COMMENTARY

Comparing Survival of a Sample to That of a Standard Population

Dianne M. Finkelstein, Alona Muzikansky, David A. Schoenfeld

Comparing groups on the basis of survival is common in medical research. Survival time data require methods that properly account for the situation when the time of death is not observed because some subjects are still alive at the end of the study (censoring). In addition, methods are required that make no assumptions about the shape of the survival time distribution (nonparametric). There are widely used methods for statistical comparison and graphic display of survival of two samples. The log-rank test *(1)* provides a comparison of the observed number of deaths in each group versus the number that would be expected if the total mortality were distributed according to the proportion in each group. These statistical comparisons are often accompanied by Kaplan–Meier curves that provide a graphic display of the distribution of survivorship over time *(2)*. This estimator, calculated from samples that are partially censored, is a monotonically non-increasing step function with steps at each observed death time. Although calculated separately for each group, these graphs are displayed simultaneously in a single plot to promote a visual comparison of survival over the entire study period.

It is often of interest to compare the survival of a single sample to that of a defined reference population. For example, when a series of patients with a rare, life-threatening disease has been collected, it may be of interest to know if the study sample is experiencing the same survival as the demographically matched standard (general) population, according to actuarial tables. This is especially of interest when the disease is curable or not usually lethal and the age of onset is late in life. It is not appropriate to use methods developed for two-sample comparisons to do this analysis, because the variance would be incorrectly calculated and thus the *P* value would be invalid.

It is possible to provide a single summary measure of the relative survival of a sample compared with a standard population by estimating the standardized mortality ratio *(3)*. However, investigators often want to report a *P* value from a statistical test that compares the two populations—essentially a one-sample log-rank test. Although such tests are published in the statistical literature *(4,5)*, medical investigators do not generally read this literature, and thus these tests are not widely known and used by this community. In fact, these articles *(4,5)* have been cited fewer than 10 times in the medical literature over the past 20 years.

Some *ad hoc* methods have been devised for a one-sample survival test. For example, one approach is to use the actuarial tables to determine the expected remaining lifetime at the age of study entry for each of the subjects in the sample and then to treat these times as exact death times of a hypothetical sample of the same number of subjects from the reference population. It is possible then to calculate a two-sample log-rank test and report the resulting *P* value. However, this test is inappropriate because the variance would be incorrectly calculated and thus the *P* value would be invalid.

Similar pitfalls arise in trying to obtain an accompanying graphical display that would appropriately represent the survival of the standard population in the same manner in which the Kaplan–Meier plot represents the sample. Because there are no methods that are widely cited in the medical literature, there is a tendency to develop *ad hoc* methods. For example, one approach is to calculate the expected remaining lifetime for each subject in the sample by using the actuarial tables matched by age, sex, and/or race. This set of numbers is then treated as exact observed death times, and the Kaplan–Meier estimator is calculated for this hypothetical population. This calculation results in a step function with the number of steps equal to the size of the sample being studied. This is not correct. As an illustration, suppose everyone in the sample began observation at the same age and, thus, would have the same expected remaining lifetime, say *s.* The survival distribution for the hypothetical matched sample representing the reference group would then have a value of 1 until *s,* at which point the curve would drop to 0. The correct method must use the entire remaining survival curve (calculated from the reference actuarial tables) for each subject.

The purpose of this commentary is to describe both the simple one-sample log-rank test that is equivalent to the standardized mortality ratio and an estimate for survivorship in the matched standard population that allows a visual comparison of survivorship of the sample and standard populations. We will discuss the issues in designing a study that will rely on one-sample methods. The software to perform analyses discussed in this commentary can be found at our Web site: http://biostatistics.mgh.harvard.edu/biostatistics/resources.html *(6)*. As an illustration, we use these methods to compare the survival of a small cohort of patients diagnosed with extra-mammary Paget's disease at Massachusetts General Hospital with the survival of the general population *(7)*.

ONE-SAMPLE LOG-RANK TEST

The data from our sample consist of intervals of time during which subjects were under observation. Using the example of the series of patients treated for extra-mammary Paget's disease, we determine the age at diagnosis and the age at which vital

Affiliation of authors: Biostatistics Center, Massachusetts General Hospital, Boston, MA.

Correspondence to: Dianne M. Finkelstein, PhD, Biostatistics Center, Massachusetts General Hospital, 50 Staniford St., Suite 560, Boston, MA 02114 (e-mail: finkel@biostat.harvard.edu).

See "Notes" following "References."

status was last confirmed (either at death or at most recent visit) for each patient. From this sample, we know the observed number of patients who have died, which we call *O,* and wish to compare this number with the number of deaths that would be expected in a comparison (standard) population with the same distribution of age and length of follow-up.

To obtain the expected number of deaths, *E,* for a standard population with the same demographic profile (race/sex) and length of follow-up, it is necessary to calculate the probability of observing a death during the same follow-up period as was available for our sample of subjects. For this calculation, it is necessary to obtain the annual death rate for each year of age during follow-up. The standardized (actuarial) tables in the National Vital Statistics Reports *(8)* provide these death rates classified by age, sex, and race. For simplicity, it is useful to apply the convention that both the diagnosis and end of follow-up occurred on a patient's birthday. Spreadsheet software can be used to facilitate the calculation of the expected number of deaths. Table 1 illustrates this calculation on a small sample of subjects similar to those in the extra-mammary Paget's disease dataset: two Caucasian females (patients 2 and 3), diagnosed at ages 74 and 78 years who lived until ages 80 and 84 years, respectively; a Caucasian male (patient 1), diagnosed at age 72 years and followed until age 89 years; and a Caucasian female (patient 4), diagnosed at 71 years and followed until her death at age 83 years. We recorded in the spreadsheet the ages ranging from the age at earliest diagnosis in the sample (71 years) to the oldest age observed at last follow-up in the sample (89 years). For each subject i (where $i = 1$–4), the probability (or hazard) of dying for each year of age t, $h_i(t)$, is obtained from the National Vital Statistics Reports within the table that is matched on the subject's sex and race. The redundancy within the data for each subject results from the fact that the actuarial table we used

gave annual death rates in 5-year categories only (e.g., 70–74 years).

To obtain the overall expected number of deaths, it is necessary to calculate the cumulative death rate at the last follow-up for all subjects in the study. For each subject, the cumulative death rate by each age t after diagnosis is

$$H_i(t) = \sum_{u=a_i}^{t-1} h_i(u), \tag{1}$$

and is calculated for each year after the age at diagnosis. This value is calculated at all years (beyond the end of follow-up for each patient) because it is needed for the survival curve estimate. The cumulative death rate is obtained by adding the death rates at each year to the cumulative death rate for the previous year. If patient 2, diagnosed at age 74 years, is used as an example, then the cumulative death rate at age 76 years is $0.023 + 0.037 = 0.060$, and the cumulative death rate at age 77 is $0.037 + 0.060 = 0.097$. The overall expected number of deaths, *E,* is found by adding the cumulative death rate at the last age of follow-up, t_i, over all *N* individuals in the sample.

$$E = \sum_{i=1}^{N} H_i(t_i). \tag{2}$$

In Table 1, the cumulative death rate at the age of last follow-up is indicated in boldface type. For this small dataset, the expected number of deaths $(E) = 1.584 + 0.208 + 0.326 + 0.466 = 2.584$. The test for equality of mortality of the sample population with that of the standard population is calculated as

$$[O - E]^2 / E. \tag{3}$$

We show in the Appendix, under the null hypothesis, that the survival of the sample population is the same as that of the

Table 1. Calculating expected number of deaths

Age i	Patient 1		Patient 2		Patient 3		Patient 4	
	Age-specific death rate*	Cumulative death rate†	Age-specific death rate	Cumulative death rate	Age-specific death rate	Cumulative death rate	Age-specific death rate	Cumulative death rate
71							0.023	
72	0.039						0.023	0.023
73	0.039	0.039					0.023	0.046
74	0.039	0.078	0.023				0.023	0.069
75	0.058	0.117	0.037	0.023			0.037	0.092
76	0.058	0.175	0.037	0.060			0.037	0.129
77	0.058	0.233	0.037	0.097			0.037	0.166
78	0.058	0.291	0.037	0.134	0.037		0.037	0.203
79	0.058	0.349	0.037	0.171	0.037	0.037	0.037	0.240
80	0.093	0.407	0.063	**0.208‡**	0.063	0.074	0.063	0.277
81	0.093	0.500	0.063	0.271	0.063	0.137	0.063	0.340
82	0.093	0.593	0.063	0.334	0.063	0.200	0.063	0.403
83	0.093	0.686	0.063	0.397	0.063	0.263	0.063	**0.466**
84	0.093	0.779	0.063	0.460	0.063	**0.326**	0.063	0.529
85	0.178	0.872	0.147	0.523	0.147	0.389	0.147	0.592
86	0.178	1.05	0.147	0.670	0.147	0.536	0.147	0.739
87	0.178	1.228	0.147	0.817	0.147	0.683	0.147	0.886
88	0.178	1.406	0.147	0.964	0.147	0.830	0.147	1.033
89	0.178	**1.584**	0.147	1.111	0.147	0.977	0.147	1.180

*Age-specific death rate $= h_i(t)$.

†Cumulative death rate $= H_i(t) = \sum_{u=a_i}^{t-1} h_i(u)$.

‡Boldface type indicates the cumulative death rate at the age of last follow-up.

matched population—the test shown in equation 3 will be distributed as a chi-square distribution with 1 *df*. Therefore, a test for equality is performed by comparing equation 3 with tabled extreme values of this distribution *(9)*.

We note that the test expressed in equation 3 using equation 2 is equivalent to the test as discussed *(4,5)*. This test has been referred to as the one-sample log-rank test *(5)*, in that it is based on the proportional hazards model, which is the same model that underlies the standard two-sample log-rank test. The mathematical development of this test is given in the Appendix. This test has the same functional form as the usual (two-sample) log-rank test. However, in this test, the expected value, *E*, is derived from actuarial tables, whereas, in the two-sample test, the expected value is derived from the combined sample under an assumption that the mortality is the same in the two populations. The variances of these two tests are quite different. The log-rank test is calculated under the assumption that both groups are samples and thus subject to variation, whereas the one-sample test takes into account that only the sample (not the standard population) would change if the study were done on a new set of patients.

RELATIONSHIP TO THE STANDARDIZED MORTALITY RATIO

Analysis of a cohort group usually entails comparison of the survival distribution for the cohort with that of the standard population. The standardized mortality ratio has been used since 1786. It is the ratio of the number of observed deaths to the number of expected deaths in a demographically matched standard population, *O/E*. The expected number of deaths, *E*, is calculated as in equation 2. The test for whether survival in the sample differs from that of the demographically matched population is given by the statistic $(O - E)^2/E$, which is compared with a chi-square distribution on 1 *df*. For small datasets, an exact test is calculated as previously described *(10)*. The depar-

ture of the standardized mortality ratio from 1 provides evidence that the observed mortality in the cohort differs from that of the standard population. The one-sample log-rank test described above is a test of whether the standardized mortality ratio is different from 1.

ESTIMATION OF EXPECTED SURVIVAL DISTRIBUTION

It is useful to provide a graph to visually compare the survivorship of the sample with that of the standard population. The graph of the survival distribution of the standard population is calculated by using the cumulative death rate for each subject (as calculated for obtaining the expected number of deaths), measured from the age at diagnosis. This is illustrated on our small dataset used in Table 1. First, we create a second spreadsheet as shown in Table 2. We then enter the cumulative death rate up to each time for each subject. In contrast to Table 1, where time was measured by chronological age of the patients, in Table 2, time, *s*, is measured from the point of diagnosis to the end of the longest follow-up time of the study. For each patient, the first entry corresponds to the cumulative death rate for the first year after diagnosis. The entry in the sth row for subject *i* is the cumulative death rate at a time *s* years after diagnosis, $H_i(a_i + s)$.

For each subject *i* in the study sample and for each year *s* after diagnosis, we can calculate the expected probability of survival beyond that year for an age-/sex-/race-matched person in the standard population. This probability is calculated by

$$e^{-H_i(a_i + s)}. \qquad [4]$$

Note that the probability of survival calculated by equation 4 is a function of the death rates calculated by equation 1. However, in contrast to the calculations for the one-sample test, which stop at the end of observation (time of censoring or death) for each subject, the calculations for the survival curve estimate are performed for every year after diagnosis, regardless of how

Table 2. Calculating survival estimate

Years on study	Patient 1 Cumulative death rate*	Patient 1 Individual survival†	Patient 2 Cumulative death rate	Patient 2 Individual survival	Patient 3 Cumulative death rate	Patient 3 Individual survival	Patient 4 Cumulative death rate	Patient 4 Individual survival	Survival estimate‡
0									1.0
1	0.039	.96	0.023	.98	0.037	.96	0.023	.98	.97
2	0.078	.92	0.060	.94	0.074	.93	0.046	.96	.94
3	0.117	.89	0.097	.91	0.137	.87	0.069	.93	.90
4	0.175	.84	0.134	.87	0.200	.82	0.092	.91	.86
5	0.233	.79	0.171	.84	0.263	.77	0.129	.88	.82
6	0.291	.75	0.208	.81	0.326	.72	0.166	.85	.78
7	0.349	.71	0.271	.76	0.389	.68	0.203	.82	.74
8	0.407	.67	0.334	.72	0.536	.59	0.240	.79	.69
9	0.500	.61	0.397	.67	0.683	.51	0.277	.76	.64
10	0.593	.55	0.460	.63	0.830	.44	0.340	.71	.58
11	0.686	.50	0.523	.59	0.977	.38	0.403	.67	.54
12	0.779	.46	0.670	.51	1.124	.32	0.466	.63	.48
13	0.872	.42	0.817	.44	1.271	.28	0.529	.59	.43
14	1.050	.35	0.964	.38	1.418	.24	0.592	.55	.38
15	1.228	.29	1.111	.33	1.565	.21	0.739	.48	.33
16	1.406	.25	1.258	.28	1.712	.18	0.886	.41	.28
17	1.584	.21	1.405	.25	1.859	.16	1.033	.36	.25
18	1.762	.17	1.552	.21	2.006	.13	1.180	.31	.21

*Cumulative death rate = $H_i(a_i + s)$.

†Individual survival = $e^{-H_i(a_i + s)}$.

‡Survival estimate = $S_0(s) = \frac{1}{N}\sum_{i=1}^{N} e^{-H_i(a_i + s)}$.

long each subject was actually followed. We note that survivorship is estimated by assuming that the censoring is independent of age at diagnosis. However, if this were not the case, alternate methods would be required. The proof for equation 4 is shown in the Appendix.

The overall expected survival rate for the study population at each year, s, after diagnosis, $S_0(s)$, is calculated by dividing the sum of the N survival rates at each time by the number of subjects (N):

$$S_0(s) = \frac{1}{N} \sum_{i=1}^{N} e^{-H_i(a_i + s)}. \qquad [5]$$

In the small example shown in Table 2, these survival estimates are calculated by averaging the individual survival rates that appear in each row and are recorded in the final column as the survival estimate. Note that the final row of Table 2 for patient 3 corresponds to the age of 95 years even though she was followed only until age 84 years. We filled in the last rows of Table 2 with the race-/sex-specific death rate in the National Vital Statistics tables for individuals grouped into a single age category of 85 years or older.

The survival function (equation 5) is calculated for all times s from 1 to the longest time from diagnosis to end of follow-up in the sample. At time 0, its value is 1. The estimate can be simultaneously plotted with the Kaplan–Meier estimate (2) for survivorship of the observed sample. Confidence intervals on the sample survivorship function can be calculated as previously published (2). We note that censorship has been accounted for in the Kaplan–Meier curve; in the standard population curve, there is no censoring.

POWER AND CONFIDENCE INTERVALS

Often comparisons with a standard population are designed to argue that the treated population has better or the same survival distribution as the standard population. When the intention is to find equivalence to the standard population, it is best to present a confidence interval for the standardized mortality ratio. If this interval includes 1, then the evidence is consistent with the conclusion that the two populations may have the same survival distribution. We note that the estimate of the standardized mortality ratio, O/E, is an estimate of the relative risk from the proportional hazards model. The 95% confidence interval on the standardized mortality ratio is given by

$$O/E + \chi^2_{1,1-\alpha/2}/2E \pm \sqrt{\chi^2_{1,1-\alpha/2}(4O + \chi^2_{1,1-\alpha/2})}/2E. \qquad [6]$$

This formula is derived in the Appendix.

If a study is undertaken to determine whether there is a difference in survival between a sample and a standard population, then the study must be designed to ensure sufficient statistical power to detect a specified relative risk, R, for death associated with the study population compared with the standard population. To accomplish this, it is necessary to determine how many deaths must be observed in the sample. To calculate the required sample size, we assume that the probability that the null hypothesis will be rejected will be set at a two-sided α level (usually .05) and need to determine how many deaths must be observed to ensure that the one-sample log-rank test will reject the null

hypothesis, given that the alternative is true with probability $1 - \beta$ (commonly 80%). This number of deaths, D, can be derived [see (11)] by

$$D = \left[\frac{(Z_{1-\alpha/2} + Z_{1-\beta}\sqrt{R})}{R - 1} \right]^2, \qquad [7]$$

where Z_u corresponds to the tabled value of the uth percentile of a standard normal distribution (9).

ILLUSTRATION: COMPARING PATIENTS TREATED FOR EXTRA-MAMMARY PAGET'S DISEASE TO THE STANDARD POPULATION

Extra-mammary Paget's disease is a relatively uncommon clinical entity. Although Paget primarily described the disease (which now bears his name) in the breast, he did mention that the same disease may also occur in other parts of the body (12). Paget's disease is a cutaneous adenocarcinoma, usually of epidermal origin and glandular differentiation, that might be associated with invasive adenocarcinoma or *in situ* adenocarcinoma of the apocrine glands or underlying visceral malignancy (13). Extra-mammary Paget's disease is histologically identical to Paget's disease of the breast. In extra-mammary Paget's disease, the epidermis is infiltrated by irregular large, pale cells that are scattered between compressed squamous epithelial cells with an otherwise normal appearance (14–17). Because extra-mammary Paget's disease is such a rare condition, it is not possible to estimate its true incidence or the frequency with which it becomes clinically malignant or coexists with a visceral carcinoma. Furthermore, the optimal management of patients with extra-mammary Paget's disease is still unclear. Complete surgical resection, with or without skin grafting, seems to be the treatment of choice for patients with extra-mammary Paget's disease. The disease, however, tends to recur with the same frequency after local excision and more extended surgery. An important issue when treating patients with extra-mammary Paget's disease is the high incidence of concurrent secondary malignancies. Given the high median age (70 years) of patients with extra-mammary Paget's disease, this observation partly reflects the approximate 1:4 and 1:3 lifetime risks for any malignancy in women and men, respectively.

Recently, investigators at Massachusetts General Hospital undertook a retrospective study of patients with extra-mammary Paget's disease who were treated at the hospital to determine the clinical course of this disease. Thirty-three consecutive patients treated at Massachusetts General Hospital between January 1971 and August 1998 for extra-mammary Paget's disease were identified from the Massachusetts General Hospital Tumor Registry. The medical records for these patients were reviewed, and information was collected on demographic features, clinical presentation, the extent and pathologic features of the disease, tumor location, type of treatment, time of recurrence, and time and cause of death. Interest was focused on time to recurrence and recurrence-free and overall survival. The investigators believed that, because this cancer occurs late in life and is of relatively low lethality, the survival experience of the 33 subjects of the report was not different from that of a demographically matched cohort of the standard population. They requested a graph showing this comparison and a statistical test of the hypothesis.

The 33 subjects in the extra-mammary Paget's disease study had a median age of 70 years (range = 52–86 years). No patient died of extra-mammary Paget's disease. However, 14 (42%) of the 33 patients with extra-mammary Paget's disease had a total of 16 concurrent secondary malignancies (five breast, three colon/rectal, three basal cell, two renal cell, one bladder, one hepatocellular, and one vulvar squamous cell carcinomas). In addition, 14 (42%) of the 33 patients had disease recurrence that required additional surgeries. The median follow-up of these subjects was 5.7 years (range = 0.5–22.16 years). In the report on this series of patients (7), the standardized mortality ratio and associated 95% confidence interval were reported. Because the confidence interval included 1, the conclusion was that the survival of the study subjects was not statistically significantly different from that of a demographically matched standard population.

The comparison of survival using the one-sample log-rank test described in this commentary results in a χ_1^2 of 1.24 with an associated $P = .266$, concurring with the conclusion that survival in this group of patients was not different from that in the age-matched standard population. Fig. 1 shows the curves that have been calculated as described in this commentary and provides visual evidence that the patients with extra-mammary Paget's disease do not experience a different survival distribution than that of the standard population. The 95% confidence intervals on the 5-, 10-, and 15-year survivorship of the sample are also provided.

DISCUSSION

It is important to consider that the power of the comparison of survival in a sample with that in a population is a function of the number of deaths. If investigators wish to test whether the population from which the sample is drawn is experiencing a
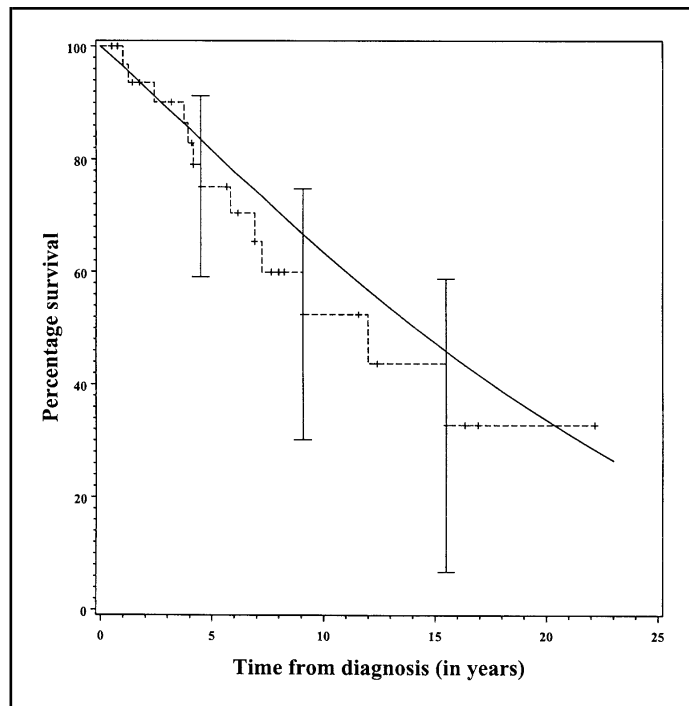
Fig. 1. Comparison of survival distribution of extra-mammary Paget's disease patients (**dashed line**) to that of the age-/sex-/race-matched general U.S. population (**solid line**).

lower (or higher) survival than that of the standard population, then they should design the study so that there will be an adequate number of observed deaths to ensure sufficient power. If, instead, the intent is to show that the population from which the sample is drawn is experiencing the same mortality as the comparison population, it may be more appropriate to provide confidence intervals on the standardized mortality ratio and base the discussion of equivalence on this interval. The test of equivalence ordinarily requires a very large study, and a large P value from a small study should not be mistakenly reported as conclusively showing survival equivalence. Finally, we note that the test described herein is based on the proportional hazards assumption and will not be appropriate if the survival curve for the sample population crosses the curve for the standard population. The curves could cross, for example, if there were a high early mortality from treatment, but the risk in survivors of this period returned to the risk in the standard population. If such a crossing occurs, it may be more appropriate to compare survival after a specific guarantee time (such as time from end of treatment) or to use a model-free method such as the Kolmogorov–Smirnov test (9).

APPENDIX

Derivation of One-Sample Log-Rank Test

For each subject, i, our data consist of a_i (the time of diagnosis), δ_i (an indicator of whether the subject died on study), and t_i (the age at most recent follow-up time or death). The survivorship for the standard population is S_O. We assume the proportional hazards model, that is, that the survivorship for our cohort is S, and if $S(a,t)$ is the probability of surviving at least t units of time conditional on survival to age a (diagnosis), then

$$S(a,t) = S_0(a,t)^{e^\beta}. \qquad [8]$$

Thus, if $\beta = 0$, then the population under study is experiencing a survivorship that cannot be distinguished from that of the standard population (of people who were alive at the age a_i). The relative risk, referred to as R above, is equivalent to e^β from the model equation 8.

To derive a test for whether survival distribution of the sample population is different from that of the standard population, we will use a score test derived under the proportional hazards model, equation 8. We note that individual i who is diagnosed at age a_i and who is alive at the last observation at t_i years of age (censored) will have an indicator for failure of $\delta_i = 0$ and will contribute a factor $S_0(a_i,t_i)^{e^\beta}$ to the likelihood. Someone who dies at t_i years of age will have an indicator for failure of $\delta_i = 1$ and will contribute the factor $-\partial/\partial t(S_0(a_i,t_i)^{e^\beta})$ to the likelihood. Therefore, the likelihood will be

$$\prod_{i=1}^{N}[e^\beta S_0(a_i,t_i)^{e^\beta - 1})]^{\delta_i}[-S_0{}'(a_i,t_i)]^{\delta_i}S_0(a_i,t_i)^{e^\beta(1 - \delta_i)}, \qquad [9]$$

where $S_0{}'(a_i,t_i)$ is the derivative of $S_0(a,t)$ with respect to t, evaluated at $a = a_i, t = t_i$. Taking the derivative of the log of the likelihood given by equation 9 with respect to β and evaluating it at $\beta = 0$, we obtain the score test of

$$\sum_i \delta_i + \log S_0(a_i,t_i). \qquad [10]$$

However, we can show that the second part of the sum in equation 10, $\Sigma_i \log S_0(a_i,t_i)$, represents the negative expected value for the num-

ber of deaths. In fact, for someone diagnosed at time a_i, the probability of surviving to t_i is

$$S_0(a_i, t_i) = \prod_{u=a_i}^{t_i - 1} (1 - h_i(u)) = e^{\sum_{u=a_i}^{t_i - 1} \log(1 - h_i(u))}. \qquad [11]$$

However, since h_i is small, $\log(1 - h_i) \approx -h_i$ (using the Taylor expansion) and thus equation 11 is approximately equal to $e^{-\sum_{u=a_i}^{t_i - 1} h_i(u)}$. However, $S_0(a_i, t_i)$ is the survival for a member of the standard population (alive at a_i) and thus $-\sum_i \log S_0(a_i, t_i)$ is the cumulative hazard, which can be approximated by the cumulative death rate for this member of the standard population. When this is added up over all i, we obtain $\sum_{i=1}^{N} \sum_{u=a_i}^{t_i - 1} h_i(u)$, which is the expected number of deaths (E). Thus, equation 10 is just the observed number of deaths minus the expected number of deaths, $O - E$.

Confidence Interval for Standardized Mortality Ratio

The confidence interval for the standardized mortality ratio is based on the fact that, regardless of the true value of β in the model described by equation 8, because O is Poisson distributed with mean $e^\beta E$, the function

$$(O - e^{\beta_0} E)^2 / (e^{\beta_0} E) \qquad [12]$$

is chi-square distributed with 1 df. Thus, the $(1 - a)\%$ confidence interval is found by setting equation 12 equal to $\chi^2_{1, 1 - \alpha/2}$ and solving the resulting quadratic equation. This solution will be the function shown in equation 6.

REFERENCES

(1) Peto R, Peto J. Asymptotically efficient rank invariant test procedures. J Royal Stat Soc A 1972;135:185–206.

(2) Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457–81.

(3) Breslow NE, Day NE. Statistical methods in cancer research. Vol. II. The design and analysis of cohort studies. Lyon (France): IARC Publications; 1987. p. 65.

(4) Gail MH, Ware JH. Comparing observed life table data with a known survival curve in the presence of random censorship. Biometrics 1979;35:385–91.

(5) Woolson RF. Rank-tests and a one-sample log-rank test for comparing observed survival-data to a standard population. Biometrics 1981;37:687–96.

(6) Massachusetts General Hospital (MGH). MGH Biostatistics. Available at: http://biostatistics.mgh.harvard.edu/biostatistics/resources.html. [Last accessed: 7/24/03.]

(7) Pierie JE, Choudry U, Muzikansky A, Finkelstein DM, Ott MJ. Prognosis and management of extramammary Paget's Disease and the association with secondary malignancies. J Am Coll Surg 2003;196:45–50.

(8) National Center for Health Statistics. National Vital Statistics Reports, Vol. 47, No. 19, June 30, 1999. Available at: http://www.cdc.gov/nchs/products/pubs/pubd/nvsr/nvsr.htm. [Last accessed: 7/24/03.]

(9) Zar JH. Biostatistical analysis. Englewood Cliffs (NJ): Prentice-Hall; 1974. p. 55.

(10) Liddell FD. Simple exact analysis of standardized mortality ratio. J Epidemiol Community Health 1984;38:85–8.

(11) Selvin S. Statistical analysis of epidemiologic data. New York (NY): Oxford University Press; 1996. p 89.

(12) Paget J. Disease of the mammary areola preceding cancer of the mammary gland. St Barth Hosp Rep 1874;10:87–9.

(13) Chanda JJ. Extramammary Paget's disease: prognosis and relationship to internal malignancy. J Am Acad Dermatol 1985;13:1009–14.

(14) Jensen SL, Sjolin KE, Shokouh-Amiri MH, Hagen K, Harling H. Paget's disease of the anal margin. Br J Surg 1988;75:1089–92.

(15) Goldblum JR, Hart WR. Vulvar Paget's disease: a clinicopathological and immunohistochemical study of 19 cases. Am J Surg Pathol 1997;21:1178–87.

(16) Goldblum JR, Hart WR. Perianal Paget's disease: a histologic and immunohistochemical study of 11 cases with and without associated rectal adenocarcinoma. Am J Surg Pathol 1998;22:170–9.

(17) Sarmiento JM, Wolff BG, Burgart LJ, Frizelle FA, Ilstrup DM. Paget's disease of the perianal region-an aggressive disease? Dis Colon Rectum 1997;40:1187–94.

NOTES