

Using SAS System to Generate Customized Kaplan-Meier Curves and its Application to Presentation in Clinical Research

Daojun Mo, Yiyong Fu

Introduction

One of the goals of clinical studies is to demonstrate the efficacy of a new treatment as compared with a standard treatment. For example, clinical investigators are concerned about whether a newly proposed surgical treatment could lead to better survival than a standard chemotherapy for advanced breast cancer patients.

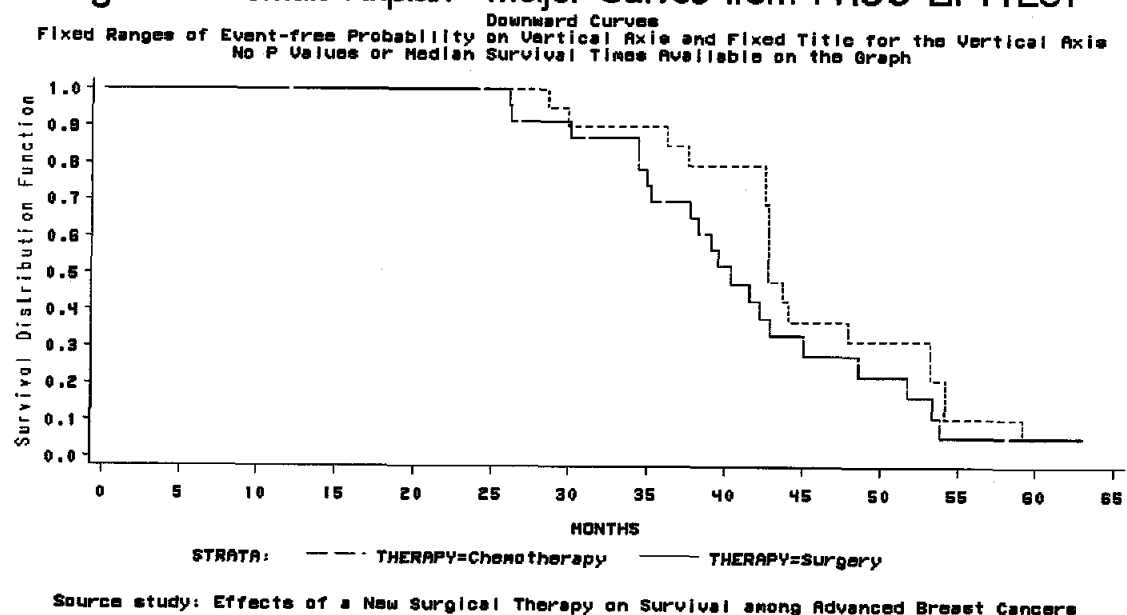
A Log-Rank test is introduced to test whether survival rates between the treatment groups are similar not only for the whole study period, but also for any arbitrary time point during the period. The statistical null hypothesis for the Log-Rank test is the equal survival time distribution between treatment groups during the whole study period. The alternative hypothesis is that survival rates differ among treatment groups at one or more time points during the study.

PROC LIFETEST in SAS generates sufficient statistical output to test the above null hypothesis and serves as an important tool for preliminary analysis of survival data.

PROC LIFETEST also outputs a *Kaplan-Meier* curve to show the graphic difference in the survival time distribution between treatment groups. The default horizontal-axis for the *Kaplan-Meier* curve from SAS output is the time since the starting point of observation (initial treatment, diagnosis date, index date of the case, date of enrollment, etc.). The default vertical-axis is the event-free probability (survival probability or survival distribution). The events of interest in clinical studies can be deaths, relapses, complications, and secondary events. The default vertical axis has fixed range of event-free probability from 0 to 1, inclusively and has fixed title – survival distribution function. The default *Kaplan-Meier* curve, if not level, goes down by its definition.

In the above advanced breast cancer example (Figure 1), the curve for the surgical group goes down faster than the chemotherapy group during the study period, thus it appears that the newly proposed surgical treatment led to a poorer survival than the standard chemotherapy.

Figure 1. Default Kaplan—Meier Curves from PROC LIFTEST



Kaplan-Meier curves are widely used in clinical studies. By retrieving four medical journals (Lancet, The New England Journal of Medicine, JAMA (Journal of American Medical Association), and CMAJ (Canadian Medical Association Journal)), which were published during February, March and Days 1-17 of April in 2002, *Kaplan-Meier* curves appeared 37 times. Since there are more than 500 medical journals globally, the estimated number of published *Kaplan-Meier* curves in these journals can be as many as 5-digit annually. Out of this huge number of the published *Kaplan-Meier* curves, many curves provide clues in treating and preventing important diseases such as AIDS, diabetes, cardiovascular diseases among human beings. Tables 1 and 2 listed some examples of the applications of *Kaplan-Meier* curves to the medical research studies.

The default *Kaplan-Meier* curves from *PROC LIFETEST* can serve as ideal presentations for the comparison of event-free probability (or survival distribution) between treatments in many published medical papers. For example, the default *Kaplan-Meier* curves are ideal in the comparison of death-free probability (or survival distribution) among advanced malignancies such as breast cancer, lung cancer, ovary cancer, and bone carcinoma, etc. In general, the default *Kaplan-Meier* curves are good enough for all disease entities, which have a wide range of event-free probability during the study period. In other words, *Kaplan-Meier* curves are appropriate for all disease entities, which are likely to experience an outcome of interest during the study period.

However, A *Kaplan-Meier* curve output from *PROC LIFETEST* may not be the most desirable graphic presentation in all the clinical studies.

The default vertical axis of *Kaplan-Meier* curve in *PROC LIFETEST* refers to the event-free probability. But the clinical investigators sometime are concerned about event probability. For example, one wants to know how it is likely for the study subjects to develop HIV infection, diabetes or breast cancer in the intervention treatments (Table 1). A default downward *Kaplan-Meier* curve from *PROC LIFETEST* may look awkward in this case. And an upward curve is more desirable. Four example studies were identified from recently published medical journals – where upward curves are actually plotted (Table 1). The horizontal axis remains the same, but the vertical axis is defined as event probability (% of study subjects with events, or cumulative incidence) in the examples. Event probability is simply defined as $1 - \text{event-free probability}$.

The default vertical axis of *Kaplan-Meier* curve in *PROC LIFETEST* has a fixed full range of event-free probability (0-1, inclusively). But clinical investigators sometime may wish to present a *Kaplan-Meier* curve with only a subset of the full range of event-free probability (e.g., a range of 0.80-1.00). A *Kaplan-Meier* curve focusing on a narrow range of event-free probability (or event probability) provides better visual effects than a graph covering a full range because we actually amplify the part of interest and ignore the irrelevant part on the same page. Since it is easy to differentiate the effects of treatments among subgroups on the amplified graph, the graph is desirable for those who want to address that their treatments (or drugs) are clinically superior to their competitors even though the expected event probabilities are low (or the event-free probabilities are high) for all treatment groups. Studies with narrow ranges of event-free probability (or event

probability) are frequently presented in the medical journals. Table 2 contains four studies where one can identify downward *Kaplan-Meier* curves with narrow ranges of even-free probability (0.80-1.00). Table 1 also includes four studies where one can identify upward curves with narrow ranges of event probability (0.00-0.35). Five of eight studies in Tables 1 and 2 plotted the observed narrow ranges of event-free probability (or event probability).

Table 1. Examples of Medical Studies, which Presented Upward Kaplan-Meier Curves, with Narrow Ranges of Event Probability

#	Study	Primary outcome measure	Study period (Horizontal axis)	Range of observed event probability** (Vertical axis)	Journal source and time of publication
1	Efficacy of three short-course regimens of zidovudine* and lamivudine* in preventing early and later transmission of HIV-1 from mother to child in Tanzania, South Africa, and Uganda (Petra study): a randomized, double-blind, placebo-controlled trial	Compare the HIV incidences among children whose mothers in different regimens of antiretroviral drugs (zidovudine and lamivudine)	72 weeks since birth	0-0.02	The Lancet, April 2002
2	Efficacy of Tamoxifen in preventing breast cancer among hysterectomised women	Compare the incidences of breast cancers between the Tamoxifen group and placebo group	108 months since randomization	0-0.21	The Lancet, March 2002
3	Risk factors for elevated HIV incidence rates among female injection drug users in Vancouver	Compare the HIV incidences among males and females	48 months since enrolment	0-0.17	CMAJ (Canadian Medical Association) Journal, April, 2002
4	Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin*	Compare the incidences of diabetes among high risk population who adjusted lifestyle, and those used metformin and those on placebo	4 years since randomization	0-0.35	The New England Journal of Medicine, February 2002

* Zidovudine and Lamivudine

* Tamoxifen

* Metformin

** Event probability

Drugs for antiretroviral interventions

A candidate drug for prevention of breast cancer

Drug for increase of pancreas cells' sensitivity to insulin. Type 2

Diabetes patients lack sensitivity to insulin

equivalent to 1 - event-free probability

Table 2. Examples of Medical Studies, which Presented Downward Kaplan-Meier Curves, with Narrow Ranges of Event-free Probability

#	Study	Primary outcome measure	Study period (Horizontal axis)	Range of observed event-free probability** (Vertical axis)	Journal source and time of publication
5	Effect of nicorandil* on coronary events in patients with stable angina: the Impact of Nicorandil in Angina (IONA) randomized trial	Compare the incidences of major coronary events between nicorandil group and placebo group	3 years since randomization	0.80-1.00	The New England Journal of Medicine, February 2002
6	Thrombolytic therapy vs primary percutaneous coronary intervention for myocardial infarction in patients presenting to hospital without on-site cardiac surgery	Compare the composite incidences of death, recurrent MI and stroke between thrombolytic therapy group and primary percutaneous coronary intervention group	6 weeks since randomization	0.82-1.00	JAMA (Journal of American Medical Association), April 2002
7	Use of a straphylococcus aureus conjugate vaccine in patients receiving hemodialysis	Compare the incidences of straphylococcus aureus bacteremia between immunized group and placebo group	110 weeks since randomization	0.94-1.00	The New England Journal of Medicine, February 2002
8	Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia	Compare the death rates between patients with major cytogenetic response* and patients with no major cytogenetic responses	22 months since start of treatment	0.87-1.00	The New England Journal of Medicine, February 2002

* Nicorandil

* Cytogenetic response

** Event-free probability

A drug is thought to have cardioprotective properties

Chemical responses at gene level in cells is thought to be related the deaths in the clinical study

equivalent to 1 – event probability

A default *Kaplan-Meier* curve from *PROC LIFETEST* does not contain P values from statistical test or median times of event (or of free event). However, clinical investigators sometime want to see if the difference in event-probability (or event-free probability) distribution between treatments is statistically significant. The option of *annotate* in the procedure *PROC LIFETEST* can pull the P values in the graph, but it is not a convenient way in the first place. Six out of eight papers in the Tables 1-2 presented P values on the graphs.

PROC LIFETEST installed on the platform such as MVS may not directly output a *Kaplan-Meier* curve to a location on the platform for later use on PC unless a graphic environment is set up in advance (in terms of correspondence with SAS Institute Inc.). An alternative way of creating a *Kaplan-Meier* curve on MVS seems necessary.

Considering the wide application of *Kaplan-Meier* curves and varied situations where customized features of *Kaplan-Meier* curves are needed in clinical studies, we developed a SAS MACRO to draw the customized *Kaplan-Meier* curves.

Programming summary and data sources

Briefly, the SAS MACRO was compiled using SAS Macro language, DATA STEP, *PROC LIFETEST* and *PROC GPLOT*. For details, see the Appendix.

To demonstrate the customized *Kaplan-Meier curves*, data were simulated to compare the effects of drugs A, B and Placebo on prevention of breast cancer in postmenopausal women. Although breast cancer is the most commonly diagnosed female malignant tumor, since the occurrences of cancer are rare events in general, the event probability (incidence rate) can be just approximately as low as 0.12 among postmenopausal women during an assumptive ten-year follow-up period. The simulated data for this population of women, where new breast cancers were diagnosed, were used to plot the *Kaplan-Meier* curves focusing on narrow ranges of event probability (or event-free probability). The generated graphs from these data are similar to the curves from the studies in the tables 1 and 2, where the customized features of graphs are not available from SAS graphic default output in *PROC LIFETEST*. However, these features may be expected to appear on the graphs by the researchers.

The data were also simulated for the example at the beginning of this paper – comparing the effects of the radical surgical treatment and conservative chemotherapy on survival in patients with advanced breast cancer. In contrast with the newly diagnosed breast cancers, the patients with advanced cancer had high event probability of death (or low event-free probability). The data for these patients were used to plot the *Kaplan-Meier* curves covering the full range of event probability (or event-free probability) (0-1, inclusively). The generated graph using these data was from SAS default output in *PROC LIFETEST*.

Results

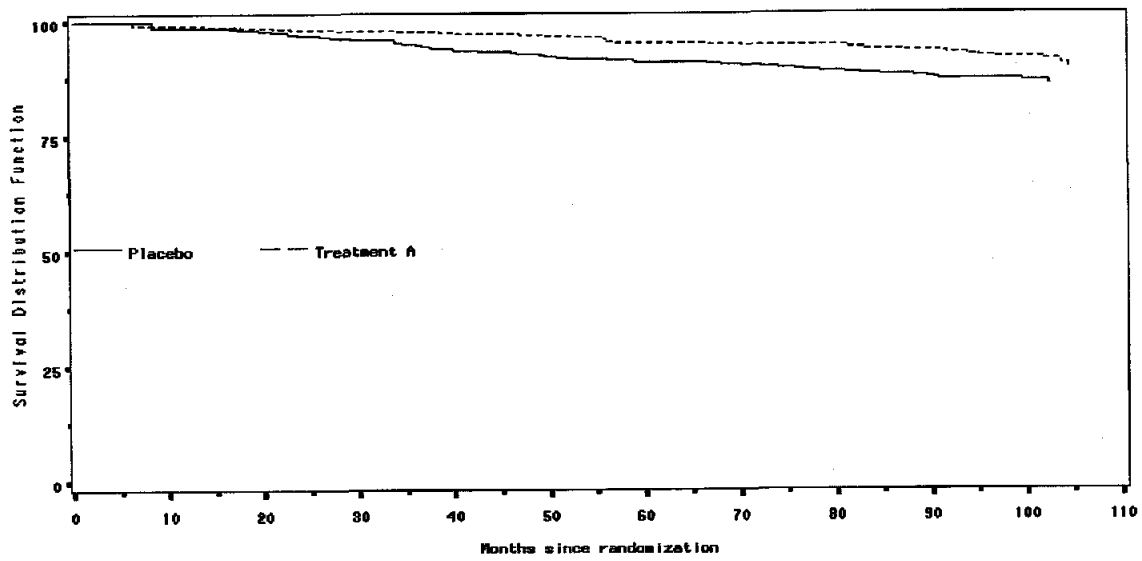
For the default *Kaplan-Meier* curve, see Figure 1.

For the customized graphs, see Figures 2-8. The customized graphs cover the features, which are not available from the default graphic output from PROC LIFETEST.

As compared with Figure 2, Figure 3 has upward curves. Figures 4 and 5 focus on the narrow ranges of event probability (or event-free probability). In addition to the features in Figures 4 and 5, Figure 6 and 7 include P values on the graph. Figure 8 extends the comparison of the event probability distribution to 3 treatment groups. Figure 8 also contains median times of event.

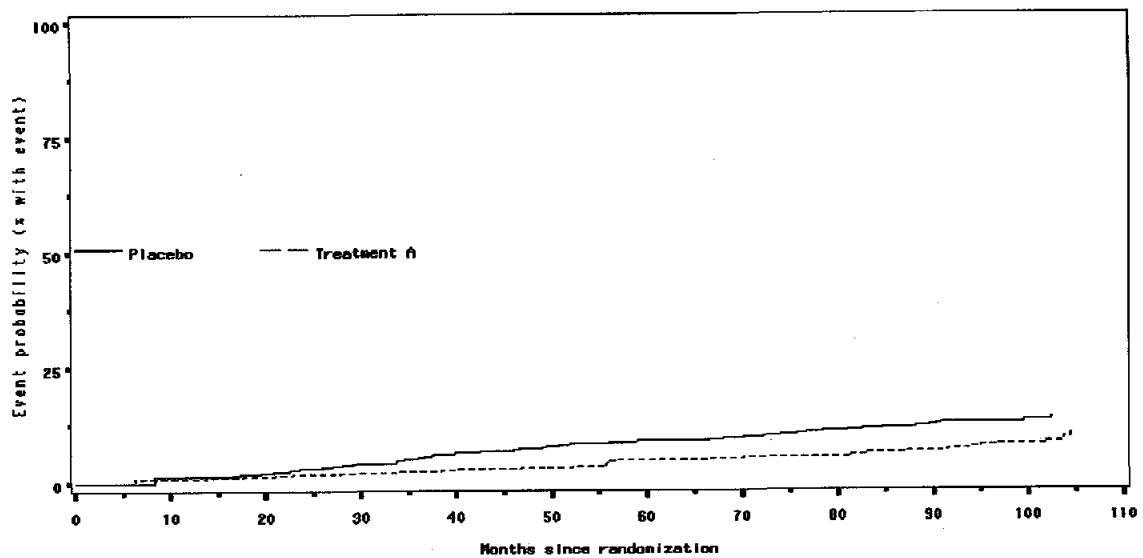
Figure 9 has the same pattern as Figure 1 because both figures used the same data. However, Figure 9 was generated by MS Excel, which is more commonly used for presentation. The SAS MACRO introduced in this paper exported the data and stored the data on PC for later plotting in Excel.

Figure 2. Kaplan-Meier Curves with Narrow Ranges of Event-free Probability
Downward Curves
A Graph Similar to the Default Graphic Output from Proc Lifetest



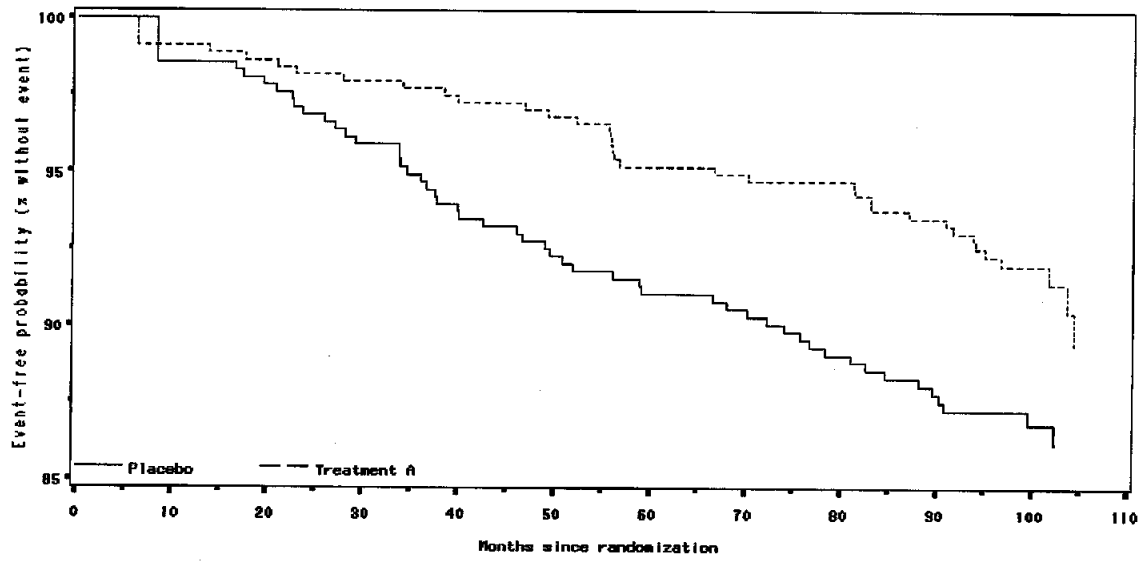
Source Study: Effects of Drugs on the Prevention of Breast Cancer among Postmenopausal Women

Figure 3. Kaplan-Meier Curves with Narrow Ranges of Event Probability
Upward Curves
Some Researchers Would Like to See Upward Curves



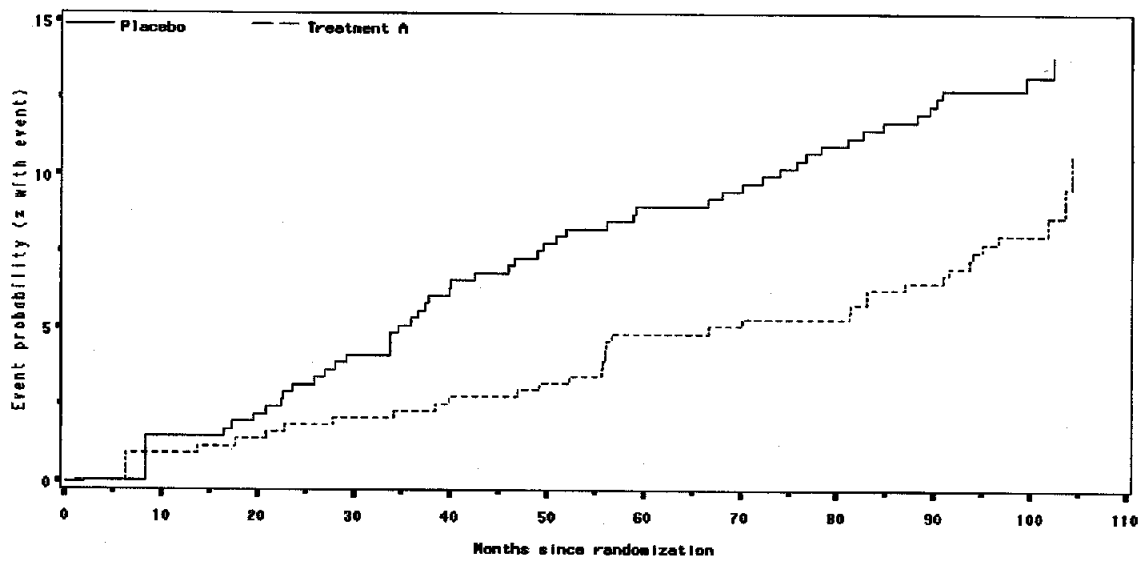
Source Study: Effects of Drugs on the Prevention of Breast Cancer among Postmenopausal Women

Figure 4. Kaplan-Meier Curves with Narrow Ranges of Event-free Probability
Downward Curves
Focus on the Narrow Ranges of Probability



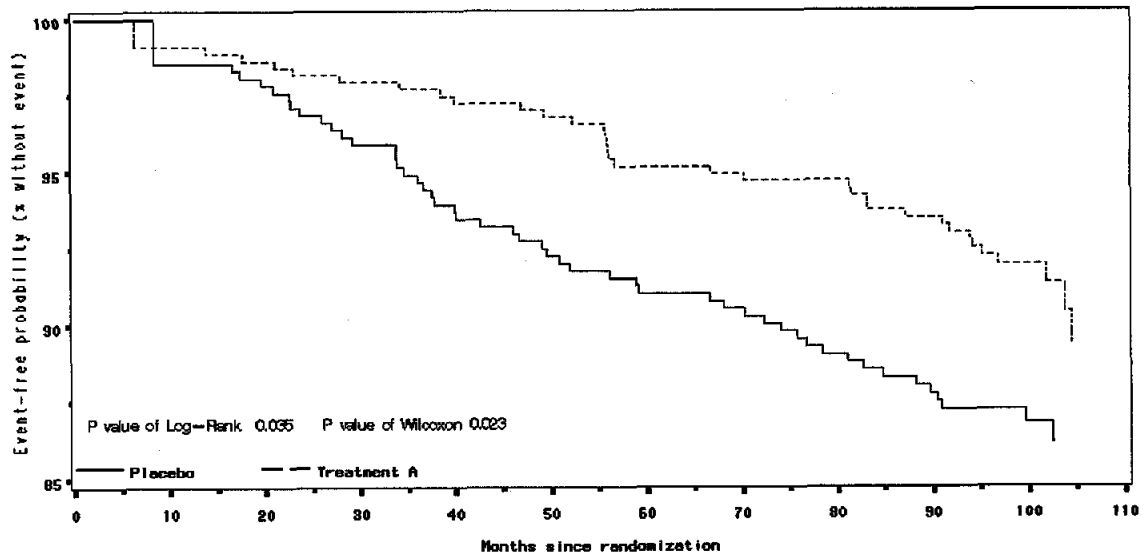
Source Study: Effects of Drugs on the Prevention of Breast Cancer among Postmenopausal Women

Figure 5. Kaplan-Meier Curves with Narrow Ranges of Event Probabilities
Upward Curves
Focus on the Narrow Ranges of Probability



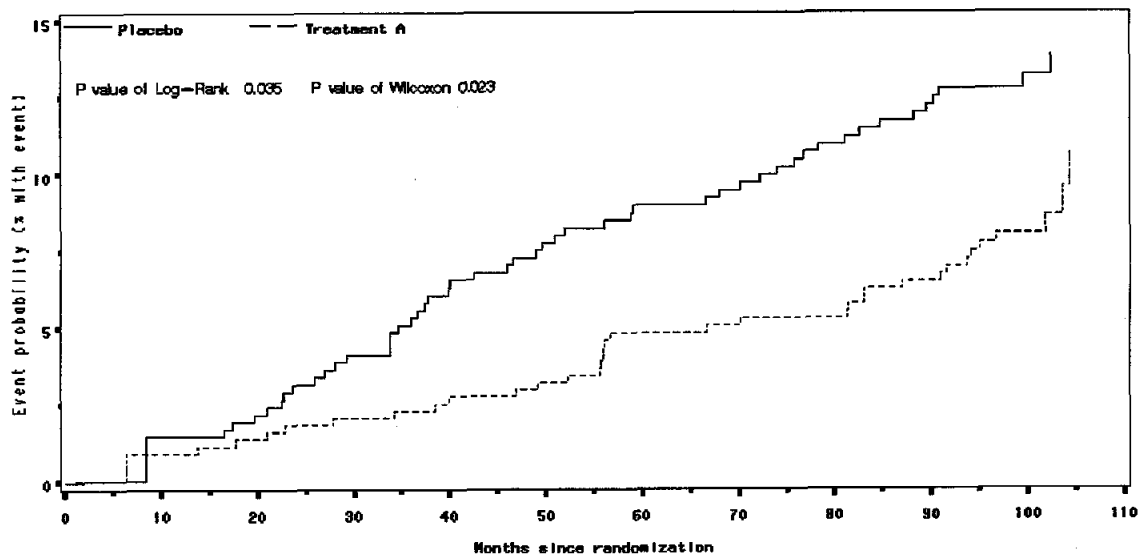
Source Study: Effects of Drugs on the Prevention of Breast Cancer among Postmenopausal Women

Figure 6. Kaplan-Meier Curve with Narrow Ranges of Event-free Probabilities
A Downward Curve
With Inserted P Values for Log-rank and Wilcoxon Statistics



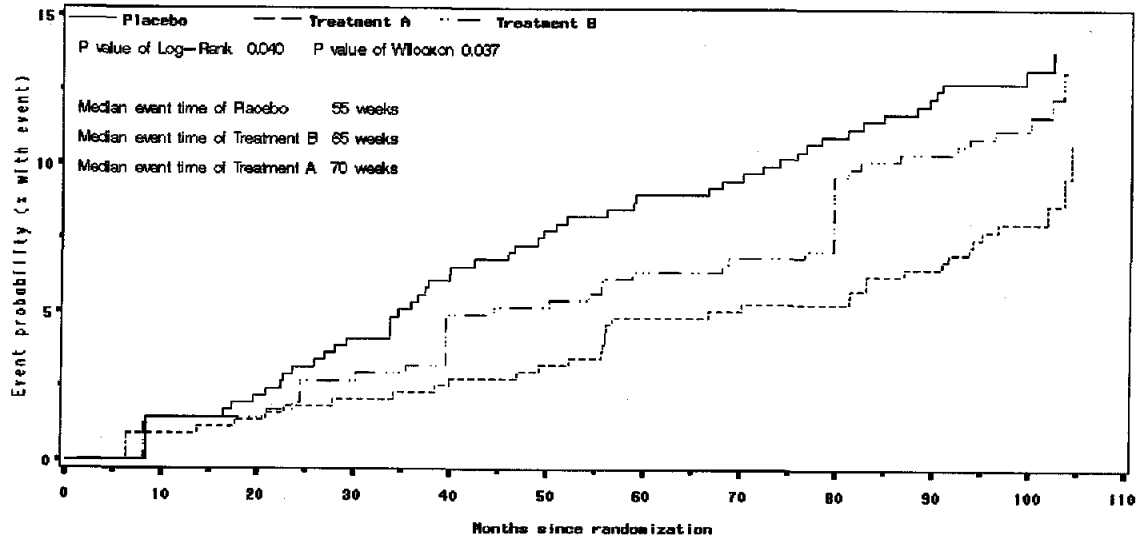
Source Study: Effects of Drugs on the Prevention of Breast Cancer among Postmenopausal Women

Figure 7. Kaplan-Meier Curves with Narrow Ranges of Event Probabilities
Upward Curves
With Inserted P Values for Log-rank and Wilcoxon Statistics



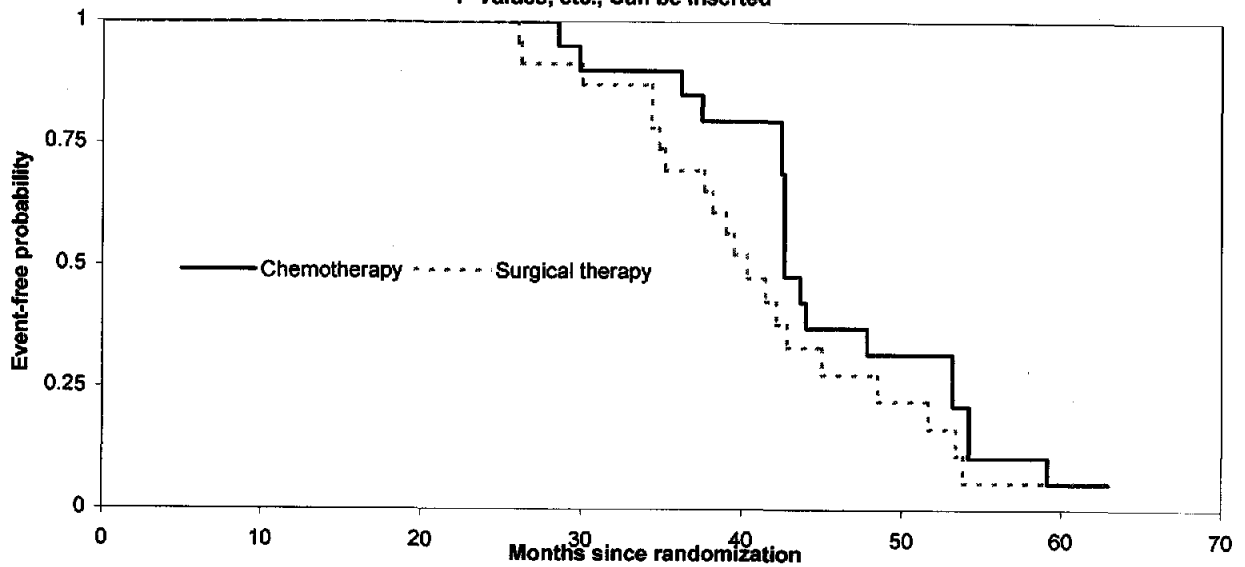
Source Study: Effects of Drugs on the Prevention of Breast Cancer among Postmenopausal Women

Figure 8. Kaplan-Meier Curves with Narrow Ranges of Event Probabilities
Upward Curves
With Inserted P Values and Median Event Times (Event-free times)
Extended to the Comparison between Multiple Groups (>2)



Source Study: Effects of Drugs on the Prevention of Breast Cancer among Postmenopausal Women

Figure 9. Kaplan-Meier Curves Generated from MS Excel
The Pattern of the Curves are the Same as Those in Figure 1
The Direction of the Curves Can be Either Upward or Downward
The Scale of the Vertical Axis Can be Adjusted
P Values, etc., Can be Inserted



Source study: Effects of a New Surgical Therapy on Survival among Advanced Breast Cancers

Brief discussion

Out of the thirty-seven *Kaplan-Meier* curves extracted from the literature, nine curves (25%) were upward (as opposed to downward by default in SAS), eighteen curves (50%) had narrow ranges of event-free probabilities (or event probabilities), and ten (27%) had their narrow ranges of event-free probabilities (or event probabilities) actually plotted. Thus, a substantial proportion of the published *Kaplan-Meier* curves used customized graphs. Since the number of published *Kaplan-Meier* curves can be huge each year, the number of published customized *Kaplan-Meier* curves can also be large.

Using the SAS macro, one can generate the *Kaplan-Meier* curves as good as the default graphs from *PROC LIFETEST* (Figures 1-8). In addition, it can generate upward *Kaplan-Meier* curves (Figures 2, 5, 7, 8), focus a subset of the full range of event probability (or event-free probability)(Figure 4-8), as well as present P values and median event time (or median event-free time) (Figures 6-8). Moreover, the SAS codes can export data to PC for plotting in more common software such as MS Excel. As well, the codes can be easily adapted to the platforms such as MVS where additional graphical environment is not available and a default *Kaplan-Meier* curves is impossible. Thus, the customized *Kaplan-Meier* curves by the MACRO in the paper can have wider application than the default *Kaplan-Meier* curves from *PROC LIFETEST*.

In addition to the *Kaplan-Meier* curves actually published in peer-reviewed journals, it is expected that this SAS MACRO can also be applied to new drug submission to FDA by pharmaceutical applicants, regulatory correspondence between FDA and pharmaceutical industries, production presentation by pharmaceutical companies, drug development studies, as well as health status reports by government in public health field.

Summary and conclusion

Kaplan-Meier curves are widely used in important clinical studies.

Customized *Kaplan-Meier* curve overcomes the limitations in the *POOC LIFETEST*, thus leading to better visual effects in some studies.

The SAS codes for customized *Kaplan-Meier* curves are simple, thus it can facilitate using SAS system to generate *Kaplan-Meier* curves in clinical studies.

Reference

1. Glenn Walker The common statistical methods for clinical research. Publication of SAS Institute. P244-267
2. Paul Allison. A practical guide to survival analysis. Publication of SAS Institute. P1-60
3. The Petra study team. Efficacy of three short-course regimens of zidovudine* and lamivudine* in preventing early and later transmission of HIV-1 from mother to child in Tanzania, South Africa, and Uganda (Petra study): a randomized, double-blinded, placebo-controlled trial. *The Lancet*, 359, 1178-86, April 2002
4. U Veronesi, P Maisonneuve, V Sachini, etc. Tamoxifen for breast Cancer among hysterectomised women. *The Lancet*, 359, 1122-24, March, 2002
5. P Spittal, K Craib, E Wood. Risk factors for elevated HIV incidence rates among female injection drug users in Vancouver. *Canadian Medical Association Journal*, 166(7) 894-99, April, 2002
6. Diabetes prevention program research group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The new England Journal of Medicine*, 346(6) 393-403, February, 2002
7. IONA study group. Effect of nicorandil* on coronary events in patients with stable angina: the Impact of Nicorandil in Angina (IONA) randomized trial. *The lancet*, 359 1269-74, April, 2002
8. T Aversano, L Aversano, E Passamani etc. Thrombolytic therapy vs primary percutaneous coronary intervention for myocardial infarction in patients presenting to hospital without on-site cardiac surgery. *JAMA* 287 (15) 1943-51, April 2002
9. H Shinefield, S Black, A Fattom, etc. Use of a staphylococcus aureus conjugate vaccine in patients receiving hemodialysis. *The new England Journal of Medicine*, 346(7) 491-96, February, 2002
10. H Kantarjian, C Sawyers, A Hochhaus, etc. Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. *The new England Journal of Medicine*, 346(9) 645-52, February, 2002

Appendix

SAS codes for customized Kaplan-Meier curves will be released at presentation.