Eur J Epidemiol (2008) 23:585–592
DOI 10.1007/s10654-008-9278-7

METHODS

Survival analysis: up from Kaplan–Meier–Greenwood

Olli S. Miettinen

Received: 16 April 2008 / Accepted: 18 July 2008 / Published online: 9 September 2008

Abstract In the type of survival analysis that now is routine, only the points of follow-up at which deaths from the cause at issue occurred make contributions to the Greenwood standard error (SE) of the survival rate's Kaplan–Meier (KM) point estimate. An equivalent of this 'KMG' analysis draws from defined subintervals of the survival period being addressed. The data on each subinterval consist of the number of deaths from the cause at issue and the amount of population–time of follow-up, d_j and T_j, together with the duration of the interval, t_j. The KM point estimate is replicated by $\exp[-\sum_j (d_j/T_j)t_j]$, and the KMG interval estimate is replicated by treating the $\{d_j\}$ as a set of point estimates of Poisson parameters $\{\lambda_j\}$, thus taking the SE of $\sum_j (d_j T_j)t_j$ to be $[\sum_j d_j (t_j/T_j)^2]^{1/2}$. In both the KMG analysis and this equivalent of it, the SE used to derive the survival rate's lower confidence limit needs to be augmented by a factor that accounts for the loss of information due to censorings subsequent to the last 'failure' in the survival period at issue. But, SE-based interval estimation of survival rate actually needs to be replaced by a first-principles counterpart of it. A suitable point of departure in this is first-principles asymptotic interval estimation of the Poisson parameter $\lambda = \sum_j \lambda_j$, if not the exact counterpart of this. A confidence limit for the survival rate can then be based on suitable augmentation or contraction of the $\{d_j\}$ set to $\{d_j^*\}$ consistent with a given limit for λ, the corresponding survival-rate limit being $\exp[-\sum_j (d_j^*/T_j)t_j]$. Suitable augmentation is constituted by an identical addition to each $d_j^{1/2}$, suitable contraction by an identical subtraction from each $d_j^{1/2} \geq 1$.

Keywords First principles · Kaplan–Meier · Nelson–Aalen · Survival analysis

O. S. Miettinen (✉)
McGill University, 1020, avenue des Pins ouest,
Montreal, QC, Canada H3A 1A2
e-mail: olli.miettinen@mcgill.ca

O. S. Miettinen
Department of Epidemiology, Biostatistics and Occupational Health, and Department of Medicine, McGill University, Montreal, QC, Canada

O. S. Miettinen
Department of Medicine, Weill Medical College, Cornell University, New York, NY, USA

Introduction: examples give rise to concern

A recent, major study addressed survival following diagnosis of lung cancer in the framework of annual CT (computed tomography) screening for the disease [1]. The interest was in stage-specific (Stage I) and overall rates of surviving lung cancer itself, conditional on not succumbing to some other cause of death; and more specifically, the interest was in the *asymptotes* of these 'cause-specific' rates of survival, viewed as functions of time since diagnosis. For, the aim in screening for a cancer is to provide for curative treatment of the disease in the context of its early, latent-stage diagnosis, and the asymptotic cause-specific survival rates served as measures of the respective rates of *curability* of the cancer, given its diagnosis in the context of the annual CT screening for it (regardless of whether the diagnosis actually resulted from the screening). The interest in these two rates was not comparative but complementary, as the overall rate of curability attainable by means of screening approximately equals the rate of curability of Stage I cases multiplied by the proportion of diagnoses that are achieved in Stage I. For the latter the point estimate was reported [1] to be 85%.

Thus, as a matter of prevailing routines, Kaplan–Meier (KM) rates of long-term survival were derived, and the confidence intervals supplementing these were based on the Greenwood formulation of the standard error (SE) of the KM survival rate. These 'KMG' statistics were derived "with the use of SAS statistical software (version 8)."

A graph in the report showed the KM survival patterns, separately, for "302 participants with Stage I cancer resected within 1 month after diagnosis" and "484 participants with lung cancer [regardless of stage at diagnosis and treatment]." The respective 10-year KMG results were "92% (95% CI, 88–95)" and "80% (95% CI, 74–85)," as given in that Fig. 2.

Those are very impressive results on the curability of lung cancer—if correct. But *are* they correct? Specifically here, are they *statistically* correct?

Relevant to answering that question are the other numbers that were associated with those results and also were given in that Fig. 2. They specified the respective numbers of study subjects that were under follow-up, whether after Stage I diagnosis or any diagnosis of lung cancer, at the beginning of each successive 1-year period after diagnosis. These numbers are shown in Table 1 here. At 9 years they were 7 and 9, respectively, and at 10 years, 1 and 2, respectively. The last deaths from lung cancer occurred in the 5th and 7th years of follow-up, respectively (Table 1).

The principal point of note about these examples statistically is that, as the KM point estimates of the survival rates were constant after the last deaths from lung cancer, so inherently also were their Greenwood SEs, even though ever fewer subjects were under follow-up, very few in the 10th year in particular. To wit, if only one of the study subjects had been under follow-up after the last death from lung cancer, not only the 10-year survival rate's point estimate but its SE also would have been the same as immediately after that death—as though the survival of a single study subject, in the absence of others, could make it certain that no one dies of lung cancer in this interval of post-diagnostic follow-up time. Just as notably, a second study subject entering this period of follow-up and dying of lung cancer in it would have replaced this implicit certainly by a very imprecise 50% point estimate of the rate of survival across this period.

Clearly, something can be seriously wrong with the KMG analysis of survival data, at least in studying a survival rate's asymptotic level. With an equivalent of the KMG analysis as the point of departure, a preferable substitute for the KMG survival analysis is introduced here.

The Kaplan–Meier–Greenwood statistics

The KMG analysis focuses on the points of follow-up time at which deaths from the cause at issue occurred. The *i*th one of these deaths occurred at a time immediately before which some number S_i of survivors still were being followed. With a total of d deaths of interest having been 'observed' to occur during the entire period of follow-up, the KM survival rate [2]—cause-specific, empirical—is the product of the rates at these points, d in number:

$$\widehat{\mathrm{SR}} = \prod_{1}^{d} (S_i - 1)/S_i = \prod_{1}^{d} \widehat{\mathrm{R}}_i.$$

The replication variance of $\widehat{\mathrm{SR}}$ is, to a first-order Taylor series approximation, equal to $(\mathrm{SR})^2$ multiplied by the variance of $\log(\widehat{\mathrm{SR}})$. The variance of $\log(\widehat{\mathrm{SR}})$, in turn, is equal to the sum of the variances of $\log(\widehat{R}_i)$, $i = 1, \ldots, d$,

Table 1 Lung-cancer survival in the I-ELCAP experience [1]

	Year of follow-up										
	1	2	3	4	5	6	7	8	9	10	11
Cohort A											
Entries	302	280	242	191	120	59	34	18	12	7	1
P–T[a]	291	261	216.5	155.5	89.5	46.5	26	15	9.5	4	
Deaths	5	4	4	3	1	0	0	0	0	0	
Cohort B											
Entries	484	433	356	280	183	90	50	28	16	9	2
P–T[b]	458.5	394.5	318	231.5	136.5	70	39	22	12.5	5.5	
Deaths	24	29	15	4	2	0	1	0	0	0	

Reported was survival of a cohort—Cohort A—of 302 subjects diagnosed with clinical Stage I disease and having its resection within 1 month from diagnosis, and of a cohort—Cohort B—consisting of all 484 subjects in whom lung cancer was diagnosed. Shown here are the reported numbers of subjects entering each of the 10 successive years of follow-up and the numbers of lung-cancer deaths in each of those years, supplemented by the corresponding amounts of population–time (P–T, in years) of follow-up

[a] ½(302 + 280) = 291; etc.

[b] ½(484 + 433) = 458.5; etc.

which to a first-order Taylor series approximation are estimable as $(1-\widehat{R}_i)/S_i\widehat{R}_i$. Thus the corresponding approximation to the SE of the KM $\widehat{\mathrm{SR}}$ is

$$\mathrm{SE} = (\widehat{\mathrm{SR}})\Big[\sum_i (1-\widehat{R}_i)/S_i\widehat{R}_i\Big]^{1/2} = (\widehat{\mathrm{SR}})\Big[\sum_i 1/S_i(S_i-1)\Big]^{1/2}.$$

This is the Greenwood SE of the KM survival rate [2, 3]. Involved in this are ML (maximum likelihood) estimates of the variances of the $\{\widehat{R}_i\}$. With the respective unbiased estimates the statistic would have been

$$\mathrm{SE} = (\widehat{\mathrm{SR}})\Big[\sum_i 1/(S_i-1)^2\Big]^{1/2}.$$

Clearly, no contributions to the SE are made by the experiences in time intervals in which no deaths from the cause at issue occurred; and thus, as for the examples here, the paucity of the experience beyond 5 years of follow-up is treated as irrelevant to the precision of the 10-year survival rate's KM point estimate.

Kaplan–Meier–Greenwood-equivalent statistics

Referring back to Table 1, for Cohort A in year 1 of follow-up the *incidence density* [4] of lung-cancer death was (approximately) 5 per $(1/2)(302+280)$y (i.e., 5 per 291 person-years) or 1.7/100y; and for years 2–10, analogously, the rates per 100y were 1.5, 1.8, 1.9, 1.1, 0.0, 0.0, 0.0, 0.0, and 0.0, respectively.

For the corresponding 10-year *cumulative incidence* [4] of lung-cancer death, conditional on not succumbing to any other cause of death, the point estimate from the Cohort A data is

$$\widehat{\mathrm{CI}}_{0,10} = 1-\exp[-(0.017+\cdots+0.011)] = 0.080.$$

Thus the cause-specific survival rate was $1-0.080 =$ 92%, as also was the corresponding KM rate [1].

For Cohort B the corresponding 10 successive incidence rates, per 100y, were (per Table 1 again) 5.2, 7.4, 4.7, 1.7, 1.5, 0.0, 2.5, 0.0, 0.0, and 0.0. These translate into a 10-year integral of 0.23 and its corresponding cause-specific survival rate of $\exp(-0.23) = 79\%$—well consistent with the reported [1] 80% for the KM rate.

Further in the spirit of the KMG analysis, confidence intervals for these rates would be based on the SEs of the primary estimates involved. These estimates here are the incidence-density integrals over the period of follow-up; that is, they are the sums $\sum_j(\widehat{\mathrm{ID}}_j)(1\mathrm{y})$ used in the calculations above. For Cohort A this integral is 0.080 (cf. above). The SE of this, given Poisson distribution of the number of deaths in each interval and ML variance-estimation (à la KMG), is $[5/(291)^2+\cdots+1/(89.5)^2]^2 = 0.021$. The corresponding SE-based 95% two-sided limits for the integral are 0.039 and 0.12. The exponentials of the negatives of these are the SE-based 95% two-sided limits for the survival rate. They are 89% and 96%, in good accord with the reported [1] KMG limits of 88% and 95%. For Cohort B the corresponding results are 74% and 85%, the same as from the KMG analysis.

In terms of this (practical) equivalent of the KMG analysis, the problem initially at issue here takes this form: For an interval of follow-up time in which the empirical $\widehat{\mathrm{ID}}_j$ $(=d_j/T_j)$ involves $d_j = 0$ as an input, this empirical number is treated as though it were the corresponding theoretical—'expected'—number; that is, the empirical $\widehat{\mathrm{ID}}_j = 0$ is treated as though all possible replications of the experience, however small the population–time (T_j), be known to reproduce this interval-specific result. This is unjustifiable; and thus, at least when there are intervals with $d_j = 0$ after the last event of interest and in those intervals ever fewer survivors still are under follow-up, a suitable modification of this KMG-equivalent approach is needed.

The Nelson–Aalen counterparts of the KMG statistics I address in Appendix 1.

Up from the Greenwood standard error

In the data from which the KMG statistics are derived, information on the follow-up of any given person ends at death from the cause at issue—or some other 'failure' of interest—or it ends at 'censoring,' meaning termination of follow-up before potential 'failure' within the survival period at issue. Censorings in this meaning result from deaths due to extraneous causes and from losses to follow-up.

Regarding the KM survival function (of follow-up time since entry into the cohort) and its SE, based on censored data, Borkowf [5] made this "fundamental point":

> The calculation of the KM survival function involves both the *estimation* of the value of that function at each time that an event [a failure] occurs, and the *decision* to carry forward that value until the next time that an event occurs, despite any censorings that may happen in the meanwhile. This carry-forward decision is an important part of the construction of the KM survival function, but it is almost always overlooked.… [W]hile it makes sense that the KM survival function should remain constant between events, its variance should increase with censorings.

So, as for the examples here, it is inherent in the KM survival-function concept that the survival rate attained in Cohort A within the 5th year of follow-up is carried

forward all the way to the end of the 10-year survival period at issue; and the same applies to Cohort B as of the last death, within the 7th year. And given that censorings (galore) occurred in these periods of constant KM survival rate, its purported SE should have increased (substantially) on account of those censorings. But the Greenwood SE remained constant.

Using the SE introduced by Peto et al. [6] as the point of departure, Borkowf [5] presented a modification of this SE such that it indeed does increase as censorings occur. His fine points aside, the basic idea was this: Right after the ith outcome event of interest, the SE can be taken to be the square root of a 'binomial' variance estimate involving the KM survival rate at that time together with a suitably defined 'effective' number of binomial observations:

$$\mathrm{SE}_i = [\widehat{\mathrm{SR}}_i(1 - \widehat{\mathrm{SR}}_i)/(S_0 - c_i)]^{1/2},$$

where S_0 is the initial number of survivors under follow-up (i.e., the size of the cohort) and c_i is the number of censorings before the ith event.

From this it follows that if c increases (toward c_{i+1}) before the next event, this interim increase in censorings from c_i to c_{i+}, increases the SE:

$$\mathrm{SE}_{i+} = [\widehat{\mathrm{SR}}_i(1 - \widehat{\mathrm{SR}}_i)/(S_0 - c_{i+})]^{1/2} \geq \mathrm{SE}_i.$$

After the last event of interest the increase can be very substantial:

In Cohort A, with $S_0 = 291$, right after the last (17th) death, in the 5th year of follow-up, $\widehat{\mathrm{SR}}_i = 0.92$ and $c_i = 291 - 17 - 89 = 185$, translating into (Borkowf) $\mathrm{SE}_i = 0.026$ and its corresponding 95% two-sided confidence interval for the survival rate ranging from 87% to 97%. At 10 years, by contrast, $c_i = 291 - 17 - 1 = 273$, so that $\mathrm{SE} = 0.063$ and the corresponding interval estimate is 80–100% (the nominal upper limit actually exceeding 100%).

Valuable though this Borkowf substitute for the Greenwood SE is, I need to make a (novel) point about its application: When censorings occur without the death of interest occurring in the survivor group still under follow-up, it is not that the amount of evidence about possible increase in the underlying rate of survival is thereby compromized, as the rate's increase is logically impossible; it is information about decrease of the survival rate that is reduced. Thus, the upper confidence limit should not go further up; only, the lower limit should go lower. The need thus is to distinguish between upper SE and lower SE:

$$\text{upper } \mathrm{SE}_{i+} = [\widehat{\mathrm{SR}}_i(1 - \widehat{\mathrm{SR}}_i)/(S_0 - c_i)]^{1/2},$$
$$\text{lower } \mathrm{SE}_{i+} = [\widehat{\mathrm{SR}}_i(1 - \widehat{\mathrm{SR}}_i)/(S_0 - c_{i+})]^{1/2},$$

where $c_i \leq c_{i+} \leq c_{i+1}$. The upper SE is constant between the ith and $(i + 1)$th event, while the lower SE may grow.

Returning to the example above, the upper SE is 0.026 from the last (17th) event, in the 5th year, all the way to 10 years, while the lower SE grows from the 0.026 to 0.063. In consequence, then, the upper two-sided 95% confidence limit remains at 97% throughout this period, while the lower limit declines from 87% to 80% (the point estimate remaining at 92%).

I might add, as a relatively minor point, that in each of these SEs, the denominator of the variance estimate should be reduced by one, to move from the ML estimate to its unbiased counterpart. By the same token, by the way, the elements $\{1/S_i(S_i - 1)\}$ in the Greenwood SE should be replaced by $\{1/(S_i - 1)^2\}$.

That Borkowf substitute for the Greenwood SE naturally has its counterpart in the KMG-equivalent framework outlined above. Borkowf's essential point I take to have been that due to censorings between the ith and $(i + 1)$th event the information about the survival rate at issue is reduced by the factor $(S_0 - c_{1+})/(S_0 - c_i)$. This means that the SE of the ID integral over the survival period at issue needs to be augmented by the square root of the inverse of this factor, with c_i the number of censorings before the last event and c_{i+} the total number of censorings in the survival period at issue. And again, this augmentation is to be applied in one direction only, here to the SE that bears on the upper confidence limit of the ID integral.

Without this correction, the lower limit of the survival rate's 96% two-sided confidence interval was derived from the Cohort A data on ID as $\exp\{-[0.080 + 1.96(0.021)\} = 89\%$. Now the ID integral's upper SE is that 0.021 multiplied by $[(291 - 185)/(291 - 273)]^{1/2} = 2.43$. As a result, 89% is replaced by 84%. This is the result also when applying that 'Borkowf factor,' 2.43, to the corresponding KMG statistics—in contrast to the actual Borkowf lower limit of 80% (cf. above).

Up from the Kaplan–Meier framework

That one of the SE-based confidence intervals above has, for a survival rate, an upper limit in excess of 100% reflects the fact that SE-based confidence intervals for survival rates are not first-principles asymptotic intervals. In the context of KM point estimation, this problem is irremediable.

The KMG-equivalent approach that was outlined above does lend itself to the modifications that provide for first-principles asymptotic interval estimation; and they even allow for an 'exact' counterpart of this.

A first-principles asymptotic interval for the survival rate can be based on the Poisson distribution of $\sum_j \widehat{\lambda}_j = \sum_j d_j$, on a first-principles asymptotic interval for the

parameter of this distribution. Based on the variance-stabilizing transformation, 95% two-sided limits for $\lambda = \sum_j \lambda_j$ are $[(\sum_j d_j)^{1/2} \mp 1.96(1/2)]^2$. These might be preferred to the solutions of $(\sum_j d_j - \lambda)^2/\lambda = (1.96)^2$.

Focusing first on the upper limit for $\sum_j \lambda_j$, the need is to identify a corresponding suitable set $\{d_j^+\}$ such that $\sum_j d_j^+$ coincides with that limit for $\sum_j \lambda_j$. Suitable augmentation of the $\{d_j\}$ to the $\{d_j^+\}$, with that constraint, involves due attention to the respective precisions in the $\{\hat{\lambda} = d_j\}$ set. These precisions are (essentially) identical upon the square-root transformation. Thus, it is appropriate to use $\{d_j^+\} = \{(d_j^{1/2} + d^+)^2\}$, where d^+ represents the suitable augmentation constant, the same for each $\{d_j^{1/2}\}$. The corresponding limit—lower—for the survival rate is the exponential of the negative of the ID integral with the $\{d_j^+\}$ in place of the $\{d_j\}$.

The lower limit for $\sum_j \lambda_j$ translates, quite analogously, to the substitute set $\{d_j^-\}$ and, based on this, to the upper limit for the survival rate. An added constraint in this naturally is that the downward adjustments—$\{d_j^- = (d_j^{1/2} - d^-)^2\}$—are applied to $d_j \geq 1$ only.

For Cohort A, 95% two-sided limits in these first-principles terms can be founded on the limits $[17^{1/2} \pm 1.96(1/2)]^2)]^2 = (26.0, 9.9)$ for $\sum_j \lambda_j$. The $\{d_j^+\}$ set involves $d^+ = 0.409$ as the augmentation of each $d_j^{1/2}$ to $(d_j^+)^{1/2}$, consistent with $\sum_j d_j^+ = 26.0$. The augmented set {7.00, …, 0.17} implies the ID integral 0.206 and, then, the lower limit exp(–0.206) = 81% for the 10-year survival rate. For the other limit, the $\{d_j^-\}$ set, consistent with $\sum_j d_j^- = 9.9$, involves $d^- = 0.453$; and the set resulting from this contraction, {3.18, …, 0.00}, translates into 96% as the survival rate's upper limit. For Cohort B the survival rate's thus-derived limits are 70% and 85%, corresponding to the $\sum_j \lambda_j$ limits 91.8% and 58.1%.

The 95% two-sided *exact* limits for $\sum_j \lambda_j$ are the solutions of $(1/3)\Pr(\sum_j D_j = \sum_j d_j|\lambda) + \Pr(\sum_j D_j < \sum_j d_j|\lambda) = (0.025, 0.975)$. (The multiplier 1/3 for the probability of the observed realization provides for the latter to be the point estimate in the meaning of 0% two-sided, or 50% one-sided, interval estimate.) Corresponding to $\sum_j d_j = 17$ in Cohort A, these limits are 26.5 and 10.1, translating into the survival-rate limits 81% and 96%. For Cohort B the exact limits of $\sum_j \lambda_j$ are 93.3 and 59.3, and the survival rate's corresponding exact limits are 69% and 85%.

The implications of this alternative to the KMG survival analysis is illustrated further by comparing the 95% two-sided confidence intervals it gives for the survival rate based on the Cohort A data at 5 years and 10 years of follow-up. In the here-introduced modified version of the KMG-equivalent, constant from 5 years to 10 years with Cohort A is, for a start, the (point and) interval estimate for the parameter of the Poisson distribution whose realization is $d = \sum_j d_j = 17$; the 95% two-sided interval is, as was noted, 9.9–26.0 in asymptotic terms (and 10.1–26.5 in exact terms). At 5 years these Poisson limits imply for the survival rate the 95% interval from 88% to 96%. The upper limit remains constant, at that 96%, from 5 years to 10 years, but the lower limit declines from that 88% (with $d^+ = 0.446$) to 81% (with $d^+ = 0.409$; cf. above). This is akin to the results with the 'semi-Borkowf' SE-based limits above, the transition in these from (87%, 97%) at 5 years to (80%, 97%) at 10 years.

Hypothesis testing in the substitute framework

When survival rate is compared between two non-overlapping cohorts (different from those used in the examples here), statistical testing of the hypothesis that there is a difference (in the abstract) is readily based on the ID-oriented data-formulation considered here. Let us denote the inputs to the component rates in the *j*th interval of time by d_{1j} together with T_{1j} for one of the cohorts, and by d_{0j} together with T_{0j} for the other one. In these terms, a suitable, asymptotically standard-Gaussian test statistic has the realization

$$z_0 = \Big(\sum_j d_{1j} - \sum_j d_j T_{1j}/T_j\Big) \Big/ \Big(\sum_j d_j T_{1j} T_{0j}/T_j^2\Big)^{1/2},$$

where $d_j = d_{ij} + d_{0j}$ and $T_j = T_{1j} + T_{0j}$.

When at issue is hypothesized difference between a single survival rate and its corresponding null value SR_0, the latter translates into its corresponding null value for the ID integral (as the negative of the log of SR_0). The latter, in turn, translates into its corresponding set $\{d_j^+\}$ or $\{d_j^-\}$ with a suitable constant adjustment of each $d_j^{1/2}$ (cf. above). The sum of these adjusted numbers constitutes what effectively is the null value, λ_0, corresponding to the empirical value $\sum_j d_j$. Thus a suitable, asymptotically standard-Gaussian test statistic for this test has the realization $z_0 = 2[(\sum_j d_j)^{1/2} - \lambda_0^{1/2}]$.

When, as in the examples here, the object of study is the level of a survival rate's asymptote, one is, in principle, concerned to test the hypothesis/premise that the asymptote indeed is reached within the survival period addressed in the survival analysis—within 10 years in the examples here. A testable prerequisite for this is that the ID late in the follow-up is (if not actually nil, then at least) lower than earlier in the follow-up.

In the Cohort A experience, the ID in years 8–10 may be contrasted with that in years 1–5. None of the events of interest occurred in that late period, while the null expected number was 17 multiplied by the binomial proportion of population–time in that period out of the total in years 1–5 and 8–10, $17[28.5/(1,013.5 + 28.5)] = 17(0.0274) = 0.466$, and its associated variance was 17(0.0274)

$(1 - 0.0274) = 0.453$. The corresponding value for a standard-Gaussian test statistic thus is $z = (0 - 0.466)/(0.453)^{1/2} = -0.69$. This represents no indication at all that ID (theoretical) late in the follow-up is lower than early in the follow-up, let alone that it has reached the asymptotic level of zero.

The attainment of the survival rate's asymptote actually is not a matter of hypothesis testing but estimation. Study of it requires setting a lower confidence limit for survival rate over that subperiod of follow-up in which the empirical $\widehat{\mathrm{ID}} = 0$ may correctly reflect the theoretical asymptotic $\widehat{\mathrm{ID}} = 0$. For the number of events over the period at issue (with ID = 0), the 95% one-sided upper confidence limit, as the solution of $(1/2)\exp(-\lambda) = 0.05$, is 2.30. Thus, over the years 8–10 in the Cohort A experience, the survival rate's corresponding limit (lower) is $\exp\{-[2.3/(15 + 9.5 + 4)\mathrm{y}]3\mathrm{y}\} = 79\%$; and for the years 6–10 it is 89%—neither one of these very close to the asymptotic 100%. The Cohort B data mean, correspondingly, that the focus needs to be on years 8–10, and that for this period the survival rate's lower limit is 84%.

For Bayesian hypothesis-testing—that is, evidence-based updating of the (subjective) probability that the hypothesis (non-null) is correct—the value of the null test statistic needs to be translated into its corresponding 'Bayes factor,' the likelihood ratio to be used in Bayes' theorem together with the prior probability (subjective) of the correctness of the hypothesis. The null test statistic's realization (z_0), however, explicitly provides the null likelihood only:

$$\mathrm{L}_0 = (2\pi)^{-1/2}\exp(-z_0^2/2).$$

The non-null counterpart of this is, in principle,

$$\mathrm{L}_1 = (2\pi)^{-1/2}\exp(-z_1^2/2),$$

where z_1 is the non-null counterpart of z_0, one in which the null expectation is replaced by its non-null counterpart, correct insofar as the hypothesis (non-null, qualitative) is correct.

While that non-null expectation generally is unknown, something is known about the non-null distribution of which z_1 is the notional realization. In particular, the median of the distribution of $|Z_1|$ is 0.67. It thus is reasonable—and suitably non-subjectivist—to use $z_1 = 0.67$ and, thus,

$$\mathrm{L}_1 = (2\pi)^{-1/2}\exp[-(0.67)^2/2];$$

and from this it follows that

$$\mathrm{LR} = \exp[(z_0^2 - 0.45)/2].$$

Reporting this LR would be a suitable substitute for reporting the p value corresponding to the z_0 value.

The data on Cohort A regarding whether the ID (theoretical) is lower in years 8–10 than in years 1–5 gave $z_0 = 0.69$ (cf. above), which translates into $\mathrm{LR} = 1.01 \simeq 1.00$; that is, the data give no support to the hypothesis/premise/impression that there is a decline in the ID (which attainment of the asymptote would imply).

Discussion: looking back at it all

The examples of survival analysis addressed here [1] had two features that jointly gave rise to concern. They involved calculation of the usual, Kaplan–Meier–Greenwood (KMG) statistics [2, 3] for point and interval estimation of the survival rate; and there was an unusual degree of censoring after the last deaths contributing the (decline in the) empirical rate of survival. It was quite unsettling to me that the standard errors (SEs) of the survival rates, and hence the confidence intervals based on these, remained unchanged over the periods of substantial censoring after the empirical survival rates had reached their respective asymptotes.

As I then came to realize, in 2005 Borkowf [5] made, emphatically, the point that a proper SE of a KM point estimate increases after the last event of interest in consequence to the censorings that occur in this period, different not only from the Greenwood SE but that of Peto et al. [6] as well; and he proceeded to show how. In the development of the substitute SE, the Peto SE was the point of departure for Borkowf, specifically its underlying idea of replacing the Greenwood SE by a binomial one: the KM survival rate can be viewed as a point estimate of a binomial proportion in conjunction with a suitably defined 'effective' number of trials, N. In the Peto SE, N is the solution of $N(\widehat{\mathrm{SR}}) = S$, where $\widehat{\mathrm{SR}}$ is the KM survival rate and S is the number of survivors under follow-up right before (*sic*) the last event of interest.

As a potentially very important modification of this, Borkowf replaced $N = S/(\widehat{\mathrm{SR}})$ by the total number of those members of the cohort whose follow-up had not been censored before the last event of interest. While the resulting Borkowf SE generally is in good accord with the Peto SE right after the last event of interest [5], it alone has the characteristic of increasing in consequence of the censorings that occur subsequent to the last event of interest. In the principal example here, after the last death of interest occurred, in the 5th year of the 10-year survival period at issue, the Borkowf SE increased by a factor of 2.4 by the end of the 10th year of follow-up.

While laudably appreciating that a proper SE increases with censorings subsequent to the last event of interest, and also showing how it increases, Borkowf failed to recognize that this increased SE is needed only in the calculation of the *lower* confidence limit of the survival rate. When (unjustifiably) applied in the calculation of the upper limit

in the principal example here, that (nominal) limit exceeds 100%.

While Borkowf thus provided for the formulation of a proper substitute for the Greenwood and Peto SEs of the KM survival rate, potentially having a major implication for the SE-based confidence interval (its lower limit), I here call for leaving behind also the KM point estimator of survival rate. My reason for this is that the KM point estimation can only be supplemented by SE-based interval estimation; that it does not lend itself to a first-principles counterpart of this.

I suggest that the point estimate of a survival rate be based on *incidence density* [4] of the death/failure at issue, on the integral of this over the survival period (as the exponential of the negative of this integral). In this framework, the total number of the events of interest can be taken to have a replication distribution of the Poisson type. First-principles asymptotic confidence limits for the parameter of this Poisson distribution can readily be derived, and these can be translated into their corresponding limits for the incidence-density integral and, through these, into the corresponding limits for the survival rate itself. By the same token, exact limits for the Poisson parameter can be derived and then translated into exact limits for the survival rate.

Once there is a preferable alternative to the KMG statistics in interval estimation of a 'cause-specific' rate of survival, hypothesis testing about survival rates also is to be brought into a new framework, not yet envisioned in the 1998 review by Barber and Jennison [7] but addressed here.

Different estimators lead to different estimates. In the principal example here (Cohort A, Table 1), the KMG analysis gave for the survival rate the point estimate 92% together with the 95% two-sided interval estimate of 88% to 95%. The approach proposed here would replace that interval by one ranging from 81% to 96%.

Now, as the data gave no indication at all that the underlying (theoretical) incidence density (ID) is anything but constant over the survival period at issue, 10 years, it would not have been unreasonable to base the survival rate's point estimate on the overall $\widehat{\mathrm{ID}}$ in place of the ones specific to one-year subintervals of the 10 years at issue; that is, on $\widehat{\mathrm{ID}} = (5 + 4 + \cdots + 0)/(291 + 261 + \cdots + 4)\mathrm{y} = 1.53/100\mathrm{y}$. On this basis the point estimate would have been $\widehat{\mathrm{SR}} = \exp[-(1.53/100\mathrm{y})10\mathrm{y}] = 86\%$; and the Poisson limits 26.0 and 9.9 corresponding to the 'observed' number of 17 'failures' would have translated into the survival rate limits 79% and 92%. Most notable in this would have been the point estimate's shift from 92% to 86%.

A salient feature of the approach proposed here is the definition of the set of intervals of follow-up time, meant to be short enough so that the incidence density within each of them is essentially constant. Good practice is to use intervals of identical widths; and it deserves note that while use of shorter intervals may add to validity, it does not take away from the efficiency of the analysis (increasing the width of the interval estimate).

All in all, it seems that there is a need to rethink the KMG routines that now permeate and dominate the available software systems for survival analysis.

Appendix 1

The Nelson–Aalen estimator

An eminent alternative to the KMG statistics is now constituted by the Nelson–Aalen (NA) statistics [8], which, like the statistics proposed here, are based on consideration of survival rate ('cause-specific') as the complement of cumulative incidence based on the integral of incidence density (ID). In the NA approach, the ID integral is derived as $\sum_j d_i/S_i$, where d_i is the number of deaths/failures at a point in follow-up time when S_i survivors were at risk for that outcome (and d_i of them experienced this event). Thus the point estimate of the survival rate is taken to be $\widehat{\mathrm{SR}} = \exp\left(-\sum_i d_i/S_i\right)$. The value of this is always somewhat higher than that of the corresponding KM estimate [8].

As a simple example, we might have $d_1/S_1 = 2/6$ together with only $d_2/S_2 = 2/4$, this in the absence of any censorings. The correct point estimate in this case may be taken to be the binomial one: $\widehat{\mathrm{SR}} = 2/6 = 0.33$. The corresponding NA estimate is $\exp[-(2/6 + 2/4)] = 0.43$, while the KM estimate is $(4/6)(2/4) = 2/6 = 0.33$.

One way to arrive at the N/A estimator, based on data in the form of $\{S_i, d_i\}$, is to focus on time elements of duration dt backward (*sic*) from each of the failure times (indexed by $i = 1, 2,\ldots$). From the ith one of these time elements the contribution to the ID integral is $(d_i/S_i\ \mathrm{d}t)\mathrm{d}t = d_i/S_i$. As only these time elements contribute to the ID integral, the latter is $\sum_i d_i/S_i$.

If, however, we consider time elements of duration 2dt and, specifically, of duration dt forward as well as backward from each of the failure times, as is natural, then the ID integral becomes $\sum_i\{d_i/[S_i\mathrm{d}t + (S_i - d_i)\mathrm{d}t]2\mathrm{d}t = \sum_i d_i/(S_i - \tfrac{1}{2}d_i)$. In the simple example above, this modification replaces the NA estimate, 0.43 (above), by 0.34 ($\simeq 2/6$).

For the here-proposed approach, the NA estimator has the virtue of suggesting that the population–time for the jth interval (with $d_j > 0$) can be derived from the usual $\{S_i, d_i\}$ data, supplemented by the timings of the $\{d_i\}$, as $T_j = t_j \sum_i d_{ij} / \sum_i d_{ij}/(S_{ij} - \tfrac{1}{2}d_{ij})$, the $\{S_{ij}, d_{ij}\}$ constituting the set falling in the jth interval of follow-up time.

References

1. The International Early Lung Cancer Action Program investigators. Survival of patients with Stage I lung cancer detected on CT screening. N Engl J Med. 2006;355(17):1763–71. doi:10.1056/NEJMoa060476.
2. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53(282):457–81. doi:10.2307/2281868.
3. Greenwood M. A report on the natural duration of cancer. In: Reports on public health and medical subjects, vol. 33. His Majesty's Stationery Office: London; 1926. p. 1–26.
4. Miettinen OS. Estimability and estimation in case-referent studies. Am J Epidemiol. 1976;103:30–6.
5. Borkowf CB. A simple hybrid variance estimator for the Kaplan–Meier survival function. Stat Med. 2005;24:827–51. doi:10.1002/sim.1960.
6. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. Br J Cancer. 1977;35(1):1–39.
7. Barber S, Jennison C. A review of inferential methods for the Kaplan-Meier estimator. In: Research Report 98: 02, Statistics Group, University of Bath, UK. 1998. http://www.maths.leeds.ac.uk/∼stuart/research/publications.html.
8. Collett D. Modelling survival data in medical research. 2nd ed. Boca Raton: Chapman & Hall/CRC; 2003.