European Journal of Cancer (2013) xxx, xxx–xxx

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Power analysis to detect time trends on population-based cancer registries data: When size really matters

Roberto Zanetti [a], Francesco Sera [b], Lidia Sacchetto [a], Jan Willem Coebergh [c], Stefano Rosso [a,*]

[a] *Piedmont Cancer Registry, CPO – Centre for Cancer Prevention, Via San Francesco da Paola 31, 10123 Torino, Italy*
[b] *UCL Institute of Child Health, 30 Guilford St, London WC1N 1EH, UK*
[c] *IKZ Integraal Kankercentrum Zuid, P.O. Box 231, 5600 AE Eindhoven, The Netherlands*

KEYWORDS
Population-based
Cancer registry
Time trend
Joinpoint regression analysis
Statistical power
Sample sizes
Population coverage

Abstract Detecting statistically significant trends in incidence with cancer registries data not only depends on the size of their covered population but also on the levels of incidence rates, duration of diagnostic period and type of temporal variation.

We simulated sample sizes of newly diagnosed cases based on a variety of plausible levels of cancer rates and scenarios of changing trends over a period of about 30 years. Each simulated set of cases was then analysed with joinpoint regression models. The power was derived as the relative frequency of the simulation runs where the *p*-value of the coefficient was less than 0.05 under the alternative model.

In case of a decreasing trend with no change of direction (join), an Annual Percentage Change (*APC*) of 1% for an average rate of 10 per 100,000 is detectable in populations of half a million inhabitants or more with a nominal power of 80%. In a model with one joinpoint followed by an increasing trend, the minimum detectable *APC* increases, and an *APC* of about 2%, can be detected only with populations of at least 2 million.

For analyses requiring a larger sample size than the actual covered population, alternative organisational strategies should be considered, such as an extension of population coverage or data pooling and merging from registries with comparable data. (i.e. when heterogeneity across merging registries is low or acceptable for the specific study question).

1. Introduction

Cancer registries in Europe increased from 12 at the beginning of the sixties [1], to more than 200 in these years [2]. This increase has incidentally been on the basis of a programmed plan of implementation, but rather

* *Corresponding author:* Tel.: +39 011 5665351; fax: +39 011 5665362.
E-mail address: stefano.rosso@cpo.it (S. Rosso).

0959-8049/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ejca.2013.10.008

resulted from local professional or environmental pressure and general/specific opportunities, stakeholders willing to invest. The emerging picture entails a huge variation in scope, methods, resources and size of the population covered.

Population-based cancer registries vary in their size of covered population from a number of small sized local registries with less than five hundred thousand inhabitants, especially located in the South of Europe, to national registries covering populations with more than 10 million inhabitants, Germany comprising all of these within their current borders. The reason for maintaining small sized registries, against the obvious observation that larger registries would allow for more statistically stable measures, was mainly local, and because, in the past, without established health information systems, the amount of manual work required for a large population was prohibitively expensive. This is also reflected in the lack of guidelines or recommendation on this subject among those issued by the European Network of Cancer Registries (ENCR) [3].

However, the improvement of technologies, and the shrinking of economic resources impose a rethinking of the 'optimal' size when programming the implementation of new cancer registries, or the adaptation or extension or merging of old ones. In fact, the coverage of Europe is far from complete, but the fulfilment of this objective requires careful planning in the attempt of using the resources at their best.

From a statistical point of view, the definition of the sample size is inherently linked to the study objectives. In our case, when dealing with institutions like cancer registries that serve several purposes and have different aims, from public health surveillance to evaluation of screening programs or evaluation of policies on cancer care [4], the first step is to precisely define the framework of application of cancer registries results.

While usually started with the aim of exploring geographical variation in cancer incidence among different populations and including potential clustering [5], researchers, clinicians and policy makers nowadays look at the cancer registries incidence and survival data with the intent of comparing health policies in different countries, and, more essentially, look at the temporal variation that such policies can induce.

Then, attention was rather driven to trend analysis, where we cut observed time periods in thin slices thus posing further challenges to the possibility of detecting real variations and differences due to either lack of power or limited observation time.

The aim of this exploration study was therefore to explore the likelihood of detecting statistically significant trends with cancer registries data, in relation to the size of their covered population, level of studied rates, observed period length and type of temporal variation.

Besides other methods (see Inset A and its references for an overview), analysis of cancer incidence trends is often approached with the so called 'joinpoint' method or joinpoint regression [6], which allows for changing direction of trends during the observed period. Furthermore, the availability of computer software [7], largely contributed to the diffusion of this method of analysis.

In this study we explored the associated power and population size needed for detecting statistically significant variation in temporal trends as emerging from joinpoint regression analysis. Of course, as mentioned before (see Inset A), it is implicit to say that joinpoint regression analysis is not the only way to look at temporal variation, but it is rather a starting point, now widely used in the routine analyses of registry data, that ignites further investigations with other tools such as age-specific analyses, age–period–cohort models and in general various and more sophisticated tools of analysis. The study findings can then be used as guidelines, among others, when establishing new cancer registries (or extending old ones), and as a warning about the conclusion that can be drawn when analysing trends.

2. Methods

2.1. Joinpoint models

In line with the work of Kim and colleagues [6], where a simulation study was performed to evaluate the size and power of the permutation test, we performed simulations choosing parameters that can reproduce the real situation of cancer incidence in Europe from cancer registries data. A brief description of methods on which joinpoint analysis is based on and the simulation models is given in Appendix.

2.2. Simulations

Firstly, we simulated samples of cases based on a variety of plausible levels of cancer rates. For all of simulations, the time period (the predictor time variable x) was from 1980 to 2008, denoting yearly time points (in our example 29 years). This period represents the period during which most of the currently active cancer registries in Europe have data (about 38 registries, and among them 12 nationals, with an overall population coverage of more than 121 million inhabitants). Each simulated set of cases was then analysed with the Joinpoint software (a batch version of the Joinpoint Regression Program, Version 3.5, was kindly provided by the Statistical Research and Application Branch, National Cancer Institute [7]), and the weighted least squares permutation test was calculated. The power was derived as the relative frequency of the simulation runs where the p-value of the coefficient was less than 0.05 under the alternative model.

Please cite this article in press as: Zanetti R. et al., Power analysis to detect time trends on population-based cancer registries data: When size really matters, *Eur J Cancer* (2013), http://dx.doi.org/10.1016/j.ejca.2013.10.008