

Evaluating Disease Management Program Effectiveness: An Introduction to Survival Analysis

ARIEL LINDEN, Dr.P.H., M.S.,¹ JOHN L. ADAMS, Ph.D.,² and NANCY ROBERTS, M.P.H.³

ABSTRACT

Currently, the most widely used method in the disease management industry for evaluating program effectiveness is the "total population approach." This model is a pretest-posttest design, with the most basic limitation being that without a control group, there may be sources of bias and/or competing extraneous confounding factors that offer plausible rationale explaining the change from baseline. Survival analysis allows for the inclusion of data from censored cases, those subjects who either "survived" the program without experiencing the event (e.g., achievement of target clinical levels, hospitalization) or left the program prematurely, due to disenrollment from the health plan or program, or were lost to follow-up. Additionally, independent variables may be included in the model to help explain the variability in the outcome measure. In order to maximize the potential of this statistical method, validity of the model and research design must be assured. This paper reviews survival analysis as an alternative, and more appropriate, approach to evaluating DM program effectiveness than the current total population approach. (Disease Management 2004;7:180-190.)

INTRODUCTION

THE DISEASE MANAGEMENT (DM) industry is currently at a crossroads in its existence. Health plan and employer group partners are no longer willing to blindly assume that positive health outcomes and cost savings are guaranteed without evidence of such, and DM programs are scrambling to demonstrate that their programs indeed work.¹⁻³

In previous papers, Linden et al⁴ described the limitations of the currently-used "total population approach" for determining DM program cost-savings effectiveness and offered time-series analysis as an alternative method

for evaluating DM programs on an aggregate basis.⁵ The current paper offers an additional application of health service research methods to the DM industry for the evaluation of program effectiveness using survival analysis on individuals and cohorts.

The advantage of survival analysis over the currently used total population approach is that it offers insight into the effect of disease process progression over time while providing the ability to measure the impact of secondary prevention techniques on these processes. More specifically, as shown in Figure 1, survival analysis can be used to accurately determine how long it takes for the DM interven-

¹Linden Consulting Group, Portland, Oregon.

²RAND Corporation, Santa Monica, California.

³Providence Health System, Portland, Oregon.

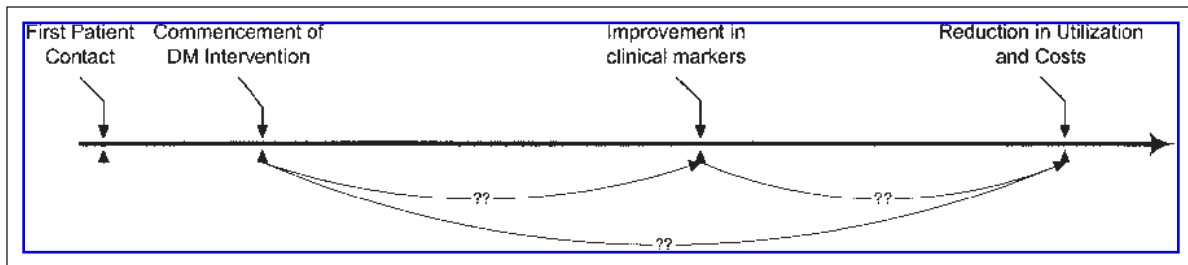


FIG. 1. The intended process of a typical disease management program. The timelines still left undetermined are between the commencement of the disease management (DM) intervention and improvement in patient-level physiological markers, and the translation of that improvement into reduced utilization.

tions to improve patient physiologic markers (e.g., HbA1c levels in diabetics, LDL levels in patients at risk for coronary artery disease, pulmonary function in asthmatics), and how long after that reductions in utilization and cost become evident.

This paper presents a non-technical introduction to survival analysis as an alternative, and more appropriate, approach to evaluating DM program effectiveness than the current total population approach. This introduction will provide DM program evaluators, both within DM firms and for those managing in-house programs, enough detail to begin using these techniques. For those organizations who purchase DM services, this paper will provide a substantive background with which to discuss alternative evaluation possibilities with their contracted vendors.

PRINCIPLES OF SURVIVAL ANALYSIS

In much medical related research, the outcome of interest is the time to an event. Typically, the event is survival or, conversely, mortality, over a given period of observation. However, any event can be selected as the endpoint, such as a hospitalization, absence, or presentation of a symptom, or reaching a threshold for a physiologic marker.⁶⁻⁸

What makes survival analysis unique among the many statistical methods is that data from patients who do not realize the “event” by the end of the study are included in the development of the model. These patients’ survival times are called *censored*, indicating that the study period ended before the event occurred, or that the patient may have been lost to fol-

low-up at some point during the study. In either case, the censored survival times are used along with the survival times of patients who ultimately experienced the event in order to construct the survival analysis model.

An important feature of survival studies is that patients are enrolled at various points during the observation period and then followed until the end of the study. As such, patients enrolled near the end of the study will be followed for a shorter period of time than those patients enrolled early on, and thus will have a lower likelihood of experiencing the event. Nonetheless, it would be incorrect to assume that because these patient data are censored they have a better or worse prognosis than those patients who experienced the event after being observed for a longer period during the study. For example, a patient who “survived” the study for one year before its termination may not have a shorter survival time than a patient who was enrolled in the study for 2 years before ultimately experiencing the event.

Another important characteristic of survival analysis is that it is prospective in nature. In other words, survival data from a selected group of patients should be followed forward in time from their inclusion in the study, even if the patients were identified retrospectively. It would be incorrect and potentially misleading to work backward from the event to enrollment.⁹

The general principles discussed thus far support survival analysis as being one of the most powerful analytic tools available for evaluating DM program effectiveness. For example, DM programs manage a population of patients that move freely in and out of the program during the contract period (analogous

to the end of a study). This is due to new enrollment or disenrollment in the health plan, recent identification of suitability based on disease state, members opting in or out of the program, or death. Using survival analysis, each of these patient's censored data can be used to develop the model for predicting time to event. Additionally, the outcome event for a DM program can be either a utilization measure (e.g., time to first hospitalization from enrollment) or a clinical indicator (e.g., time to receipt of a lipid panel from first nursing intervention). Also, as will be discussed later in the paper, survival analysis allows for a comparison between cohorts, and the inclusion of explanatory variables to assist in determining which specific patient or program characteristics are related to better outcomes.

SURVIVAL ANALYSIS MODELS

The two most widely used models for performing survival analysis are the Kaplan-Meier method and the Cox proportional hazards regression model. The Kaplan-Meier method is suitable if a simple model is required with no additional needed for explaining the variation in survival time vis-à-vis independent variables. The Cox regression model is similar in form and function to ordinary regression models. This technique allows the analyst to introduce independent variables and interaction terms to help explain the variation of the survival time function, making it a robust model for theoretical and practical applications. In this section, both models will be briefly described and illustrated using an example.

Kaplan-Meier method

Using the Kaplan-Meier method,¹⁰ the probability of surviving to the end of each observation period (e.g., days, months, years) is estimated for each patient surviving at the beginning of that period. This procedure is labeled *conditional probability* because the estimate of surviving to the end of the current observation period is based on the condition that the patient survived the last period. Censored patients contribute to the probability of sur-

living each observation period for which they are followed in the study, while patients who experience the "event" will decrease the probability of survival for the next observation period. The term *cumulative survival probability* refers to the product of all the conditional probabilities of surviving each observation period and is symbolized by the notation $S(t)$.

As an example of how the Kaplan-Meier method is executed, we will use data from Linden and Schweitzer.⁸ In this study, newly enrolled HMO Medicare members completed a health risk assessment called the P_{RA} test¹¹ where the results classified them as being either at low or high risk of hospitalization. A Kaplan-Meier survival analysis was performed to determine what the probability of hospitalization was for the low- and high-risk cohorts over the course of the following 25-month period.

Table 1 presents a sample of eight of the 2920 members who experienced a hospitalization during the study (there were 10,749 censored cases). As illustrated, members' conditional survival probability decreased with time, indicating a higher probability of hospitalization. Similarly, members whose P_{RA} score classified them as high risk had a lower conditional survival probability (and thus a higher likelihood of being hospitalized) than those members classified as low risk.

Figure 2 graphically displays the estimated [1 – cumulative survival probability] curves, plotted for both the low- and high-risk subjects in the study ($N = 13,152$ and 517 , for the low- and high-risk cohorts respectively). Subtracting the cumulative survival probability from 1 provides a more meaningful presentation of the likelihood of hospitalization (as opposed to the probability of not experiencing a hospitalization if we used the standard cumulative survival probability curve).¹² Each curve is illustrated as a step function, with each step corresponding to times at which a hospitalization was observed. The times of the censored data are indicated by X markers. Upon visual inspection we see that the probability of hospitalization in the high-risk group is about twice that of the low-risk group, reaching approximately 50% at month 25 of the study.

TABLE 1. SAMPLE SURVIVAL DATA

<i>P_{RA}</i> score [high risk (1)/low risk (0)]	Status [event (1)/censored (0)]	Survival time (month 5)	Conditional probability	Standard error(SE)
0	1	1	0.99133	0.00081
1	1	1	0.96712	0.00784
0	1	12	0.84672	0.00314
1	1	12	0.63056	0.02123
0	1	18	0.79799	0.00364
1	1	18	0.54865	0.02253
0	1	23	0.77074	0.00422
1	1	23	0.49669	0.02600

Data from Linden and Schweitzer.⁸

In order to make an inference about the population from which these cohorts of patients were drawn, an analyst can choose among three different statistical tests: The log-rank (or Mantel-Cox) test, the Breslow (or generalized Wilcoxon) test and the Tarone-Ware test.¹³ Each of these tests compares the number of hospitalizations that were observed to the number that were expected, which is calculated from the number of surviving members and the number of members who had the hospitalization at each period in the study.

Tables 2 and 3 provide summary statistics and results of the statistical tests performed on these data. As shown, about 80% of the low-risk cohort was censored (did not experience the “event”), whereas only 54% of the high-risk cohort was censored, indicating that nearly half of the high-risk group was hospitalized. All three statistical tests were highly significant, inferring that indeed Medicare members who were classified as high risk on the P_{RA} test had a significantly higher probability of being hospitalized than those who were classified as low

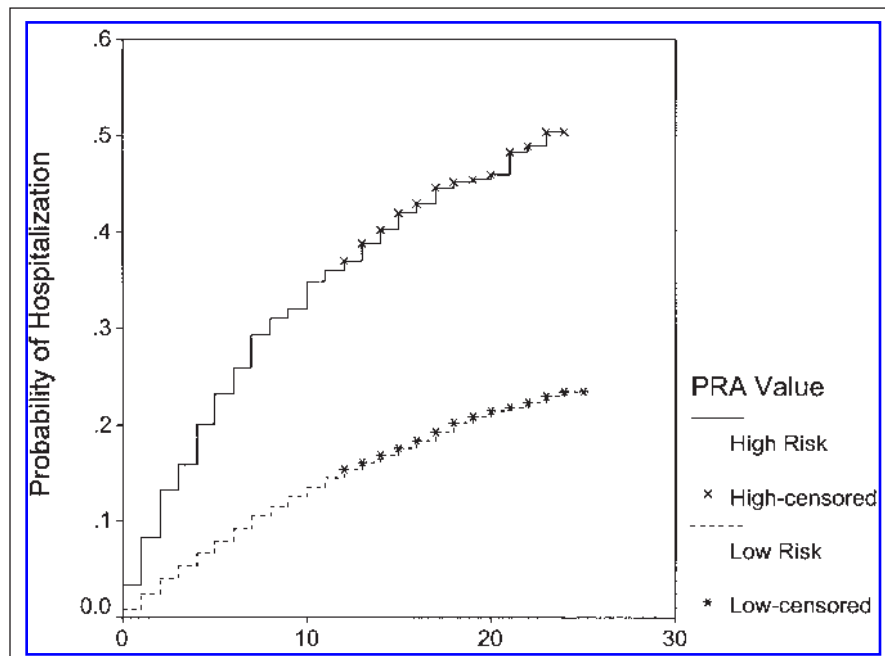


FIG. 2. Cumulative survival probability curves, plotted for both the low- and high-risk subjects. As illustrated, there is an approximate twofold likelihood of hospitalization in the high-risk cohort over the low-risk group.⁸

TABLE 2. SUMMARY STATISTICS

<i>P_{RA}</i> score [high risk (1)/low risk (0)]	Total	Number of events	Number censored	Percent censored
0	13,152	2683	10,469	79.60
1	517	237	280	54.16
Total	13,669	2920	10,749	78.64

Data from Linden and Schweitzer.⁸

risk. All analyses presented in this section were performed using SPSS for Windows, Release 11.0.1 (2001).

The Cox proportional hazards regression model

The Cox regression model¹⁴ is arguably the most widely used method for analyzing survival data.¹⁵ The appeal is in its similarity to standard regression methods, in which a dependent or outcome variable is explained by one or more independent variables. Since most analysts are familiar with the techniques used for model development in regression, the transition to use of the Cox model is seamless.

To illustrate the application of the Cox regression method, additional unpublished data from Linden and Schweitzer⁸ will be used. Specifically, we will add age and gender as independent variables to assess their contribution to the model. Age is given as a continuous variable and gender is presented as a dichotomous variable with 0 representing males and 1 representing females.

Table 4 presents the results of the statistical analysis for these data. Both age and gender were found to be significant variables in the model. Each unit increase in a patient's age is expected to increase the risk of hospitalization by 2.6%, while being female reduces the risk of hospitalization by nearly 8%.

TABLE 3. STATISTICAL TESTS ON SURVIVAL DATA

Test	Statistic	df	Significance
Log rank	239.46	1	<0.00001
Breslow	245.02	1	<0.00001
Tarone-Ware	244.08	1	<0.00001

Data from Linden and Schweitzer.⁸

In any regression analysis, it is extremely important to test for colinearity between independent variables. In other words, if two variables, each of which has shown a significant correlation with the survival time, are also strongly correlated with each other, the results of the model will be difficult to interpret. Therefore a matrix should be produced that tests the correlation between independent variables.¹⁶ In the present model no correlation was detected between the two variables; age and gender.

Figure 3 illustrates the survival plots for the Cox regression model based on the mean values of age (74.3 years) and gender (59% female). As expected, the curves are identical to those using the Kaplan-Meier technique. The lower level of probability of hospitalization for the two curves in this figure is due to the fact that the plots were based on the mean values of the independent variables used in the model. The Appendix provides a more detailed description of this analysis and a discussion of the concerns that must be addressed in using the Cox model.

APPLYING SURVIVAL ANALYSIS TO DISEASE MANAGEMENT PROGRAM EVALUATION

Thus far in the paper, the process for developing and analyzing survival analyses has been presented. As stated early on and depicted in Figure 1, what DM programs currently lack is the understanding of the timeline of program intervention impact on patient physiologic markers and the ultimate reduction in utilization that stems from that. The models used in this paper for demonstration purposes also provide a good illustration of the

X^2 model	df	p model	Variables	Regression coefficient (B)	SE(B)	p variables	Exp(B)	95% CI for Exp(B)
118.72	2	<0.0001	Age	0.25	0.002	<0.000	1.026	1.021–1.030
			Gender	−0.088	0.039	<0.023	0.916	0.849–0.988

Data from Linden and Schweitzer.⁸

types of analyses that should be performed for evaluating DM program effectiveness. In this section we will provide some ideas for the practical application of this analysis to evaluate and/or improve a DM program's effectiveness in several areas: validation of risk stratification methods, tailoring of interventions to specific patient characteristics, understanding the timing between intervention, clinical impact and utilization impact, and ultimately determining the degree of program intervention impact on utilization measures.

In the HMO Medicare health risk assessment example used earlier, the models validated the risk stratification tool by demonstrating that

the high-risk group had double the probability of the low-risk group of being hospitalized within a 25-month period. DM programs typically use certain claims-based criteria for stratifying patients into risk categories for having future utilization events (e.g., hospitalization, ED visits). Survival analysis can assess the validity of those algorithms in properly identifying patients at different risk levels.

Once the risk assessment tool has been validated (which would be indicated by a clinically meaningful or statistically significant difference in the probability of an acute event by risk level), the second analysis should be performed. Here, the timeline between the com-

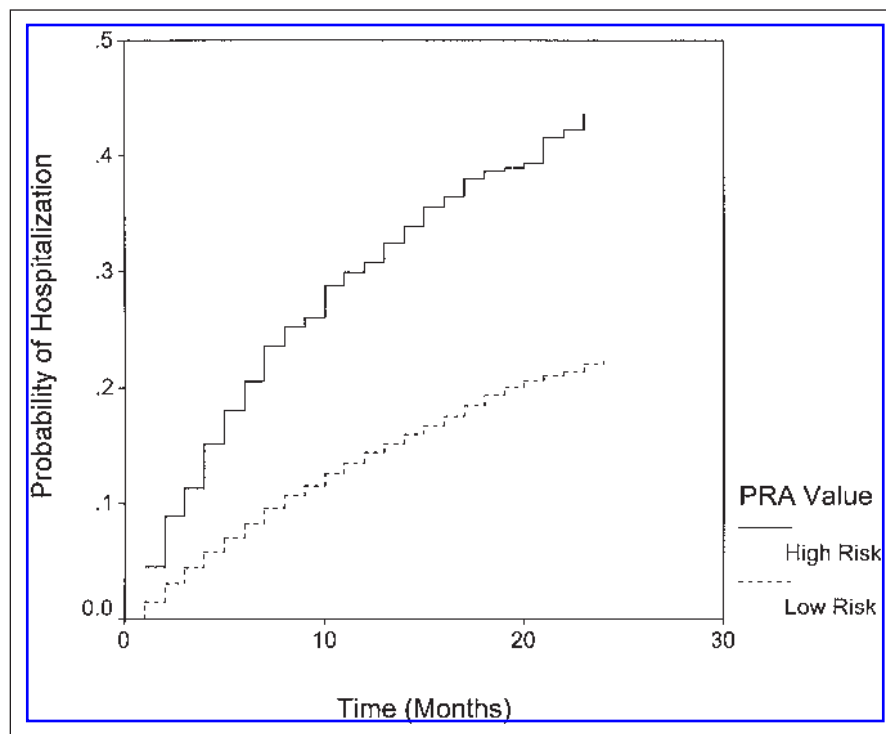


FIG. 3. Cumulative survival probability curves from Cox regression, plotted for both the low- and high-risk subjects at mean values for age (74.3 years) and gender (59% female).

mencement of the intervention and the achievement of the physiologic or clinical target can be ascertained. For example, survival analysis can determine the median number of months that it takes for diabetics to achieve glycosylated hemoglobin (HbA1c) levels at or below 7%. These data can also be stratified by risk level to ascertain if differences are evident in these cohorts. One caveat is that for many DM programs such clinical data will only be available for patients enrolled in the "high risk" intervention. Either relevant clinical data should be collected across cohorts for this type of longitudinal analysis, or different clinical measures chosen that can be readily collected via common databases, etc, so as to reduce the reliance on patient self-reporting.

The importance of knowing the timeline to reach clinical targets is so that (1) a DM program can more accurately set performance target timeframes for achieving target levels, and (2) to assist in understanding the relationship between enrollment (timeline and penetration level) and achievement of the target clinical levels. The assumption here is that these clinical measures are proxies of positive long-term health outcomes. Additionally, independent variables may indicate a correlation with survival time and provide insight into possible avenues to pursue for DM program modification or development, etc.

The next analysis should clarify the timeline and probability of hospitalization from the point at which the patient achieves the clinical or physiologic target. Again, these data should be stratified by risk level in order to assess the effect of the intervention at the different levels of patient illness and severity. As mentioned above, this type of analysis may be limited if clinical data is only available for "high risk" patients.

For evaluating DM program effectiveness, the probability of hospitalization should be compared between the DM program intervention group and a control group. If the program consists of an intensive nursing intervention and a more passive option consisting of quarterly mailings and updates, a three level comparison can be conducted (high, low and control). If the program is indeed successful in improving clinical outcomes of care and low-

ering utilization, survival analysis should be considered an excellent tool for proving it.

IMPORTANT CONSIDERATIONS FOR USING SURVIVAL ANALYSIS

There are several issues that must be considered and/or addressed before choosing survival analysis as a DM program evaluation tool. Some of the considerations are statistical in nature, while others have to do with ensuring that the appropriate research methods are followed.

Statistical issues

Risk estimates determined by the model may be unreliable if too few events were realized. To reduce the likelihood of detecting spurious associations between explanatory variables and the outcome measure, it is recommended that the number of independent variables be no more than 10% of the number of events.²⁰ The study presented in this paper for demonstration purposes had 2920 events, well above the required number.

The method used for including predictors in the model also requires attention. If many variables are available to the analyst, a mechanical method available in most statistical software packages can weed out the insignificant variables. The forward stepwise technique enters each variable into the equation and assesses its contribution to the model by appraising its significance level and its effect on the other variables. The backwards stepwise routine begins with all the potential variables in the model, and eliminates variables one-by-one based on their significance level and effect on other variables. It is entirely possible, however, that these two techniques will elicit different model parameters.²¹ Therefore, the analyst should always rely on clinical judgment in determining which variables ultimately should be included or excluded in order to maximize the predictive power of the model. Similarly, independent variables should be checked for interaction. Interaction occurs if the effect of one explanatory variable on the outcome event depends on the level of another variable. A con-

sequence of this may be an increased likelihood of overfitting the model and drawing conclusions based on spurious results. Factorial analysis may be used to ferret out interactions if they exist.

As described in the previous section, every model should be evaluated to assure that assumptions of proportionality have not been violated. If these assumptions are not checked, estimates of risk by strata may be incorrect.¹⁷ One method that may correct for this problem, if variable changes over time appear to be impacting the model, is the use of time-dependent variables, as described previously.

Model validation is an integral element of the evaluation process. Comparing actual individual patient outcomes to those predicted by the model on a prospective basis usually does this. However, the more ubiquitous method employed by analysts is by way of the split-sample testing technique.²² This method involves splitting the original data into two sets, the first being used to develop the model, and the second set used to validate it. Again, comparing actual outcomes to what was predicted does this.

Methodological issues

Irrespective of the type of statistical modeling technique used to measure outcomes, there are common methodological principles that must be followed to ensure the validity of DM program outcomes.

In using Cox regression modeling for evaluating DM program effectiveness, one assumes that most, if not all, sources of bias and/or competing extraneous confounding factors that offer plausible alternative explanations for differences between cohorts have been adjusted for.^{23,24}

Selection bias poses a major threat to the validity of the program outcomes if evaluated by the Cox regression model. By stratifying cohorts into intervention and control groups for comparison, it is imperative that both groups have a comparable mix of subjects. In other words, all subjects should have the same risk profile and equal chance for inclusion in the intervention group. Since DM programs function with the “intent to treat” approach, members suitable for the program but excluded because

of benefit coverage or other factors extraneous to the program can be used as controls. The ideal approach for controlling for selection bias is by using random samples for the intervention group to be compared with a random sample for the “suitable but not enrolled” group. However, this may be difficult to do in practice.

Another equally and related bias is that of regression to the mean. Also referred to as statistical regression, this concept suggests that, without the effect of the intervention, members with high costs and utilization in the baseline year will tend to cost less and use fewer services in the following year (a move toward the mean). Conversely, members using few services in the baseline year will use more services and accrue higher costs in the subsequent year.⁴ Most algorithms used for assigning risk levels initially to members in DM programs are based on the level of utilization in the prior year. Therefore, a patient who had either a visit to the emergency department or a hospitalization would be classified as high risk while members identified strictly by their prescription drug data may be classified as low risk. If the outcome measure used in the Cox regression is the probability of hospitalization, stratified by risk level, it stands to reason that some members may regress toward the mean and reduce the power of the model to detect true program effects. That said, the Cox regression model just might be the best tool available for testing the predictive value of the DM program’s risk stratification algorithm. While these are the main threats to validity, several others need to be controlled for. The reader is referred to Linden et al,⁴ Campbell and Stanley,²³ and Cook and Campbell²⁴ for a comprehensive discussion on the topic.

CONCLUSION

This paper has described in some detail the application of survival analysis to the evaluation of DM program effectiveness. The advantage of this method over more widely used statistical models is the inclusion of censored data. This is an important factor, since DM programs typically have high turnover rates, due to health

plan enrollment/disenrollment as well as DM program attrition. Additionally, Cox regression allows for identification of the independent variables that may help explain the variability in the outcome variable—probability of achieving target clinical or physiological levels of control, hospitalization, or other endpoints of interest. For these methods to be successful it is crucial that threats to validity of the evaluation design be reduced, if not eliminated altogether. Additionally, it is imperative that the model be assessed for validity as well, to ensure that the analyst does not draw conclusions based on spurious results.

APPENDIX

The Cox proportional hazards regression model

A key difference between Cox and linear models is the ability to control for the effects of

censored cases. Moreover, the quantity of interest in Cox regression is the *hazard* function, which may be described as the risk that the event will occur for a subject in an observation period given the subject did not have the event before then. A high hazard function indicates a high event rate (low survival probability) and conversely, a low hazard indicates a low event rate (high survival probability). Additionally, Cox regression assumes that the hazards between any two subjects are proportional over time (hence, the name *proportional hazards regression*) with the proportion being a function of the explanatory variables. This assumption needs to be tested to ensure the model's reliability.

Table 4 presents the results of Cox regression analysis using data from Linden and Schweitzer⁸ with age and gender as variables. The overall fit of the Cox regression model is tested via the likelihood ratio test. Specifically,

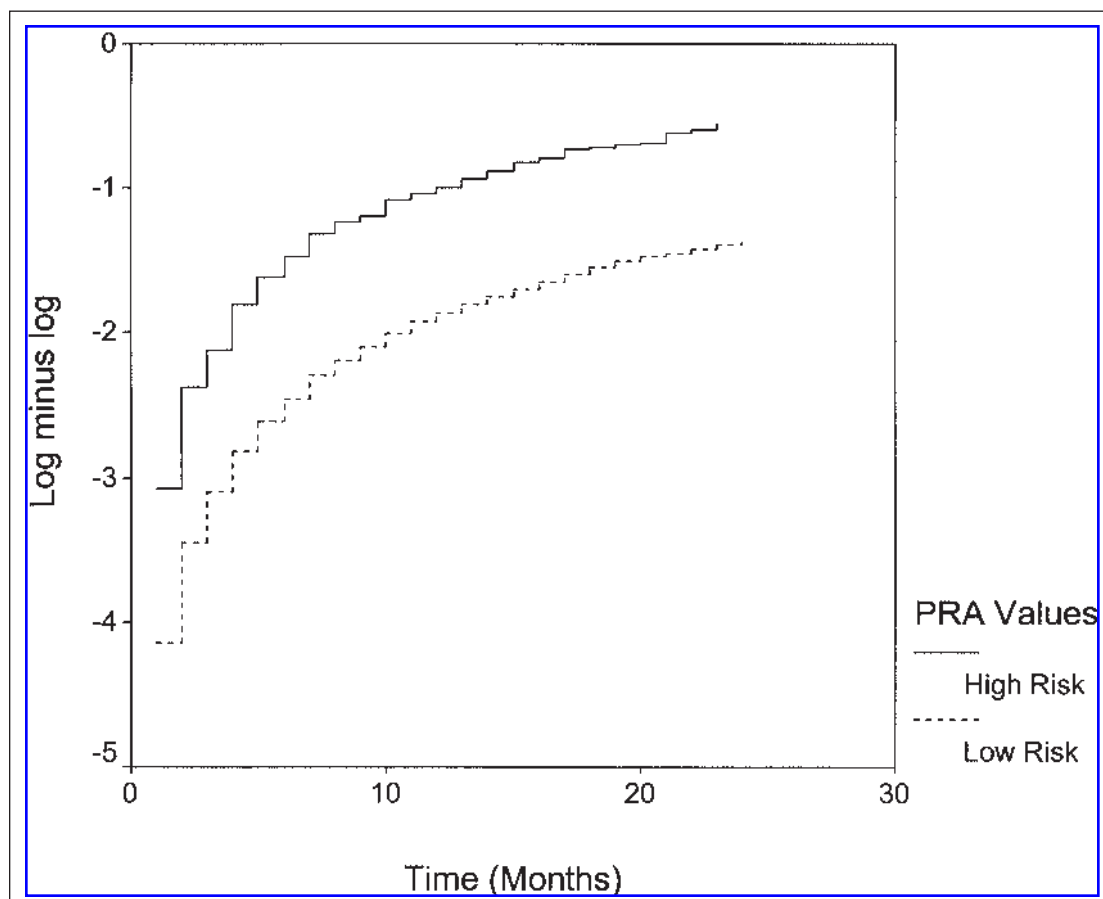


FIG. 4. Log minus log (LML) function at mean values for age (74.3 years) and gender (59% female). As illustrated, both curves appear to be proportional over time, indicating that the assumption of proportional hazards has not been violated.

the significance of the model is based on the ratio between the likelihood that the variables show no correlation with survival time (all regression coefficients = 0) and the likelihood of the regression coefficients estimated by the model. The smaller the ratio, the better the model actually fits the data. The high χ^2 statistic with two degrees of freedom (df) indicates the model is highly significant.

B is the estimated regression coefficient. As shown, for each unit increase in age, there is a 0.25 increase in the log hazard (or a higher probability of hospitalization). Similarly, being a female is associated with a 0.88 decrease in log hazard (or a lower likelihood of hospitalization) as compared to males. Both age and gender were determined to be significant variables in the model determined by the Wald statistic (<0.0001 and 0.023 for age and gender respectively). $\text{Exp}(B)$ estimates the percentage change in risk with each unit change in the given X variable. Each unit increase in a patient's age is expected to increase the risk of hospitalization by 2.6% (since $\text{Exp}(B) = 1.026$). Similarly, being female reduces the risk of hospitalization by nearly 8% (since $\text{Exp}(B) = 0.916$). 95% confidence intervals are also provided for the relative risk (hazard) predicted by each of these two regression coefficients.

In Cox regression models it is extremely important to assure that the assumption of proportionality is not violated.¹⁷⁻¹⁹ This is analogous to tests of nonlinearity in ordinary regression methods. The simple explanation of proportionality is that the ratio of hazards between any two patients over time should be constant and parallel. Figure 4 illustrates a log-minus-log function, holding the mean of both age and sex constant over time. As shown, the curves for low and high risk P_{RA} values appear proportional over the duration of the study period, indicating that assumption of proportionality was not violated.

The model described heretofore assumes that independent variables are constant over time. In some situations however, especially over a long observation period, variables can change. An example in disease management is if a patient begins the program in the low intervention group but is then persuaded to participate in the high level nursing intervention at some future point along the continuum. It is

possible to account for these occurrences by creating a *time-dependent* variable which is an interaction term formed by the product of X and a function of the survival time variable of the model.¹⁴

The mathematical model and explanation of parameters in the Cox regression is presented below. The simple Cox Proportional Hazard Model takes this form:

$$\lambda(t) = [\lambda_0(t)]\exp^{(BX)} \quad (1)$$

where $\lambda(t)$ is the hazard at observation time t , \exp is the base of the natural logarithm, B is the regression coefficient, X is the independent variable, $\lambda_0(t)$ is the baseline or underlying hazard function when X is set to zero. The baseline hazard is analogous to the "intercept" in standard regression models that can be considered a "reference" hazard.

As a final note, all analyses presented in this section were performed using SPSS for Windows, Release 11.0.1 (2001).

REFERENCES

1. Carroll J. Health plans demand proof that DM saves them money. *Managed Care* 2000;9:25-30.
2. Diamond F. DM's motivation factor can skew study results. *Managed Care* 1999;8:45-50.
3. Johnson A. Measuring DM's net effect is harder than you might think. *Managed Care* 2003;6:28-32.
4. Linden A, Adams J, Roberts N. An assessment of the total population approach for evaluating disease management program effectiveness. *Dis Manage* 2003;6:93-102.
5. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: an introduction to time series analysis. *Dis Manage* 2003;6:243-255.
6. Olivarius N, Beck-Neilsen H, Andreasen AH, et al. Randomized controlled trial of structured personal care of type 2 diabetes mellitus. *BMJ* 2001;323:970.
7. Murchie P, Campbell NC, Ritchie LD, et al. Secondary prevention clinics for coronary heart disease: four year follow up of a randomized controlled trial in primary care. *BMJ* 2003;326:84.
8. Linden A, Schweitzer SO. Applying survival analysis to health risk assessment data to predict time to first hospitalization. *AHSRHP Annu Mtg* 2001;18:26.
9. Altman DG, Bland MJ. Time to event (survival) data. *BMJ* 1998;317:468-469.
10. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. *J Am Statist Assoc* 1958; 53:457-481.
11. Boulton C, Dowd B, McCaffrey D, et al. Screening elders

- for risk of hospital admission. *J Am Geriatr Soc* 1993;41:811–817.
12. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br J Cancer* 1977;35:1–39.
 13. SPSS for Windows, Rel. 11.0.1. Chicago: SPSS Inc., 2001.
 14. Cox DR. Regression models and life tables (with discussion). *J R Statist Soc B* 1972;34:187–220.
 15. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993;118:201–210.
 16. Draper NR, Smith H. *Applied regression analysis*. New York: Wiley, 1981.
 17. Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. New York: Wiley, 1980.
 18. Andersen PK. Testing goodness-of-fit of Cox regression and life model. *Biometrics* 1982;38:67–77.
 19. Kay R. Goodness-of-fit methods for the proportional hazards regression model: a review. *Rev Epidemiol Sante Publ* 1984;32:185–198.
 20. Harrell FE Jr, Lee KL, Matchar DB, et al. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep* 1985;69:1071–1077.
 21. Lee KL, McNeer JF, Starmer CF, et al. Clinical judgment and statistics—lessons from a simulated randomized trial in coronary heart disease. *Circulation* 1980;61:508–515.
 22. Christensen E. Multivariate survival analysis using Cox's regression model. *Hepatology* 1987;7:1346–1358.
 23. Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1966.
 24. Cook TD, Campbell DT. *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally, 1979.

Address reprint requests to:
Ariel Linden, Dr.P.H., M.S.
Linden Consulting
6208 NE Chestnut St.
Hillsboro, OR 97124

E-mail: ariellinden@yahoo.com