ELSEVIER

European Journal of Cancer 41 (2005) 438–444

European Journal of Cancer

www.ejconline.com

Period analysis of cancer patient survival in datasets from which the month of diagnosis has been removed

Hermann Brenner *, Volker Arndt

Department of Epidemiology, German Centre for Research on Ageing, Bergheimer Strasse 20, D-69115 Heidelberg, Germany

Received 21 September 2004; received in revised form 8 October 2004; accepted 19 November 2004
Available online 8 January 2005

Abstract

Up-to-date monitoring of long-term survival is an important task of population-based cancer registries. Period analysis, a new method of survival analysis introduced a few years ago, has been shown to be particularly useful for that purpose. The "classical" period analysis uses a life-table approach which requires both the year and month of diagnosis for implementation in pertinent software programs. However, an increasing number of cancer registries remove the month of diagnosis from their datasets, mainly to ensure the highest possible protection against re-identification of patients. In this paper, we present modifications of period analysis that allow the application of this technique, while almost completely preserving its advantages, in datasets without the month of diagnosis. The modified techniques are illustrated and evaluated using examples from the Surveillance, Epidemiology, and End Results (SEER) programme of the United States (US) National Cancer Institute (NCI), which also has removed month of diagnosis from its most recently released public use database.

Keywords: Cancer registries; Statistical methods; Survival; Month of diagnosis

1. Introduction

Up-to-date monitoring of long-term survival is an important task of population-based cancer registries [1]. Period analysis, a new method of survival analysis introduced a few years ago [2], has been shown to be particularly useful for that purpose, as it enables detection of time trends in long-term survival rates long before they can be disclosed by traditional survival analysis [3–5]. The principle of period analysis is to restrict the survival experience to be included in the analysis to some recent time period, such as some recent calendar year, which is achieved by left truncation of observations at the beginning of that period in addition to the commonly employed right censoring at its end.

The "classical" period analysis as introduced by Brenner and Gefeller in 1996 [2] uses a life-table approach which requires both the year and month of diagnosis for implementation in pertinent software programs [6,7]. Recently, an increasing number of cancer registries started to remove the month of diagnosis from their datasets, mainly to ensure the highest possible protection against re-identification of patients. For example, both the year and month of diagnosis had been included in former releases of data from the Surveillance, Epidemiology, and End Results (SEER) programme of the United States (US) National Cancer Institute (NCI), whereas the month of diagnosis was removed, for the first time, from the 1973 to 2001 dataset released in April 2004 [8]. Similarly, the European database of the Automated Childhood Cancer Information System (ACCIS) includes only the year, and not the month of diagnosis [9]. Although these datasets still include the survival time of patients in months, this information, along with

* Corresponding author. Tel.: +49 6221 548140; fax: +49 6221 548142.
E-mail address: Brenner@dzfa.uni-heidelberg.de (H. Brenner).

0959-8049/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2004.11.018

calendar year of diagnosis, is not sufficient to determine the calendar year of death (or of end of follow-up among censored patients) which would be required for a classical period analysis.

In this paper, we propose and evaluate modifications of period analysis that still allow application of this useful technique in datasets in which the month of diagnosis has been removed.

2. Patients and methods

2.1. The "classical" period analysis

The classical period analysis is illustrated by the following example. Let us assume that one wished to obtain an up-to-date estimate of 10-year survival for patients with some form of cancer using a database of cancer registry data that includes incident cases as well as follow-up of patients with respect to vital status up to and including the year 2000.

With a traditional cohort analysis, 10-year survival might have been obtained from patients who could have been observed for 10 years following diagnosis by the end of 2000, i.e., patients diagnosed in 1990 or earlier. To increase the precision of estimates, one might have pooled data from several consecutive years of diagnosis, such as the years 1988–1990, as indicated by the solid frame in Fig. 1. However, this approach would not have reflected improvements in prognosis achieved in the 1990s, e.g., by advancements in early detection or therapy.

By contrast, a period analysis for the period of 1998–2000 would have exclusively reflected the survival experience in these three recent years. With this approach, survival during the 1st year following diagnosis would have been obtained from patients diagnosed in 1997–2000, survival in the 2nd year following diagnosis would have been obtained from patients diagnosed in 1996–1999, and so on, until conditional survival in the 10th year following diagnosis, which would have been obtained from patients diagnosed between 1988 and 1991 (dashed frame in Fig. 1). These conditional survival rates by year following diagnosis would then be multiplied to come up with a period estimate of 10-year survival.

2.2. A modified period analysis

As mentioned previously, the year of diagnosis is not sufficient for the classical period approach, as it does not allow an unequivocal attribution of deaths to calendar years, even if the exact survival time is known. However, deaths by year of follow-up can be unequivocally attributed to a pair of calendar years, as shown in Fig. 2. For example, it is clear that a death during the 10th year

Years of Diagnosis	Years of Follow-up												
	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
1988	1	1/2	2/3	3/4	4/5	5/6	6/7	7/8	8/9	9/10	10		
1989		1	1/2	2/3	3/4	4/5	5/6	6/7	7/8	8/9	9/10	10	
1990			1	1/2	2/3	3/4	4/5	5/6	6/7	7/8	8/9	9/10	10
1991				1	1/2	2/3	3/4	4/5	5/6	6/7	7/8	8/9	9/10
1992					1	1/2	2/3	3/4	4/5	5/6	6/7	7/8	8/9
1993						1	1/2	2/3	3/4	4/5	5/6	6/7	7/8
1994							1	1/2	2/3	3/4	4/5	5/6	6/7
1995								1	1/2	2/3	3/4	4/5	5/6
1996									1	1/2	2/3	3/4	4/5
1997										1	1/2	2/3	3/4
1998											1	1/2	2/3
1999												1	1/2
2000													1

Fig. 1. Data included in a traditional cohort analysis of 10-year survival for a cohort of patients diagnosed in 1988–1990 (solid frame) and in a classical period analysis of 10-year survival for the 1998–2000 calendar period (dashed frame). The numbers within the cells indicate the years of follow-up after diagnosis.