Cancer Epidemiology 36 (2012) 116–121

Contents lists available at SciVerse ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Joint use of epidemiological and hospital medico-administrative data to estimate prevalence. Application to French data on breast cancer

Marc Colonna [a,b,*], Nicolas Mitton [a], Anne-Marie Schott [c,d], Laurent Remontet [e,f,g,h], Frédéric Olive [i], Frédéric Gomez [j], Jean Iwaz [e,f,g,h], Stéphanie Polazzi [c,d], Nadine Bossard [e,f,g,h], Béatrice Trombert [k]

[a] Registre des Cancers de l'Isère, F-38043, Grenoble, France
[b] Réseau des Registres de Cancers FRANCIM, F-31000 Toulouse, France
[c] Hospices Civils de Lyon, Pole Information Médicale Evaluation Recherche, F-69003 Lyon, France
[d] Université de Lyon, EA 4129 Laboratoire Santé Individu Société, F-69500 Bron, France
[e] Hospices Civils de Lyon, Service de Biostatistique, F-69003 Lyon, France
[f] Université de Lyon, F-69000 Lyon, France
[g] Université Lyon 1, F-69100 Villeurbanne, France
[h] CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Biotatistique-Santé, F-69100 Villeurbanne, France
[i] Département d'Information Médicale, Centre Hospitalier Universitaire de Grenoble, F-38043 Grenoble, France
[j] Centre Léon Bérard, Département d'Information Médicale, F-69008 Lyon, France
[k] Service de Santé Publique et d'Information Médicale, CHU de Saint-Etienne, F-42100 Saint Etienne, France

ARTICLE INFO

Article history:
Received 30 July 2011
Received in revised form 1 December 2011
Accepted 4 December 2011
Available online 26 December 2011

Keywords:
Epidemiology
Health status indicators
Incidence
Prevalence
Breast neoplasms
Registries
Hospital information systems

ABSTRACT

Background Estimate complete, limited-duration, and hospital prevalence of breast cancer in a French Département covered by a population-based cancer registry and in whole France using complementary information sources. *Methods*: Incidence data from a cancer registry, national incidence estimations for France, mortality data, and hospital medico-administrative data were used to estimate the three prevalence indices. The methods included a modelling of epidemiological data and a specific process of data extraction from medico-administrative databases. *Results*: Limited-duration prevalence at 33 years was a proxy for complete prevalence only in patients aged less than 70 years. In 2007 and in women older than 15 years, the limited-duration prevalence at 33 years rate per 100,000 women was estimated at 2372 for Département Isère and 2354 for whole France. The latter rate corresponded to 613,000 women. The highest rate corresponded to women aged 65–74 years (6161 per 100,000 in whole France). About one third of the 33-year limited-duration prevalence cases were diagnosed five years before and about one fourth were hospitalized for breast-cancer-related care (i.e., hospital prevalence). In 2007, the rate of hospitalized women was 557 per 100,000 in whole France. Among the 120,310 women hospitalized for breast-cancer-related care in 2007, about 13% were diagnosed before 2004. *Conclusion*: Limited-duration prevalence (long- and short-term), and hospital prevalence are complementary indices of cancer prevalence. Their efficient direct or indirect estimations are essential to reflect the burden of the disease and forecast median- and long-term medical, economic, and social patient needs, especially after the initial treatment.

1. Introduction

In French women, breast cancer represents more than one third the incident cancer cases and accounts for nearly one fifth of cancer deaths. In 2005, there were nearly 50,000 new cases and more than 11,000 deaths [1]. However, breast cancer has generally a good prognosis; the annual number of deaths is much lower than the annual number of incident cases.

Together with cancer incidence, cancer prevalence is a major epidemiological indicator; it corresponds to the number of living persons with a previous diagnosis of cancer [2]; this is the complete prevalence. Another indicator is the limited-duration prevalence (LDP(x)). LDP(x) represents the proportion of people

* Corresponding author at: Registre des Cancers de l'Isère – CHU de Grenoble – Pavillon E, BP 217 – F-38043, Grenoble Cedex 9 – France. Tel.: +33 476 907 610; fax: +33 476 418 700.
E-mail addresses: mcolonna.registre@wanadoo.fr (M. Colonna), nmitton.registre@orange.fr (N. Mitton), anne-marie.schott-pethelaz@chu-lyon.fr (A.-M. Schott), laurent.remontet@chu-lyon.fr (L. Remontet), FOlive@chu-grenoble.fr (F. Olive), GOMEZ@lyon.fnclcc.fr (F. Gomez), jean.iwaz@chu-lyon.fr (J. Iwaz), stephanie.polazzi@chu-lyon.fr (S. Polazzi), nadine.bossard@chu-lyon.fr (N. Bossard), trombert@univ-st-etienne.fr (B. Trombert).

1877-7821/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.canep.2011.12.001

alive on a certain day who had a diagnosis of the disease within the past *x* years [3–5]. Under some conditions, the LDP may be used to estimate the complete prevalence. A third indicator, we name hospital prevalence, is the number of patients who accessed the health system for cancer-related care within a given period.

These indicators may provide interesting clinical information. For example, complete prevalence estimates the number of persons who need or will need cancer-specific treatment (patients with first cancer or relapse) or posttreatment surveillance, or physical, psychological, or social support (remitting or cured patients). Also, for a number of cancer types, one-, three-, and five-year LDP estimate, respectively, the number of patients under initial treatment, the number of patients under clinical surveillance, and the number of cured patients [4]. However, the use of a five-year LDP is not suited for breast cancer, for example, because the survival curve does not reach or show any plateau, even 10 years after diagnosis [6,7]. In breast cancer, the LDP at five years or more represents then a group of patients still under surveillance because still at risk of relapse.

The present article describes the methods used to estimate these indicators and present their results for breast cancer in Isère (a French Département covered by a cancer registry) and in whole metropolitan France (not covered by a national cancer registry). With appropriate adaptation, these indicators may be used for other cancer sites or geographical areas to estimate the burden of the disease; thus, various medical, psychosocial, and economic needs.

2. Materials and methods

2.1. Data

Depending on the type of prevalence one wishes to estimate for Département Isère or whole France, incidence data from the cancer registry of Isère, incidence data for whole France, mortality and medico-administrative data for Isère or whole France may be required.

2.2. Breast cancer incidence data in Isère

These data are provided by the Cancer Registry of Isère (nearly 1,212,000 inhabitants over 7431 km^2). Case collection started in 1979 and, over the 1979–2007 period, the registry has recorded 17,602 incident breast cancer cases and 4522 deaths from breast cancer. In an attempt to cover a longer period (1975–2007), the observed incidence for 1975–1978 was extrapolated using a modelling that allowed for age and cohort [1].

Patients' vital status is crucial to estimate the LDP. This was sought for 13,448 incident cases diagnosed over the 1989–2007 period. The completeness rate of case follow-up was >95%; at January 1, 2008, only 632 patients were lost to follow-up.

2.3. Mortality from breast cancer

In France, the CépiDc (Centre d'épidémiologie sur les causes médicales de Décès) is a public research department in charge of collecting the causes of death over the whole territory. For the period 1975–2007, mortality is available per year of death, Département of residence, sex, age at death, and cancer site.

2.4. Incidence data for whole France

In the absence of a national cancer registry, the national cancer incidence can be only estimated. Estimation was made using incidence and mortality data from existing cancer registries; the extrapolation to the national level was made using national mortality data [1,8]. Breast cancer incidence for whole France over the 1975–2007 period was estimated using an approach already published by Belot et al. [1] with data on incidence available from various cancer registries over the same period and the corresponding data on mortality from breast cancer.

2.5. Hospital medico-administrative data

In France, the collection of hospital medico-administrative data was generalized to public and private hospitals in 1998. Its importance in funding hospitals was heightened in 2004. One asset of these data (nearly 20 million recordings per year) is that they are collected, coded, and structured according to national standardized rules. This database describes each hospital stay by patient's age, sex, residence-zone code, hospital code, disease(s) treated, and medical or surgical procedures used. A single anonymized patient identifier is used since 2001; it allows chaining patient stays across time and the whole national territory [9]. Between 2003 and 2004, a change in the patient's identifier coding procedure caused a break in chaining patient stays. This restricted the hospital medico-administrative data we used to the 2004–2007 period (nearly 85 million recordings, cancer and non-cancer).

The estimation of hospital prevalence in Isère considered only Isère residents at diagnosis. The estimation of the national hospital prevalence used all the recordings of the medico-administrative database.

3. Methods

3.1. Complete prevalence

Complete prevalence is the number (or rate among the population) of persons affected or having being affected by a disease and still living at a given date, whatever the date of diagnosis. By definition, especially in breast cancer, the estimation of complete prevalence requires an exhaustive recording of cases over a long period to allow counting patients with remote diagnoses. This was often a serious constraint upon the estimation of complete prevalence from cancer registries (except the Connecticut Cancer Registry that started in 1935 [10] – reliable registration since 1940 – and some registries of Nordic countries that started circa 1950 [11]). Here, the estimation of complete prevalence was based on the theoretical relationship between prevalence, incidence, and specific mortality [12]. This relationship may be written:

$$p(x,u) = \frac{CI(x,u) - CM(x,u)}{1 - CM(x,u)}$$

where $p(x, u)$ is the probability for a patient aged x born year u of having a diagnosis of cancer before age x. This probability corresponds to a cross-sectional prevalence at time $t = u + x$. Expression $CI(x, u)$ is the net risk of cancer and corresponds, for a generation born year u, to the cumulative incidence between age 0 and x. For this specific generation, $CM(x, u)$ is the net risk of death from cancer and corresponds to the cumulative mortality between age 0 and x. These cumulative risks of incidence and mortality may be estimated using the specific rates [13]. The specific rates of incidence and mortality are modelled by age-cohort Poisson regression models [14] taking into account the effects of age x and birth cohort u, thus:

$$\log \lambda_{x,u} = \alpha(x) + \gamma(u)$$

$\lambda_{x,u}$ corresponds to the incidence or mortality of patients aged x born year u; x, $\alpha(x)$, and $\gamma(u)$ are smoothing splines [8]. This