Cancer Epidemiology 56 (2018) 97–105

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

Time trends and short term projections of cancer prevalence in France

M. Colonna[a,b,*], O. Boussari[c], A. Cowppli-Bony[b,d], P. Delafosse[a,b], G. Romain[c], P. Grosclaude[b,e], V. Jooste[b,c], French Network of Cancer Registries (FRANCIM)[1]

[a] Isere Cancer Registry, CHU Grenoble-Alpes, F-38043, Grenoble, France
[b] FRANCIM Network, F-31073, Toulouse, France
[c] Digestive Cancer Registry of Burgundy, INSERM UMR 1231, University Hospital of Dijon, University of Burgundy, F-21079, Dijon, France
[d] Loire-Atlantique Vendée Cancer Registry, F-44093, Nantes, France
[e] Tarn Cancer Registry, Claudius Regaud Institute, IUCT-O, F-31059, Toulouse, France

ARTICLE INFO

Keywords:
Prevalence
Short-term projections
Projection scenarios
Time trends
Flexible models

ABSTRACT

Background: This study analyzes time trends in cancer prevalence in France and provides short-term projections up to the year 2017. The 15-year prevalence for 24 cancers was estimated from the French cancer registries network (FRANCIM) incidence and survival data.

Method: We estimated prevalence using the $P = I \times S$ relationship, with flexible modeling of incidence and survival. Based on observations of the incidence and survival up to 2010, different scenarios for evolution up to 2017 were studied, combining stable and dynamic incidence and survival. The determinants of variations in prevalence (incidence, survival and demography) were quantified.

Results: At the end of 2017, an estimated 1,396,000 men and 1,359,000 women having had cancer in the previous 15 years were alive, respectively 5.4% and 4.8% of the population. Twelve percent had been diagnosed in the preceding year and 23% between 10 and 15 years. Between 2010 and 2017, changes in incidence and survival depended on the cancer site. The effect of the demographic change was null for those under age 65, whereas above age 65, the contribution of this factor was 20% in men and 17% in women at 15 years. The different projection scenarios led to very different estimates for some cancers for which incidence strongly varied in the last decades.

Conclusion: Prevalent cases are numerous in a country such as France, where incidence and survival are high. Due to the sensitivity of prevalence to changes in incidence and survival, we recommend that the results of projections are presented under different scenarios. We propose a robust and flexible prevalence estimate.

1. Introduction

Prevalence, estimating the number of people who have had cancer within a given delay since diagnosis, allows an estimation of the needs for therapeutic care and monitoring. Prevalence also gives information on the number of people who may suffer prolonged psychological and physical health conditions or experience social difficulties such as problems in professional (re)insertion, or in obtaining loans from banks, etc. [1–5]. While the complete prevalence counts the number of people alive having a diagnosis of cancer, whatever the delay since the diagnosis, the partial prevalence [6] breaks up the whole group according to the length of this delay.

In this article, we focus on partial prevalence measured over the 15 years following the diagnosis. Using a perspective of up to 15-years period makes it possible to quantify the importance of different groups defined according to the delay since the diagnosis and in particular to distinguish persons still in the course from initial treatment and those who can be considered as cured [2,7–10].

Partial prevalence can be obtained by counting incident cases still alive at a given date [2,4,6,8,9,11]. This method requires a follow-up of incident cases, the prevalence depending on both the new cases of the disease (inflow) and survival [12]. Prevalence estimates are therefore often based on old data. Partial prevalence has also been estimated from models that include age, period and cohort effects for incidence [13] and parametric models for relative survival [14]. The prevalence estimate for recent years (i.e. for the current year) must take into account changes in incidence and survival, and sometimes make assumptions about the evolution of these indicators.

* Corresponding author at: Registre du cancer de l'Isère, Pavillon E – CHU Grenoble, BP 217, F-38043, Grenoble Cedex 9, France.
E-mail address: mcolonna.registre@wanadoo.fr (M. Colonna).
[1] See Appendix A.

https://doi.org/10.1016/j.canep.2018.08.001
Received 6 April 2018; Received in revised form 7 August 2018; Accepted 8 August 2018

The objective of this article is to estimate the partial prevalence (from 1 to 15 years) of adults with cancer (age ≥15 years at diagnosis) in France at the end of 2017, from the French registries' data using flexible parametric modeling of both incidence [15] and overall survival [16]. Based on observations of the incidence and survival up to 2010, different scenarios of evolution up to 2017 are studied. Variations in prevalence between 2010 and 2017, and their determinants of this variation (incidence, survival and demography) are quantified.

2. Materials and methods

The study includes all incident cases of cancer (except non-melanoma skin cancers) as described by Belot et al. [15] and collected by 22 population-based cancer registries participating in the French network of cancer registries (Francim) common database. In 2013, the last year of incidence available at the time of the study, these registries covered 17% of the metropolitan population (11 million people).

More than 530,000 incident cases diagnosed between 1989 and 2010 have been followed-up for vital status until June 2013. The proportion of lost to follow-up was 2.1% for solid tumors [17] and 1.9% for hematologic malignancies [18]. According to these publications, several cancers (acute leukemia, central nervous system, esophagus, lung, pancreas, and liver cancer) had poor prognosis (5-year age-standardized net survival < 33%). In contrast, very good prognosis (age-standardized net survival ≥80% at 5 years) was found for Hodgkin lymphoma, skin melanoma, prostate, breast, thyroid, and testis cancers. Other cancers were considered as having good prognosis (66% ≤ age-standardized net survival < 79%) or having moderate prognosis (33% ≤ age-standardized net survival < 66%) [17,18].

The demographic data, coming from "Omphale" projections provided by INSEE (National Institute of Statistics and Economic Studies) were used to estimate the number of incident and prevalent cases.

2.1. General principle of prevalence estimation

The partial prevalence (P) has been estimated from a combination of incidence (I) and survival (S) values using the relationship $P = I \times S$ [19]. More specifically, prevalence at the end of year y was calculated as a follow-up (survival) of incident cases by cohort:

$$P(k, y, d) = \sum_{i=1}^{d} I(k - i + 1, y - i + 1) \times S(k - i + 1, y - i + 1, i - 0.5)$$

Where k is age and d the time since diagnosis.

2.2. Estimation of incidence

In France, the national incidence was estimated by correcting the incidence in the area covered by the registries by the mortality ratio between France as a whole and the area covered by the cancer registries, ie [15]:

$$I_{FR} = I_{ZR} \times M_{FR}/M_{ZR} = (I_{ZR} / M_{ZR}) \times M_{FR}$$

Where FR corresponds to France, ZR to the area covered by the registries, I to incidence and M to mortality.

The mortality in France, the mortality and the incidence in the registries' area were modeled separately using age-period-cohort models. The age and cohort variables were introduced as smoothed splines [20] and the period p as p^2 (age*cohort interaction) [15]. This approach provides a flexible modeling strategy for age, period and cohort and allows incidence projections to be made. A more flexible model has been specifically used for breast and prostate cancer. A more flexible model has been used specifically for breast and prostate cancer. Indeed these two cancer sites present complex trends that require more flexible effects than the simple p^2 approach. Multidimensional penalized splines (ie tensor product of age and period) have thus been used; this model allows one to take into account complex effects as well as complex interaction between age and period [21].

2.3. Modelling of overall survival

The logarithm of the cumulative mortality rate H(.) was estimated using the flexible model proposed by Royston & Lambert [16] parameterized as follows:

$$\ln[H(t|age, y)] = s(\ln(t), d_t) + s(age, d_{ag}) + s(y, d_{an}) + s(\ln(t), d_{tvc,ag}) \times age + s(\ln(t), d_{tvc,an}) \times y + age*y$$

Where t is the time elapsed since diagnosis, *age* the age at diagnosis, y the year of diagnosis, *s(.)* a restricted cubic spline of degree of freedom d_k ($\equiv d_k$-1 internal nodes) determined using the Akaike Information Criterion, except for d_t that was set to 4. The 4th and 5th terms corresponded to time dependent effects of respectively age and year of diagnosis. The last term corresponded to the interaction between the linear effects of calendar year and age.

The STPM2 command of STATA© was used [22]. Overall survival was deduced according to the formula S(.) = exp(-H(.))

2.4. Determinants of variation in the number of prevalent cases estimated at the end of 2010 and 2017

The number of prevalent cases changes over time as a result of three separate factors: incidence, survival and demography (ageing and population size). The respective contributions of these factors were evaluated using the method described by De Angelis et al. [7]. The general principle is to decompose the overall variation Δ that corresponds to the difference between the number of prevalent cases estimated at the end of year t and that of year t + k. The overall variation is considered as the sum of the 3 sources of variations:

- The variation related to the demographic changes between t and t + k (size and age structure of the population), noted Δ_d. This variation is calculated as the difference in the number of prevalent cases estimated at the end of year t and at the end of year t + k and obtained by applying the prevalence of year t respectively to the population of year t and to the population of year t + 1.
- The variation related to the evolution of the survival, noted Δ_s. This variation corresponds to the difference in the number of prevalent cases estimated at the end of year t + k and the number of prevalent cases at the end of year t + k estimated by considering a stabilized survival from year t.
- The variation related to the evolution of the incidence, noted Δ_i. This variation corresponds to the difference in the number of prevalent cases estimated at the end of year t + k by considering stabilized survival from year t and the same number obtained by applying the prevalence at the end of year t to the population of the year t + k.

Thus overall variation $\Delta = \Delta_d + \Delta_s + \Delta_i$ [7]. The calculations are carried out by age before adding to the estimate the variation between t and t + k. We also define the net variation in prevalence as $\Delta_n = \Delta - \Delta_d$, which corresponds to the variation attributed to epidemiological factors. The objective is to highlight the specific effect of incidence and survival in the prevalence variation.

Finally, the contribution of each component is evaluated in terms of percentage, namely D = Δ / number of prevalent cases in the year t and $D_k = \Delta_k$ / number of prevalent cases in the year t (k = i, s, d,n).

2.5. Scenarios for projections

Estimating prevalence in 2017 required projections of incidence and survival over the period 2011-2017. Two scenarios for incidence: stable