

## Paper 201-28

## An Approach to Displaying Predicted Survival Data Based on the Level of a Continuous Covariate

Cristine Allmer, Mayo Clinic, Rochester, MN  
Daniel J. Sargent, Mayo Clinic, Rochester, MN

### Abstract

Kaplan-Meier estimates are widely used for plotting survival curves over time, for one or more strata. Graphical display of survival estimates based on a continuous covariate are more problematic. We propose a simple approach to plotting survival estimates at a specified point in time, based on the value of a continuous explanatory variable. This approach allows for the adjustment of survival estimates by any set of covariates. We illustrate this approach using Cox's proportional hazard model to calculate 5 year survival estimates for a North Central Cancer Treatment Group (NCCTG) retrospective colorectal cancer study. This paper presents a comparison between the standard methods and the new approach, and is intended for an audience with intermediate SAS skills. A basic knowledge of the PHREG and GPLOT procedures is assumed.

### Introduction

The purpose of survival analysis is to model the underlying distribution of the event-time variable, and to assess the dependence of the event-time variable on a set of independent variables.

For categorical variables, Kaplan-Meier estimates are the basic tool for computing and plotting unadjusted survival estimates. The log-rank test is used to compare survival distributions within specified strata. The SAS procedure used is PROC LIFETEST. For continuous variables, Cox (1972) regression models are used to test for association of continuous explanatory variable with survival. The model allows for stratification, and adjustment for other covariates. PROC PHREG is the standard SAS procedure used to test for statistically significant differences between strata, and to test for an association between the dependent and independent variable(s). However, a graphical display of survival estimates by the level of a continuous variable of interest is not standard. One approach is to smooth the Martingale residuals to look for functional form (Figure 1); additional notes can be consulted in Therneau et al (1990).

### Motivation

The motivations for this approach are several. Plots are easier to read and interpret than tables. Additionally, important information may be lost by grouping or categorizing a continuous explanatory variable, and cut-off points for a particular covariate may not be consistently used across studies. Finally, inferences from analysis based on cut-off points are always subject to suspicions of 'data dredging'.

Furthermore, as shown in our example (Figure 2A and 2B), different sets of cut-off points may lead to different results. For example, in the data set that we will examine in this paper, the choice of 2 different sets of cut-off points results in two very different log rank p-values and two different interpretations from the Kaplan-Meier curves. When using a first set of cut-off points (Figure 2a), we obtained a p-value of 0.007, and could conclude that the patients in the 11+ mm group have much greater local recurrence rate than the patients in the other two groups. However, when using the second set of cut-off points, there is no statistically significant difference between the three groups (p-value = 0.17), and the local recurrence rate curves look more similar for all three groups.

### New Method

We propose an alternative approach to visualizing the effect of a continuous covariate on estimated survival. Our approach allows for adjustment by any other covariates that may be of interest.

The method that we propose is as follows. First, using the PHREG procedure, calculate an estimated survivor function for each combination of the explanatory variable values present in the data set. Because all relevant explanatory variables can be included in this model, these estimates are adjusted for all of the covariates of interest. Next, select the estimated survival probability at any desired time point (for example 5 years) for each person in the study. Based on these estimates for each individual, a graphical illustration is obtained by plotting the survival probabilities at the desired time point for each value of the covariate in question (Figure 3). A smoothing method, such as a spline, is then used to provide an estimate of the impact of the covariate on the survival rate at the time point of interest over the whole range of covariate values (Figure 4).

### Example study design

We illustrate this new approach using the data for 680 patients (Stocchi et al, 2001) randomized to three Phase III North Central Cancer Treatment Group (NCCTG) colorectal clinical trials: NCCTG 794651 (190 patients), NCCTG 864751 (252 patients), and NCCTG 904751 (238 patients).

The goal of this re-analysis was to investigate the effect of surgery related variables on local recurrence of the cancer and overall survival. Rectal cancer variables analyzed were surgeon variability, surgery procedure performed, radial and distal resection margin, adherence of tumor, type of resection if adherent, intra-

operative inadvertent perforation, number of nodes retrieved, T stage, nodal involvement, and radial spread of tumor. Here we will focus on the radial margin of the surgery, which is the amount of tissue beyond the tumor removed by the surgeon that was confirmed by a pathologist to be tumor free.

In these trials, patients were randomized to one of five different treatment arms. In addition to surgery and postoperative radiation therapy, most patients received one of four different chemotherapy regimens. The treatment received has previously been shown to influence patient survival, therefore it was important to adjust for these variables in the analysis.

### Example SAS code

Here we present the code necessary to generate the graphics that we describe. First, run a Cox model using the PROC PHREG using the baseline option. This creates a new SAS data set that contains the survival estimates at event times for each pattern of explanatory variable values given the covariates data set. In our example, we model the probability of having a local recurrence (i.e. `rec_stat = 1`) adjusting for the treatment arm (A, B, C, D or E) and free radial margin (`f_margin`). We output the recurrence-free estimates (`loc_rec`) to a data set called `estimate`. We calculated the time to local recurrence (`rec_time`) as days from study registration.

```
PROC PHREG DATA = RECUR;
  MODEL REC_TIME * REC_ST(1) = ARM_A ARM_B
    ARM_C ARM_D F_MARGIN;
  BASELINE OUT = ESTIMATE SURVIVAL = LOC_REC
    COVARIATES = RECUR;
```

For this study, we are interested in the 5-year local recurrence rate. However, none of the 680 patients in our cohort had a local recurrence or been censored at exactly 5 years (i.e. 1,825 days) from study registration. The closest two time points were 1402 days and 2185 days. We created two data sets containing the estimates for these two time points, and used interpolation to calculate the 5-year local recurrence estimate.

```
DATA FIRST;
  SET ESTIMATE;
  IF PG_TIME = 1402;
  ORDER = _N_;
  RECURR1 = LOC_REC;
  PGTIME1 = PG_TIME;
  DROP LOC_REC PG_TIME;

DATA SECOND;
  SET ESTIMATE;
  IF PG_TIME = 2185;
  ORDER = _N_ - 1;
  RECURR2 = LOC_REC;
  PGTIME2 = PG_TIME;
  DROP LOC_REC PG_TIME;
```

Next we merged the two data sets created into one data set containing all the point estimates needed for the graphical display. Also, create a common variable for treatment arm out of the binary variables needed for the Cox model, to be used in the plot.

```
DATA FINAL;
  MERGE FIRST SECOND;
  BY ORDER;
  LOC_REC = ( 1 - ( RECURR1 - ( ( RECURR1 - RECURR2 )
    * .54 ) ) ) * 100;
  IF ARM_A = 1 THEN ARM = 1;
  ELSE IF ARM_B = 1 THEN ARM = 2;
  ELSE IF ARM_C = 1 THEN ARM = 3;
  ELSE IF ARM_D = 1 THEN ARM = 4;
  ELSE ARM = 5;
```

Finally, we plot the survival estimates by the level of the covariate of interest, which in our case is free radial margin. Following is the code used to create the plot of all the individual estimates (Figure 3).

```
GOPTIONS RESET = GLOBAL CBACK=WHITE
  HTEXT = 1.8 FTEXT = SWISSB;
  TITLE1 ' FIGURE 3: NEW APPROACH: ADJUSTED
  ESTIMATE PER PATIENT ';
  AXIS1 LABEL = ( ' FREE RADIAL MARGIN, MM ' )
  OFFSET = ( 3, 2 );
  AXIS2 LABEL = ( A = 90 ' LOCAL RECURRENCE
  RATE ' ) OFFSET = ( 0, 2 );
  SYMBOL I = SM60S V = NONE L = 1 C = BLACK
  W = 6 H = 1.1 ;
  PROC GPLOT DATA = FINAL;
  PLOT LOC_REC * F_MARGIN / VAXIS = AXIS2
  HAXIS = AXIS1 FRAME;
  NOTE HEIGHT = 1.8 F = SWISSB MOVE = ( 67, 32 )
  ' P = 0.002 ';
```

To plot the survival estimates for each patient (Figure 4), replace the symbol statement above by the following statement:

```
SYMBOL I = NONE V = DOT C = BLACK W = 6 ;
```

### Conclusions

The method discussed in this paper makes it possible for the reader to visualize the effect of a continuous covariate on a time-to-event outcome of interest, and allows for adjustment by any other set of covariates. For our study example, one can see that the patients that have a small free radial margin have a much higher 5-year local recurrence rate than the patients that have a large free radial margin. However, our method allows the viewer to detect greater differences between patients with 0 mm and 20 mm free radial margin, but little or no difference between patients with 30 mm and 50 mm free radial margin. Also this method allows us to see a higher variability in 5-year local recurrence rates for patients with 0-20 mm free radial margin than in patients with a 30-50 mm free radial margin. One limitation of this method is that it provides display of the association between a variable of interest and an outcome at a single time point. However, in many settings there are standard benchmarks, such as 5-year survival, that will be easily interpreted and familiar to the clinical audience. In addition, even though the estimate is presented at a single time point, the data for the entire survival curve has been used in the generation of the estimate.

This method is easily codable in SAS for programmers with a basic understanding of PROC PHREG and PROC GPLOT. This approach can be very useful to those working in the pharmaceutical or medical fields.

## References

Cox, D.R., (1972), "Regression models and life-tables (with discussion)," *Journal of the Royal Statistical Society, Series B*, **34**, 187-220.

Kaplan, E., and Meyer, P., (1958), "Nonparametric estimation from incomplete observation," *Journal of American Statistical Association*, **53**, 457-81.

Stocchi, L., Nelson, H., Sargent, J. D., O'Connell, J.M., Tepper, E.J., and Beart, R., (2001), "Impact of Surgical and Pathologic Variables in Rectal Cancer: A United States Community and Cooperative Group Report," *Journal of Medical Oncology*, **19**, 3895-3902.

Therneau, T.M., Grambsch, P.M., and Fleming, T.R., (1990), "Martingale-based residuals and survival models," *Biometrika*, **77**, 147-160.

## Contact Information

Your comments and questions are valued and encouraged. Please feel free to contact the authors at:

Cristine Allmer  
Mayo Medical Center  
Cancer Center Statistics  
200 First Street SW  
Rochester MN, 55902  
Phone: (507) 266-4624  
Fax: (507) 266-2477  
Email: [allmer@mayo.edu](mailto:allmer@mayo.edu)

Figure 1: Martingale Residual: Unadjusted Estimate

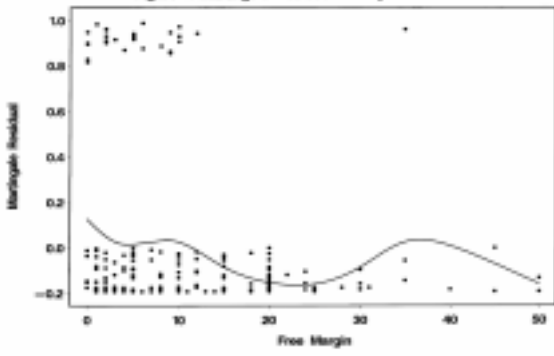


Figure 3: New Approach: Adjusted Estimate per Patient

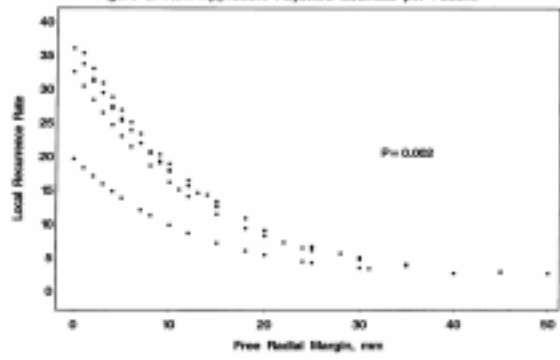


Figure 2A: Kaplan-Meier Estimate

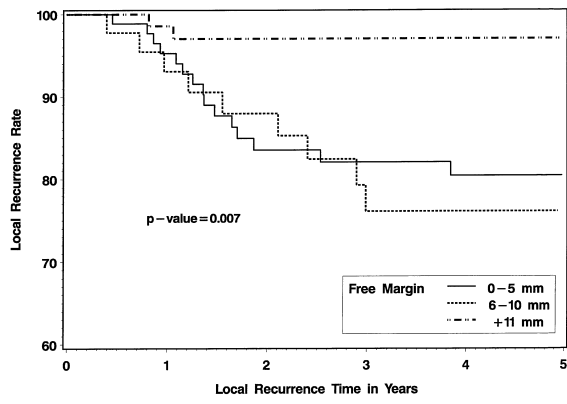


Figure 4: New Approach: Adjusted Estimate

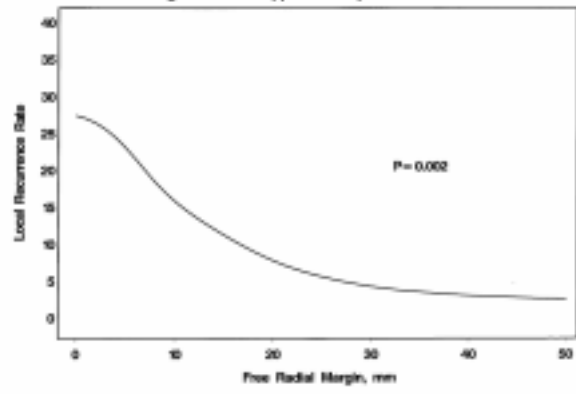


Figure 2B: Kaplan-Meier Estimate

