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.

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

Fig. 2. 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 modified period analysis of 10-year survival within the 1997–2000 calendar period (dashed frame). The numbers within the cells indicate the years of follow-up after diagnosis

following diagnosis of a patient diagnosed in 1988 occurred at some time during 1997/1998. Hence, a modified period analysis for (overlapping) pairs of calendar years 1997/1998, 1998/1999, and 1999/2000 would still be feasible, as indicated by the dashed frame in Fig. 2. The result should be similar, but not equivalent to a classical period analysis for the period of 1998–2000. Like the latter, the modified period analysis covers all of the patients' survival experience in 1998 and 1999. However, some part of the patients' survival experience in 1997 is also included, whereas the survival experience of patients in 2000 is only partly included. Hence, survival rates obtained with this analysis, would be expected to be slightly (i.e., approximately half a year) less up-to date than the classical period survival rates, but still much more up-to-date than the traditional cohort survival rates. A traditional cohort analysis of 10-year survival for the 1988–1990 cohort could be realised with this data structure (Fig. 2, solid frame) in the same way as in cases with complete dates of diagnosis (Fig. 1, solid frame).

2.3. Other alternatives

An even simpler alternative approach to come up with period estimates of survival in datasets lacking month of diagnosis would be to impute a common month for each patient. June or July would initially appear to be natural choices for that month, as they should represent the "median" month of diagnosis for the cancer patients. However, December might also be considered as an alternative, as this would ensure that follow-up is complete and vital status is known for all patients until the closing date of follow-up (except for those patients who were lost to follow-up anyway), because the date of diagnosis is not artificially advanced to an earlier date for any of the patients.

2.4. Empirical evaluation

We evaluate and compare the different approaches for cancer patients in the US using the SEER 1973–2000 database [10]. This database includes records of cancer patients diagnosed in 1973–2000 and followed with respect to vital status until the end of 2000 from nine population-based cancer registries in the US (Connecticut, New Mexico, Utah, Iowa, Hawaii, Atlanta, Detroit, Seattle-Paget Sound, and San Francisco-Oakland) covering a population of approximately 30 million people. The SEER programme is the most authoritative source of information on cancer incidence and survival in the US, and is considered a model for quality among cancer registries around the world. Quality control has been an integral part of SEER since its inception. Every

year, studies are conducted in the SEER areas to evaluate the quality and completeness of the data being reported (SEERs' standard for case ascertainment is 98%).

The 1973–2000 SEER database was released in April 2003 and contains both the year and month of diagnosis for cancer patients [10]. Hence it allows us to estimate traditional 10-year cohort survival rates for the 1988–1990 cohorts of patients, as well as both classical 10-year period survival rates for the 1998–2000 period and modified 10-year period survival rates for 1997–2000, as shown in Figs. 1 and 2.

In this analysis, all three types of estimates were derived for patients with a first diagnosis of either one of 22 common forms of cancer (which together account for approximately 90% of all cancers). In addition, classical period analyses for the 1998–2000 period were repeated after imputation of either June or December as the month of diagnosis to assess the performance of these alternative potential analytical strategies in situations with missing information on the month of diagnosis.

The analysis was restricted to patients above the age of 15 years because childhood cancers, which account for less than 1% of all cancers, differ from adulthood cancers in many respects. Furthermore, patients with a truly unknown month of diagnosis (0.6%) or length of follow-up (1.5%) were excluded, as were patients who were reported to the registries by death certificate only (1.2%) or by autopsy only (0.3%).

To address both total and cancer-related mortality, both absolute and relative survival rates are presented. Relative survival rates indicate so-called net survival of patients with cancer. They can be interpreted as expected survival rates of cancer patients in the hypothetical situation in which cancer is the only cause of death [11]. Relative survival rates are calculated as the ratio of absolute survival rates of cancer patients divided by the expected survival rates for a group with similar age and gender distribution from the general population. Expected survival was estimated according to calendar year, age (single years), gender, and race from population life-tables according to Hakulinen's method [12]. For the years 1986–1995, US life-tables for the year 1990 provided from the SEER public use database were used [10], for the years 1996–2000, US life-tables for the year 2000 were used [13]. Standard errors were calculated according to Greenwood's method [14]. However, as standard errors were generally very small (<1% for 205 of 220 survival estimates shown in this paper, <1.5% for all cases) given the very large numbers of patients included in all analyses, standard errors are not individually reported in this paper.

All analyses were done with the SAS statistical software package version 8.2 using a previously published macro for both cohort and classical period analyses [7], with minor adaptations to realise the modified period analysis shown in Fig. 2.

3. Results

Table 1 shows the total numbers of patients (diagnosed in 1988–2000) by cancer site included in this analysis, as well as the various estimates of absolute 10-year survival rates. The most common form of cancer was prostate cancer, followed by breast cancer, lung cancer and colon cancer.

10-year absolute survival for the cohorts of patients diagnosed in 1988–1990 strongly varied according to the cancer site, ranging from 90.5% for testicular cancer to 1.5% for pancreatic cancer. According to the classical period analysis for the 1998–2000 period using the full information on the month of diagnosis (''true month of diagnosis''), 10-year absolute survival increased for most cancer sites in the 1990s. The strongest increase was seen for prostate cancer (52.8% vs. 39.9%). For many other cancers, somewhat less pronounced, but still relevant improvements were disclosed by the period analysis (e.g. oral cavity cancer: 34.4% vs. 30.1%, rectum cancer: 36.3% vs. 32.7%, breast cancer: 61.8% vs. 58.7%, ovarian cancer: 38.5% vs. 35.0%, kidney cancer: 38.6% vs. 35.3%, Non-Hodgkin's lymphoma: 35.3% vs. 29.4%). Unfortunately, no major improvements were achieved for some cancers that have a very dismal prognosis, including lung and pancreatic cancers.

In general, the modified period analysis covering the survival experiences of patients in 1997–2000 as described above, which would still be feasible if no information on months of diagnosis was included in the dataset, provided very similar, and, in most cases, almost identical estimates of 10-year absolute survival rates as the classical period analysis for the 1998–2000 period. Hence the major improvements achieved in the 1990s could have been disclosed by this modified method almost as well as with the classical period analysis. Only for prostate cancer would the large increase not have been captured as completely by the modified period analysis.

Of the two alternative approaches with an imputed month of diagnosis, the one using December also provided very similar and often essentially identical results as the classical period analysis including a known (true) month of diagnosis. By contrast, 10-year absolute survival was generally somewhat overestimated for all patients if June was imputed as the month of diagnosis.

10-year relative survival estimates are generally considerably higher than 10-year absolute survival rates, particularly for those cancers which occur predominantly at older ages, such as cancers of the prostate, colon, rectum, corpus uteri, and urinary bladder (see Table 2). As a result, the differences between the various survival estimates are, in absolute terms, also somewhat

Table 1
10-year absolute survival estimates for cohorts of patients diagnosed in 1988–1990, "classical" period estimates for the 1998–2000 period (see Fig. 1) with the true month of diagnosis, or imputing June or December as the month of diagnosis, and modified period estimates for the 1997–2000 period (see Fig. 2)

Cancer site	n[a]	10-year absolute survival (%)				
		Cohort analysis 1988–1990	Classical period analysis 1998–2000			Modified period analysis 1997–2000
			Month of diagnosis			
			True	"June"	"December"	
Oral cavity	28,969	30.1	34.4	35.9	34.3	34.2
Oesophagus	11,192	4.5	6.5	7.6	6.7	6.1
Stomach	22,240	10.5	11.3	12.6	11.4	11.2
Colon	97,008	32.9	33.3	35.2	33.2	33.0
Rectum	39,205	32.7	36.3	38.0	36.5	36.0
Pancreas	27,288	1.5	2.0	2.5	2.3	2.2
Larynx	11,294	37.2	36.4	38.2	36.0	36.0
Lung	161,293	6.3	6.9	7.9	7.0	7.0
Melanoma	41,195	69.9	70.7	71.6	71.0	70.5
Breast	184,532	58.7	61.8	63.4	61.9	61.4
Cervix	14,296	58.6	59.8	61.4	59.5	59.5
Corpus	35,782	61.7	62.4	64.0	63.1	62.7
Ovaries	24,339	35.0	38.5	40.4	38.7	38.8
Prostate	201,578	39.9	52.8	54.6	53.2	51.6
Testis	8,607	90.5	91.0	91.7	91.3	90.6
Urinary bladder	50,215	44.1	44.2	46.3	44.6	44.3
Kidneys	25,483	35.3	38.6	40.6	39.5	38.7
Hodgkin's lymphoma	16,380	16.8	19.0	20.4	19.6	19.2
Non-Hodgkin's lymphoma	48,111	29.4	35.3	36.8	35.0	34.3
Myeloma	13,954	7.0	9.4	10.7	9.8	9.3
Lymphocytic leukaemia	12,603	29.5	31.7	34.0	31.8	31.8
Myeloid leukaemia	13,001	8.9	11.2	12.6	11.7	11.1
Difference from 1988 to 1990 cohort estimate						
Maximum			+12.9	+14.7	+13.3	+11.7
Minimum			−0.8	+1.0	−1.2	−1.2
Median			+2.1	+3.35	+2.25	+1.9
Difference from period analysis with true month of diagnosis						
Maximum				+2.3	+0.9	+0.3
Minimum				+0.5	−0.4	−1.2
Median				+1.55	+0.2	−0.15
Within ±0.5%				1 of 22	19 of 22	21 of 22

[a] n = total number of patients in 1988–2000 included in the analyses.

higher than the differences seen for the 10-year absolute survival rates. However, otherwise, the relative performance of the various analytical strategies is essentially the same: the often substantial increase in the 1990s is disclosed by the modified period analysis and by the classical period analysis using December as the month of diagnosis almost as well as by the classical period analysis with information on the true month of diagnosis, whereas imputing June as the month of diagnosis in the classical period analysis again led to higher 10-year survival estimates in all cases.

4. Discussion

In this manuscript, we provide alternative approaches to the classical period analysis of survival to enable application of the period analysis methodology even in situations in which the month of diagnosis is not included in the dataset. Our empirical evaluation using cancer registry data of the SEER programme illustrates that two of these approaches are, for practical purposes, essentially equivalent to a classical period analysis. However, the apparently obvious approach of imputing June for the missing month of diagnosis provided higher survival estimates and should not be used.

The latter finding may initially be surprising, as June or July presumably represent the median month of diagnosis for cancer patients diagnosed in any given year. One might therefore expect that imputing June or July would introduce only minor imprecision. However, a more careful look at the implications reveals that this apparently obvious reasoning is flawed. For the "shift" of some patients with a true month of diagnosis between July and December to a diagnosis in June will, for some patients, a shift their death from the first calendar year

Table 2
10-year relative survival estimates for cohorts of patients diagnosed in 1988–1990, "classical" period estimates for the 1998–2000 period (see Fig. 1) with the true month of diagnosis, or imputing June or December as the month of diagnosis, and modified period estimates for the 1997–2000 period (see Fig. 2)

Cancer site	n[a]	10-year relative survival (%)				
		Cohort analysis 1988–1990	Classical period analysis 1998–2000			Modified period analysis 1997–2000
			Month of diagnosis			
			True	"June"	"December"	
Oral cavity	28,969	41.8	47.2	49.4	47.2	47.0
Oesophagus	11,192	6.9	10.0	11.8	10.3	9.4
Stomach	22,240	17.6	19.0	21.2	19.1	19.0
Colon	97,008	57.1	57.2	60.5	57.0	56.9
Rectum	39,205	52.6	56.5	59.2	56.8	56.2
Pancreas	27,288	2.5	3.3	4.2	3.8	3.7
Larynx	11,294	53.3	51.8	54.4	51.3	51.1
Lung	161,293	9.6	10.5	12.2	10.7	10.6
Melanoma	41,195	85.5	87.2	88.4	87.6	87.3
Breast	184,532	76.8	80.5	82.5	80.6	80.1
Cervix	14,296	67.0	68.2	70.0	67.8	67.9
Corpus	35,782	82.7	84.0	86.2	84.9	84.3
Ovaries	24,339	45.2	50.2	52.6	50.4	50.6
Prostate	201,578	78.7	92.4	95.7	93.1	90.1
Testis	8,607	94.2	94.6	95.3	94.9	94.1
Urinary bladder	50,215	74.6	73.6	77.2	74.3	74.0
Kidneys	25,483	51.2	55.3	58.1	56.6	55.5
Hodgkin's lymphoma	16,380	21.4	24.2	26.0	25.1	24.4
Non-Hodgkin's lymphoma	48,111	41.5	49.7	51.8	49.3	48.5
Myeloma	13,954	11.4	15.2	17.3	15.8	15.0
Lymphocytic leukaemia	12,603	46.8	48.7	52.3	48.9	49.3
Myeloid leukaemia	13,001	13.3	16.5	18.5	17.3	16.5
Difference from 1988 to 1990 cohort estimate						
Maximum			+13.7	+17.0	+14.4	+11.4
Minimum			−1.5	+1.1	−2.0	−2.2
Median			+2.35	+4.75	+2.8	+2.5
Difference from period analysis with the true month of diagnosis						
Maximum				+3.6	+1.3	+0.6
Minimum				+0.7	−0.5	−2.3
Median				+2.15	+0.3	−0.1
Within ±0.5%				0 of 22	15 of 22	17 of 22

[a] n = total number of patients in 1988–2000 included in the analyses.

included in the period analysis to the preceding year. If, as in our example and as in most applications of period analysis, the last calendar year included in the period analysis is also the last year of follow-up, this "loss of deaths" will not be compensated for by deaths shifted from the calendar year following the period of interest into that period, which explains the overestimation of survival for the most recent calendar period. Although the shift of some patients with a true month of diagnosis between January and May to a diagnosis in June will also, for some patients, shift their death from the preceding calendar year to the first calendar year included in the period of analysis, this shift is approximately compensated for by an analogous shift of deaths from the last calendar year included in the period analysis to the following year (not included in the analysis).

Analogously, the shift of most patients with a true month of diagnosis between January and November to an imputed month of diagnosis in December will imply for some patients a shift of their death from the preceding calendar year to the first calendar year included in the period of analysis. However, this shift is approximately compensated for by an analogous shift of deaths from the last calendar year included in the period analysis to the following year (not included in the analysis), which explains why this approach generally gives results that are quite similar (although not necessarily exactly identical) to those obtained in a classical period analysis. Note that the specific month of diagnosis imputed would be of little relevance for period estimates for earlier periods (not including the last year of follow-up) and entirely irrelevant for cohort estimates for cohorts with complete follow-up.

The modified type of period analysis also gives results very close to those obtained in a classical period analysis, even though they tend to be, on average, slightly lower. This very minor difference, which is negligible for practical purposes in most cases, is explained by a

delay of approximately half a year in the monitoring of long-term survival compared with the classical period analysis of survival, as outlined in Fig. 2. Given that classical period analysis has been shown to advance detection of time trends in 5-, 10-, 15- and 20-year survival by almost 5, 10, 15 and 20 years compared with traditional cohort analysis [4], this delay is quite minor, and only slightly diminishes the advantages of period analysis in timeliness of survival estimates. Nevertheless, this slight loss of timeliness of survival estimates has to be kept in mind and weighed against the reasons given for the removal of the month of diagnosis from datasets which are mainly confidentiality concerns.

Our empirical evaluation found the performance of the modified period analysis and the classical period analysis using December as the month of diagnosis to be approximately similar. Which of these methods one might wish to apply may therefore be partly a matter of taste. If one does not feel too comfortable with imputing a "pseudomonth" of diagnosis, one might prefer to use the method herein denoted the modified period analysis. The "imputation method" has the advantage that the same computer programs can be used as for the classical period analysis (e.g. [5,6]), without the need to switch to yet another program code.

In summary, our analysis provides solutions on how to almost completely preserve the advantages of period analysis of cancer patient survival, even in situations in which a classical period analysis cannot be carried out due to lack of information on the month of diagnosis. Since the month of diagnosis is increasingly being removed from cancer registry datasets, these alternative solutions should be helpful in an increasing number of settings.

Conflict of interest statement

None declared.

References

1. Parkin DM, Hakulinen T. Analysis of survival. In: Jensen OM, Parkin DM, Maclennan R, Muir CS, Skeet RG, eds. *Cancer registration. Principles and methods*. IARC Scientific Publications No. 95. Lyon, International Agency for Research on Cancer, 1991, pp. 159–176.
2. Brenner H, Gefeller O. An alternative approach to monitoring of cancer patient survival. *Cancer* 1996, **78**, 2004–2010.
3. Brenner H, Hakulinen T. Up-to-date survival curves of patients with cancer by period analysis. *J Clin Oncol* 2002, **20**, 826–832.
4. Brenner H, Hakulinen T. Advanced detection of time trends in long-term cancer patient survival: experience from 50 years of cancer registration in Finland. *Am J Epidemiol* 2002, **156**, 566–577.
5. Talbäck M, Stenbeck M, Rosén M. Up-to-date long-term survival of cancer patients: an evaluation of period analysis on Swedish Cancer Registry data. *Eur J Cancer* 2004, **40**, 1361–1372.
6. Brenner H, Gefeller O, Hakulinen T. A computer program for period analysis of survival. *Eur J Cancer* 2002, **38**, 690–695.
7. Arndt V, Talbäck M, Gefeller O, Hakulinen T, Brenner H. Modification of SAS macros for more efficient analysis of relative survival rates. *Eur J Cancer* 2004, **40**, 778–779.
8. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Public-Use Data (1973–2001), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2004, based on the November 2003 submission.
9. ACCIS Automated Childhood Cancer Information System [Online]. Available from: http://www-dep.iarc.fr/accis/about.htm. Accessed September 16, 2004.
10. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Public-Use Data (1973-2000), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2003, based on the November 2002 submission.
11. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. *Natl Cancer Inst Monogr* 1961, **6**, 101–121.
12. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982, **38**, 933–942.
13. Arias E. United States life tables, 2000. *Natl Vital Stat Rep* 2002, **51**, 1–39.
14. Greenwood M. *A report on the natural duration of cancer*. London, Ministry of Health, HMSO, 1926.