Selecting the minimum prediction base of historical data to perform cancer predictions: the GoF-optimal method

Joan Valls[1], Gerard Castellà[1],Tadeusz Dyba[2] and Ramon Clèries[3]

[1]jvalls@irblleida.cat, Unit of Biostatistics and Epidemiology, Biomedical Research Institute of Lleida, Lleida, Catalonia, Spain

[2]European Commission, DG Joint Research Centre, Institute for Health and Consumer Protection, Public Health – Cancer Policy Support, Ispra, Italy.

[3]Plan for Oncology of the Catalan Government. IDIBELL, Hospital Duran i Reynals. L'Hospitalet de Llobregat , Catalonia, Spain

Background: Predicting the future burden of cancer is a key issue for health services planning, being the crux of this problem the choice of the method for selecting a predictive model and a prediction base. Most epidemiological studies make an arbitrary choice of the base of prediction as the last 5, 10 or even 20 years, but an assessment of this choice is not usually provided. A method, named here Goodness-of-Fit optimal (GoF-optimal), is presented to determine the minimum prediction base of historical data with the goal to perform accurate 5-year predictions of the number of new cancer cases or deaths.

Methods: An empirical ex-post evaluation exercise for cancer mortality data in Spain and cancer incidence in Finland using simple linear and log-linear Poisson models was performed. Prediction bases were considered within the time periods 1951–2006 in Spain and 1975-2007 in Finland, and then predictions were made for 37 and 33 single years in these periods, respectively. The performance of three fixed different prediction bases (last 5, 10, and 20 years of historical data) was compared to that of the prediction base determined by the GoF-optimal method. The coverage (COV) of the 95% prediction interval and the discrepancy ratio (DR) were calculated to assess the accuracy of the predictions.

Results: The results showed that (i) models using the prediction base selected through GoF-optimal method reached the highest COV and the lowest DR and (ii) the best alternative strategy to GoF-optimal was the one using the base of prediction of 5-years.

Conclusions: The GoF-optimal approach can be used as a selection criterion in order to find an adequate base of prediction.

Keywords: cancer epidemiology, Poisson models, cancer predictions.

AMS: 62J12, 62P10

Reference: Valls J, Castellà G, Dyba T, Clèries R. Selecting the minimum prediction base of historical data to perform 5-year predictions of the cancer burden: The GoF-optimal method. Cancer Epidemiol. 2015, in press