PharmaSUG 2026 - Paper AS-422

Linear versus Log-Log Confidence Intervals for Kaplan-Meier Survival Estimates Statistical Rationale and Practical Implementation Using SAS® 9.4

Prayag K. Shah, Revolution Medicines

Abstract

Kaplan-Meier (KM) methods are routinely used to analyze oncology time-to-event endpoints such as overall survival (OS), progression-free survival (PFS), and duration of response (DoR). Although the Kaplan-Meier point estimate is invariant to the confidence interval (CI) method, the uncertainty around that estimate depends directly on how the interval is constructed. Linear confidence intervals, derived on the raw survival scale using Greenwood's variance, can extend beyond the admissible probability range and can be misleading when survival probabilities are near 0 or 1 or when the number of observed events is small. Log-log confidence intervals instead apply a transformation that stabilizes variance, improves the normal approximation, and yields asymmetric confidence limits that remain within valid probability bounds. This paper presents the statistical rationale for linear and log-log intervals, a compact manual numerical illustration, and an empirical comparison using SAS 9.4 PROC LIFETEST with ADaM ADTTE datasets for OS and PFS. The empirical results show that linear intervals frequently hit the upper boundary of 1.0, whereas log-log intervals remain bounded and more informative. These findings support the routine use of log-log confidence intervals for production survival reporting.

1. Introduction

Kaplan–Meier survival analysis is a standard tool for evaluating oncology endpoints such as OS, PFS, and DoR. In ADaM, these endpoints are commonly represented using the ADTTE structure, where analysis time and censoring status are stored in standardized variables and processed in production workflows through procedures such as PROC LIFETEST. The Kaplan-Meier estimate itself is straightforward, the greater practical challenge is how to represent uncertainty around that estimate in a statistically appropriate and interpretable way.

The Kaplan–Meier estimator of the survival function is

$$\hat{S}(t) = \prod_{i:t_i \le t} \left(1 - \frac{d_i}{n_i}\right),$$

where $d_i$is the number of events and $n_i$is the number at risk at time t_i. Its variance is typically estimated using Greenwood's formula,

$$\widehat{\mathrm{Var}}\{\hat{S}(t)\} = \hat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)}.$$

This variance is not constant across the survival curve because it depends on both the number of events and the level of the survival estimate itself. In simple terms, the uncertainty of $\hat{S}(t)$does not behave the same way when survival is close to 1, around the middle of the curve, or close to 0. That matters because the most familiar linear confidence interval uses a normal approximation directly on the probability scale:

$$\hat{S}(t) \pm z_{\alpha/2}\, SE\{\hat{S}(t)\}.$$

This interval is symmetric around $\hat{S}(t)$, but survival probabilities are bounded between 0 and 1. As a result, linear intervals can exceed those bounds, especially with heavy censoring, limited events, or survival estimates near the boundaries. SAS documentation explicitly notes that transformed confidence intervals often perform better than ordinary linear intervals and provides CONFTYPE= options to compute these transformed limits.

One of the most useful transformations is the log-log transformation,

$$\theta = \log\{-\log(\hat{S}(t))\}.$$

A confidence interval is constructed on the transformed scale and then back-transformed to the survival scale. The practical advantage is twofold. First, the interval remains within the admissible probability range. Second, the transformation stabilizes variance and improves the normal approximation, especially when the survival estimate is near 0 or 1 or when the number of events is small. Borgan and Liestøl showed that transformed confidence intervals, including the log-minus-log form, can substantially improve small-sample performance over untransformed intervals.

The purpose of this paper is therefore not to compare different Kaplan-Meier estimates, since those are identical, but to show how the choice of confidence interval method changes the uncertainty statement attached to the same survival estimate. The paper combines a manual numerical illustration with empirical SAS 9.4 results from ADTTE OS and PFS data and concludes with practical implications for survival analyses.

2. Methods

2.1 Statistical framework

In this paper, uncertainty is represented through pointwise confidence intervals around the Kaplan-Meier survival estimate. A confidence interval is not only a range; it is the formal expression of how variable the estimate would be across repeated samples. If that uncertainty is represented poorly, the interval can appear more precise than the data justify.

For the linear method, the confidence interval is

$$\hat{S}(t) \pm z_{\alpha/2}\, SE\{\hat{S}(t)\}.$$

This method assumes that uncertainty is approximately symmetric on the raw probability scale. That assumption is often weak in survival settings, because the Kaplan-Meier estimate is bounded and because its variance changes with the level of survival and the number of failures. When survival is close to 1 early in follow-up, the upper part of a linear interval can be artificially compressed by truncation at 1.0, creating an impression of high precision that is not fully supported by the data.

For the log-log method, the transformation

$$\theta = \log\{-\log(\hat{S}(t))\}$$

is applied first. Using the delta method, the standard error is computed on the transformed scale, and the interval is then back-transformed:

$$[\hat{S}(t)]^{\exp(z_{\alpha/2}\hat{\tau}(t))} \le S(t) \le [\hat{S}(t)]^{\exp(-z_{\alpha/2}\hat{\tau}(t))}$$

Variance stabilization here means that the transformed estimator has a variance that behaves more uniformly across the range of survival probabilities. In practice, this gives a better normal approximation and therefore more reliable confidence intervals, especially when event counts are low or survival is close to the probability boundaries.

2.2 Manual numerical illustration

To make the distinction concrete, consider the following values from the empirical analysis (value taken from ADTTE.PARAM= OS for Xanomeline low dose):

$$\hat{S}(t) = 0.9824, SE\{\hat{S}(t)\} = 0.0174.$$

For a 95% confidence interval, $z_{\alpha/2} = 1.96$.

For the linear method,

$$0.9824 \pm 1.96 \times 0.0174 = (0.9483,\ 1.0165)$$

For the log-log method, first compute

$$\theta = \log\{-\log(0.9824)\} = -4.0342.$$

The transformed standard error is

$$SE^* = \frac{0.0174}{0.9824 \times |\log(0.9824)|} = 1.0006.$$

The confidence interval on the transformed scale is

$$-4.0342 \pm 1.96 \times 1.0006 = (-5.9954,\ -2.0729).$$

Back-transformation gives

$$S_L(t) = \exp\{-\exp(-2.0729)\} = 0.8817$$
$$S_U(t) = \exp\{-\exp(-5.9954)\} = 0.9975$$

Thus, the log-log interval is $(0.8817, 0.9975)$

The two intervals are based on the same Kaplan-Meier estimate, but they differ materially in interpretation. The linear interval reaches beyond the probability boundary and must be truncated, whereas the log-log interval remains bounded and asymmetric.

2.3 Empirical SAS implementation

The empirical example used SAS 9.4 PROC LIFETEST on ADaM ADTTE-style datasets for ADTTE.PARAM= OS, and for PFS. The dummy ADTTE dataset was first obtained in R using the pharmaverseadam package, and the xportr package was then used to write the dataset to a transport (.xpt) file suitable for CDISC-compliant SDTM/ADaM workflows. After the XPT file was generated, the survival analyses were performed in SAS 9.4. The working variables were:

- AVAL = time-to-event in days
- CNSR = censoring indicator, where 0 = event and 1 = censored
- ARMCD = treatment arm

The treatment groups were Placebo, Xanomeline High Dose, and Xanomeline Low Dose. Confidence intervals were generated using both CONFTYPE=LINEAR and CONFTYPE=LOGLOG. The implementation was automated through a macro that generated the survival plots, OUTSURV= data sets, and comparison tables for each endpoint. The sample SAS code shows the PROC LIFETEST method used. SAS documentation confirms that PROC LIFETEST computes pointwise confidence limits according to the requested transformation and that transformed intervals are available through the CONFTYPE= option.
(Note: In SAS versions after 9.1, log-log is the default method for constructing survival confidence intervals, reflecting its improved statistical properties)

3. Results

3.1 Overall survival

The OS dataset contained 254 observations, with only 3 observed failures and 251 censored observations, corresponding to an overall censoring rate of 98.82%. By treatment arm, Placebo had 2 events among 86 subjects, Xanomeline High Dose had 0 events among 84 subjects, and Xanomeline Low Dose had 1 event among 84 subjects. These data create the classic high-survival, low-event setting in which the choice of confidence interval method becomes especially important.

At Day 12 in the Placebo arm, the Kaplan-Meier estimate was 0.9882. The linear confidence interval was (0.9653, 1.0), whereas the log-log interval was (0.9194, 0.9983). At Day 175 in the Placebo arm, the survival estimate was 0.9714, with a linear interval of (0.9318, 1.0) and a log-log interval of (0.8884, 0.9929). In the Xanomeline Low Dose arm at Day 61, the survival estimate was 0.9824, the linear interval was (0.9483,1.0), and the log-log interval was (0.8819, 0.9975).

Across all OS examples, the Kaplan-Meier estimate is identical across methods, but the linear upper limit repeatedly reaches 1.0, whereas the log-log upper limit remains below 1. The lower log-log limit is also meaningfully lower, showing greater uncertainty than the linear interval suggests.

Table 1: Linear Vs LogLog Confidence Limits for Overall Survival

Arm	Analysis Day	Survival Estimate	Linear LCL	Linear UCL	LogLog LCL	LogLog UCL
Placebo	12	0.9882	0.9653	1.0	0.9194	0.9983
	175	0.9714	0.9318	1.0	0.8884	0.9929
Xanomeline Low	61	0.9824	0.9483	1.0	0.8819	0.9975

3.2 Progression-free survival

The PFS dataset also contained 254 observations, with 6 observed failures and 248 censored observations, corresponding to an overall censoring rate of 97.64%. By treatment arm, Placebo had 3 events, Xanomeline High Dose had 2 events, and Xanomeline Low Dose had 1 event.

At Day 12 in the Placebo arm, the survival estimate was 0.8. The linear interval was (0.4493,1.0), whereas the log–log interval was (0.2038, 0.9691). At Day 64 in the Placebo arm, the survival estimate was 0.6, with a linear interval of (0.1705, 1.0) and a log-log interval of (0.1257, 0.8817). At Day 43 in the Xanomeline High Dose arm, the survival estimate was 0.6666, with a linear interval of (0.1332, 1.0) and a log-log interval of (0.0540, 0.9452). The most extreme example occurred in the Xanomeline Low Dose arm at Day 61, where the survival estimate was 0.5. The linear interval expanded to (0.0, 1.0), while the log-log interval remained (0.0059, 0.9104).

Table 2: Linear Vs LogLog Confidence Limits for Progression-Free Survival

Arm	Analysis Day	Survival Estimate	Linear LCL	Linear UCL	LogLog LCL	LogLog UCL
Placebo	12	0.8	0.4493	1.0	0.2038	0.9691
	64	0.6	0.1705	1.0	0.1257	0.8817
Xanomeline High	43	0.6666	0.1332	1.0	0.0540	0.9452
Xanomeline Low	61	0.5	0.0	1.0	0.0059	0.9104

Figure 1 displays the pointwise confidence limits for progression-free survival in the Xanomeline Low Dose arm, comparing linear and log–log methods. This PFS result is particularly important for programmers and statisticians because it shows the practical consequence of method choice. The linear confidence interval spans the full probability range at the event time and is therefore minimally informative, whereas the log–log interval remains bounded and asymmetric, conveying useful inferential content.

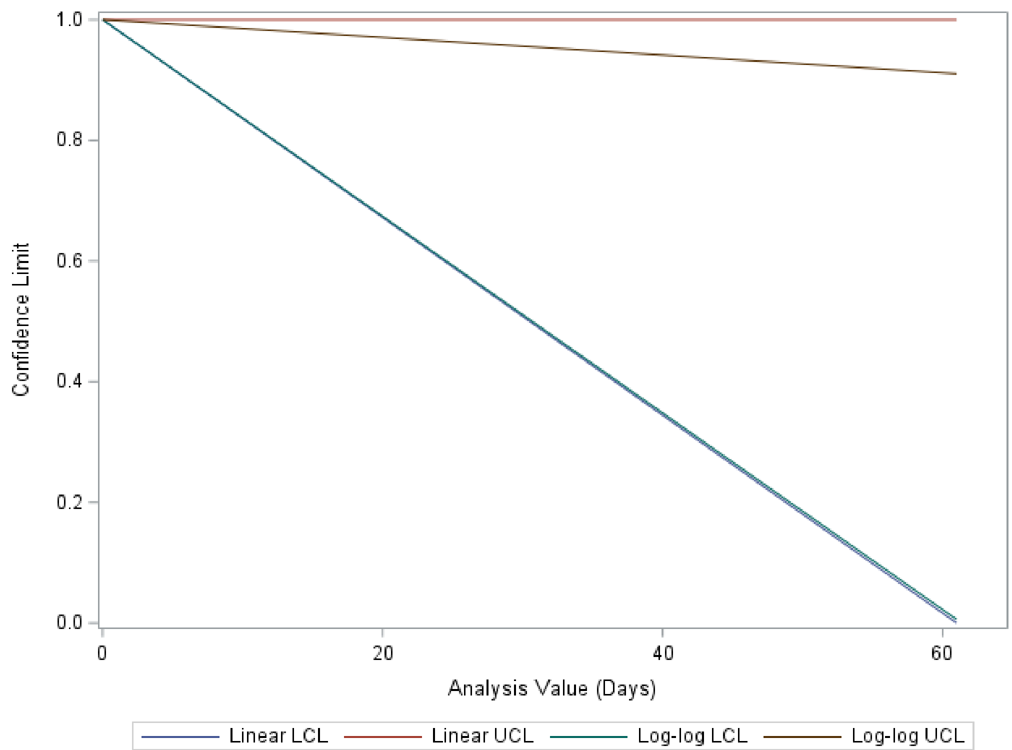

Fig 1: Linear vs LogLog Upper and Lower Confidence Limits difference

4. Discussion and Implications

The manual and empirical results lead to the same conclusion. Linear and log–log confidence intervals do not change the Kaplan-Meier estimate, but they do change the uncertainty attached to that estimate.

The practical weakness of the linear interval is not only that it can exceed 0 or 1. The deeper problem is that it assumes uncertainty is symmetric on the survival scale even when the sampling behavior of the Kaplan-Meier estimator is not symmetric. This is most evident when survival probabilities are high and the number of events is small. In those settings, linear intervals can appear deceptively tight because truncation at 1.0 compresses the upper part of the interval. That is not actual precision, it is an artifact of imposing a probability boundary on an interval that was constructed on an inappropriate scale.

The log-log interval is preferable because it addresses both the boundary issue and the variance issue. By transforming the survival function to $\log\{-\log(S)\}$, the method makes the variance more stable across the range of survival probabilities and improves the normal approximation used to construct the interval. When transformed back, the resulting confidence limits remain bounded and naturally asymmetric. This asymmetry is not a disadvantage, it is a more realistic reflection of the true sampling uncertainty of the survival estimate. Borgan and Liestøl showed that transformed confidence intervals can substantially improve small-sample performance relative to non-transformed intervals, which is directly relevant when the effective sample size is driven by few events rather than by total subject count.

In ADTTE workflows, programmers often focus on generating the survival estimate and neglect the type of confidence interval chosen. The present results shows the impact statistically. In heavily censored oncology datasets, the confidence interval method can materially change how the reliability of the same survival estimate is interpreted. Therefore, programmers should treat CONFTYPE= as an analytic choice, not merely a formatting option.

For regulatory-quality time-to-event reporting, log–log confidence intervals provide a more defensible default.

5. Conclusion

This paper compared linear and log–log confidence intervals for Kaplan-Meier survival estimates from both a theoretical and an empirical SAS 9.4 perspective. The Kaplan–Meier point estimate is identical under both methods, but the confidence intervals differ in statistical behavior and interpretability.

The manual example showed that a linear confidence interval can extend beyond 1 and require truncation, whereas the corresponding log-log interval remains bounded and asymmetric. The empirical OS and PFS examples confirmed the same pattern. In sparse-event, heavily censored settings, linear intervals repeatedly reached the upper boundary of 1.0 and, in one PFS example, expanded to the full interval [0,1]. Log–log intervals remained bounded and retained informative structure.

For advanced statistical programming practice, the implication is direct: when implementing survival analyses in SAS 9.4, particularly for oncology ADTTE workflows, log–log confidence intervals should generally be preferred over linear intervals for pointwise inference.

References

1. Greenwood, M. (1926). *The natural duration of cancer*. Reports on Public Health and Medical Subjects.
2. Borgan, Ø., & Liestøl, K. (1990). *A note on confidence intervals and bands for the survival function*. Scandinavian Journal of Statistics.
3. SAS Communities. *CONFTYPE=LOGLOG or LINEAR in PROC LIFETEST?* Solved: CONFTYPE = LOGLOG or LINEAR in PROC LIFETEST? - SAS Support Communities
4. SAS Support Knowledge Base. *Note 12983 – Confidence limits in the OUTSURV= data set are log-log transformed limits beginning in SAS 9.1*. 12983 - Confidence limits in the OUTSURV= data set are log-log transformed limits beginning in SAS 9.1
5. SAS/STAT® User's Guide. *PROC LIFETEST – Pointwise Confidence Limits*. SAS Help Center: Pointwise Confidence Limits in the OUTSURV= Data Set

6. Brookmeyer, R., & Crowley, J. (1982). *A confidence interval for the median survival time*. Biometrics, 38(1), 29–41.
7. Gautier F, Bundfuss S, Mancini E, Patil L, Shuliar V (2025). *pharmaverseadam: ADaM Test Data for the 'Pharmaverse' Family of Packages*. R package version 1.2.0. Available at: GitHub - pharmaverse/pharmaverseadam: Test ADaM data for use with pharmaverse · GitHub
8. Miller E et al. (2026). *xportr: Tools to build CDISC compliant data sets and write SAS XPT files*. GitHub repository. Available at: GitHub - atorus-research/xportr: Tools to build CDISC compliant data sets and check for CDISC compliance. · GitHub

Sample SAS Code

```
******************START of comparison*********************************;

%macro survival (data=,);

*********For OS/PFS Linear *******;

proc lifetest data=adtte_&data outsurv=outsurv_&data._linear

              conftype=linear

              plots=survival(cb=hw atrisk=0 to &maxtime by 30) timelist=&_s;

  time aval*cnsr(1); /* 1=censored */

  strata armcd;

run;

*********For OS/PFS LogLog *******;

ods output SurvivalPlot=loglog_ci_&data;

proc lifetest data=adtte_&data outsurv=outsurv_&data._loglog

              conftype=loglog

              plots=survival(cb=hw atrisk=0 to &maxtime by 30) timelist=&_s;

  time aval*cnsr(1);

  strata armcd;

run;

%mend survival;

%survival (data=os);

%survival (data=pfs);

******************************END of Comparison******************************;
```

Contact Information

Prayag Shah
Revolution Medicines, Inc.
Email: prshah@revmed.com