European Journal of Cancer 41 (2005) 1788–1793

European Journal of Cancer

www.ejconline.com

Reduction in selective under-ascertainment bias in population-based estimates of cancer patient survival by age adjustment

Hermann Brenner [a,*], Timo Hakulinen [b]

[a] Department of Epidemiology, German Centre for Research on Ageing, Bergheimer Strasse 20, D-69115 Heidelberg, Germany
[b] Finnish Cancer Registry, Institute for Statistical and Epidemiological Cancer Research, Liisankatu 21 B, FIN-00170 Helsinki, Finland

Received 24 February 2005; received in revised form 23 March 2005; accepted 29 March 2005
Available online 25 July 2005

Abstract

Selective under-ascertainment of cancer patients by cancer registries may bias population-based estimates of cancer patient survival. As prognosis is strongly age related for many forms of cancer, age adjustment of cancer survival rates, which is primarily used to enhance the validity of comparative analyses between cancer populations, might also effectively reduce this type of bias. We empirically assessed this potential "side-effect" of age adjustment using data from the Finnish Cancer Registry. Analyses of five-year absolute and relative survival for patients diagnosed in 1990–1994, and age adjustment to the age structure of patients diagnosed in 1985–1989, were used as examples. Various patterns of selective under-ascertainment were simulated, and the bias in crude and age adjusted five-year survival rates was compared. Age adjusted estimates were less biased in most scenarios, which may be an additional argument for application of age adjustment in the analysis and reporting of population-based cancer survival rates.

Keywords: Bias; Cancer registries; Statistical methods; Survival

1. Introduction

Monitoring cancer incidence and survival are important tasks of population-based cancer registries. High levels of completeness of case ascertainment are an essential prerequisite for valid estimates of cancer incidence. Under-ascertainment of cases, which is of potential concern, albeit to a strongly varying degree, in virtually all population-based cancer registries will always lead to erroneously low estimates of cancer incidence. The impact of case-under-ascertainment on estimates of survival is somewhat more complex. While no bias is expected if under-ascertainment of patients is independent of their prognosis, survival may be underestimated or overestimated if patients with a poor prognosis are less or more likely missed by the registry than patients with good prognosis [1].

It has been shown repeatedly that selective under-ascertainment of cancer patients may indeed bias estimates of cancer patient survival [2] and may invalidate comparisons of cancer patient survival between populations [3,4]. For example, patients with a poor prognosis may have a lower chance of notification to a cancer registry, as they may have only few, if any contacts, with the health care system before they die. Furthermore, prognosis is strongly age related for many forms of cancer [5–8], and differential under-ascertainment of cases may thus go along with a distortion of the age distribution of cases. This distortion may be overcome by age adjustment of cancer survival rates. We therefore

* 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 © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2005.03.029

hypothesised that age adjustment, which is primarily used to enhance the validity of comparative analyses between cancer populations with different age structure (*e.g.*, in international comparisons or in time trend analyses of cancer patient survival), might also reduce bias due to differential under-ascertainment of cases. The aim of this study was to empirically assess this potential "side-effect" of age adjustment in various scenarios of differential under-ascertainment of cancer cases.

2. Patients and methods

2.1. Database

Our analysis is based on data from the nationwide Finnish Cancer Registry. The Finnish Cancer Registry that covers the whole population of Finland (≈5.3 million people) is well known for its high levels of data quality and completeness [9]. The registry obtains information from many different sources, including hospitals, physicians outside hospitals, and pathologic and cytologic laboratories. Notification of cancer cases to the registry is mandatory by law, and studies have shown that the registry achieves close to 100% completeness. This data set allows comparison of true estimates of cancer patient survival with those obtained by simulating various patterns and levels of incomplete case ascertainment. Mortality follow-up is performed by annual record linkage of registry data with death records, using a personal identification number as key.

For our empirical evaluation, we included patients with a first diagnosis of one of eight common forms of cancer from 1990 to 1994 aged 15 or older. The survival experience within five years following diagnosis was analysed in these patients. The selected cancers included the four cancers most commonly diagnosed in Finland in 1990–1994 (cancers of the breast, lung, prostate, and colon), as well as four common forms of cancer with a relatively strong age gradient in survival (cancers of the ovaries, brain, thyroid gland, and lymphomas). Patients notified to the Finnish Cancer Registry by death certificate only (0.8%) or by autopsy only (2.1%) were excluded from the analyses, as were patients with missing information on month of diagnosis (0.1%).

2.2. Statistical analysis

We first calculated estimates of five year absolute and relative survival based on all cancer patients notified to the Finnish Cancer Registry meeting the above mentioned inclusion criteria. Next, we calculated expected estimates of five year absolute and relative survival assuming the following types of selective under-ascertainment of cancer patients: (a) selective under-ascertainment of patients dying within five years following diagnosis (regardless of their age) by 20%; (b) selective under-ascertainment of patients surviving five years following diagnosis (regardless of their age) by 20%; and (c) selective under-ascertainment of patients above 75 years of age (regardless whether or not they died within five years from diagnosis) by 50%.

Scenarios (a) and (b), thus, reflect extreme prototypes of selective under-ascertainment of patients with good or poor prognosis. These scenarios necessarily result in underestimation or overestimation of survival, respectively, and they may go along with potential distortions in the age distribution of patients if prognosis differs between age groups. Scenario (c) reflects an extreme prototype of exclusively age-related under-ascertainment. Such age-related under-ascertainment of cases may result, for example, from under-diagnosis of cancer in the elderly. This type of under-ascertainment necessarily results in distortions of the age distribution of patients, which may go along with potential under- or overestimation of survival if prognosis differs between older and younger patients.

For each scenario, the expected survival estimates were derived by weighted survival analyses, in which a weight of 0.8 (under-ascertainment type (a) and (b)) or 0.5 (under-ascertainment type (c)) was assigned to the patients in the groups affected by under-ascertainment, and a weight of 1 was assigned to all other patients. The weighting was employed by a recently introduced algorithm, which was primarily developed for age adjustment of cancer survival rates, but which can also be used for other applications of weighted survival analyses [10].

For each scenario, we calculated a crude and an age-adjusted estimate of five-year absolute and relative survival. Relative survival [11] was calculated according to Hakulinen's method [12]. For illustration, age adjustment was made to the (site specific) age distribution of cancer patients notified to the Finnish Cancer Registry in 1985–1989. Such age adjustment would typically be considered, for example, in a time series analysis comparing survival of cancer patients diagnosed in 1990–1994 and in 1985–1989. In all scenarios, the latter was assumed to be determined correctly, *i.e.*, with no selective under-ascertainment of cancer patients was assumed for the 1985–1989 period. In this way, the difference in the adjusted survival estimates obtained with complete case ascertainment in 1990–1994 and the adjusted survival estimates expected in the scenarios with selective under-ascertainment of cases would also reflect the bias in a time series analysis in the (entirely hypothetical) situation of a reduction of completeness of cancer ascertainment between both time periods.

Age adjustment was made according to the method recently introduced by Brenner *et al.* [10], using 5 age categories (15–44, 45–54, 55–64, 65–74, and 75+). Due to the very rare occurrence of prostate cancer at younger ages, age groups 15–54, 55–64, 65–74 75–84, and 85+ were used for this cancer (analogous age categorisations have been used in previous studies, such as the EUROCARE study [13]).

All analyses were done with the SAS statistical software package using the macros "periodh" and "adperiodh" [10,14,15].

3. Results

Table 1 shows the numbers and the age distribution of patients diagnosed in 1990–1994 in Finland with one of the eight forms of cancer considered in this analysis. With more than 13,000 cases, breast cancer was the most common form of cancer, followed by lung cancer, prostate cancer and colon cancer. Although somewhat less common, the other four forms of cancer were still represented by between 1487 and 2446 patients. Patients with prostate cancer and colon cancer included the highest proportions of older patients. Patients with breast cancer and with ovarian cancer were more uniformly distributed across the five age groups, while the proportion of younger patients was highest among patients with cancers of the brain and the thyroid gland.

Five-year absolute and relative survival of cancer patients by age group is shown in Table 2. With few exceptions, there is a strong decrease of five-year absolute survival with age. The age gradient is most pronounced for cancers of the ovaries, brain, thyroid gland and lymphomas. By contrast, a major decline in five-year survival with age is restricted to the oldest age groups for cancers of the breast and prostate (age groups 65+ and 75+, respectively). The age specific relative survival rates are higher than the absolute survival rates, and the differences increase with increasing age. Nevertheless, five-year relative survival rates also decrease with age for most forms of cancer, but the age gradient is generally much less pronounced than for the absolute survival rates. Among women with breast cancer and men with prostate cancer, relative survival is highest in the intermediate age groups.

Table 3 shows crude and adjusted five-year absolute survival with complete case ascertainment and expected estimates of crude and adjusted five-year absolute survival assuming selective under-ascertainment of patients dying within five years or of patients surviving at least five years after diagnosis by 20%. The true crude five-year survival rates ranged from 7.2% (lung cancer) to 79.8% (thyroid cancer). They are overestimated by between 1.63% units and 5.53% units in situations with selective under-ascertainment of 20% of patients dying within five years and underestimated by between 0.74% and 3.80% units in situations with selective under-ascertainment of

Table 1
Age distribution (% by age group) of patients diagnosed with common forms of cancer above 14 years of age in Finland in 1990–1994

	n[a]	Age group				
		15–44	45–54	55–64	65–74	75+
Colon	5313	5.4	7.9	16.8	28.4	41.5
Lung	10,305	1.9	7.5	22.0	40.8	27.8
Breast	13,572	12.7	25.3	21.7	19.9	20.5
Ovaries	2268	10.5	16.1	23.6	27.1	22.8
Prostate[b]	8535	1.4	12.4	37.7	39.0	9.5
Brain	1567	31.2	17.7	19.3	20.6	11.1
Thyroid gland	1487	34.7	20.6	15.7	15.1	14.0
Lymphomas	2446	13.4	14.7	19.8	25.1	27.0

[a] Total number of patients in 1990–1994.
[b] Age groups are 15–54, 55–64, 65–74, 75–84, and 85+.

Table 2
Five-year absolute and relative survival by age group of patients diagnosed with common forms of cancer in Finland 1990–1994

	Five-year absolute survival					Five-year relative survival				
	Age group					Age group				
	15–44	45–54	55–64	65–74	75+	15–44	45–54	55–64	65–74	75+
Colon	71.4	60.1	50.8	43.1	28.3	72.1	61.5	54.2	50.9	49.2
Lung	14.7	10.3	11.1	7.4	2.5	14.9	10.7	12.1	9.1	4.5
Breast	78.7	84.8	83.6	71.4	43.8	79.3	86.0	86.5	79.8	71.5
Ovaries	67.6	49.7	36.2	25.6	11.8	68.0	50.5	37.5	28.4	18.6
Prostate[a]	43.2	57.1	56.0	38.0	15.9	45.0	63.2	71.1	66.8	51.9
Brain	64.6	31.6	15.2	5.0	5.2	65.2	32.4	16.2	5.8	8.3
Thyroid gland	98.4	94.8	83.7	59.4	29.3	99.1	96.5	87.8	67.5	48.0
Lymphomas	73.7	62.7	50.0	32.5	14.7	74.5	64.3	53.4	38.4	25.0

[a] Age groups are 15–54, 55–64, 65–74, 75–84, and 85+.

Table 3
Crude and adjusted five-year absolute survival with complete case ascertainment and expected estimates of crude and adjusted five-year absolute survival, assuming selective under-ascertainment of patients dying within five years or of patients surviving at least five years after diagnosis by 20%

Cancer site	Complete case ascertainment		Under-ascertainment of dying patients by 20%				Under-ascertainment of surviving patients by 20%			
	Crude	Adjusted[a]	Crude	Bias	Adjusted[a]	Bias	Crude	Bias	Adjusted[a]	Bias
Colon	41.14	41.54	46.58	+5.44	46.61	+5.07	37.36	−3.78	38.14	−3.40
Lung	7.22	7.59	8.85	+1.63	9.26	+1.67	6.48	−0.74	6.82	−0.77
Breast	72.71	71.74	76.89	+4.18	75.52	+3.71	69.08	−3.63	68.81	−2.93
Ovaries	33.23	33.71	38.32	+5.09	38.11	+4.40	29.43	−3.80	30.47	−3.24
Prostate	45.13	45.40	50.66	+5.53	50.57	+5.17	42.26	−2.87	42.78	−2.62
Brain	30.29	32.26	35.15	+4.86	35.72	+3.46	27.05	−3.24	29.85	−2.41
Thyroid gland	79.81	80.38	83.17	+3.36	82.53	+2.15	76.30	−3.51	78.62	−1.76
Lymphomas	41.09	40.93	46.53	+5.44	45.32	+4.39	37.45	−3.64	38.16	−2.77

[a] Adjusted to the age structure of patients diagnosed in 1985–1989 according to the method by Brenner *et al.* [10].

20% of patients surviving at least five years following diagnosis.

The true age adjusted five-year survival rates are generally quite similar to the crude ones, due to the similarity of the age distribution of patients diagnosed in 1985–1989 and 1990–1994 for most forms of cancer. Only for brain cancer, the true adjusted five-year survival rate is ≈2% units higher than the true crude five-year survival rate as the proportion of younger patients, who have a much better prognosis, was somewhat higher in 1985–1989 than in 1990–1994. Although the adjusted survival rates are also biased by selective under-ascertainment of cancer patients, the bias is smaller than for the crude survival estimates in most cases. The differences in the bias between crude and adjusted survival rates strongly varied by cancer site, ranging from negligible or modest differences to reduction of bias by approximately half for patients with thyroid cancer, whose prognosis strongly varied by age.

With overestimation of crude relative survival up to 7.57% units and underestimation up to 4.48% units, the bias due to selective under-ascertainment of cases is even larger for relative than for absolute survival rates (see Table 4). Again, the bias may be reduced to some degree by age adjustment, albeit to a lesser extent than for the absolute survival rates. The differences in the biases are very small for cancers with only minor to moderate variation in age specific relative survival rates (e.g. lung cancer) or for cancers with no clear trend in age specific relative survival rates (e.g. prostate cancer) but the bias may still be more than one third smaller for cancers with a strong age gradient in relative survival (e.g. thyroid cancer).

Selective under-ascertainment of patients aged 75 years or older by 50% (regardless of their prognosis) may also lead to substantial overestimation of crude five-year absolute survival rates, ranging from 0.75% units to 4.12% units for the eight cancers assessed in this analysis (see Table 5). The reason for that pattern is the poorer prognosis of patients in this age group, leading to stronger under-ascertainment of dying patients than of surviving patients. This type of bias could be fully overcome by age adjustment. The same also applies to relative survival rates, which overall would be less affected by selective under-ascertainment of older patients, than absolute survival rates even in the crude analysis.

4. Discussion

In this paper, we assessed the potential bias in survival estimates resulting from various types of selective

Table 4
Crude and adjusted five-year relative survival with complete case ascertainment and expected estimates of crude and adjusted five-year relative survival, assuming selective under-ascertainment of patients dying within five years or of patients surviving at least five years after diagnosis by 20%

Cancer site	Complete case ascertainment		Under-ascertainment of dying patients by 20%				Under-ascertainment of surviving patients by 20%			
	Crude	Adjusted[a]	Crude	Bias	Adjusted[a]	Bias	Crude	Bias	Adjusted[a]	Bias
Colon	53.31	53.55	59.92	+6.61	59.98	+6.43	48.83	−4.48	49.24	−4.31
Lung	9.18	9.50	11.24	+2.06	11.59	+2.09	8.26	−0.92	8.54	−0.96
Breast	81.62	81.27	85.79	+4.17	85.46	+4.19	78.19	−3.43	78.04	−3.23
Ovaries	37.73	38.20	43.27	+5.54	43.16	+4.96	33.58	−4.15	34.55	−3.65
Prostate	66.52	66.50	74.09	+7.57	74.00	+7.50	62.72	−3.80	62.73	−3.77
Brain	33.34	35.22	38.48	+5.14	38.99	+3.77	29.87	−3.47	32.60	−2.62
Thyroid gland	87.10	87.49	90.09	+2.99	89.75	+2.26	84.08	−3.02	85.64	−1.85
Lymphomas	49.18	49.14	55.19	+6.01	54.35	+5.21	45.18	−4.00	45.85	−3.29

[a] Adjusted to the age structure of patients diagnosed in 1985–1989 according to the method by Brenner *et al.* [10].

Table 5
Expected estimates of crude and adjusted five-year absolute and relative survival assuming selective under-ascertainment of patients aged 75 years or older by 50%

	Five-year absolute survival				Five-year relative survival			
	Under-ascertainment of patients ⩾75 years by 50%				Under-ascertainment of patients ⩾75 years by 50%			
	Crude	Bias	Adjusted[a]	Bias	Crude	Bias	Adjusted[a]	Bias
Colon	44.49	+3.35	41.54	–	54.19	+0.88	53.55	–
Lung	7.97	+0.75	7.59	–	9.73	+0.55	9.50	–
Breast	76.01	+3.30	71.74	–	82.51	+0.89	81.27	–
Ovaries	35.97	+2.74	33.71	–	39.49	+1.76	38.20	–
Prostate	48.81	+3.68	45.40	–	67.08	+0.56	66.50	–
Brain	31.77	+1.48	32.26	–	34.34	+1.00	35.22	–
Thyroid gland	83.61	+3.80	80.38	–	89.11	+2.01	87.49	–
Lymphomas	45.21	+4.12	40.93	–	51.81	+2.63	49.14	–

[a] Adjusted to the age structure of patients diagnosed in 1985–1989 according to the method by Brenner *et al.* [10].

under-ascertainment of cancer cases, and we demonstrate that this bias may often be substantially smaller for age adjusted cancer survival rates than for crude cancer survival rates.

The types of under-ascertainment assessed in this paper are special prototypes, which are unlikely to occur in "pure form" in practice. However, assessment of such prototypes is helpful to demonstrate and understand the mechanisms and magnitude of biases resulting from various sources. In practice, combinations of the prototypes addressed in this paper would be most likely. In particular, a combination of selective under-ascertainment of dying patients (as assessed in Tables 3 and 4) and of older patients (as assessed in Table 5) would often appear to be the most likely pattern. In such situations, the extent of bias reduction would be expected to lie between the values derived for both prototypes. As bias reduction may already be substantial for some cancers even if under-ascertainment of cases was solely related to prognosis (regardless of age), and bias would be entirely overcome if under-ascertainment was strictly age related (regardless of prognosis), age adjustment appears to be a useful tool to reduce bias due to selective under-ascertainment of cases in general.

Clearly, the main reason for age adjustment of cancer survival rates is to correct for distortions from differential age distribution of cancer cases when comparing cancer survival rates between countries or regions, or in time series analyses of cancer survival. Our results imply, however, that age adjustment may also have the very welcome additional "side-effect" to help correct for bias from various degrees of under-ascertainment of cases. This may be another argument for the more widespread application of age adjustment in analyses of cancer survival.

When looking at our results, the following limitations must be kept in mind. To focus on the main principles and due to space limitation, results were shown for eight common forms of cancer only. However, we obtained remarkably similar patterns and results in analogous analyses for a large variety of other common forms of cancer. Similarly, results were shown for five-year survival rates only, because these are the most commonly reported outcome measures of patients with cancer. Again, very similar patterns were observed in analogous analyses of 10-year survival among patients diagnosed in 1985–1989. In these analyses, the bias reduction by age adjustment tended to be even stronger. Finally, results were only shown for traditional, "cohort-wise" analysis of survival. There is no obvious reason, however, why the observed patterns should not apply to period analysis as well, which was introduced a few years ago to derive more up to date estimates of cancer patient survival [16]. The only major difference may be, that under-ascertainment of cases in recent years is often the main concern in period analysis, whereas cohort analysis would be more affected by under-ascertainment of cases in earlier years of registration as recently shown [1].

In summary, our analysis suggests that reduced susceptibility to bias due to selective under-ascertainment of cases is an additional argument for application of age adjustment of cancer survival rates. This argument may not only apply to comparative analyses of cancer survival between registries (where age adjustment is often routinely applied anyway), but also when results from a single registry are reported, be it in form of time series analyses or for a single time period.

Conflict of interest statement

None declared.

Acknowledgement

The work of Hermann Brenner was partly supported by a grant from the German Cancer Foundation

(Deutsche Krebshilfe), Project No. 70-3166-Br 5. Timo Hakulinen's work was supported by a grant from the Academy of Finland.

References

1. Brenner H, Hakulinen T. Population-based monitoring of cancer patient survival in situations with imperfect completeness of cancer registration. *Brit J Cancer* 2005, **92**, 576–579.
2. Stotter A, Bright N, Silcocks PB, *et al*. Effect of improved data collection on breast cancer incidence and survival: reconciliation of a registry with a clinical database. *BMJ* 2000, **32**, 214.
3. Monnet E, Faivre J, Raymond L, *et al*. Comparability of colorectal cancer survival data in three European population-based registries. *Eur J Cancer Prev* 1998, **7**, 127–134.
4. Prior P, Woodman CB, Collins S. International differences in survival from colon cancer: more effective care *versus* less complete cancer registration. *Br J Surg* 1998, **85**, 101–104.
5. Kant AK, Glover C, Horm J, *et al*. Does cancer survival differ for older patients? *Cancer* 1992, **70**, 2734–2740.
6. Vercelli M, Quaglia A, Casella C, *et al*., EUROCARE Working Group. Relative survival in elderly cancer patients in Europe. *Eur J Cancer* 1998, **34**, 2264–2270.
7. Vercelli M, Capocaccia R, Quaglia A, *et al*., EUROCARE Working Group. Relative survival in elderly European cancer patients: evidence for health care inequalities. *Crit Rev Oncol Hematol* 2000, **35**, 161–179.
8. Brenner H, Arndt V. Recent increase in cancer survival according to age: higher survival in all age groups, but widening age gradient. *Cancer Causes Control* 2004, **15**, 903–910.
9. Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population-based cancer registry. Experience in Finland. *Acta Oncol* 1994, **33**, 365–369.
10. Brenner H, Arndt V, Gefeller O, *et al*. An alternative approach to age adjustment of cancer survival rates. *Eur J Cancer* 2004, **40**, 2317–2322.
11. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. *Monogr Natl Cancer Inst* 1961, **6**, 101–121.
12. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982, **39**, 933–942.
13. Berrino F, Capocaccia R, Coleman MP, Estève J, Gatta G, Hakulinen T, Micheli A, Sant M, Verdecchia A, eds. Survival of cancer patients in Europe: the EUROCARE-3 study. *Ann Oncol* 2003, **14** (Suppl. 5), v1–v155.
14. Brenner H, Hakulinen T, Gefeller O. Computational realization of period analysis for monitoring cancer patient survival. *Epidemiology* 2002, **13**, 611–612.
15. Arndt V, Talbäck M, Gefeller O, *et al*. Modification of SAS macros for a more efficient analysis of relative survival rates. *Eur J Cancer* 2004, **40**, 778–779.
16. Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. *Cancer* 1996, **78**, 2004–2010.