Chapter 3

Loss-adjusted hospital and population-based survival of cancer patients

Ganesh B, Swaminathan R, Mathew A, Sankaranarayanan R and Hakama M

Abstract

This chapter presents formulae that methodologically adjust for losses, and gives examples describing magnitude of bias in survival estimates without such adjustment. Loss-adjusted survival is estimated under the assumption that survival of patients lost to follow-up is the same as that for patients with known follow-up time and similar characteristics of different prognostic factors at first entry. The observed number of losses to follow-up is then relocated into expected numbers of death and survivors on this basis. Standard methods, such as the actuarial one, are then applied with the sum of observed and expected outcome events. A total of 336 hospital series of treated new breast cancer cases from Mumbai with 24% lost to follow-up revealed a substantial bias of 7 per cent units for 3-year survival estimated with (54%) and without (61%) loss-adjustment. Stepwise adjustment of losses established that increasing the number of prognostic factors explained the bias better. Population-based series comprising 13 371 cases of top ranking cancers from Chennai, with loss to follow-up ranging from 7–24%, revealed negligible bias, ranging from 0–2% in 5-year survival by the loss-adjusted approach for different cancers. Data source seems to affect the need for loss-adjustment, and the loss-adjusted approach is recommended when hospital-based cancer registry data of a low- or medium-resource country are used to evaluate the outcome of cancer patients.

Introduction

Cancer survival is the main indicator of outcome of cancer health services or treatment, and an important component in maintaining cancer control activities [1]. Cancer registries have long served as potential sources of data for estimating survival. Hospital-based cancer registries usually report survival of a selected series of treated patients that are registered in a hospital or group of hospitals without specific coverage of geographical area or background population. On the other hand, population-based cancer registries, which include all incident cases treated or not from a specific geographical area, usually report average survival in specific regions. Cancer survival reported from both settings may have different perspectives, but estimation of survival rates is routinely done using standard life table approaches such as the actuarial [2] or Kaplan-Meier [3] methods.

The actuarial method [2] of estimating survival by follow-up time allows utilization of all information independent of the length of follow-up of an individual patient, so that even recently diagnosed patients contribute to long-term survival. Patients who have a potential follow-up shorter than the time of the maximum estimated survival are "censored" cases. Censored cases are usually withdrawals, surviving at date of last follow-up: this date can be either individual for each patient or a common closing date for all patients. However, censorship in terms of losses to follow-up takes place if follow-up fails before this potential withdrawal. There is a qualitative difference between these two groups of censored cases.

Losses to follow-up may cause major bias. This holds true if the losses are common and correlated with the patient prognosis or survival. In most low- or medium-resource countries, such losses are common due to deficiencies in health infrastructure and recording of health statistics. The losses are also likely to be related to the patient's prognosis: low social status is related to lack of continuous patient surveillance; extent of disease is related to the motivation of follow-up, etc. Hence, this correlation, explained by information on prognostic factors, can be utilized to correct survival estimates.

A method to estimate loss-adjusted survival rates corrected, for possible bias due to losses to follow-up, is described here through two examples, one each from hospital-based and population-based registry settings. The loss-adjusted survival results are compared with the crude actuarial estimate to demonstrate the magnitude of bias.

Methods

Follow-up

Follow-up was carried out by passive and active methods. The passive approach was by data linkage either with patients' records on regular follow-up at the outpatient clinic and/or with mortality data from the vital statistics division. The active approach was by contacting the patients or their families directly by means of postal/telephone/e-mail/house visit enquiries for information on survival status.

Determinants of loss to follow-up or survival

Categorical factors (like age, sex, literacy status, tumour stage, treatment, etc.), each with reference and subcategory levels, that have the potential to influence either follow-up (complete or lost to follow-up) or survival (alive or dead) were first determined by using test of proportions (univariate only), logistic regression (unifactorial or multifactorial) or Cox proportional-hazard model (univariate or multi-factorial using survival time information). A differential pattern of follow-up or survival outcome, either between factors or within subcategories of factors, would indicate an association of non-random nature.

Estimation of loss-adjusted survival rate - stratified method

The life table method estimates annual survival during a given follow-up year by specifying four types of events including the outcome experienced by the patient: surviving throughout the year; dying (outcome) during the year; withdrawn alive, where patient was known to be alive at closing date of follow-up; and loss to follow-up, where the known survival time terminates during the follow-up year, but before closing date. Unlike traditional survival analysis, which grouped withdrawals and losses together, the proposed method for estimating loss-adjusted survival differentiated the two. For the time being, methods are developed for potential follow-up time of all subjects equalling the time for which survival is estimated. In other words, potential follow-up time for all cases would have to be five years to estimate 5-year loss-adjusted survival rate.

Every prognostic stratum is composed of a unique combination of subcategories of all identified determinants of follow-up or survival. In the estimation of loss-adjusted survival, it is assumed that those lost to follow-up in specific prognostic stratum have the same probability of death as others still remaining under observation and belonging to the same stratum. At any given follow-up time, the observed numbers of losses to follow-up in each stratum are relocated into expected numbers of deaths, withdrawals and survivors on the basis of observed survival in those without loss to follow-up in the same stratum. The actuarial method, or any other, is then applied to the sum of observed and expected events.

In the follow-up interval i in prognostic stratum j, there will be n_{ij} patients alive at beginning of interval, of whom d_{ij} will die, w_{ij} will be withdrawn alive and l_{ij} will be lost to follow-up during the interval. Since potential follow-up exceeds i intervals for all patients, w_{ij} = 0. The number with complete follow-up, n'_{ij} , is then given by:

$$n'_{ij} = n_{ij} - l_{ij.}$$

The proportion dying with complete follow-up, q'_{ij} given the prognostic factors $x_i,.....x_k$, is first estimated for patients not lost to follow-up, n'_{ij} , in the interval i:

$$q'_{ij} = \frac{d_{ij}}{n'_{ij}}.$$

The expected number of deaths in patients lost for follow-up in interval i is:

$$d'_{ij} = q'_{ij}\ l_{ij}$$

and the expected proportion of deaths in the n_{ij} cases is:

$$q_{ij} = \frac{(d_{ij} + d'_{ij})}{n_{ij}}$$

$$= \frac{D_{ij}}{n_{ij}}.$$

The procedure is repeated for the next interval ($i = i$ +1) as follows:

$n'_{(i+1)j} = n_{ij} - D_{ij} - l_{(i+1)j}$ and with

$l'_{(i+1)j} = l_{(i+1)j} + l_{ij} - d'_{(i+1)j}$ with

$d'_{(i+1)j} = q'_{(i+1)j}\ l'_{(i+1)j}$ and for the other prognostic strata.

Accumulating over prognostic strata will result in an annual loss-adjusted rate:

$$q_i(Loss\ \ Adjusted) = \frac{\sum_j D_{ij}}{\sum_j n_{ij}}$$

and the cumulative loss-adjusted survival probability is:

$$P_i(Loss\ \ Adjusted) = (1-q_1)(1-q_2).....(1-q_i).$$

Logistic regression approach to estimate expected deaths among loss to follow-up

The correction of bias in survival estimation adjusted for loss to follow-up is optimal when it is determined by including as many factors as possible. An increase in number of determinants (factors with subcategories) of follow-up or survival would result in a corresponding increase in the number of prognostic strata. Cross-tabulation of all of these factors simultaneously would require adequate sample size to keep a majority of prognostic strata non-empty. Adjusting all factors simultaneously by logistic regression is a simplification of the computational procedure to estimate expected deaths among lost to follow-up and offers maximal effect in reducing the bias.

The proportion dying in the $\sum_j n'_{ij}$ patients followed completely during the interval is:

$$q'_i = \frac{\exp(\mu'_i)}{(1+\exp(\mu'_i))}$$

where

$$\mu'_i = \beta_{0i} + \beta_{1i}x_{1i} + \beta_{2i}x_{2i} + ... + \beta_{ki}x_{ki}$$

is a linear combination of the determinant or prognostic factors. The above methods are described in detail elsewhere [4,5].

Other approaches

Loss-adjusted survival can also be estimated using the Kaplan-Meier approach [6]. Stratum-specific expected deaths are estimated and the Kaplan-Meier curve is corrected at time points when the expected deaths occur.

Results

Example 1: Hospital-based cancer registry series

A total of 336 new cases of female breast cancer cases that were diagnosed and received complete treatment at Tata Memorial Hospital, Mumbai (Bombay), India, in 1985 and followed-up until 1988 formed the study population. These cases were allocated to 64 strata involving four factors associated with follow-up or prognosis: age (in completed years: <45, 45–54, 55–64, 65+ years); stage of disease (TNM staging classification: I, II, III, IV); type of treatment (chemotherapy: without, with); place of residence (Mumbai: residents, non-residents). Outcome event with respect to follow-up was loss to follow-up <3 years from diagnosis, and outcome event for loss-adjusted survival was death due to any cause.

Patients below 55 years of age comprised 65%, with an overall mean of 49 years (Table 1). There was an equal distribution of resident and non-resident patients from Mumbai city. A majority were diagnosed in stage II (48%) followed by stage III (37%) of the disease. About 58% of the patients were treated with either surgery or radiotherapy or in combination but not with chemotherapy, while the remaining 42% were treated with chemotherapy either alone or in combination with other modalities. Differential pattern of proportion (%) or risk (odds ratio) of loss to follow-up by different prognostic factor categories was forthcoming. The proportion of patients lost to follow-up was not very different between subcategories of age and type of treatment, with 0 to 30% increased risk over corresponding reference categories that was statistically not significant. The proportion lost to follow-up was doubled among non-residents versus residents of Mumbai, with two- to three-fold increased risk that was statistically significant. The risk was two to three times higher among stage III or IV patients and 50% higher among stage II compared to stage I patients, but not statistically significant (Table 1). The findings suggest an association between these prognostic factors and loss to follow-up.

The data was further analysed to estimate loss-adjusted survival by stratification of two or three factors at a time and by logistic regression approaches. Survival was estimated at the end of 3-year follow-up by actuarial method without and with adjustment for loss to follow-up (Table 2). The 3-year survival obtained by loss-adjustment showed lower survival compared to rates obtained by standard actuarial assumption without specific adjustment for loss to follow-up. The bias in survival estimation is represented as the difference in per cent units of survival rates (%) without and with loss-

adjustment for each factor. This varied from 5.4 for patients aged 55 to 64 years to 8.6 for those aged <45 years. The bias was lesser among Mumbai residents (3.2) than non-residents (8.8). Three-year loss-adjusted survival was higher among residents (56.2%) than non-residents (54.4%), but this was the opposite for corresponding survival figures without loss-adjustment (59.4% and 63.2%), respectively. A decrease in survival (Table 2) and increase in proportion of lost to follow-up (Table 1) with severity of disease was forthcoming, which indicated a positive association between risk of dying and loss to follow-up in all disease stages. Loss-adjusted survival was greater in stage I patients, but lesser in other stages, compared to respective survival estimates without loss-adjustment. Following the elimination of bias by loss-adjustment, the difference in loss-adjusted survival between stages I and III patients increased from 51 per cent units to 61 per cent units (Table 2). The proportion of deaths in the chemotherapy group was twofold more than in the non-chemotherapy group. The comparison between actuarial and loss-adjusted survival showed that the adjusted unbiased difference between the two groups was bigger (43 per cent units) than the unadjusted ones (38 per cent units).

The variable extent of bias in survival estimation that could be elicited in the presence of loss to follow-up by utilizing information from one to four prognostic factors is shown stepwise for all cases in Table 3. The unadjusted actuarial 3-year survival was 61%. The loss-adjustment yielded a decrease of 7 per cent units in survival when all four prognostic factors were considered simultaneously by logistic regression method. The stepwise introduction of each of the prognostic factors into the adjustment procedure, by stratified method of estimating loss-adjusted survival, increased the correction of bias as follows: 1.7 per cent units when adjusted only for residential status; 2 per cent units when age was added; 3.8 per cent units when stage was added to the previous two factors; and 4.7 per cent units when all factors were adjusted.

Example 2: Case series from Chennai population-based cancer registry

A total of 13 371 cases comprising cancers of the uterine cervix (3134), female breast (1923), stomach (1845), oesophagus (1403), lung (1237), mouth (1202), lymphomas (768), tongue (670), leukaemias (668), and of ovary (521) ranked within the top ten in

Table 1. Number and proportion (%) of patients and losses at 3 years and risk (odds ratio) of loss to follow-up with 95% confidence interval by patient characteristics among female breast cancer patients diagnosed in Tata Memorial Hospital, Mumbai, India, in 1985 and followed through 1988

Patient characteristics	Patients (n=336)		Lost to follow-up (n=80; 24%)		Odds ratio (95% CI)
	Number	%[a]	Number	%[b]	
Age at diagnosis					
≤ 44 years	101	30	22	22	1.0*
45–54	117	35	29	25	1.2 (0.6–2.3)
55–64	77	23	19	25	1.2 (0.6–2.5)
65+ years	41	12	10	24	(0.5–2.9)
Residential status (Mumbai city)					
Residents	169	50	26	15	1.0*
Non-residents	167	50	54	32	2.6 (1.5–4.6)[$]
Stage of disease (TNM summary)					
I	29	9	4	14	1.0*
II	160	48	30	19	1.5 (0.5–5.6)
III	126	37	40	32	2.9 (0.9–10.6)
IV	21	6	6	29	2.5 (0.5–12.9)
Treatment					
With chemotherapy	194	58	42	22	1.0*
Without chemotherapy	142	42	38	27	1.3 (0.8–2.3)

[a] Percentage of total breast cancer cases;
[b] Percentage of total cases in respective categories;
CI: Confidence interval;
** Reference category;*
[$] p=0.05.

Table 2. Number and proportion (%) of patients and deaths and comparison of 3-year survival with and without adjustment for loss to follow-up by patient characteristics among female breast cancer patients diagnosed in Tata Memorial Hospital, Mumbai, India, in 1985 and followed through 1988

Patient characteristics	Number of patients	Deaths		3-year survival %	
		Number	%[a]	Actuarial assumption	Loss-adjusted by logistic regression*
Age at diagnosis					
≤ 44 years	101	34	34	60.1	51.5
45–54	117	42	36	56.7	48.7
55–64	77	20	26	67.7	62.3
65+ years	41	12	29	65.4	58.5
Residential status (Mumbai city)					
Residents	169	60	35	59.4	56.2
Non-residents	167	48	29	63.2	54.4
Stage of disease (TNM summary)					
I	29	2	9	92.2	93.2
II	160	36	22	74.4	71.2
III	126	55	44	41.2	31.8
IV	21	15	71	0.0	0.0
Treatment					
With chemotherapy	194	39	20	76.6	71.2
Without chemotherapy	142	69	48	38.1	28.2

[a] *Percentage of total cases in respective categories;*
* *Adjusted for other factors in the table.*

Table 3. Comparison of 3-year survival without loss-adjustment by actuarial assumption, stepwise loss- adjustment of factors using stratified method and loss-adjustment using all factors together by logistic regression for all female breast cancer patients diagnosed in Tata Memorial Hospital, Mumbai, India, in 1985 and followed through 1988

Loss-adjustedment of factors	3-year survival %
Without loss-adjustment and using actuarial assumption only	61.2
Loss-adjustment done by stratification	
Residential status	59.5
Residential status and age at diagnosis	59.2
Residential status, age at diagnosis and stage of disease	57.4
Residential status, age at diagnosis, stage of disease and treatment	56.5
Loss-adjustment done by logistic regression	
Residential status, age at diagnosis, stage of disease and treatment	54.5

the Population-Based Cancer Registry, Chennai, India, during 1990–1996 and followed-up until 2001 formed the study population.

The determinants of loss to follow-up at less than 5 years from diagnosis for each site were identified using Cox proportional-hazard model by following the method outlined in Chapter 2 of this publication. Five-year loss-adjusted absolute survival of patients through stratified method was estimated by allocating cases to 128 strata defined by 4 factors (with reference and subcategories): age at diagnosis (<45, 45–54, 55–64 and 65+ years); literacy status based on years of education (Nil, ≤ 5, 6–12 and ≥ 12 years); clinical extent of disease as a surrogate for tumour stage (localized, regional, distant metastasis and unknown); treatment status (no or unknown and yes). Outcome event was death due to any cause.

Table 4 gives the proportion of cases lost to follow-up and comparison of 5-year absolute survival estimated with and without adjustment for loss to follow-up for each cancer site. The losses ranged between 7% (oesophagus) and 24% (ovary) for different sites. Loss-adjusted survival was consistently lesser than the corresponding unadjusted estimate for all sites. Bias in survival estimation in the presence of non-random loss to follow-up, expressed in terms of absolute difference between survival (%) estimates obtained with and without loss-adjustment was minimal, ranging between 0.2 to 1.7 per cent units for different cancer sites.

Discussion

The success of cancer treatment is, as a rule, measured by survival. Population-based survival reflects the availability, development of and accessibility to cancer health services in a region. Survival based on hospital series reflects the impact of clinical services specific to the hospital. In both instances, high-level completeness of ascertainment of mortality data is an important prerequisite, and when such completeness cannot be assured, survival rates should be carefully interpreted [7,8].

Conventionally, estimation of survival was done using life table approaches by either actuarial [2] or Kaplan-Meier [3] methods. Both methods utilize observed survival time independently of whether it ends at the death of a patient. Patients withdrawn alive at closing date provide censored information that is unbiased, since closing date is independent from probability of death. If this is not true, Hakulinen [9] and Brenner [10] give means to adjust for withdrawal pattern and to correct for effects of improvement of survival by time.

Losses to follow-up because of reasons other than closing date (e.g., migration) are often few in developed countries and are dealt with identically as withdrawals. This is not justified if the losses are many and are correlated with risk of death. Distance from clinical care facility increases the likelihood of not undergoing a follow-up examination, as does serious morbidity and poverty. The factors in failure to obtain follow-up data are the same. Therefore, it is likely that patients lost to follow-up have poor prognosis and could not be compared with those under follow-up and surveillance. The direction in bias may also be the other way: those lost to follow-up have a better survival than those under follow-up, as was shown in our example on stage I breast cancer hospital series patients.

Our example from a hospital series shows that the bias due to losses may be substantial. Mathew[6] showed similar differences by applying loss-adjustment in the Kaplan-Meier survival method for hospital series ovarian cancer patients. Much of the original deficiencies in the hospital data were, however, removed by active follow-up using a postcard enquiring the vital status of patient. Only marginal adjustment effect appeared after the enquiry. However, in the example involving breast cancer hospital series, a large bias still existed after such attempts of active follow-up. On the other hand, the example involving population-based series of several cancers revealed negligible bias. In both

Table 4. Number of incident cases, proportion (%) lost to follow-up and comparison of 5-year absolute survival with and without loss-adjustment for top-ranking cancers in a population-based cancer registry, Chennai, during 1990–1996 and followed through 2001

Cancer site/type	Number of incident cases	Lost to follow-up %	5-year survival %		Absolute difference in survival
			No loss-adjustment	With loss-adjustment	
Cervix	3134	21.8	52.1	50.4	1.7
Breast	1923	20.7	39.5	39.1	0.4
Stomach	1845	8.0	9.4	8.7	0.7
Oesophagus	1403	6.7	7.7	7.5	0.2
Lung	1237	7.8	8.2	8.1	0.1
Mouth	1202	11.6	30.1	29.1	1.0
Lymphomas	768	11.5	26.5	25.6	0.9
Tongue	670	13.0	20.2	18.9	1.3
Leukaemias	668	8.2	19.8	19.2	0.6
Ovary	521	24.0	25.7	24.2	1.5

instances, loss-adjusted survival was lesser than the actuarial estimate without adjustment indicating that under-ascertainment of deaths among loss to follow-up cases may be the problem. Most population-based cancer registries are based on systems that integrate linkage or collection of mortality data as a routine and hence result in small differential bias only [5]. Hence, the data source seems to affect the need for loss-adjustment, and the problem may be more substantial in hospital-based cancer registries and clinical series. The loss-adjusted approach is likely to be useful especially when hospital-based cancer registry data of a low- or medium-resource country are used to evaluate the outcomes of cancer patients.

One may conclude that if routine follow-up is poor, the first priority is to increase the actual follow-up visits on humanitarian and scientific grounds. The second is to improve the data by instituting rigorous active follow-up measures. The improvement of data by these means may indirectly improve routine follow-up activity. Analytical methods to correct the survival data with adjusting for losses are to be used in surveillance and evaluation and in scientific comparisons. However, such means do not directly improve human health, but have the potential to improve the organization itself.

References

1. WHO/IARC. *Cancer Statistics: Report of a WHO/IARC Expert Committee. WHO Technical Report Series No. 632*. World Health Organization, Geneva, 1979.

2. Cutler SJ, Ederer F. Maximum utilization of the life table method in analysing survival. *J Chron Dis* 1958;8: 699–712.

3. Kaplan EL and Meier P. Non parametric estimation from incomplete observation. *J Amer Stat Assoc* 1958;53:457–481.

4. Ganesh B. *Effect of loss to follow-up in estimating survival rates. Acta Universitatis Tamperensis, Ser A, Vol 440*. University of Tampere, Tampere, 1995.

5. Sriamporn S, Swaminathan R, Parkin DM, Kamsa-Ard S, Hakama M. Loss-adjusted survival of cervix cancer in Khon Kaen, Northeast Thailand. *B J Cancer* 2004;91(1):106–110.

6. Mathew A. Removing bias in cancer survival estimates by active follow-up and information on determinants of loss to follow-up. Acta Universitatis Tamperensis, Ser A, vol 525. University of Tampere, Tampere, 1996.

7. Brenner H, Hakulinen T. Implications of incomplete registration of deaths on long-term survival estimates from population-based cancer registries. *Int J Cancer 2009*;125:432–437.

8. Swaminathan R, Rama R, Shanta V. Lack of active follow-up of cancer patients in Chennai, India: implications for population-based survival estimates. *Bull World Health Organ*. 2008;86:509–515.

9. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics*. 1982;38: 933–942.

10. Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. *Cancer*. 1996;78: 2004–2010.