is thus necessary to report the performance evaluation results of the first application of the completeness index method to obtain complete prevalence of rare cancers in Europe. First, we performed a validation analysis of the results using the data of some of the CRs included in the estimation process, e.g. Norway and Scotland that have been in operation for a long time. Second, we evaluated how sensitive are the complete prevalence estimates to the assumption that a registration length of 15 years allows to detect all the prevalent cases, considering the pool of the 22 RARECARE CRs used in the estimation process. For details on the modelling problems faced and solutions adopted, the reader should refer to Supplementary material.

For rare cancers, main factors influencing the performance of the completeness index method are the same as for common cancers: age distribution of incidence and lethality of the cancer. For rare cancers occurring mainly in the elderly, with low survival rates and consequently a restricted number of long-term survivors,

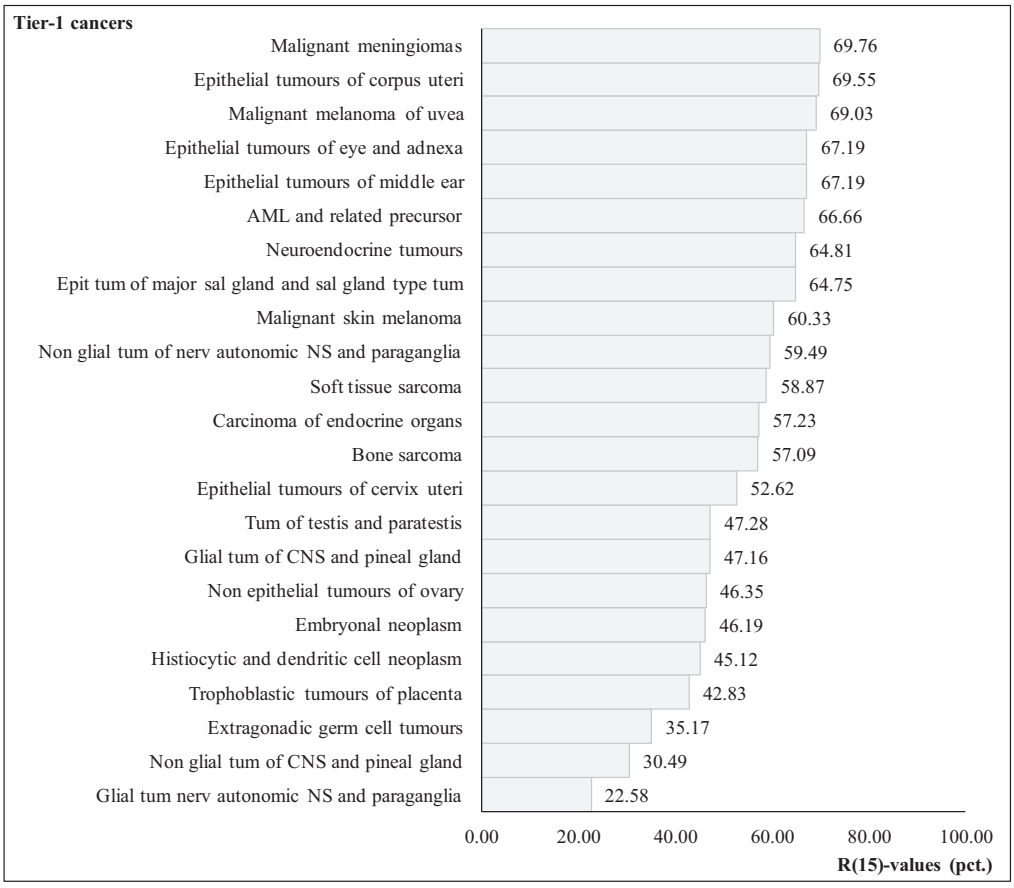


Fig. 3. *R*(15)-values lower than 70%. Tier-1 cancers of the RARECARE list.

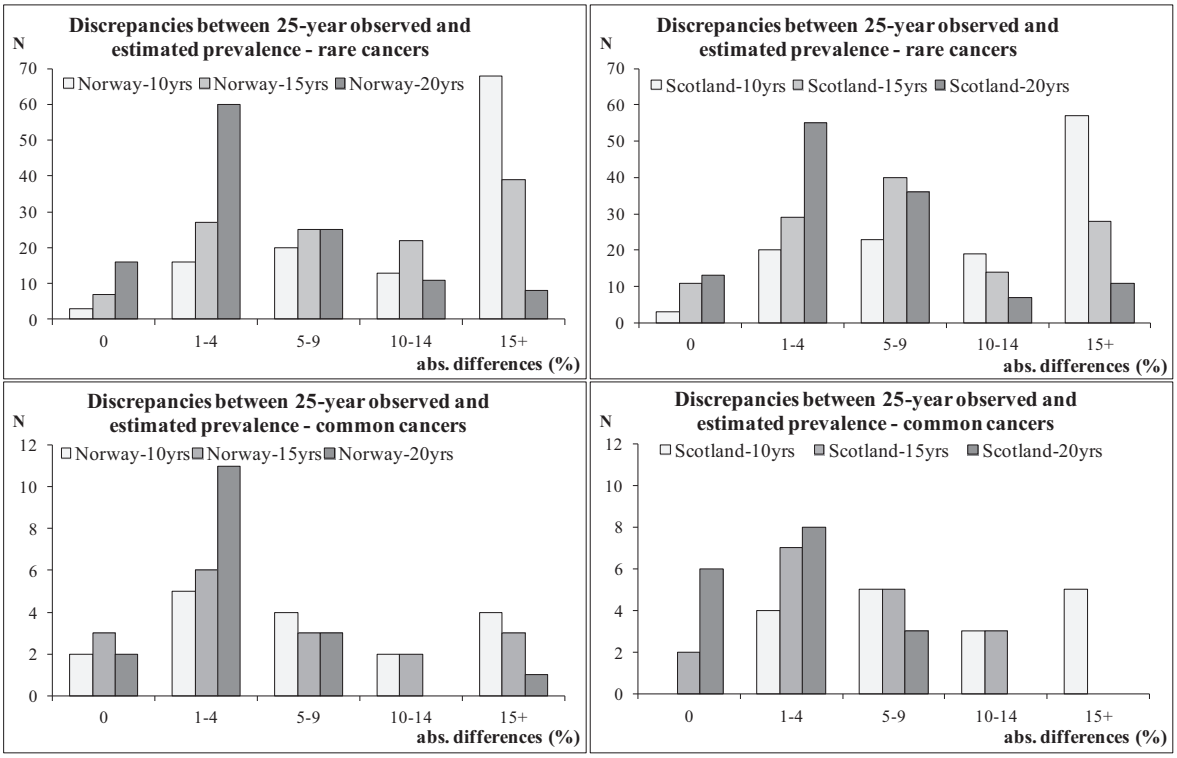


Fig. 4. Absolute percentage differences between the 25-year observed and estimated prevalence for the RARECARE list of cancers by rarity and length of the registration period (10, 15, 20) for Norway and Scotland.

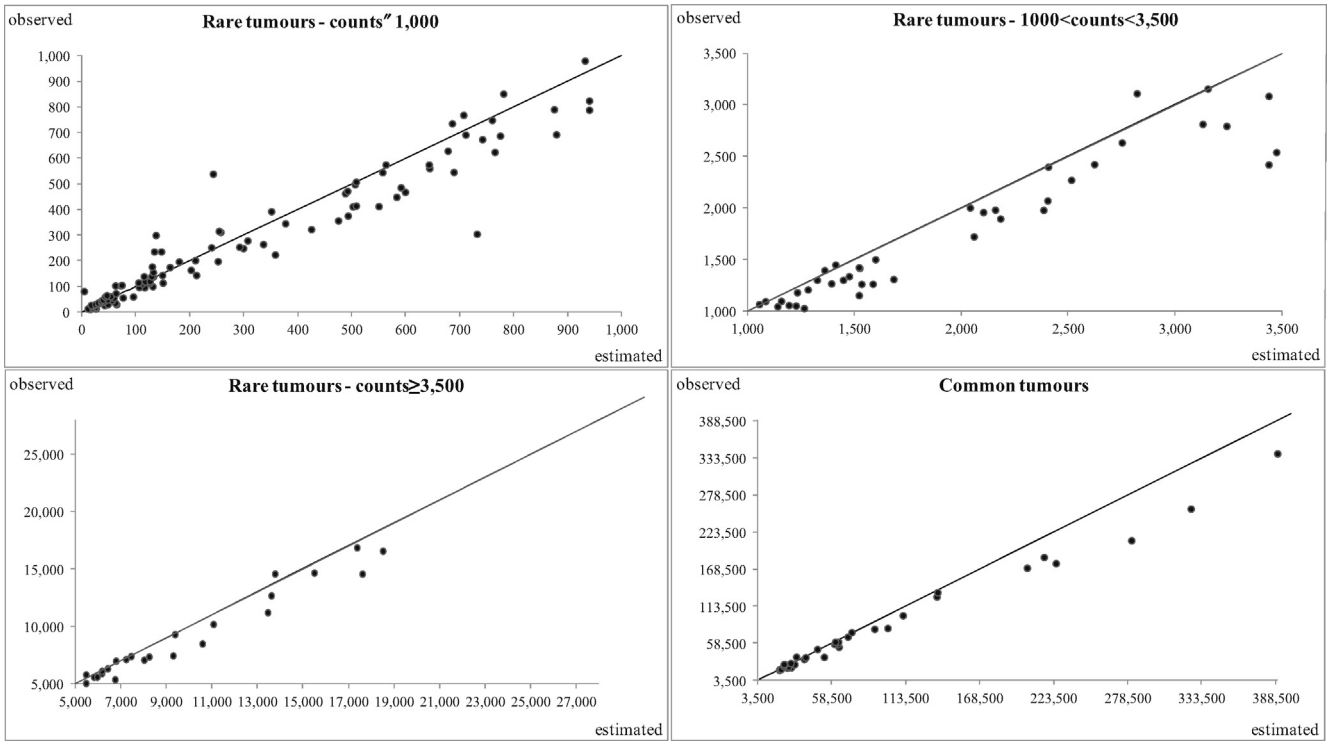
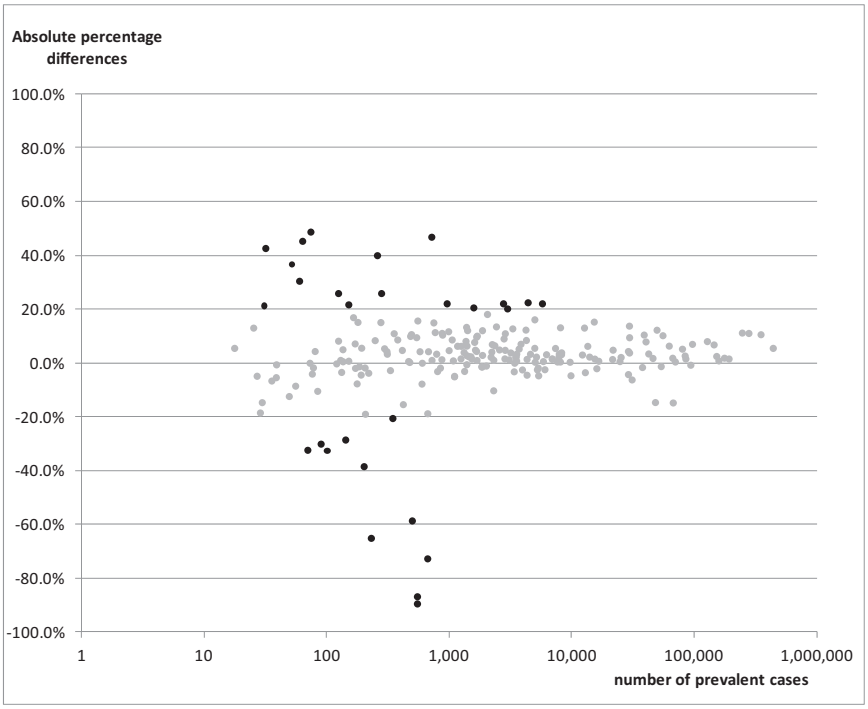


Fig. 5. (a–d) Observed and estimated 15-year prevalence for rare and common cancers of the RARECARE list.

such as cancers of pancreas, liver and IBT, gallbladder and EBT [11] the completeness index is above 0.9. For rare cancers with good prognosis (testicular and paratesticular cancers) and/or incidence more concentrated at the younger ages (tumours of cervix and corpus uteri) [11], the completeness index is below 0.7 indicating that 15 years of follow up are insufficient to detect all prevalent cases. As expected, validation analysis for Norway and Scotland CRs shows that difference between observed and estimated 25-year prevalence decreases with increasing registration length. For a restricted subgroup of rare cancers with low incidence and survival, the completeness indices were not able to adequately correct the observed prevalence even considering a registration period of 20 years. Sensitivity analysis of complete prevalence showed, as expected, results in agreement with literature [24]: a



In black, cancer sites with discrepancies greater than 20.

Fig. 6. Discrepancies between the estimates of complete prevalence based on different length of registration period by number of prevalent cases. RARECARE cancer sites with more than10 observed prevalent cases in 1998–2002. In black, cancer sites with discrepancies greater than 20.

negative difference (underestimation) for tier-1 common cancers with decreasing incidence over time such as epithelial tumour of stomach (6.2), cervical (14.8) and male lung cancers (1.3), and a positive difference (overestimation) for cancers with increasing incidence over time like epithelial tumour of prostate (1.6), skin (10.6), of breast (5.5), of colon (0.9), and for skin melanoma (8.0). This is also the case for the tier-2 rare cancers pertaining to the above tier-1 cancers. We did not include the time covariate in incidence models for two reasons. First, the number of cases is not sufficient to be explored by a trend for most rare cancers considered in the analysis. Second, due to the continuous improvement of the registration practices and to the recent introduction of the ICD-O-3 classification which started being used on a large scale from 2000 on, a trend analysis risks to detect improvement in coding practices instead of temporal variation in incidence rates. This could be the case of some ovarian cancers of borderline malignancy which changed from invasive, malignant behaviour (code 3) in ICD-O-2 to uncertain behaviour (code 1) in ICD-O-3 contributing to a decrease in incidence in Europe since 2000 [25].

Despite the high quality and completeness of the RARECARE database [20] some differences in the accurateness of cancer diagnosis, the completeness of cancer patients registration and follow-up can occur across European CRs included in the present work. These differences might result in a biased estimation of both incidence and survival of cancer, therefore we performed an additional sensitivity analysis of the completeness indices to changes of $\pm 20\%$ in relative survival and incidence parameters to check the robustness of the $R(15)$-values (Supplementary material).

The pool of 22 CRs used in the calculation of the observed 15-year prevalence is a fairly representative sample of the total European population [20]. Due to the scarcity of epidemiological data on rare cancers, it is not possible to assess to what extent this assumption may be true. However, comparisons between incidence for common cancers obtained on the base of the RARECARE database [11] and those reported by GLOBOCAN [26] considered the best available, suggest that this assumption is acceptable. The evaluation analysis of complete prevalence and the sensitivity analysis of the indices for possible variations in survival and incidence levels performed in this work represent an important experience for further prevalence studies. The procedures used in these analyses could be implemented in the Comprev software to achieve the greatest possible robustness and generalizability of results. RARECARE is the largest project on rare cancers conducted to date. This project provided estimates of complete prevalence for rare cancers in Europe, a measure which represents a major health problem but it has not been adequately estimated. If the existing European definition of rare diseases were used (prevalence $< 50/100{,}000$), rare cancers would be 24% of total cancer prevalence as estimated by RARECARE. Interested readers should refer to Gatta et al. [11] for an in-depth description of prevalence estimates for rare and common cancers defined by RARECARE (Table 2, page 2496). Prevalence has been included in the priority list of Cancer Health indicators thank to the ongoing European Partnership for Action Against Cancer [27] and the European Cancer Health Indicator Project [28]. Improving health care programs for cancer survivors is a public health priority and prevalence data which provides important information in this field should be regularly asked to Member States and included in the EU health statistics.

Acknowledgments

This research was supported by the European Commission through the Executive Agency for Health and Consumers (Grant No. 2006113). Sincere thanks are due to Dr. Silvia Francisci (Department of Cancer Epidemiology, Istituto Superiore di Sanità) for her valuable advice and thoughtful participation in the discussions during the revision of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.canep.2013.08.001.

References

[1] Feldman AR, Kessler L, Myers MH, Naughton MD. The prevalence of cancer. Estimates based on the Connecticut tumor registry. N Engl J Med 1986;315:1394–7.
[2] Hakama M, Hakulinen T, Teppo L, Saxen E. Incidence, mortality and prevalence as indicators of the cancer problem. Cancer 1975;36:2227–31.
[3] Adami HO, Gunnarson T, Sparén P, Eklund G. The prevalence of cancer in Sweden 1984. Acta Oncol 1989;28:463–70.
[4] Gail MH, Kessler L, Midthune D, Scoppa S. Two approaches for estimating disease prevalence from population-based registries of incidence and total mortality. Biometrics 1999;55(4):1137–44.
[5] Colonna M, Hedelin G, Estève J, et al. National cancer prevalence estimation in France. Int J Cancer 1987;2000:301–4.
[6] 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;8:201–16.
[7] Capocaccia R, De Angelis R. Estimating the completeness of prevalence based on cancer registry data. Stat Med 1997;16:425–40.
[8] Corazziari I, Mariotto A, Capocaccia R. Correcting the completeness bias of observed prevalence. Tumori 1999;85:370–81.
[9] Capocaccia R, Colonna M, Corazziari I, R.De Angelis, Francisci S, Micheli A, et al. Measuring cancer prevalence in Europe: the EUROPREVAL project. Ann Oncol 2002;13(6):831–9.
[10] Merrill RM, Feuer EJ, Capocaccia R, Mariotto A. Cancer prevalence estimates based on tumor registry data in the SEER program. Int J Epidemiol 2000;29:197–207.
[11] Gatta G, van der Zwan JM, Casali PG, Siesling S, Dei Tos AP, Kunkler I, et al. Rare cancers are not so rare: the rare cancer burden in Europe. Eur J Cancer 2011;47:2493–511.
[12] Pearce N. What does the odds ratio estimate in a case-control study? Int J Epidemiol 1993;22:1189–92.
[13] Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 1954;8:1–12.
[14] Cook PJ, Doll R, Fellingham A. A mathematical model for the age distribution of cancer in man. Int J Cancer 1969;4:93L 112.
[15] De Angelis R, Capocaccia R, Hakulinen T, et al. Mixture models for cancer survival analysis: application to population-based data with covariates. Stat Med 1999;18:441–54.
[16] SAS software, version 9.1.3. Cary, NC, USA: SAS – Statistical Analysis System Institute Inc, 2003.
[17] NCI. Seer*Stat – surveillance, epidemiology, and end results software. Program of the National Cancer Institute; 2010, Version 7.0.2 http://seer.cancer.gov/seerstat/.
[18] Hakulinen T. Cancer survival corrected for heterogeneity in patient withdraw-al. Biometrics 1982;38:933–42.
[19] NCI. ComPrev – complete prevalence software. Program of the National Cancer Institute; 2005, Version 1.0 http://surveillance.cancer.gov/comprev/.
[20] De Angelis R, Francisci S, Baili P, et al. The EUROCARE-4 database on cancer survival in Europe: data standardization, quality control and methods of statistical analysis. Eur J Cancer 2009;45:909–30.
[21] Forman D, Stockton D, Møller H, Quinn M, Babb P, De Angelis R, et al. Cancer prevalence in the UK: results from the EUROPREVAL study. Ann Oncol 2003;14(4):648–54.
[22] Capocaccia R, Colonna M, Corazziari I, De Angelis R, Francisci S, Micheli A, et al. Measuring cancer prevalence in Europe: the EUROPREVAL project. Ann Oncol 2002;13:831–9.
[23] Möller T, Anderson H, Aareleid T, Hakulinen T, Storm H, Tryggvadottir L, et al. Cancer prevalence in Northern Europe: results from the EUROPREVAL study. Ann Oncol 2003;14(4):946–57.
[24] Cancer incidence in five continents annual dataset; 2013, http://ci5.iarc.fr/CI5plus/ci5plus.htm (accessed 20.05.13).
[25] Ovarian cancer incidence statistics; 2013, http://www.cancerresearchuk.org/cancer-info/cancerstats/types/ovary/incidence/ (accessed 25.05.13).
[26] Ferlay J, Bray F, Pisani P, Parkin DM. GLOBOCAN 2002 cancer incidence, mortality and prevalence worldwide IARC CancerBase 2004 No. 5, version 2.0. Lyon: IARC Press, 2004.
[27] The European partnership for action against cancer (EPAAC); 2013, http://www.epaac.eu/ (accessed 4.05.13).
[28] European cancer health indicator project; 2013, http://www.tumori.net/eurochip/ (accessed 4.05.13).