Reviewer’s report

Title: Characteristics of people living in Italy after a cancer diagnosis in 2010 and projections to 2020

Version: 0 Date: 09 Oct 2017

Reviewer: Sandra Mallone

Reviewer’s report:

Study overview

Title

"Characteristics of people living in Italy after a cancer diagnosis in 2010 and projections to 2020"

Background

Improved survival and population ageing imply a progressive increase in tumour prevalence. Tumour prevalence regardless the time since diagnosis - i.e. complete prevalence - is less frequently estimated than limited duration prevalence. Estimates and projections of complete tumour prevalence are necessary to help clinicians and health care planners in improving long-term care of patients and appropriately allocating health care resources.

Please, revise the quality of the text.

In particular:

row 116 'previous five years' with respect to?

row 119 'wherease and', is it 'wherease'?

rows 120-123 'this indicator' does the sentence refer to complete prevalence? It's not clear

The aims reported are:

* "to provide an updated description of the number of people living in Italy in 2010 (please clarify updated with respect to…; in 2010 or at 1st January 2010?) after a cancer diagnosis, for all cancers combined and for a selection of cancer types by sex, age, and time since diagnosis".

* "projections of cancer prevalence in Italy are presented up to the year 2020".
Methods

Please reshape these paragraphs in order to let the reader better understand the method/data used and its complexity. A possible example could be the following

Study design and data sources

This is a descriptive analysis of individual data collected during the period 1976-2009 by Italian population-based cancer registries (CR) identified, according to the Italian legislation, as collectors of personal data for surveillance purpose without explicit individual consent. The approval of a research ethic committee is for this reason not required.

Case ascertainment and follow-up of vital status provided by the CRs (is missing in the text). Please add a very brief description or a proper reference.

All tumours except non melanoma skin cancer (more than one primary tumour are considered) and a selected list of 34 tumour/combination of tumours (only the 1st tumour is considered) defined on the basis of the ICD-10 classification, are included.

Benign-uncertain-in situ urinary bladder cancers are included also.

The ICD-O-3 morphology codes were used to define specific subtypes. Please briefly explain which subtypes.

Statistical methods

The clinical and demographic characteristics of the persons registered with a diagnosis of cancers in the Italian CRs are used to:

1. estimate how many of them are still alive at 1st January 2010 regardless the time since diagnosis -i.e. complete prevalence count- by type of tumour, sex and age group;

2. estimate the prevalence proportion in Italy at 2010 for each type/group of tumour, by sex, age;

3. obtain estimates of the complete prevalence (count and proportion) at 1st January 2015 and 2020;

4. disentangle the complete prevalence estimates (count and proportion) by time since diagnosis and describe the changing over time of these estimates.

1. Complete prevalence count is calculated on the basis of population-based cancer registries (CRs) data through the completeness index method
the maximum observed prevalence (7- to 34-year prevalence) is calculated for each of the 27 CRs with at least 7 years of cancer registration as of 31/12/2009 and follow-up of vital status of December 31, 2013; with seerstat?

a set of completeness indices by tumour type, sex, age group, time since diagnosis (7- to 34- as for the maximum observed prevalence?) are obtained by statistical modelling incidence/survival data drawn from 8 population-based Italian CRs with at least 18 years of cancer registration before 2010 (18,19); with COMPREV (17)

the above indices are used as correction factors of the maximum observed prevalence in each CR (17);

2. For each type of tumour, the number of prevalent cases in Italy at 2010 (at 1st January 2010?) are obtained by multiplying the corrected prevalence proportion, stratified by sex and age, by the corresponding Italian population (at 1st January 2010?). What about the adjustment of the prevalence proportion?

Completeness Indices - Statistical modelling of CRs data

Relative survival and incidence functions were estimated by means of parametric models within the period 1985-2011 for survival and 1985-2009 for incidence.

The survival model was a parametric cure model assuming that a proportion of individuals with cancer were bound to die (fatal cases) with a survival following a Weibull distribution, while the remaining proportion (cured fraction) had the same mortality rate as that of the general population with the same age and gender stratification (14,20). The parameters of the survival model were estimated by cancer type, sex, and age class (0-14, 15-44, 45-54, 55-64, 65-74, 75+ years). A period effect was included on the hazard of dying of cancer. Incidence data were categorised according to cancer type, sex, five-year age group, and birth cohort (<1899, 1900-1904,…, 2005-2009). With SAS? Please, add a ref.

A sixth degree polynomial age-cohort model of crude incidence rates was fitted through the SAS logistic procedure for each cancer type and sex (21).

Refs. 20 and 21 are different as regards the CRs data/period used. Please clarify

3. and 4. Projections at 2015, 2020 are obtained assuming that complete prevalence proportions follow a linear function based on the trend estimated for the last 3 calendar years (2007-2009). Please briefly explain how this function/trend is estimated or add a proper reference.

Results

In general, I would suggest the authors to put more emphasis on the projections at 2020 rather than on the 2010 estimates and to divide the paragraph as follows
Prevalence estimates at 2010.

In Italy, in 2010, there are more than 2.6 million of persons (4.6% of the Italian population) alive with a previous diagnosis of tumour (appendix 2, tables 1,2).

Prevalence proportions increase with age depending on sex: 4%, 8%, 11% and 2%, 6%, 13% at age 45-54, 55-64, 65-74 for females and males respectively (table 2).

The most frequent tumours in terms of prevalence are prostate (25% of prevalent cases at 1st jan 2010), bladder (16%), colorectal (16%) tumours for males and breast (42%), colorectal (12%), and endometrial (7%) tumours for females (tables 1,2).

Overall, 59% and 21% (60% and 20% in the paper) of the persons alive at 2010 in Italy with a tumour received a diagnosis at least 5 and 15 years before respectively (figure 1). Prevalence distribution at 2010 by time since diagnosis depends on tumour type: 7% and 39% persons are still alive with a diagnosis of cervical and lung tumours received 2 years before. Considering the same type of tumour, the percentage of persons diagnosed at least 15 years before are 59% and 13% respectively (figure 1).

Please explain, why these prevalence estimates are a bit different respect to the ones reported in ref. 21.

Prevalence projections at 2020.

The number of persons alive with a tumour diagnosis will increase by 37% in 2020, in Italy (from 2.6 to 3.6 million of persons). Female breast (0.8 millions of persons) and prostate (0.5 millions of persons) tumours will be the most frequent type of cancer in terms of prevalence. The largest increases are foreseen for prostate (+85%) and for thyroid cancers (+79%); the smallest for stomach, ovary, larynx in females.

Stomach in females (-10%) and cervical tumours (-13%) show a decline in prevalence. Please try to discuss this result.

22% and 0.7% of the population aged ≥75 and <45 respectively will have had a previous cancer diagnosis (table 4), being thyroid tumour the most frequent in terms of prevalence.

For all tumours, the largest changing in prevalence proportions from 2006 to 2020 will be observed for persons diagnosed with cancer at least 20 years before the calendar year considered (Figure 2).

Discussion
Please, reshape this paragraph discussing -very briefly- the 2010 estimates then the projections/variation over time; clarify the reasons why the 2010 results are different from the ones showed in ref. 21.

Row 237: '...among prevalent cancer cases at? at 2010…'

Row 265: we adjusted estimated proportions in cancer registry areas for the age distribution of the whole Italian population. Please clarify and add this information in the paragraph 'Statistical methods'.

Row 280-284 '…the hypothesis that complete prevalence at 2020 can be predicted by a linear function of calendar year as regressor variable is supported by empirical evidence, consistently showing an approximate linear trend in recent years [5,8,21]. Please clarify the way refs. 5,8 support this issues. Ref.21?

Notably, the use of a longer period (5 calendar years) to estimate linear slope did not materially modify the estimates (ref?).

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal