Adjusting Survival Curves for Confounders: A Review and a New Method

F. Javier Nieto[1] and Josef Coresh[1,2]

When reporting results from survival analysis, investigators often present crude Kaplan-Meier survival curves and adjusted relative hazards from the Cox proportional hazards model. Occasionally, the investigators will also provide a graphical representation of adjusted survival curves based on regression estimates and the average covariate values in the study groups. In this paper, the authors review the limitations of this approach and examine alternative approaches to obtaining adjusted survival curves that have been proposed. Furthermore, a new method to obtain multivariate adjusted survival curves is described. This method is based on direct adjustment of the observed conditional probability of survival at the time of each event. When an unexposed group is used as a standard for adjusting an exposed group, the survival curve in the exposed group is adjusted to the covariate distribution among the unexposed at the time of the event. This method has the advantage over the average covariate method of allowing for the possibility that the adjusted survival curves differ in shape. The method can handle multiple fixed or time-dependent categorical covariates as well as left truncated data, and it allows for estimation of confidence intervals. The authors have written a macro in SAS language that produces the adjusted survival estimates and graphs. This macro is available on request and can be downloaded through the World Wide Web. *Am J Epidemiol* 1996;143:1059–68.

confounding factors (epidemiology); data display; epidemiologic methods; graphical analysis; proportional hazards models; prospective studies; regression analysis; survival analysis

The results of follow-up studies are often analyzed with the use of survival analysis techniques. The graphical display of the Kaplan-Meier survival curves (1) and their statistical comparison with the logrank test are usually adequate in randomized clinical trials with a sufficiently large sample size and balanced groups. However, in observational studies, it is often necessary to carry out adjusted comparison of the survival experience of the different groups, taking into account unbalanced distribution of confounders. The most commonly used method for multivariate comparison of survival curves is the Cox proportional hazards model (2), a semi-parametric regression technique that assumes a constant relative hazard throughout the entire follow-up period. The adjusted hazard ratios obtained from the Cox regression are usually presented in tabular form, although adjusted survival curves can also be obtained (average covariate method). In this

paper, we review the limitations of this method, as well as some alternative methods to obtain adjusted survival estimates published in the literature in recent years (3–8). Finally, we present a new method for a graphical comparison of survival curves while controlling for important covariates. This method is based on calculation of adjusted conditional probability of the event at each failure time using direct adjustment. The reference population for the adjustment is the distribution of covariates in the comparison group at each precise failure time. The method allows for adjustment for fixed or time-dependent categorical covariates. We have written a macro in the SAS language (SAS Institute Inc., Cary, North Carolina) for obtaining the adjusted survival curves using the proposed method. The program is available on request by mail from the corresponding author (please send an IBM-compatible formatted, high density, 3.5" diskette) or by electronic mail (Internet: jnieto@phnet.sph.jhu.edu). It can also be downloaded from the World Wide Web at http://www.jhu.edu/~welchc/software.html. In addition to the new method we propose, the program will run some of the other methods discussed in this paper.

THE AVERAGE COVARIATE METHOD

This method for estimating adjusted survival curves relies on applying the parameter estimates obtained

from multivariate regression analysis, usually the Cox proportional hazards regression (2), to the average value of the covariates of interest in the groups being compared. Methods to estimate the baseline survival $S_0(t)$ from Cox regressions have been described (2, 9). The adjusted survival function for a group with average covariate values Z_i can be obtained by raising the baseline survival to the $e^{Z_i'\beta}$ power. Alternatively, other parametric models such as the Weibull model can be used (10).

This method of estimating the adjusted survival has been used in biomedical papers with increasing frequency in recent years. Its popularity stems from the broad use and acceptance of the Cox regression model (11, 12) and, at least in part, from its availability in some standard statistical packages, including SAS (PROC PHREG with BASELINE COVARIATES statement) (13). The limitations of this method, however, have been repeatedly discussed (4, 10, 14) and are worth reviewing:

1) For categorical covariates, the meaning of the *adjusted* survival for individuals with the average covariate value is quite difficult to explain (4, 10). For example, if one wants to adjust to a group consisting of 50 percent males (coded as 0) and 50 percent females (coded as 1), one would calculate the survival for individuals whose sex equals 0.5, which is meaningless at the individual level. An even more serious problem is that the method is not equivalent to estimating the survival of a group which is half male and half female. The reason is that the method is actually based on calculating the *average hazard*, i.e., the hazard for the average individual, which is not the same as the *average survival* estimated from a heterogeneous group of individuals, as illustrated in the following example. Let us assume that the death rate (hazard) of a given condition is 500/1,000 per year among males and zero among females (e.g., a condition that only affects males). Figure 1 shows the survival for a group with an average hazard of 250/1,000 per year, which is analogous to what would be obtained from the average covariate method. Notice that this method produces a senseless curve, going below a 50 percent survival, despite the fact that the hazard is zero for half of the population. The correct average survival curve (also shown in figure 1) has an asymptotic cumulative survival of 0.5. As Thomsen et al. (14) showed, the problem with the average covariate method and analogous procedures such as the method proposed by Neuberger et al. (15) is that the average hazard does not take into account the heterogeneity within the sample and does not have the same time-dependence as the individuals' survival. The frail die first, leading
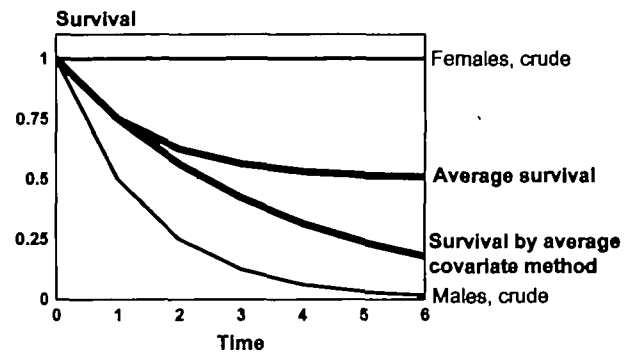
FIGURE 1. Hypothetical example of survival curves of a condition that affects 500/1,000 men per year, but does not affect women. Also shown are the survival curve based on calculating the average hazard for a group composed of an equal number of men and women, i.e., 250/1,000 per year (average covariate method), as well as the average survival, asymptotic to a cumulative survival of 0.5.

to a relative reduction of population mortality over time compared with the pattern of any individual (14).

2) The method calculates the adjusted survival for all groups based on one common baseline survival $S_0(t)$. Because of the proportionality assumption underlying the Cox model, the adjusted survival curves are forced to be powers of each other. If the proportionality assumption does not hold, the adjusted curves can be misleading. If the proportionality assumption does hold, the graphical depiction of the curves does not add much to the adjusted relative hazard estimate combined with the crude Kaplan-Meier curve of the total group.

3) When time-dependent covariates are used, the calculation of the adjusted survival curves is greatly complicated. (SAS's PROC PHREG, for example, does not produce adjusted survival curves if time-dependent covariates are used (13, p. 19).)

ALTERNATIVE METHODS

Direct adjustment based on parametric survival function

Makuch (3) and Chang et al. (4) independently proposed a new method that overcomes the first limitation of the average covariate method discussed above. In essence, the survival curve for each individual or each level of the covariates is calculated using the Cox model (or some other parametric model (10)). The average survival is then calculated as a weighted average of the individual survival curves, with weights proportional to the number of individuals at each level of the covariates in the entire sample at baseline. This method, also known as the *corrected group prognosis method* (4), is analogous to the calculation of covariate-adjusted rates using the logistic regression in a prospective study (16, 17). A SAS computer pro-

gram for its application has been made available (17) and methods to estimate the variance have been described (18).

This method is a clear improvement over the average covariate method. Its applicability, however, still relies on the validity of the assumptions underlying the models used for the estimation of the individual or strata-specific survival function. We are not aware of an application of this method to a situation with time-dependent covariates.

Adjustment based on non-parametric survival estimates

Hankey and Myers (5) proposed a method to obtain adjusted survival curves based on actuarial life table analyses (follow-up time broken up in intervals) and categorical covariates. For each interval, the adjusted odds ratio of the event is obtained using the Mantel-Haenszel method (19). Based on the adjusted odds ratio and on the observed survival in a group taken as reference, the adjusted survival for the study group is then calculated by applying the adjusted odds ratio to the observed odds in the study group. A limitation of this method is that the estimates for the late time intervals could be rather unstable if the sample size is small (4, 5). In addition, this method can not be used for exact survival time (Kaplan-Meier).

A direct adjustment method based on the Kaplan-Meier survival estimates has been proposed by Kramar and Com-Nougué (6). Given some categorical covariate(s), these authors proposed calculating a weighted average of the strata-specific Kaplan-Meier estimates, weighting according to the baseline proportion of the study population in each stratum. A similar approach has been recently proposed by Cupples et al. (7), who have used it to calculate age-adjusted Kaplan-Meier curves based on categorical age strata of the study populations at baseline (20, 21).

These methods are straightforward, non-parametric, and allow for an estimation of the predicted survival curve of a given group while controlling for baseline imbalances in the distribution of significant confounders. However, these methods can not account for changes in the distribution of covariates in the study group during the follow-up (see below) and do not allow for time-dependent covariates.

Generalized Kaplan-Meier estimator

This method was proposed by Amato (8) and can be described as the survival analog of the Mantel-Haenszel estimate of the pooled odds ratio for stratified data (19). In calculating the cumulative survival, this method takes the product of weighted conditional

probabilities of survival. In calculating the latter, one takes one minus the ratio of a weighted average of the number of events to a weighted average of the number of individuals at risk. As in the previous methods, the weights are based on the size of the different strata at the beginning of the follow-up. This method has considerable merit and offers a substantial theoretical improvement over merely averaging the strata-specific Kaplan-Meier curves (6, 7) because it allows for the calculation of the adjusted survival function even after one of the strata has no subjects left. However, by adjusting to the baseline covariate distribution, this method has similar limitations as the above methods, namely that it does not take into account changes in the covariate distribution during follow-up and is not suitable for time-dependent covariates.

SURVIVAL BASED ON ADJUSTED CONDITIONAL PROBABILITIES

Let us assume that we have a group of individuals exposed to a suspected risk factor and another group of unexposed individuals. A variable X is created, with value 1 for the exposed and value 0 for the unexposed. Suppose we wish to compare the survival experience of these groups while adjusting for a dichotomous covariate Z ($Z = 1$ if the characteristic is present, 0 if absent). The overall goal of our method is to obtain an adjusted cumulative survival curve for the exposed group, representing the survival in the exposed individuals if they had the same covariate distribution as the unexposed throughout the entire follow-up period. At each time when an event occurs in the exposed group, we calculate the conditional probability of the event among exposed individuals and use direct adjustment to adjust it to the covariate (Z) distribution of the unexposed group at that time. The outline of the procedure is as follows:

1) Similar to the Kaplan-Meier estimator, the follow-up time for all events and censored observations are ordered from shortest to longest.

2) At each time t_i when an event occurs among the observations included in the exposed group ($X = 1$), all individuals in the exposed group alive and not censored before t_i are counted. Thus, at time t_i, we have n_{11i} and n_{10i} exposed individuals ($X = 1$) in each stratum of the covariate ($Z = 1$ and $Z = 0$, respectively), with a_{11i} and a_{10i} denoting the number of events in each group, respectively (typically, either a_{11i} or a_{10i} has a value of one and the other is zero, although ties can occur).

3) The stratified conditional probabilities in the exposed group are calculated by the usual formulas:

$$q_{11i} = \frac{a_{11i}}{n_{11i}} \qquad q_{10i} = \frac{a_{10i}}{n_{10i}}. \qquad (1)$$

4) The number of individuals that are still under observation at time t_i in the unexposed (reference) group is obtained: n_{01i} and n_{00i} in each stratum of the covariate ($Z = 1$ and $Z = 0$, respectively).

5) The adjusted conditional probability in the exposed group at time t_i, $q_{1i}{}^*$, is obtained by direct adjustment using the covariate distribution in the unexposed group as the standard, i.e., as a weighted average of the stratified conditional probabilities (step 3), using as weights the stratum-specific distribution in the reference group (step 4):

$$q_{1i}^* = \frac{q_{11i}n_{01i} + q_{10i}n_{00i}}{n_{01i} + n_{00i}}. \qquad (2)$$

6) The product of the conditional adjusted survival probabilities provides an estimate of the adjusted cumulative survival in the exposed group, using as standard the covariate distribution of the reference population (the unexposed) at any given point in time:

$$S_1^*(t) = \prod_{t_i \leq t}(1 - q_{1i}^*). \qquad (3)$$

7) The resulting adjusted survival curve for the exposed group could be presented in the same graph as the Kaplan-Meier curves for both exposed and unexposed groups to obtain a visual impression of the effect of adjustment in the comparison between the groups (see examples following).

Appendix 1 shows the derivation for the variance estimator of the natural log of the adjusted survival function (equation 7). (The use of this variance estimator to obtain the 95 percent confidence intervals of the adjusted survival estimates is illustrated in example 2 below.)

As the simulations described in Appendix 2 show, in the case of proportional hazards, the relative distance between the adjusted survival curve in the exposed and the observed survival in the unexposed (both corresponding to a population with the same covariate composition at each time t_i) converges to the theoretically expected relative cumulative hazard.

As shown in Appendix 2 and the examples below, the method can be extended to multiple categorical covariates. Notice that this method can handle time-dependent covariates, because the standard population and adjusted conditional probability of the event are estimated at each point in time (see example 2).

EXAMPLES

Example 1: Survival after Ewing's sarcoma comparing different treatment regimens

This example was used by Makuch (3) to illustrate his method of calculating adjusted survival. The data come from a study comparing the disease-free survival of Ewing's sarcoma patients treated with an aggressive new treatment (S4) and those treated with early treatment regimens (S1–S3) (22). The corresponding Kaplan-Meier curves (figure 2) suggest an improved survival in patients in the S4 group. The unadjusted relative hazard estimated from the Cox model was 0.53 (95 percent confidence interval (CI) 0.30–0.96). However, the S4 group had more individuals with low lactic acid dehydrogenase (LDH) level at baseline. Low LDH levels are associated with better prognosis. Thus, when adjusting for baseline LDH (categorically defined as high or low), there was no longer a significant survival difference. If anything, the estimated survival using the Cox model was worse for the S4 group in this study (adjusted relative hazard estimate = 1.12, 95 percent CI 0.59–2.11).

Figure 3 shows the adjusted survival curves obtained using the average covariate method with estimates from the Cox proportional hazards model. These curves were obtained using SAS's PROC PHREG, including the BASELINE COVARIATES statement (13). The covariate "high LDH" is set to 0.41, the overall proportion with high LDH in the study population. Notice that there is practically no difference between the two curves, which are approximately parallel, with steps at the same time points, as a result of the constraints of the model. Notice also that the survival is slightly worse in the S4 group, in correspondence to the adjusted relative hazard of 1.12.

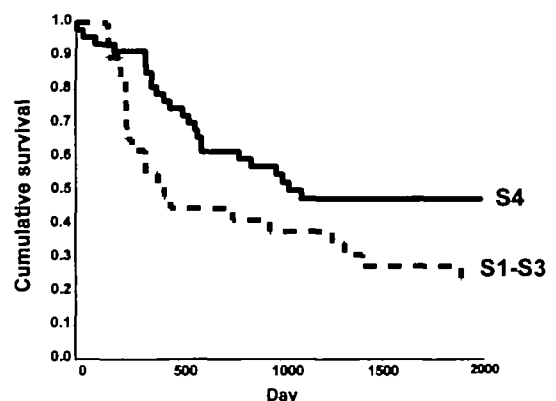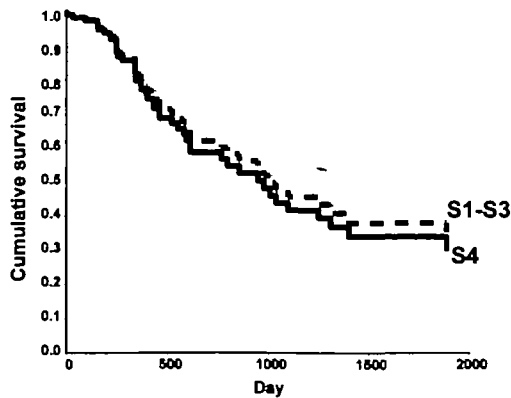Figure 4 shows the result of applying our method to the same data. The curves presented are the crude

FIGURE 2. Cumulative survival (Kaplan-Meier estimates) among Ewing's sarcoma patients treated with an aggressive new treatment (S4) and those treated with early treatment regimens (S1–S3). Sources: references 3 and 22.

FIGURE 3. Lactic acid dehydrogenase (LDH)-adjusted survival curves based on the Cox proportional hazards model comparing Ewing's sarcoma patients treated with an aggressive new treatment (S4) and those treated with early treatment regimens (S1–S3); Cox model estimate of the relative hazard = 1.12, comparing S4 with S1–S3.
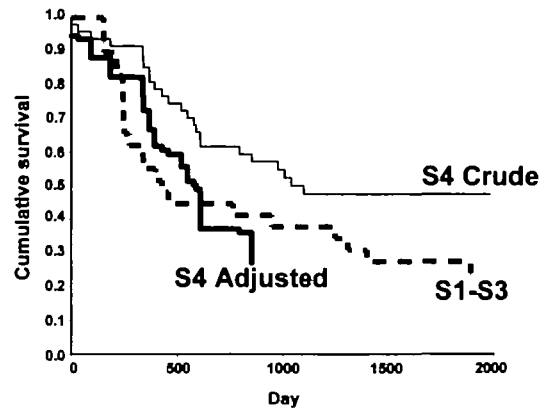
FIGURE 4. Crude survival curve in Ewing's sarcoma patients treated with early treatment regimens (S1–S3), and lactic acid dehydrogenase (LDH)-adjusted survival among S4-treated patients based on conditional probabilities adjusted for the LDH distribution among S1–S3-treated patients. (For comparison, the crude survival curve for the S4 group is shown as a thin solid line. The crude lines are identical to those presented in figure 2.)

Kaplan-Meier curves (those shown in figure 2) as well as the adjusted curve for the new treatment group (S4), adjusted for the covariate (LDH) composition in the early treatment group (S1–S3). The adjusted curve for S4 and the comparable curve for S1–S3 cross twice and almost overlap, suggesting that there is no difference in survival once pretreatment LDH levels are taken into account. In contrast to the Cox adjusted survival curves (figure 3), the curves in figure 4 are not forced to be powers of each other.

Example 2: Coronary heart disease incidence according to baseline coffee intake in the Johns Hopkins Precursors Study cohort

A previous article (23) on the association between reported baseline coffee intake and coronary disease incidence during the follow-up of the Johns Hopkins Precursors Study cohort described the study methodology in detail. Briefly, students who matriculated into the graduating classes of 1948–1964 of the Johns Hopkins University School of Medicine, provided baseline information on coffee intake, cigarette smoking, and other health-related data, as well as a serum sample. This cohort has been followed to December 1992 with yearly questionnaires to assess the occurrence of coronary disease and to update smoking information. In addition, vital status has been ascertained by contacting family members and searching the National Death Index.

Of the 1,040 white men who provided baseline information on coffee intake, 921 also had serum cholesterol and smoking data. This analysis compares the 330 participants (36 percent) who reported drinking an average of ≥3 cups/day of coffee with the 591 participants who reported drinking <3 cups/day.

Among the 330 drinkers of ≥3 cups/day, 48 (14.5 percent) developed coronary disease during follow-up, compared with 38 out of 591 drinkers of <3 cups/day (6.4 percent). Univariate Cox proportional regression analyses resulted in an estimated crude relative hazard of 2.4 (table 1). An obvious concern in interpreting the apparent increased risk in coffee drinkers is the possibility of confounding by smoking (the proportion smoking at baseline was 64 percent among drinkers of ≥3 cups/day versus 39 percent among the drinkers of <3 cups/day). As shown in table 1 and in a previous paper (23), even after adjusting for baseline or time-dependent smoking and baseline cholesterol levels, an elevated risk associated with coffee intake is still evident.

Figure 5 shows the survival curves, adjusted for the baseline smoking status and baseline cholesterol level, using the average covariate method based on Cox model estimates. (Survival curves adjusted for time-dependent smoking are not obtainable using SAS (13, p. 19).)

Figure 6 shows the crude survival curves as well as the adjusted survival for coffee drinkers, using the baseline cholesterol (categorized in three levels) and the distribution of current smoking at each time (time-dependent smoking) among non-coffee drinkers as reference. The figure also shows the 95 percent confidence interval for the adjusted curve at four points in time.

DISCUSSION

Despite the fact that methods to obtain adjusted survival curves are available, there is a surprising lack

TABLE 1. Crude and adjusted relative hazards* of coronary heart disease risk associated with baseline coffee consumption of 3 cups/day compared with <3 cups/day: the Johns Hopkins Precursors Study, 1948–1992

|	Relative hazard	95% confidence interval
Crude	2.40	1.57–3.67
Adjusted for baseline smoking†	2.05	1.32–3.18
Adjusted for time-dependent smoking†,‡	2.28	1.48–3.51
Adjusted for baseline smoking† and baseline cholesterol§	2.00	1.29–3.01
Adjusted for time-dependent smoking†,‡ and baseline cholesterol‖	2.23	1.45–3.43

* Obtained from Cox proportional hazards regression.
† Current cigarette smoking, yes vs. no.
‡ Smoking at the time of the latest questionnaire.
§ Serum cholesterol categorized as <200 mg/dl and ≥200 mg/dl.
‖ Serum cholesterol categorized as <200 mg/dl, 200–240 mg/dl, and ≥240 mg/dl.
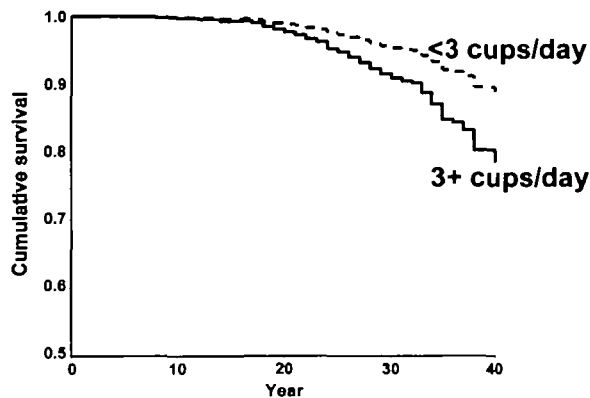
FIGURE 5. Survival free of coronary disease adjusted for baseline smoking and cholesterol level, based on the Cox proportional hazards model, in drinkers of ≥3 cups/day of coffee and drinkers of <3 cups/day. These curves correspond to a relative hazard of 2.00 (table 1), and have been calculated for the average value of the covariate in the study population (proportion smoking = 0.48; proportion with hypercholesterolemia (serum cholesterol ≥200 mg/dl) = 0.36). The Johns Hopkins Precursors Study, 1948–1992.
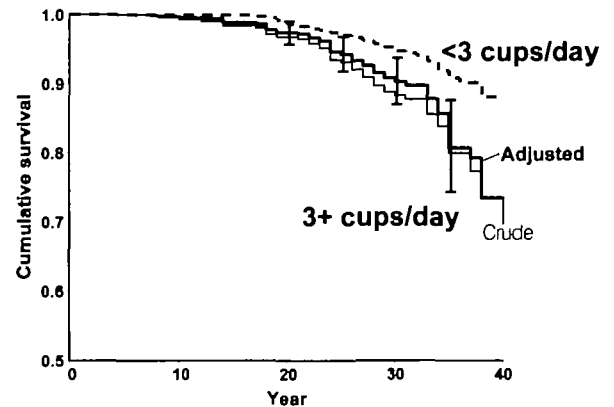
FIGURE 6. Survival free of coronary disease among drinkers of <3 cups/day of coffee compared with the adjusted survival among drinkers of ≥3 cups/day at baseline (thick solid line). The latter has been adjusted to the baseline serum cholesterol and time-dependent smoking characteristics of non-coffee drinkers or light coffee drinkers (<3 cups/day) using the method described in this paper. Smoking and cholesterol were categorized as in the last model in table 1. The bars represent the 95% confidence intervals for the adjusted survival estimates among drinkers of ≥3 cups/day at 20, 25, 30, and 35 years of follow-up. (The crude survival curve in that group is represented by a thin solid line.) The Johns Hopkins Precursors Study, 1948–1992.

of discussion about the application of such methods in the epidemiologic literature. When adjusted survival curves are presented, authors seem to lean toward the Cox model-based average covariate method (24), in spite of its theoretical problems and practical limitations. This method is readily available in commercial software (e.g., SAS), and we are concerned that some users of this application may not be aware of its limitations. The average covariate method is useful in illustrating the impact of an increased hazard due to an exposure of interest on a given survival curve. As we discussed previously, the method is not useful in predicting the survival of a heterogeneous group of individuals. The "alternative" methods described above (3–8) are much more useful in the latter situation.

However, these methods are rarely used in the epidemiologic literature, perhaps because most investigators are unaware of their existence, are unable to implement them, or feel intimidated by the relatively high mathematical sophistication of some of the papers that describe them.

The method we propose adjusts the conditional probability of survival at each event time to the covariate distribution of a standard population at that time. In the examples we present above, we illustrate its application using the unexposed group as the reference standard population. This can be interpreted as the survival of the exposed group if its members had

the covariate composition of the unexposed group throughout the follow-up period. Alternatively, the adjusted curve may be interpreted as the survival of the unexposed group if it experienced the added risk of exposed individuals.

It is important to note that this survival curve is a theoretical curve because it applies the survival probabilities of one group (the exposed) to the covariate composition of another group (the unexposed). This is useful for isolating the effect of an exposure of interest while controlling for confounding by covariates at all points in time. However, this is not the same as the survival of an exposed group with the same baseline composition as the unexposed group. For example, if an exposure, such as coffee consumption, leads to increased risk of mortality, it will result in preferential depletion in the exposed group compared with the unexposed group of individuals with covariates (e.g., smoking) which put them at increased risk of mortality. Thus, even if the two groups were comparable at baseline, the elevated risk due to exposure combined with the interindividual heterogeneity will result in a lack of comparability over time. Therefore, the proposed method is most useful in illustrating the impact on survival of the added hazard of an exposure, X, when completely isolated from a set of covariates, Z. In this sense, this method is analogous to the average covariate method and is most useful in looking for etiologic associations that are independent of a set of covariates. In fact, models such as the Cox proportional hazards which model the relation between a set of covariates and the underlying hazard inherently do the same thing but make stronger assumptions.

The advantage of the proposed method over the average covariate method based on the Cox model is that it does not impose one underlying shape on the survival curves that are being compared. This can be important in the exploratory survival analysis of a situation where strong confounding is suspected and one would like to examine the proportionality assumption visually. In addition, the proposed method is inherently time dependent and thus naturally allows for incorporation of time-dependent covariates. As a corollary, the method can also handle left truncated data (not shown in this paper). This is helpful when using age as the time scale in cohorts with a heterogeneous age distribution at entry. A further advantage over the average covariate method is that standard errors of the adjusted survival estimates can be easily estimated, as described above.

We need to emphasize, however, that if one is interested in estimating the survival experience of a given group adjusted to have certain characteristics at baseline, methods that average the survival estimates

according to a fixed baseline standard (3, 4, 6, 7) will be more appropriate than our method.

Another drawback of the method proposed here is that it can not adjust for continuous covariates, unless they are categorized. Furthermore, the adjusted survival curve stops when one of the strata defined by the covariates runs out of observations in the exposed group (so that one of the probabilities in equation 1 is missing) while there are still observations left in that stratum among the exposed, i.e., when equation 2 can not be calculated. (See, for example, figure 4 and Appendix 2.) This could be an issue, particularly when trying to adjust for many covariates. However, in many epidemiologic studies with fairly large populations and relatively rare events, this would not be a major limitation. Furthermore, even if the curve stops before the end of the follow-up period, one could still obtain an adequate assessment of the effect of adjustment in most instances (see example 1, figure 4).

The possible lack of reliability of each of these adjusted conditional probability estimates at each step, particularly at the end of the follow-up time, is analogous to the instability of each of the steps of the unadjusted Kaplan-Meier curve when the data are sparse. The strength of this method, as well as of the Kaplan-Meier method, relies on the interpretation of the entire curve rather than on the individual steps. Our method has a non-parametric flavor, in correspondence with the spirit of the Kaplan-Meier estimates.

The method that we propose provides an alternative view of the data, allowing for a visual, non-parametric assessment of the influence of covariates on the comparison of survival in two or more groups (see examples above). It could be used as an intermediate exploratory tool between the examination of the Kaplan-Meier curves and multivariate regression models. This method will help the investigators to assess the relevance of the adjusted relative hazard (i.e., the adequacy of the proportional hazards assumption) and of the survival curves estimated by means of the Cox or alternative parametric survival models.

REFERENCES

1. Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457–81.
2. Cox DR. Regression models and life-tables (with discussion). J R Stat Soc [B] 1972;34:187–220.
3. Makuch RW. Adjusted survival curve estimation using covariates. J Chronic Dis 1982;35:437–43.
4. Chang IM, Gelman R, Pagano M. Corrected group prognostic curves and summary statistics. J Chronic Dis 1982;35:668–74.
5. Hankey BF, Myers MH. Evaluating differences in survival between two groups of patients. J Chronic Dis 1971;24:523–31.
6. Kramar A, Com-Nougué C. Estimation des courbes de survie ajustées. Rev Epidémiol Santé Publique 1990;38:149–52.
7. Cupples LA, Gagnon DR, Ramaswamy R, et al. Age-adjusted survival curves with application in the Framingham Study. Stat Med 1995;14:1731–44.
8. Amato DA. A generalized Kaplan-Meier estimator for heterogeneous populations. Commun Statist Theory Meth 1988;17:263–86.
9. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: Wiley & Sons, Inc, 1980.
10. Grouven U, Bender R, Schultz A, et al. Application of adjusted survival curves to renal transplant data. Methods Inf Med 1992;31:210–14.
11. Armitage P, Berry G. Statistical methods in medical research. 2nd ed. Oxford: Blackwell Scientific Publications, 1987.
12. Breslow NE, Day NE. Statistical methods in cancer research. Vol 2. The design and analysis of cohort studies. (IARC scientific publications no. 32). Lyon: International Agency for Research on Cancer, 1987.
13. SAS Institute Inc. SAS technical report P-217. SAS/STAT software: PHREG procedure, version 6. Cary, NC: SAS Institute Inc, 1991.
14. Thomsen BL, Keiding N, Altman DG. A note on the calculation of expected survival, illustrated by the survival of liver transplant patients. Stat Med 1991;10:733–8.
15. Neuberger J, Altman DG, Christensen E, et al. Use of a prognostic index in evaluation of liver transplantation for primary biliary cirrhosis. Transplantation 1986;4:713–16.
16. Lee J. Covariance adjustment of rates based on the multiple logistic regression model. J Chronic Dis 1981;35:437–43.
17. Lee J, Yoshizawa C, Wilkens L, et al. Covariance adjustment of survival curves based on Cox's proportional hazards regression model. Comput Appl Biosci 1992;8:23–7.
18. Gail MH, Byar DP. Variance calculations for direct adjusted survival curves, with applications to testing for no treatment effect. Biom J 1986;28:587–99.
19. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. JNCI 1959;22:719–48.
20. Kiely DK, Wolf PA, Cupples LA, et al. Physical activity and stroke risk: the Framingham Study. Am J Epidemiol 1994;140:608–20.
21. Gillman MW, Cupples LA, Gagnon D, et al. Protective effect of fruits and vegetables on development of stroke in men. JAMA 1995;273:1113–17.
22. Glaubiger DL, Makuch R, Schwartz J. Determination of prognostic factors and their influence on therapeutic results in patients with Ewing's sarcoma. Cancer 1980;45:2213–19.
23. Klag MJ, Mead LA, LaCroix AZ, et al. Coffee intake and coronary heart disease. Ann Epidemiol 1994;4:425–33.
24. Altman DG. Practical statistics for medical research. London: Chapman & Hall, 1991.
25. Greenwood M. A report on the natural duration of cancer. Rep Public Health Med Subjects 1926;33:1–26.

APPENDIX 1

Variance estimator

The adjusted conditional probability of the event in the exposed group ($X = 1$) at time t_i (equation 2) after controlling for covariates with k strata can be expressed in a more general way as:

$$q_{1i}^* = \sum_{j=1}^{k} w_{ji} q_{1ji}, \qquad (4)$$

where

$$w_{ji} = \frac{n_{0ji}}{\sum\limits_{j=1}^{k} n_{0ji}}$$

and n_{0ji} is the number of people in stratum j at time t_i in the comparison group (the unexposed), while q_{1ji} is the conditional probability of the event among the exposed ($X = 1$), in stratum j, at time i (equation 1).

Taking the natural log of the adjusted survival estimate (equation 3) and substituting equation 4 into equation 3:

$$\log S_1^*(t_i) = -\Lambda(t_i) = \sum_{t_i \leq t} \ln\left(1 - \sum_{j=1}^{k} w_{ji} q_{1ji}\right). \qquad (5)$$

Because the failure probabilities in the different time points are independent, the variance of equation 5 (the negative cumulative hazard) can be written as:

$$\text{Var}(\Lambda(t_i)) = \sum_{t_i \le t} \text{Var}[\ln(1 - \sum_{j=1}^{k} w_{ji}q_{1ji})]. \tag{6}$$

By the delta method, because the different strata are independent:

$$\text{Var}(\Lambda(t_i)) = \sum_{t_i \le t} \text{Var}(\sum_{j=1}^{k} w_{ji}q_{1ji}) \frac{\partial}{\partial(\sum w_{ji}q_{1ji})}[\ln(1 - \sum_{j=1}^{k} w_{ji}q_{1ji})]^2_{E(\Sigma q_{1ji})}$$

$$= \sum_{t_i \le t} \sum_{j=1}^{k} w_{ji}^2 \left[\frac{(1 - q_{1ji})q_{1ji}}{n_{1ji}} \right] \left[\left(\frac{1}{1 - \sum_{j=1}^{k} w_{ji}q_{1ji}} \right)(-1) \right]^2_{E(\Sigma q_{1ji})} .$$

Substituting q_{1ji} for $E(q_{1ji})$, as is done in Greenwood's formula (25), we get:

$$\text{Var}(\Lambda(t_i)) = \sum_{t_i \le t} \sum_{j=1}^{k} w_{ji}^2 \left[\frac{(1 - q_{1ji})q_{1ji}}{n_{1ji}} \right] \frac{1}{(1 - \sum_{j=1}^{k} w_{ji}q_{1ji})^2}. \tag{7}$$

It can be shown that equation 7 reduces to Greenwood's formula (25) when the survival curve in the exposed is adjusted to itself, i.e., when the weights w_{ji} in equation 4 are calculated using n_{1ji} rather than n_{0ji}.

Using a normal approximation and equation 5, the 95 percent confidence intervals for the adjusted survival estimate (equation 3) can be obtained as:

$$\exp \left(- \Lambda(t_i) \pm 1.96 \times \sqrt{\text{Var}(\Lambda(t_i))} \right). \tag{8}$$

APPENDIX 2

Simulations

We conducted a set of Monte Carlo simulations ($n = 500$) based on the following model. The study population for each iteration comprised 750 individuals, or 375 exposed ($X = 1$) and 375 unexposed ($X = 0$). The joint distribution of X and two dichotomous covariates (Z_1 and Z_2) were as follows: 195 ($X = 0, Z_1 = 0, Z_2 = 0$), 55 ($X = 0, Z_1 = 0, Z_2 = 1$), 55 ($X = 0, Z_1 = 1, Z_2 = 0$), 70 ($X = 0, Z_1 = 1, Z_2 = 1$), 55 ($X = 1, Z_1 = 0, Z_2 = 0$), 70 ($X = 1, Z_1 = 0, Z_2 = 1$), 70 ($X = 1, Z_1 = 1, Z_2 = 0$), and 180 ($X = 1, Z_1 = 1, Z_2 = 1$). Note that both Z_1 and Z_2 are related to X (odds ratio $= 4$, for each Z_i and X while ignoring the other Z). The survival times were set to follow a Weibull distribution with scale, $\lambda = \exp \{(\beta X + \beta_1 Z_1 + \beta_2 Z_2)/\gamma\}$, where $\beta = \ln(2)$, $\beta_1 = \ln(3)$, $\beta_2 = \ln(1.5)$, and shape parameter, $\gamma = 2$. Note that this model

corresponds to a hazard $h(t) = 2 \times t \times \exp (\beta X + \beta_1 Z_1 + \beta_2 Z_2)$, which nicely fits the Cox proportional hazards model of $h(t) = h_0(t) \times \exp (\beta X + \beta_1 Z_1 + \beta_2 Z_2)$.

At each iteration, the survival times for each of the 750 individuals were randomly generated from the corresponding density function. The adjusted survival curves in the exposed were obtained at each cycle using the method described above and the unexposed as the comparison group. Figure A1 shows the results of one simulated study population in a graph of the crude survival curves as well as the survival curve for the exposed ($X = 1$) adjusted to the covariate distribution among the unexposed. The adjusted curve has been truncated at the first time when one of the strata runs out of exposed individuals (one of the q_{ji} is missing while the corresponding $n_{0ji} > 0$), so that equation 2 is undefined.
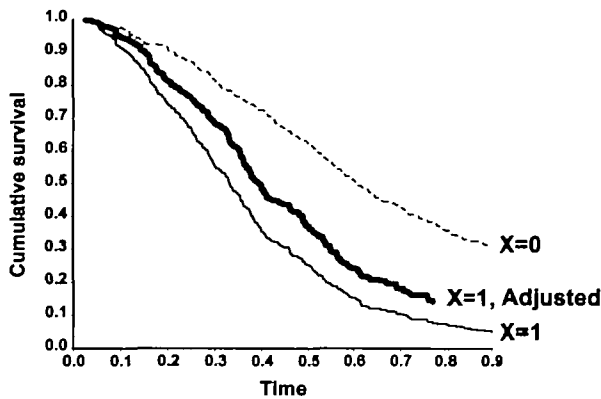
FIGURE A1. Crude and adjusted survival curves for one iteration of the simulations described in the text. The adjusted survival for the exposed group ($X = 1$) is adjusted to the covariate (Z) distribution in the unexposed.
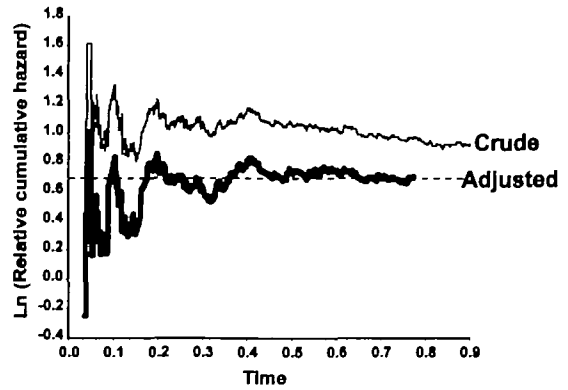
FIGURE A2. Difference in the ln (cumulative relative hazard) (estimated by the difference in the complementary log function) for the survival curves shown in figure A1. The thin line corresponds to the difference between the two crude curves (exposed minus unexposed). The thick line corresponds to the difference between the adjusted curve in the exposed and the unexposed. The broken horizontal line is set at 0.693 (natural logarithm of 2.0).

Theoretically, the survival function can be expressed as a function of the cumulative hazard (Λ):

$$S(t) = e^{-\Lambda(t)}.$$

Thus, the ratio of the natural logarithm of two survival functions [$S_1(t)$ and $S_0(t)$] equals the relative cumulative hazard at time t [RH(t)]:

$$\frac{\ln(S_1(t))}{\ln(S_0(t))} = \frac{-\Lambda_1(t)}{-\Lambda_0(t)} = \text{RH}(t).$$

Taking log at both sides:

$$\ln\left(\frac{-\ln(S_1(t))}{-\ln(S_0(t))}\right) = \ln[-\ln(S_1(t))] - \ln[-\ln(S_0(t))]$$

$$= \ln[\text{RH}(t)].$$

Thus, the difference between the $\ln(-\ln)$ functions of the two survival curves is the natural log of the relative cumulative hazard. Figure A2 displays these relative cumulative hazard functions comparing the crude and adjusted survival curves from the single population represented in figure A1. The curve in bold corresponds to the adjusted comparison. In correspondence with the theoretical expectation, and after some initial random fluctuation, the curve converges around the value 0.693, which is the natural log of 2 (the modeled relative hazard for X).

Table A1 shows the results of the 500 simulations based on the above model. In accordance with the proportionality built into the simulation, the mean Cox regression coefficient is very close to the expected (log 2 = 0.693). The difference in $\ln(-\ln)$ functions empirically obtained from the curves adjusted using our method also tend to converge around the true value. The initial random fluctuations seen in figure A2 are

TABLE A1. Summary of 500 simulations each with 750 individuals† (see text).

		ln (relative cumulative hazard) for X			
	Time	No.	Mean	SD‡	Percentiles 5, 95
Survival based on adjusted conditional probabilities	0.1	500	0.779	0.499	0.011, 1.652
	0.2	500	0.690	0.251	0.289, 1.124
	0.3	500	0.681	0.173	0.378, 0.942
	0.4	500	0.690	0.137	0.482, 0.907
	0.5	500	0.698	0.125	0.502, 0.904
	0.6	500	0.699	0.121	0.510, 0.911
	0.7	437§	0.689	0.119	0.468, 0.864
	0.8	173§	0.654	0.100	0.510, 0.807
Cox regression coefficient		500	0.699	0.087	0.556, 0.840
Theoretical expectation			0.693		

† Survival simulated based on the Weibull distribution with $\lambda = \exp\{[\ln(2)]X + \ln(3)Z_1 + \ln(1.5)Z_2]/2\}$ and $\gamma = 2$.
‡ SD, standard deviation.
§ Some adjusted survival curves stopped before this time (see text).

reflected in relatively large standard deviations in the earlier times. By time = 0.5, the empirical estimates become rather stable, although, as expected, the standard deviation of the estimate at any point in time is always slightly larger than the standard deviation obtained from the overall estimate from the Cox regression, which inherently assumes the underlying hazards to be proportional.