

A Sensitivity Analysis Paradigm for Randomized Trials with Informative Censoring

Daniel Scharfstein
Department of Biostatistics
Johns Hopkins University
dscharf@jhsph.edu



Statistical Principles for Clinical Trials (FDA Guidance - E9)

- “It is important to evaluate the **robustness** of the results and primary conclusions of the trial.”
- Robustness refers to “the **sensitivity** of the overall conclusions to various limitations of the data, assumptions, and analytic approaches to data analysis



FDA Critical Path Initiative

- ❑ FDA's *Critical Path* white-paper "calls for a joint effort of industry, academia, and the FDA to identify key problems and develop targeted solutions."
- ❑ In response to this document, the Office of Biostatistics at CDER has identified **missing data** as one of these key areas (Robert O'Neill, personal communication).



FDA: Office of Biostatistics at CDER (Robert O'Neill)

- ❑ “Virtually every drug/disease area in clinical trials has problems with patient data that are missing because patients dropped out, died, withdrew due to toxicity or aggravation with the trial, failed to complete forms, and other reasons.”
- ❑ “The success and failure of a trial and its interpretation often depends on how these missing data are dealt with at either the planning or analysis stage.”
- ❑ “Current statistical methodologies proposed need to be evaluated for their ability to address **informative treatment-related** missing data ...”



FDA: Office of Biostatistics at CDER (Robert O'Neill)

- ❑ "... a consensus on missing data approaches needs to be developed in order to minimize the impact of failed studies and remove obstacles to ambiguous interpretation of product efficacy and safety conclusions."
- ❑ "There are **no** established set of **diagnostics** to evaluate the severity or impact of missing data ..."
- ❑ "there is no easily available computer software ..."



Topiramate Study 106

- ❑ Double-blind, randomized trial to evaluate the effectiveness of low (50 mg/day) vs. high (400 mg/day) dose topiramate in reducing the time to first seizure in pediatric and adult subjects with newly diagnosed or recurrent epilepsy.



Completion Status

Completion Status	TPM 50 (N=234)		TPM 400 (N =236)		Total (N=470)	
	n	%	n	%	n	%
Completed	195	83	161	68	356	76
Completed at DB phase termination ^a	105	45	112	47	217	46
Completed by having a seizure	90	38	49	21	139	30
Withdrew	39	17	75	32	114	24
Adverse event	13	6	40	17	53	11
Subject choice	9	4	13	6	22	5
Lost to follow-up	9	4	10	4	19	4
Other	8	3	12	5	20	4

^a Subjects who were active in the study at the time of termination of the double-blind (DB) phase, i.e., 6 months after randomization of the last subject.

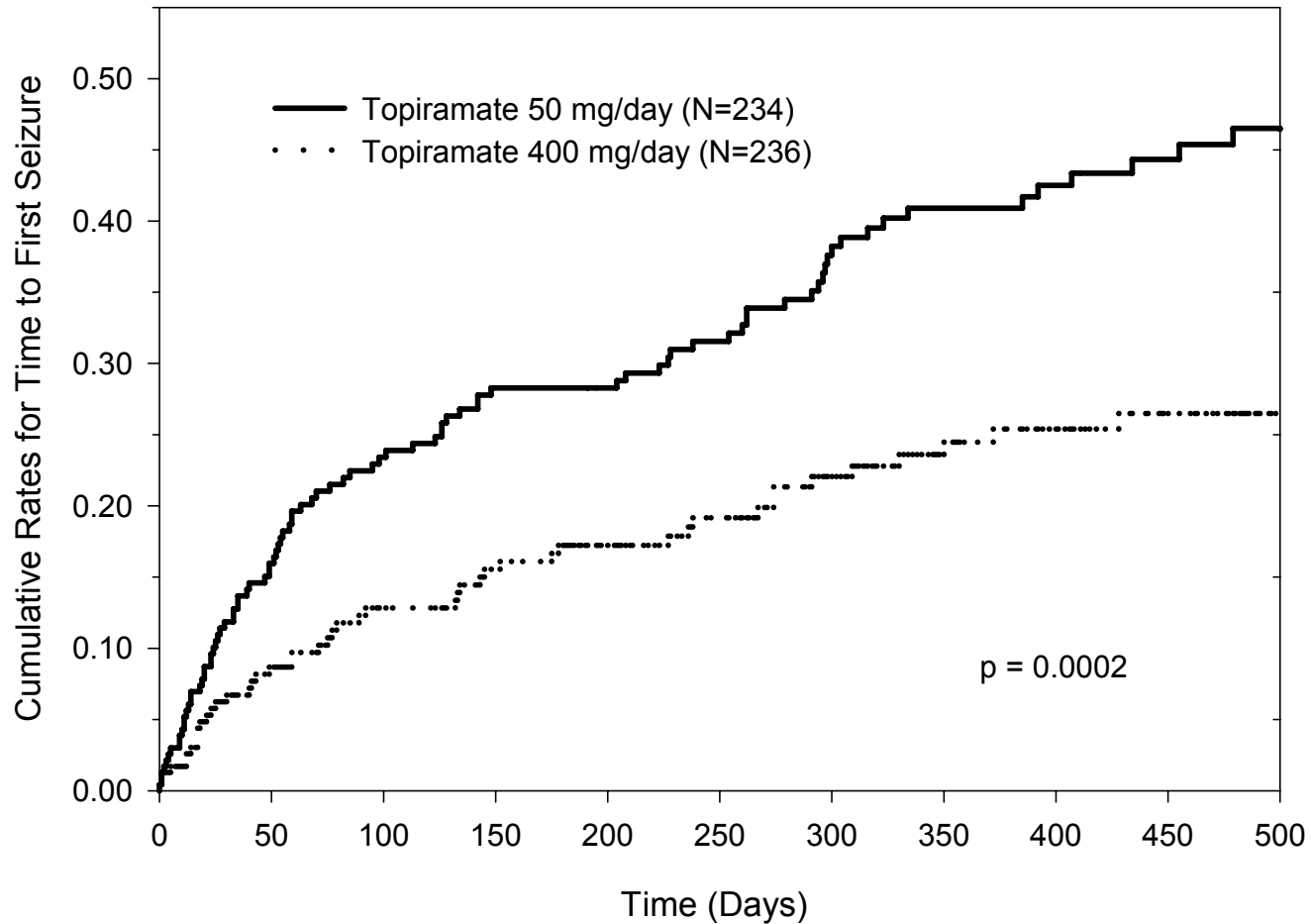


Competing Causes of Censoring

- ❑ Premature withdrawal
 - ❑ Subjects were not followed after withdrawal.
- ❑ Administrative loss due to study termination.



Independent Censoring Analysis



Concern

- ❑ The results of the log rank test might be biased and difficult to interpret due to missing efficacy data from a high number of dropouts.
- ❑ An analysis which considers everyone who dropped out of the trial as having a seizure event is not statistically significance at the 0.05 level.
 - ❑ Too conservative; not clinically reasonable



Question

- How should the “**robustness** of the results and primary conclusions of the trial” be evaluated?



Proposal

- ❑ Specify a class of models, indexed by **scientifically interpretable, non-identifiable** parameters that express deviations from independent censoring.
- ❑ Draw inference about treatment effects over a range of these parameters, especially those that are considered plausible by scientific experts.



The Statistical Methodology



Informative Censoring

- ❑ If premature withdrawal is related to the underlying time to first seizure, classical survival analysis techniques, which assume independent censoring, will be biased.
- ❑ Since the failure time and censoring times are never observed jointly, the dependence between these times is not empirically verifiable.
- ❑ To analyze these data, assumptions about this dependence are required.



Notation

T^* = days to first seizure

C_1^* = days to premature withdrawal

C_2^* = days to administrative censoring

We assume that the time from the start of randomization until the date of analysis is 500 days.

$T = \min(T^*, 500 + \epsilon)$

$C_1 = \min(C_1^*, 500 + \epsilon)$

$C_2 = \min(C_2^*, 500 + \epsilon)$



Notation

T^* = days to first seizure

C_1^* = days to premature withdrawal

C_2^* = days to administrative censoring

We assume that the time from the start of randomization until the date of analysis is 500 days.

$$T = \min(T^*, 500 + \epsilon)$$

$$C_1 = \min(C_1^*, 500 + \epsilon)$$

$$C_2 = \min(C_2^*, 500 + \epsilon)$$

For ties, we assume that seizure occurs prior to premature withdrawal which occurs prior to administrative censoring.



Observed Data

Each subject is observed until

$$X = \min(T, C_1, C_2)$$

$$\Delta = \begin{cases} 0 & \text{if } T \text{ is observed} \\ 1 & \text{if } C_1 \text{ is observed} \\ 2 & \text{if } C_2 \text{ is observed} \end{cases}$$

The observable data for a subject is $O = (X, \Delta)$.

We assume that we observe n i.i.d. copies of O .

$$\mathbf{O} = \{O_i = (X_i, \Delta_i) : i = 1, \dots, n\}$$



The Bucket

If $X = 500 + \epsilon$, then we know that the time to first seizure, T^* , is longer than 500 days.

We do not know when. In this case, we say that T^* falls in "the bucket."



Goals

1. Estimate the distribution of time to first seizure.
2. Test the null hypothesis of no difference between the treatment-specific failure time distributions.



Sensitivity Analysis Models

- Our approach is based on assuming separate cause-specific hazards models for the two types of censoring mechanisms.



Administrative Censoring

For administrative censoring, we will assume a model of the form:

$$\lambda_{C_2}^\dagger(t|T, T > t, C_1 > t) = \psi_2(t) \quad t \in [0, 500]$$

This model assumes that, for subjects who at risk for administrative censoring at time t , there is no relationship between T and censoring just after t .

That is, administrative censoring is assumed to be non-informative.



Premature Censoring

For censoring due to premature withdrawal, we will assume a model of the form:

$$\lambda_{C_1}^\dagger(t|T, T > t, C_2 \geq t) = \psi_1(t) \exp\{q(t, T)\} \quad t \in [0, T)$$

Specification of the non-identifiable function $q(t, T)$ is equivalent to quantifying, for those who remain at risk at time t , the dependence on a hazard ratio scale between T and censoring due to premature withdrawal just after time t .



Premature Censoring

We refer to the function $q(t, T)$ as a *censoring bias function*.

The censoring bias function will be non-constant when there exists unmeasured factors which are both correlated with censoring due to premature withdrawal and failure.

If the function is constant (e.g., 0), then there is no dependence relationship and censoring due to premature withdrawal is non-informative.

Since the censoring bias function is not empirically verifiable without auxiliary information, perform sensitivity analysis.



Parameterization of Censoring Bias Function

$$q(t, T; \alpha) = \alpha(T - t)$$



Parameterization of Censoring Bias Function

$$q(t, T; \beta) = \alpha \{ I(T < 500 + \epsilon)(T - t) + I(T = 500 + \epsilon)(\tau - t) \} / 365$$



Interpretation of Censoring Bias Parameters

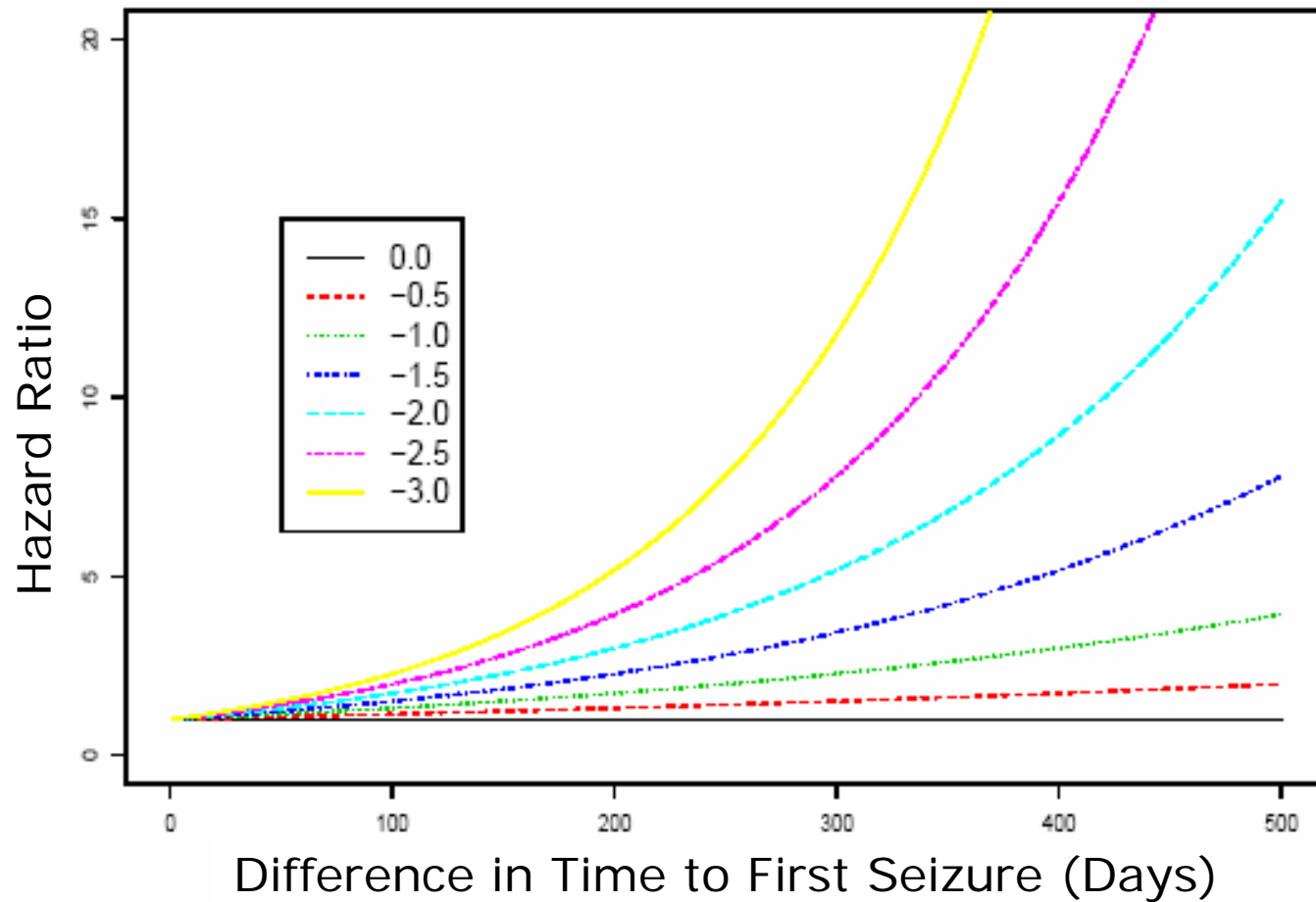
For two subjects, who

- are at risk at day t
- would experience a seizure between t and 500 days
- whose time of first seizure differ by d days

the quantity $\exp(\alpha d/365)$ is interpreted as the hazard ratio of premature withdrawal at day t .



Hazard ratio for premature censoring for subjects with varying differences between time to first seizure and α



Interpretation of Censoring Bias Parameters

Consider two subjects at risk at day t .

The first subject would not fail before 500 days. Thus, he falls in the "bucket."

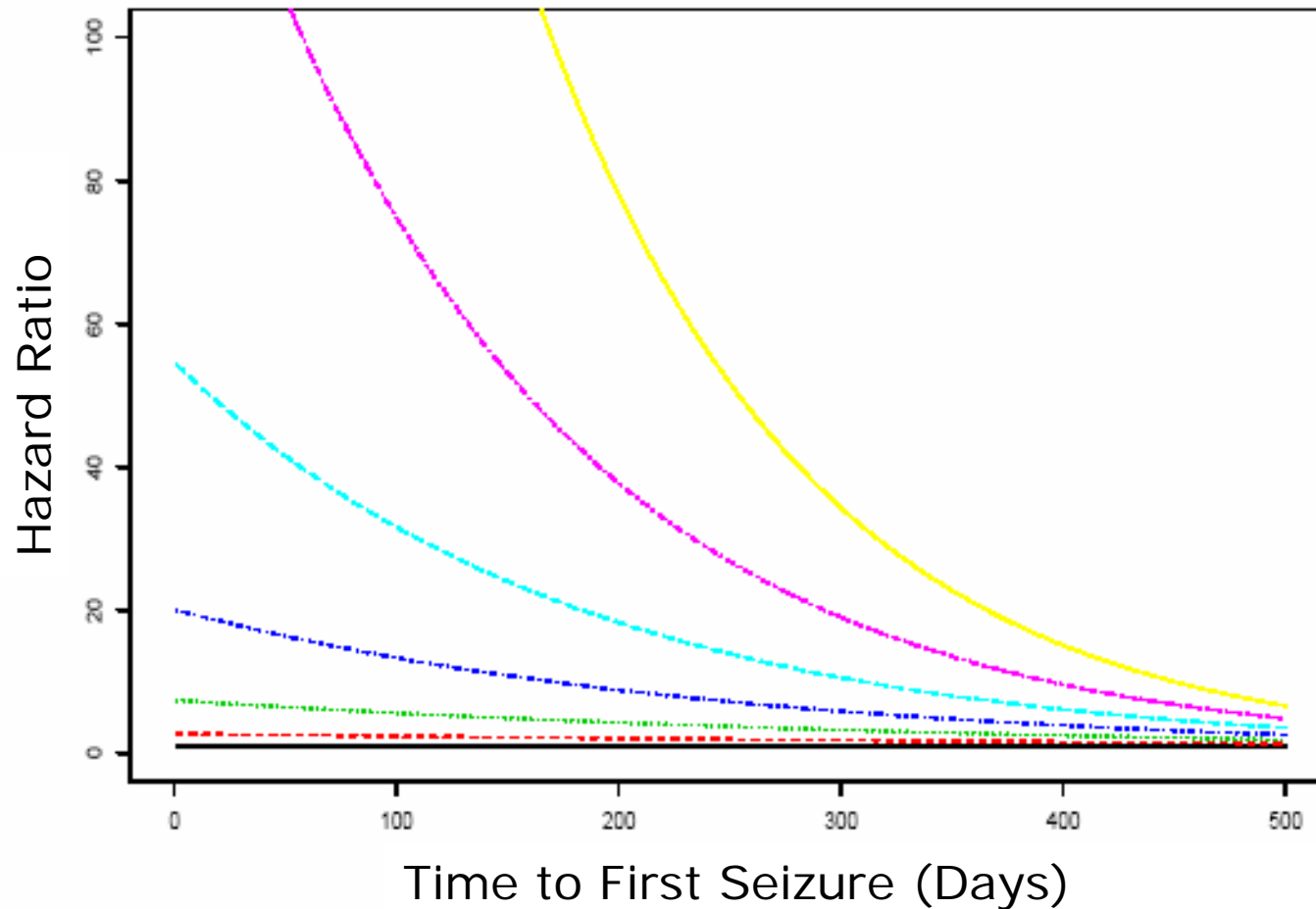
The second subject would failure just after day t .

The quantity $\exp(\alpha(\tau - t)/365)$ is the hazard ratio of premature withdrawal at day t comparing the first to the second type of subject.

Together, the choice of α and τ indicate the degree to which subjects who would experience the event before and after time 500 days are different than one another with respect to their risk of withdrawing prematurely.



Hazard ratio for premature censoring for subjects who would not fail before 500 days as comparing to subjects with varying times of first seizure (x-axis) and varying α



Interpretation of Censoring Bias Parameters

$\alpha > 0$ implies that subjects with longer times to events are more likely to be censored at any time t than subjects who would have shorter times to event.

$\alpha < 0$ implies that subjects with longer times to events are less likely to be censored at any time t than subjects who would have shorter times to event.



Estimator of Seizure Time Distribution

When $\alpha = 0$, our estimator of the distribution of time to first seizure is identically equivalent to the Kaplan-Meier estimator (the result of the primary analysis, assuming non-informative censoring).

When $\alpha = -\infty$, then our estimator is equivalent to the Kaplan-Meier estimator, where each prematurely censored observation is treated as a seizure occurring at the next observed seizure time

When $\alpha = +\infty$, then our estimator is equivalent to Kaplan-Meier estimator, where each prematurely censored observation is treated as have a seizure time in the "bucket".



Logrank Statistic

Under the null hypothesis of no treatment effect on the time to first seizure,

$$\int_0^{500} g(t) \{d\Lambda^{(1)}(t) - d\Lambda^{(2)}(t)\} = 0.$$

When $\alpha^{(1)} = \alpha^{(2)} = 0$ (i.e., non-informative censoring in both treatment groups), the numerator of the logrank statistic is

$$\int_0^{500} \frac{Y^{(1)}(t)Y^{(2)}(t)}{Y^{(1)}(t) + Y^{(2)}(t)} \{d\hat{\Lambda}^{(1)}(t) - d\hat{\Lambda}^{(2)}(t)\}.$$

Under H_0 , this statistic should be close to zero.

If this statistic is “sufficiently” large in absolute value then there is “evidence” against the null hypothesis. P-values are constructed using large-sample normal theory.



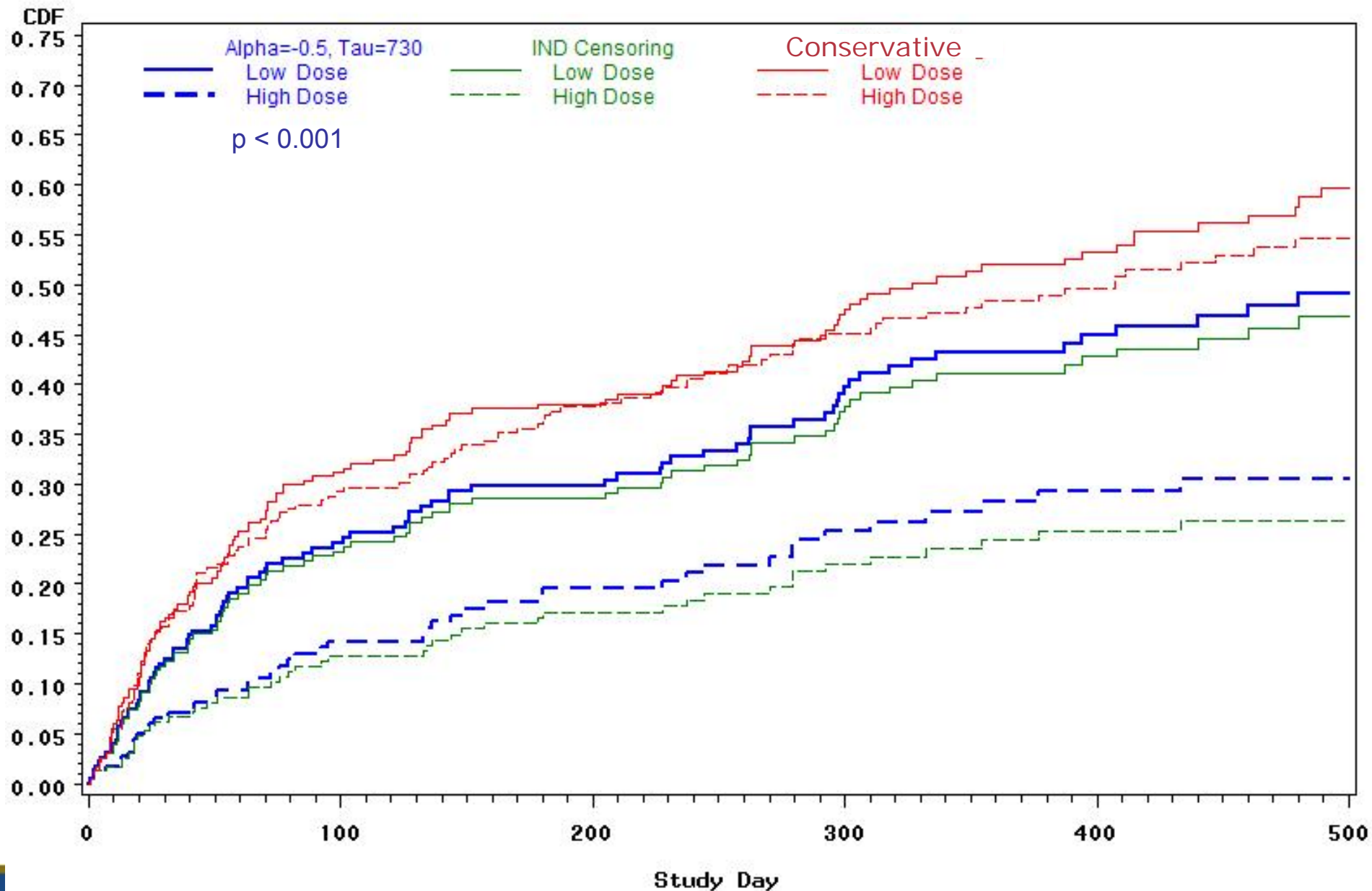
Logrank-Type Statistic

We used the same formula, but substituted the treatment-specific cumulative hazard function estimators, under the specified α 's.

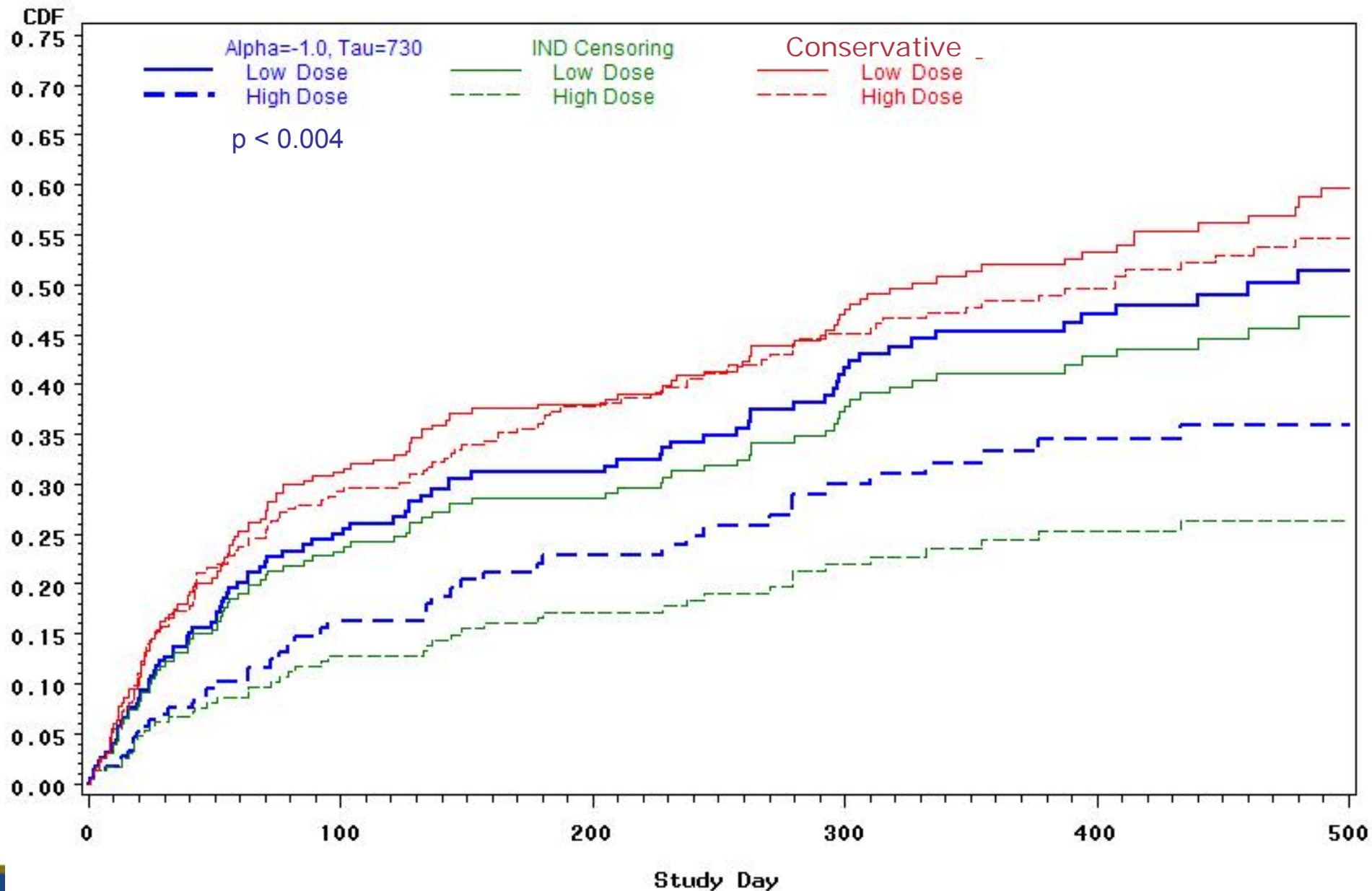
To construct p-values, we used parametric bootstrap to approximate the distribution of our statistic under H_0 .



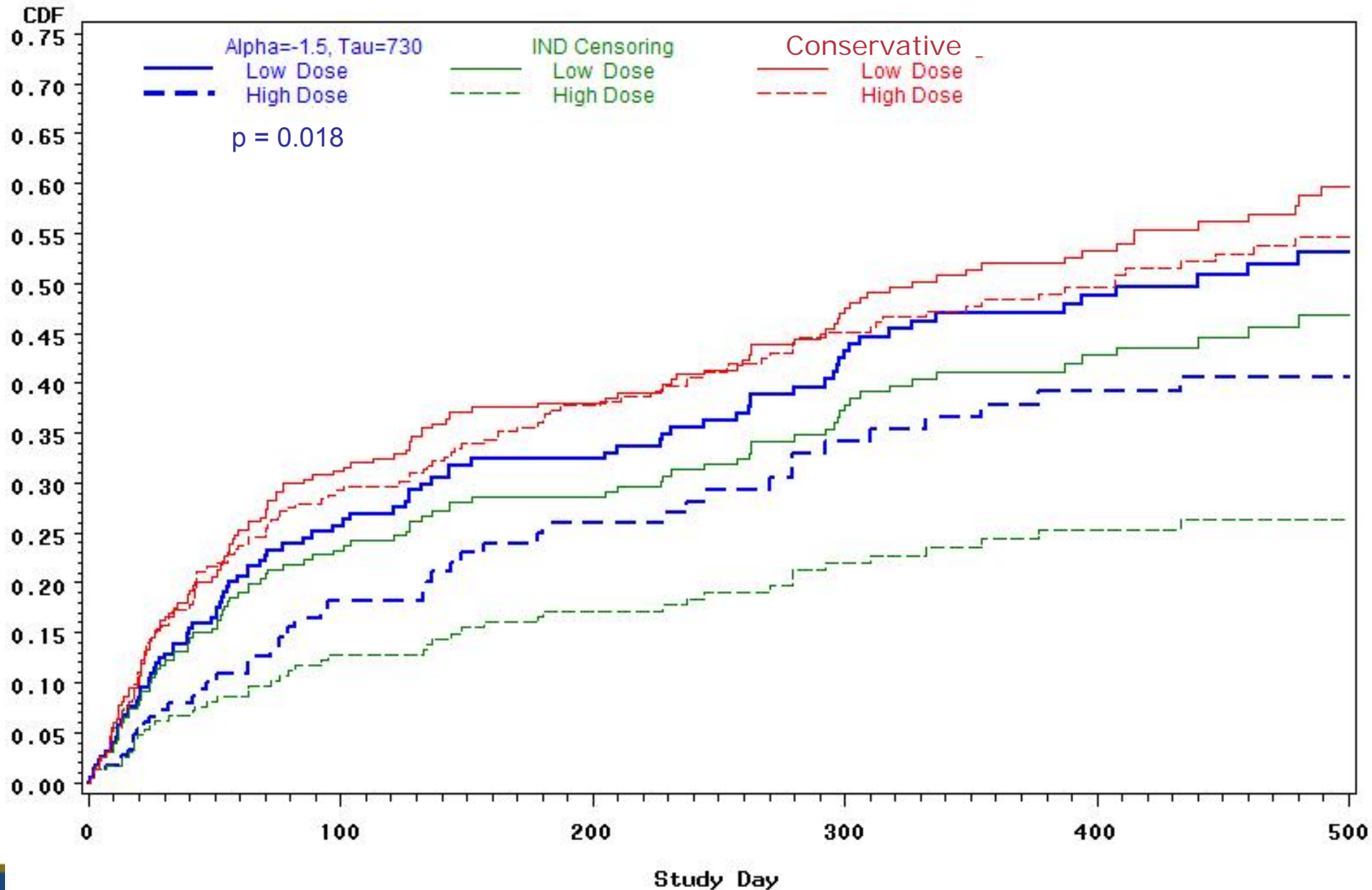
Epilepsy Event CDFs: Alpha=-0.5, Tau=730



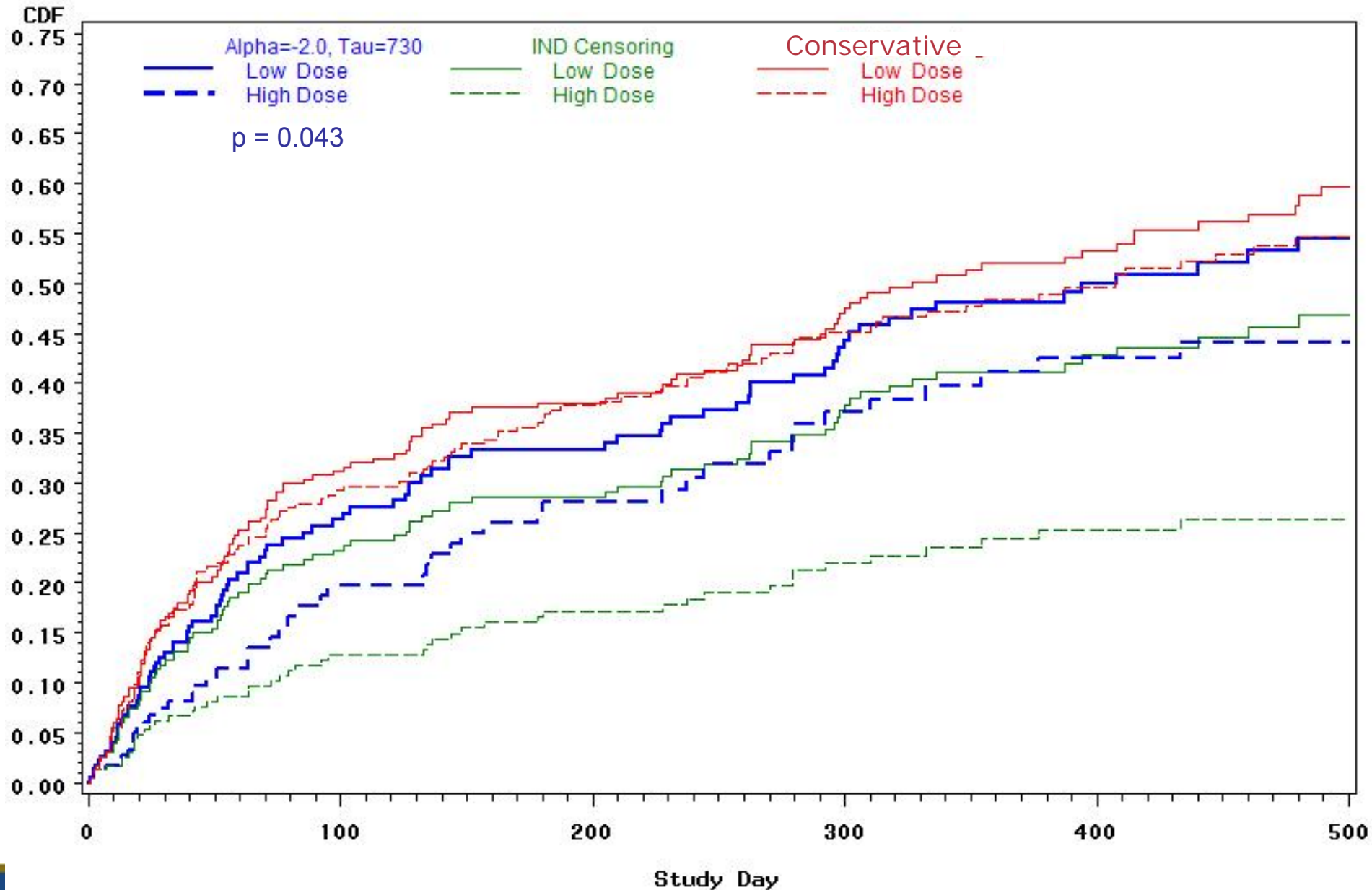
Epilepsy Event CDFs: Alpha=-1.0, Tau=730



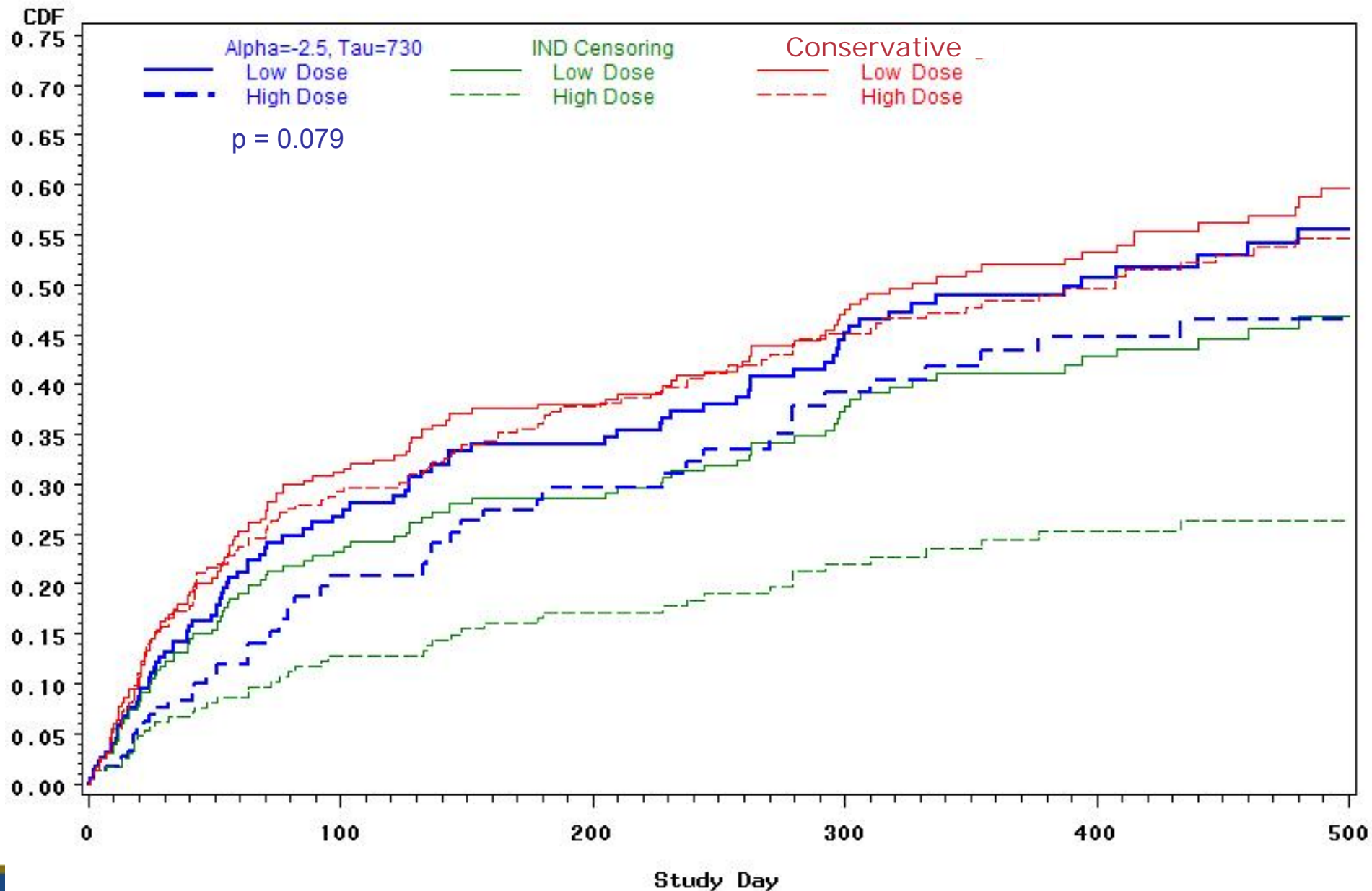
Epilepsy Event CDFs: Alpha=-1.5, Tau=730



Epilepsy Event CDFs: Alpha=-2.0, Tau=730



Epilepsy Event CDFs: Alpha=-2.5, Tau=730



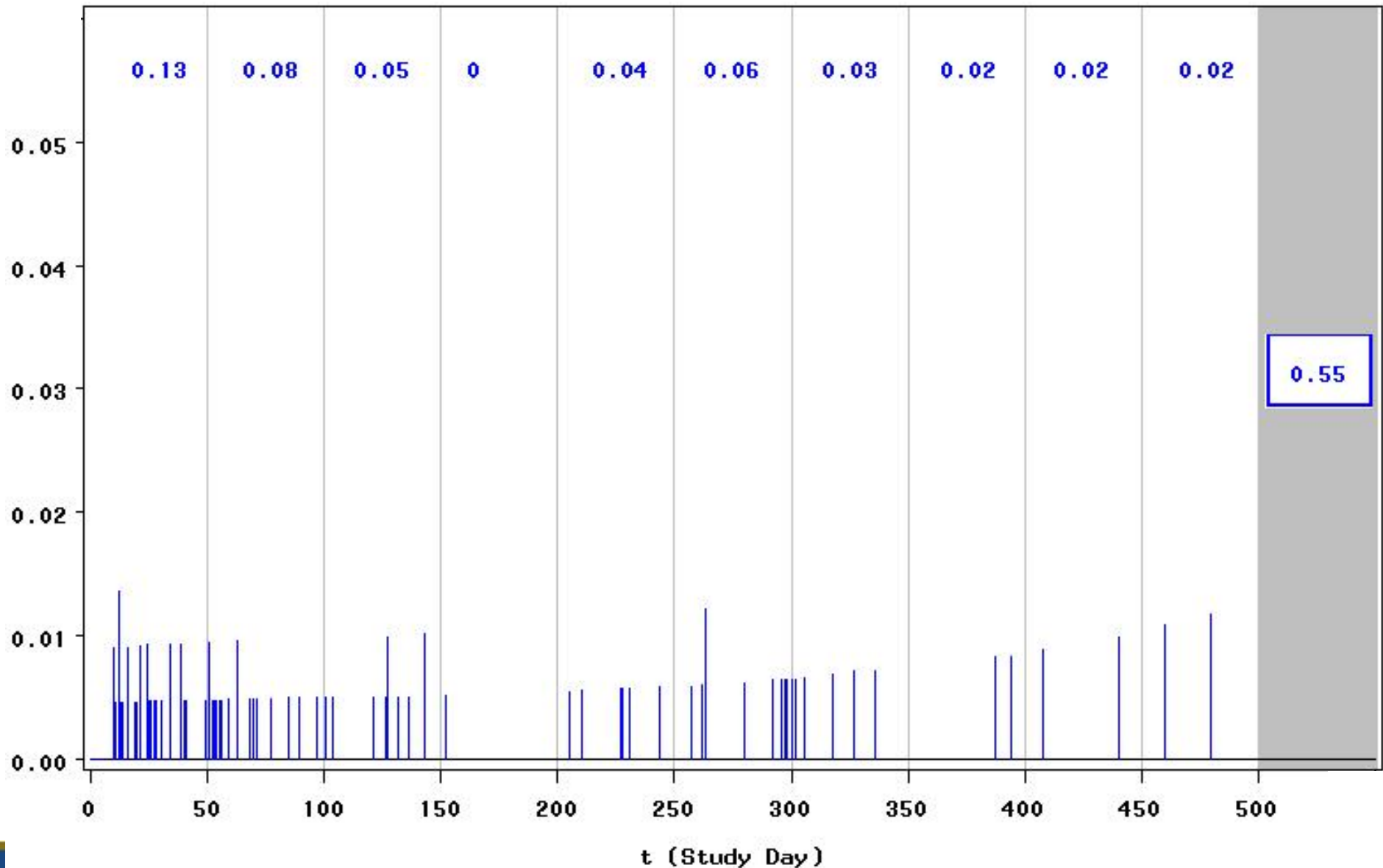
Seizure Time Distribution for a Prematurely Censored Subject

Low Dose Group
Censored on Day 8



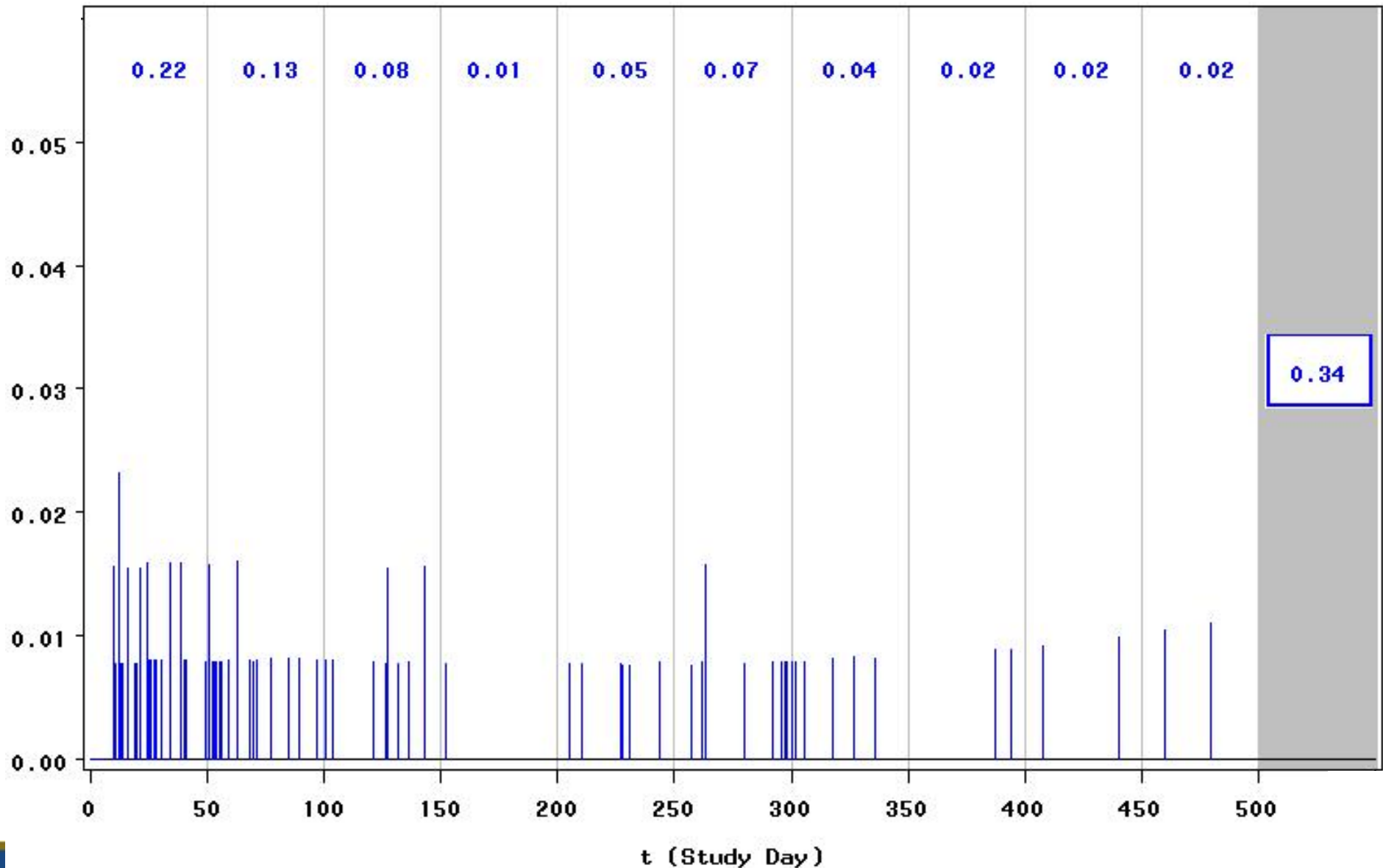
Density Function of T Given Prematurely Censored at day=8

$\Pr(T=t \mid C1=8)$ for $\alpha=0$, $\tau=730$, $\text{trt}=1$



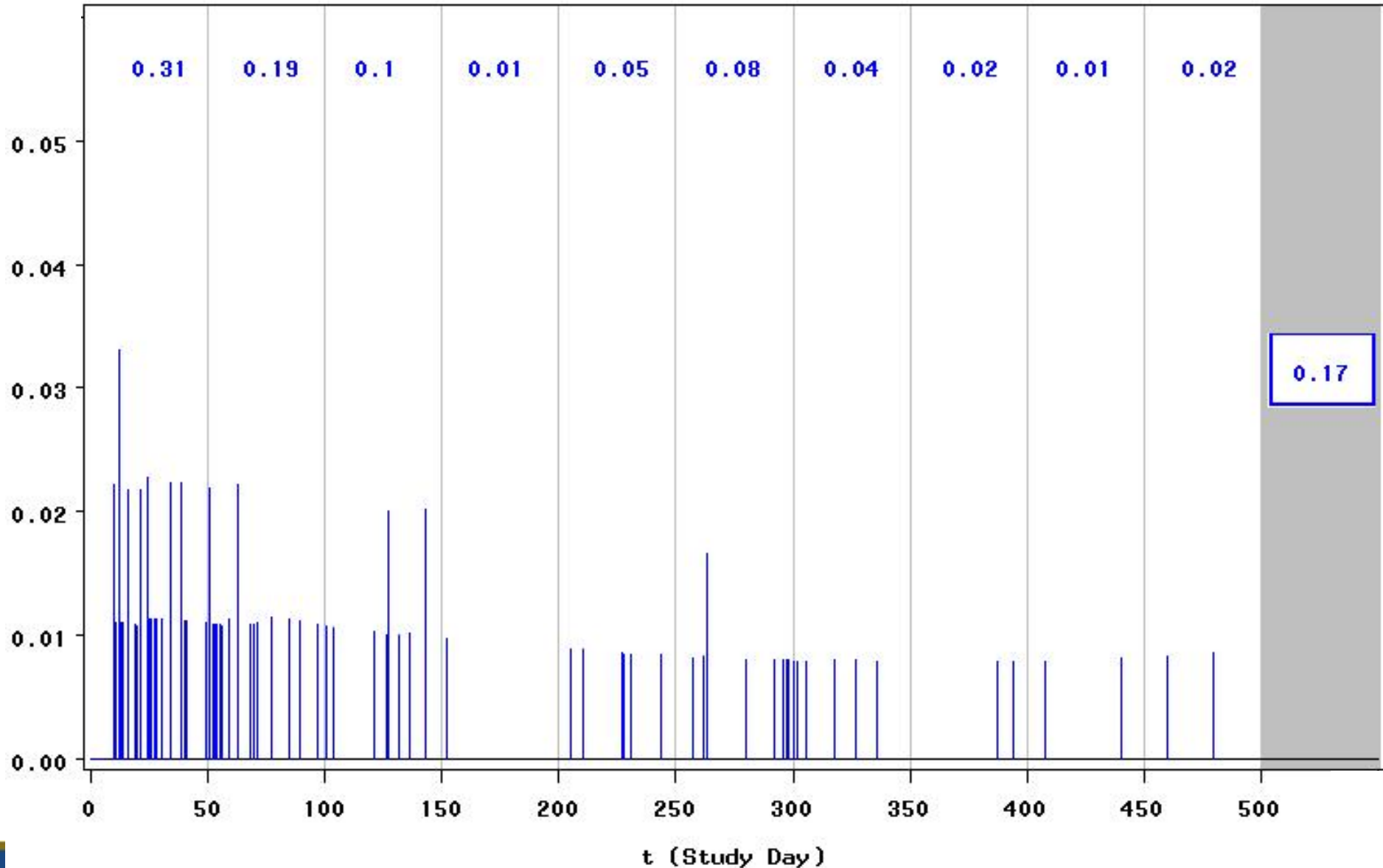
Density Function of T Given Prematurely Censored at day=8

$\Pr(T=t \mid C1=8)$ for $\alpha=-0.5$, $\tau=730$, $\text{trt}=1$



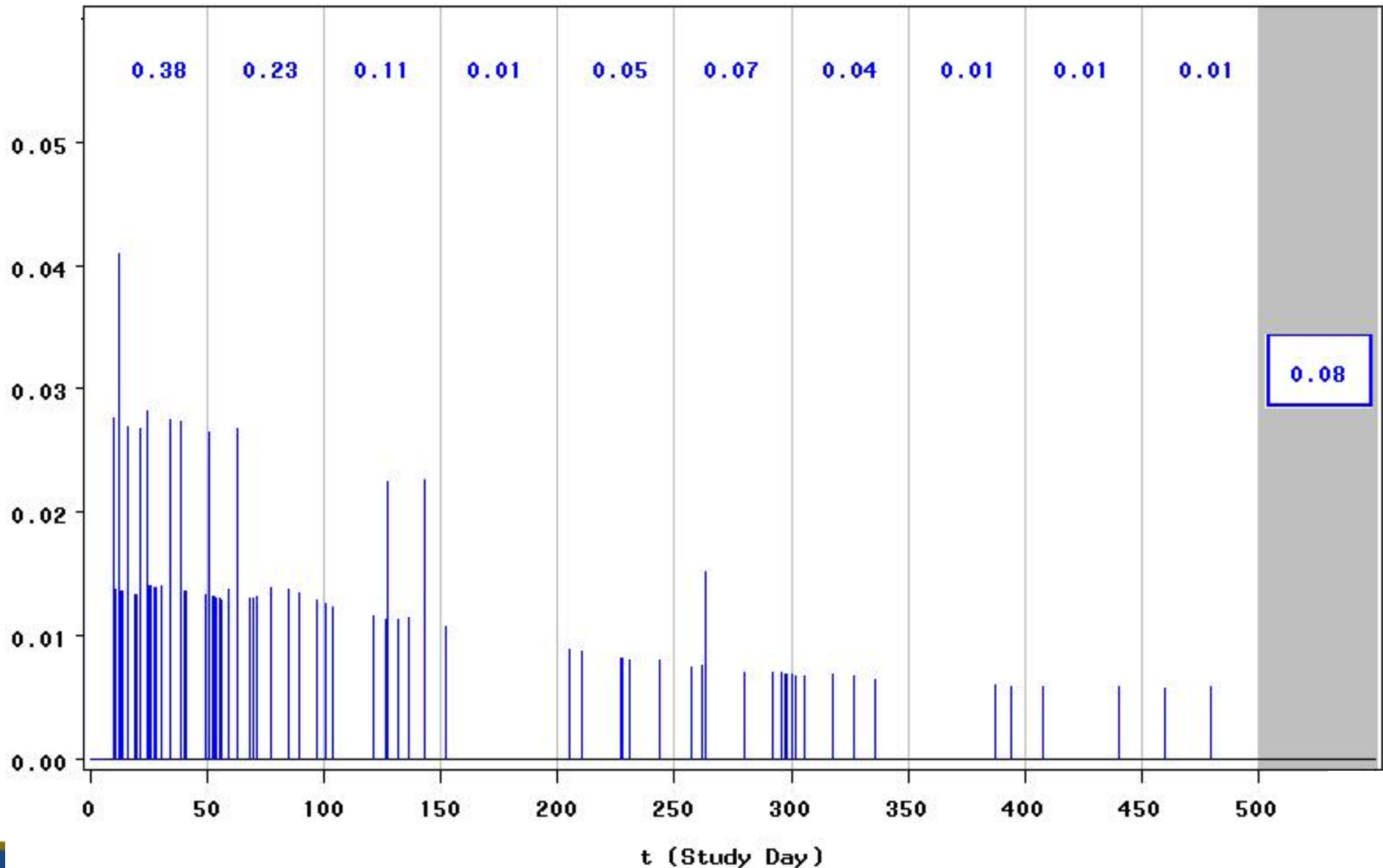
Density Function of T Given Prematurely Censored at day=8

$\Pr(T=t \mid C1=8)$ for $\alpha=-1$, $\tau=730$, $\text{trt}=1$



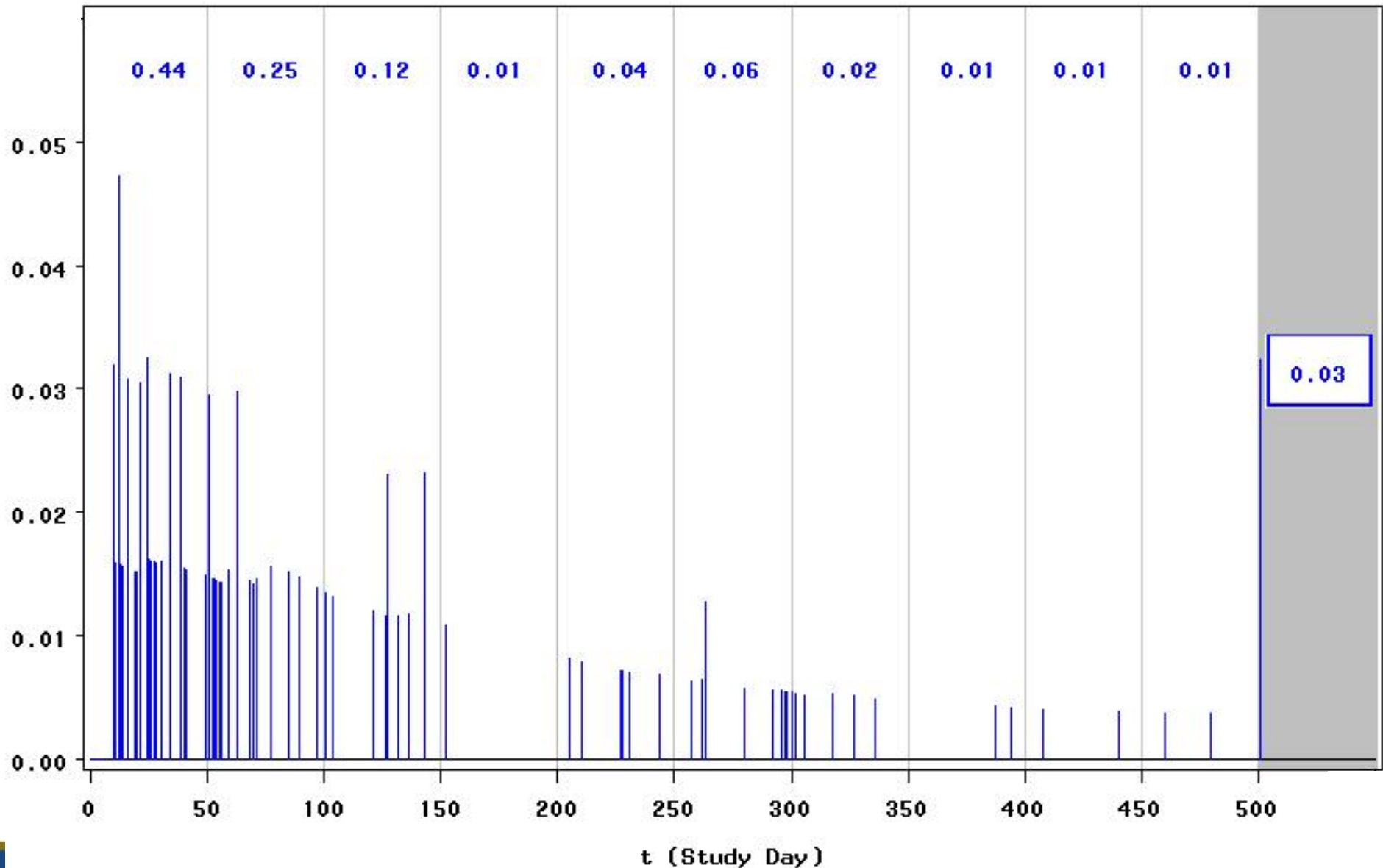
Density Function of T Given Prematurely Censored at day=8

$\Pr(T=t \mid C1=8)$ for $\alpha=-1.5$, $\tau=730$, $\text{trt}=1$



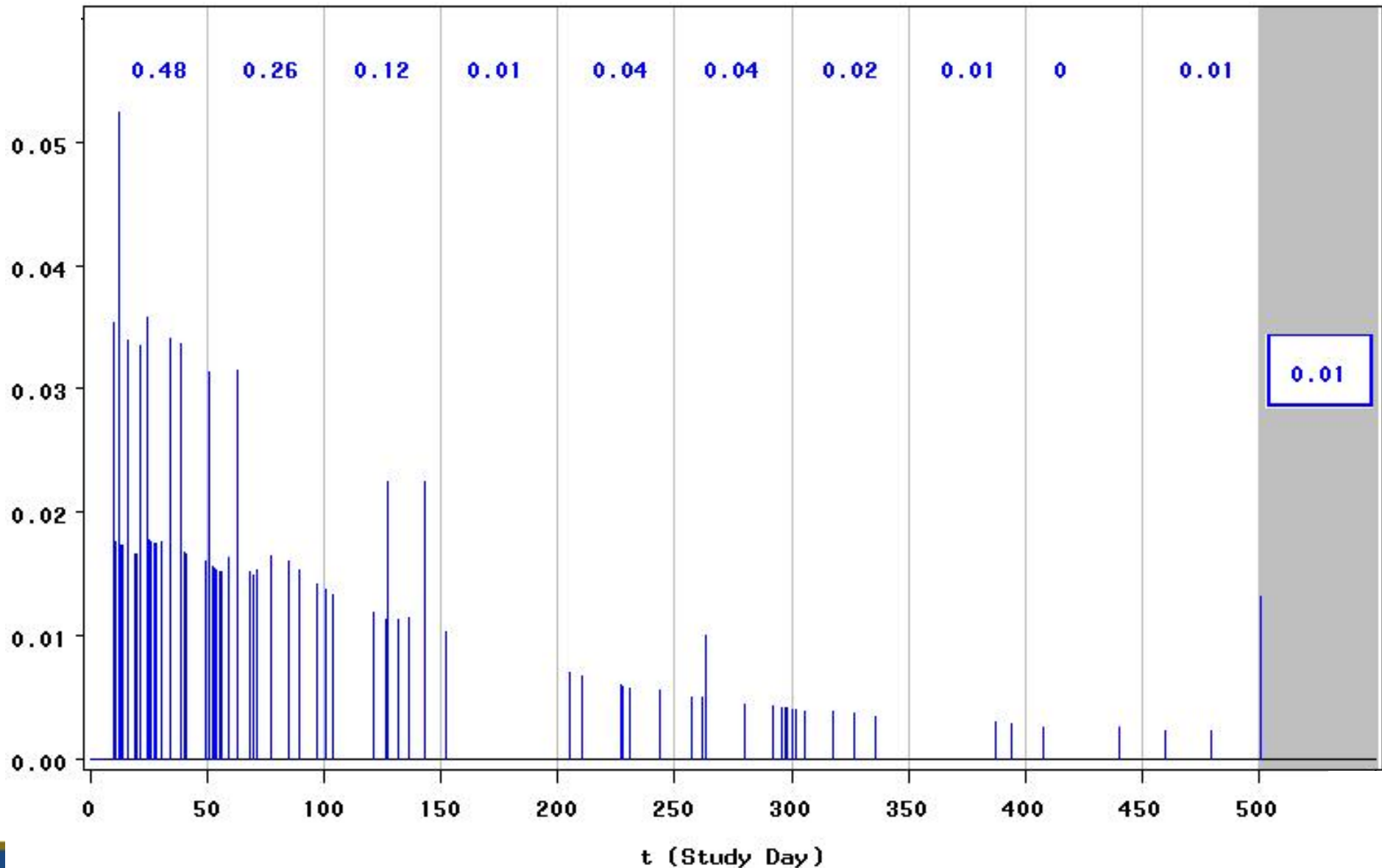
Density Function of T Given Prematurely Censored at day=8

$\Pr(T=t \mid C1=8)$ for $\alpha=-2.0$, $\tau=730$, $\text{trt}=1$



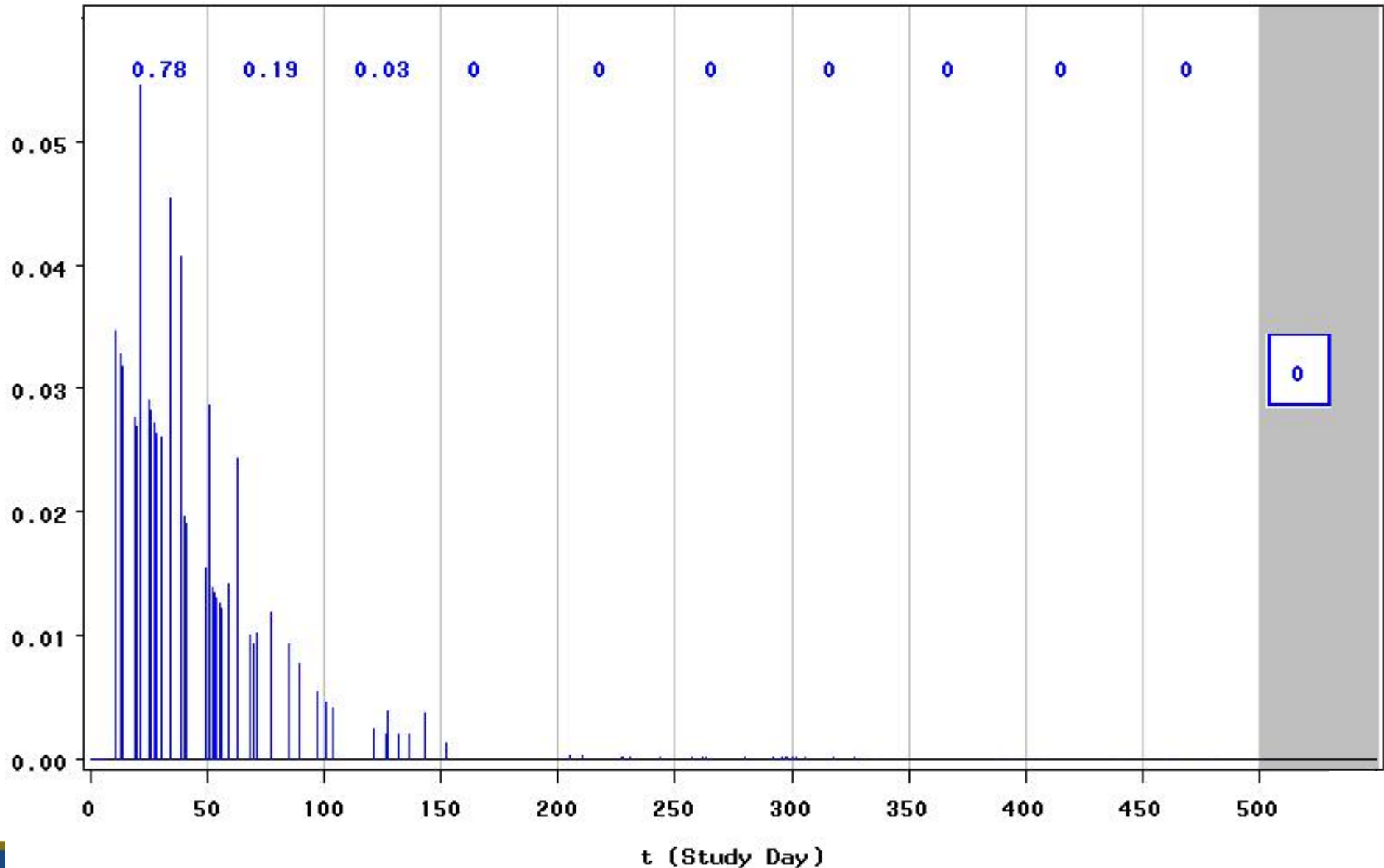
Density Function of T Given Prematurely Censored at day=8

$\Pr(T=t \mid C1=8)$ for $\alpha=-2.5$, $\tau=730$, $\text{trt}=1$



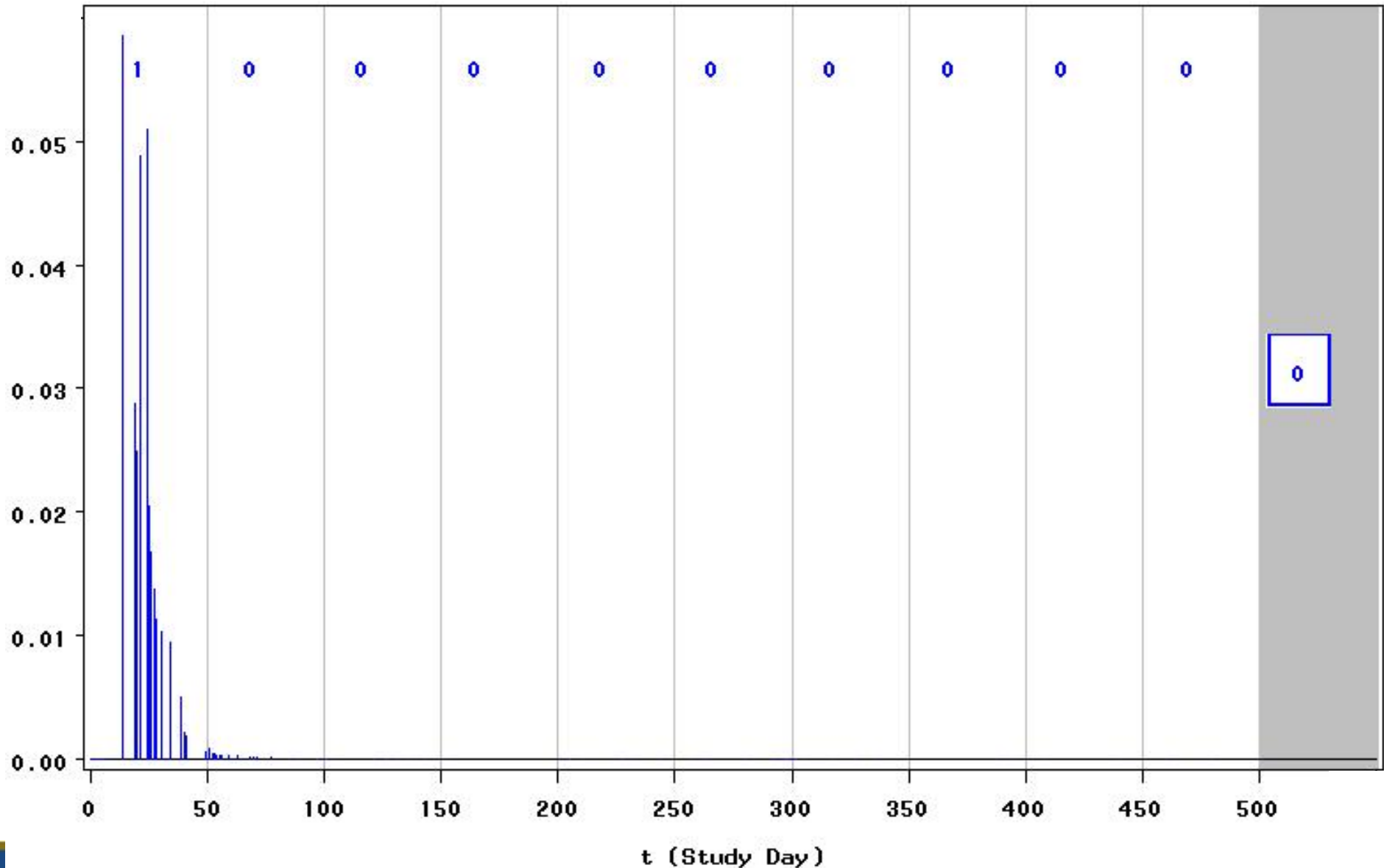
Density Function of T Given Prematurely Censored at day=8

$\Pr(T=t \mid C1=8)$ for $\alpha=-10$, $\tau=730$, $\text{trt}=1$



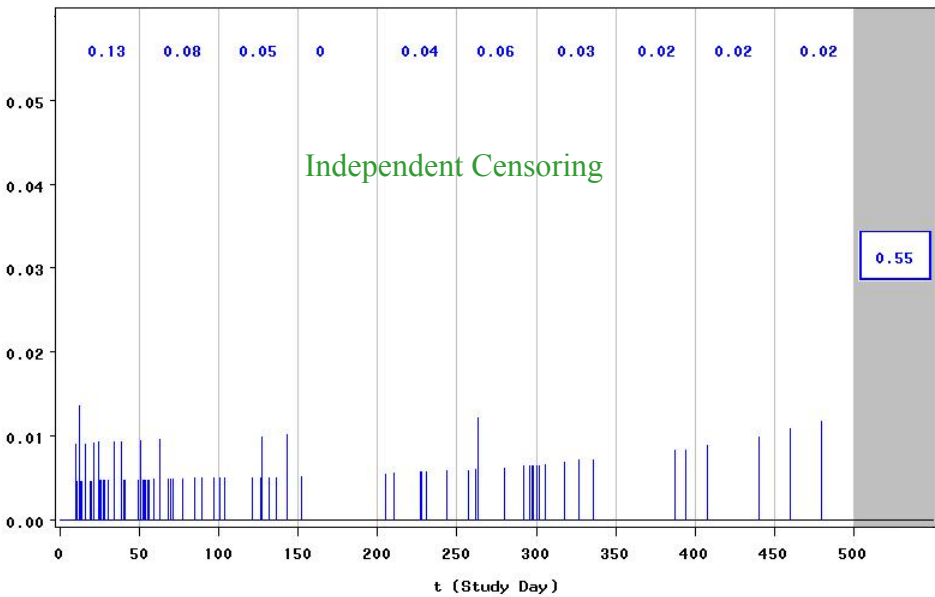
Density Function of T Given Prematurely Censored at day=8

$\Pr(T=t \mid C1=8)$ for $\alpha=-50$, $\tau=730$, $\text{trt}=1$

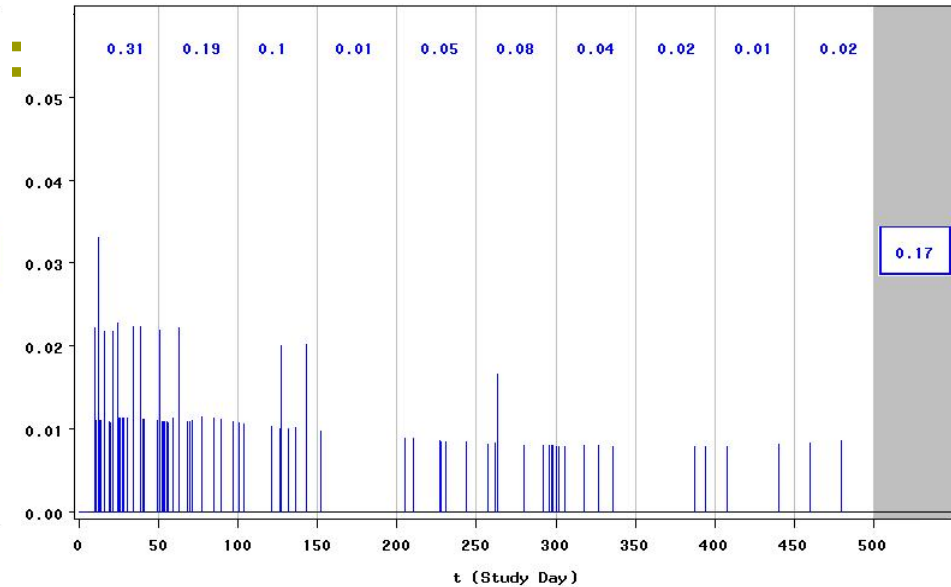


Pr(T=t | Low Dose, C₁=8 days)

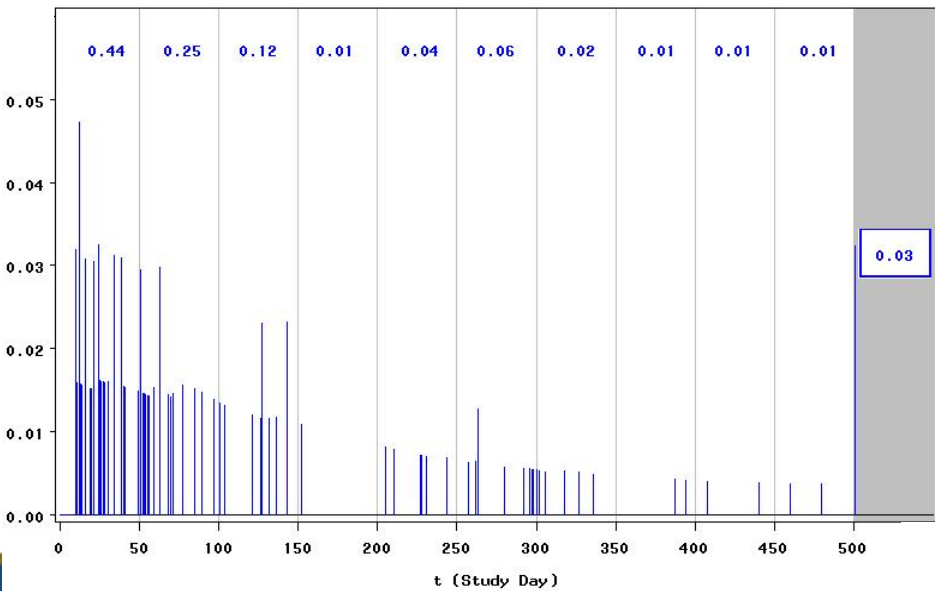
$\alpha = 0, \tau = 730$



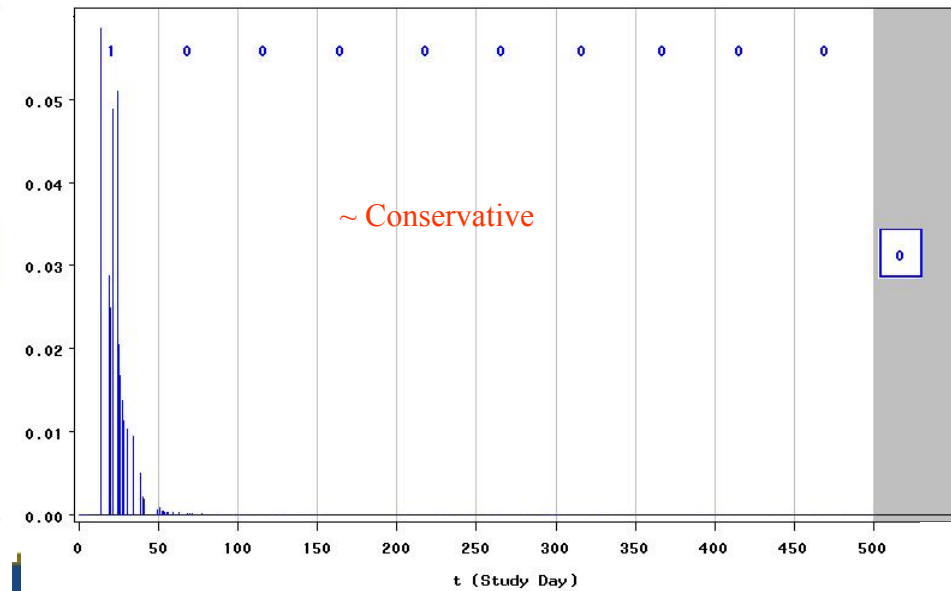
$\alpha = -1, \tau = 730$



$\alpha = -2, \tau = 730$



$\alpha = -50, \tau = 730$



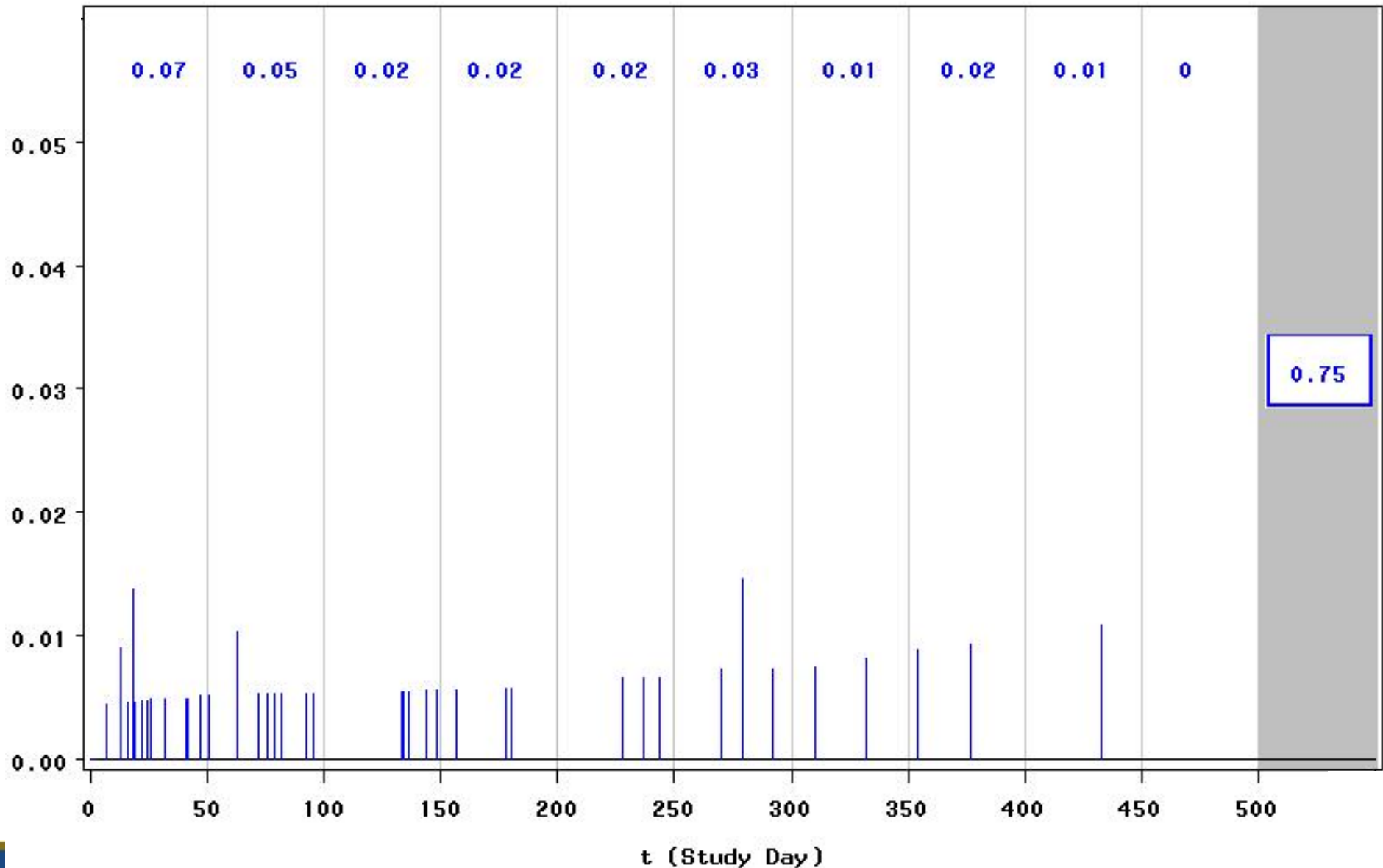
Seizure Time Distribution for a Prematurely Censored Subject

High Dose Group
Censored on Day 4



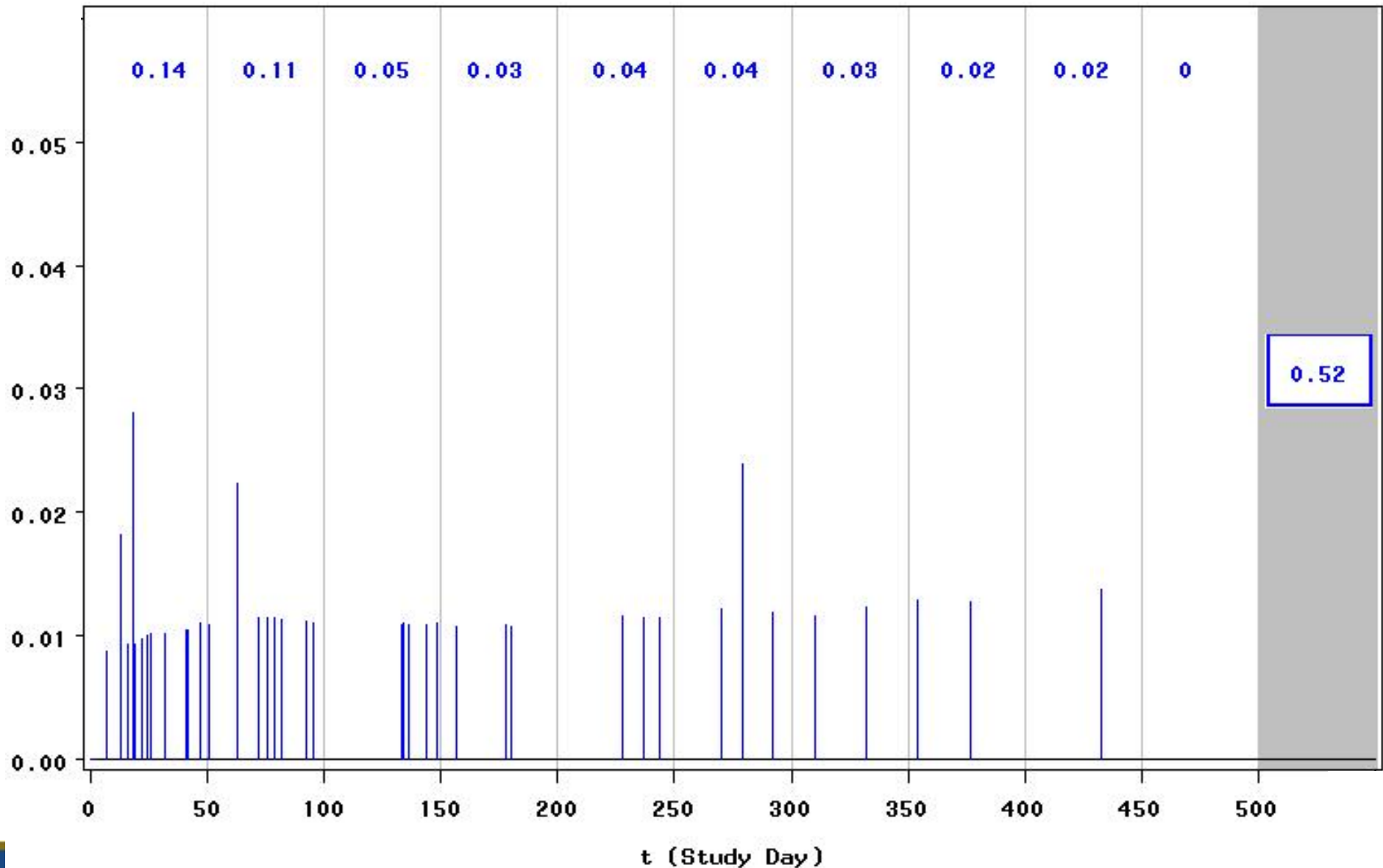
Density Function of T Given Prematurely Censored at day=4

$\Pr(T=t \mid C1=4)$ for $\alpha=0$, $\tau=730$, $\text{trt}=2$



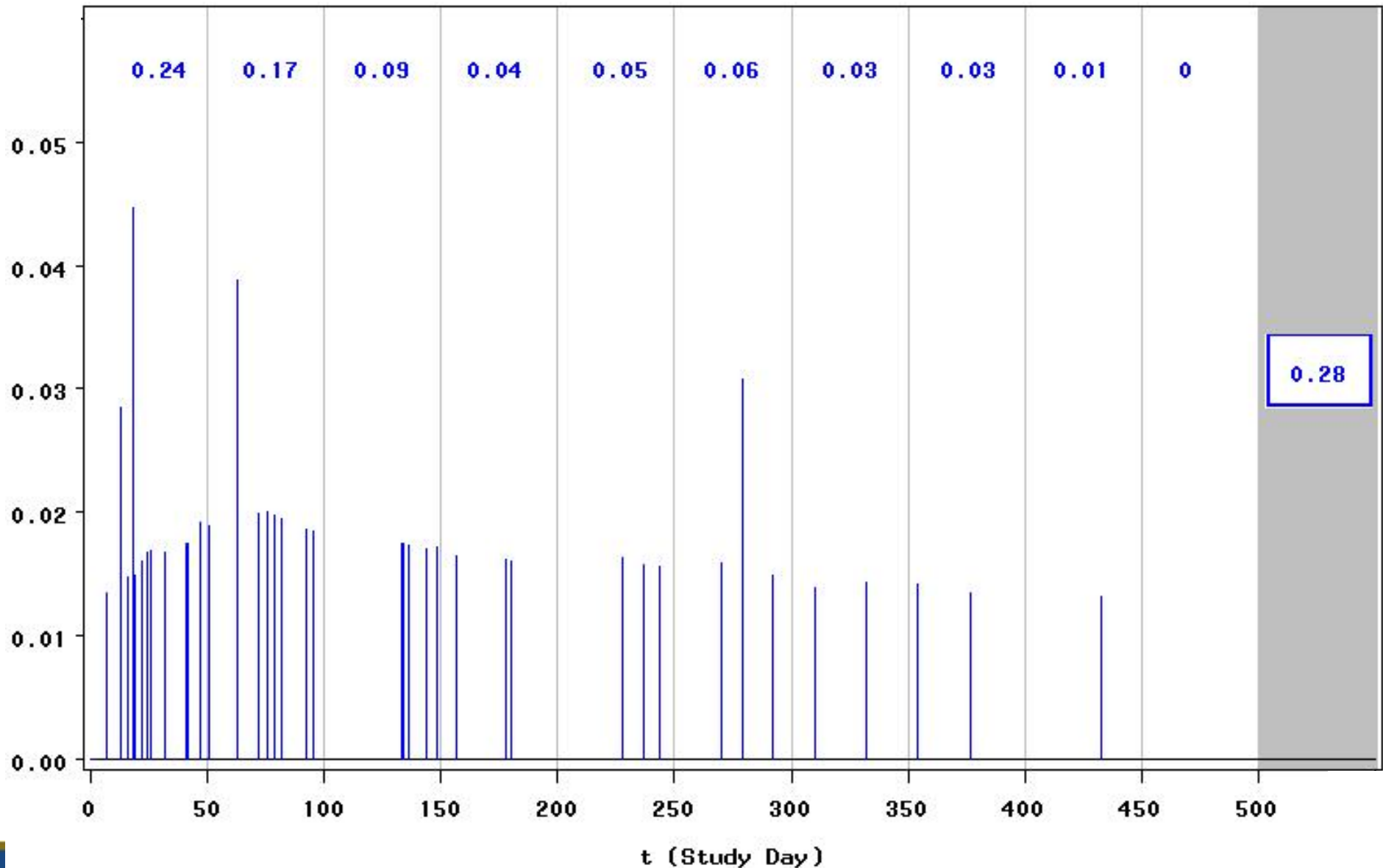
Density Function of T Given Prematurely Censored at day=4

$\Pr(T=t \mid C1=4)$ for $\alpha=-0.5$, $\tau=730$, $\text{trt}=2$



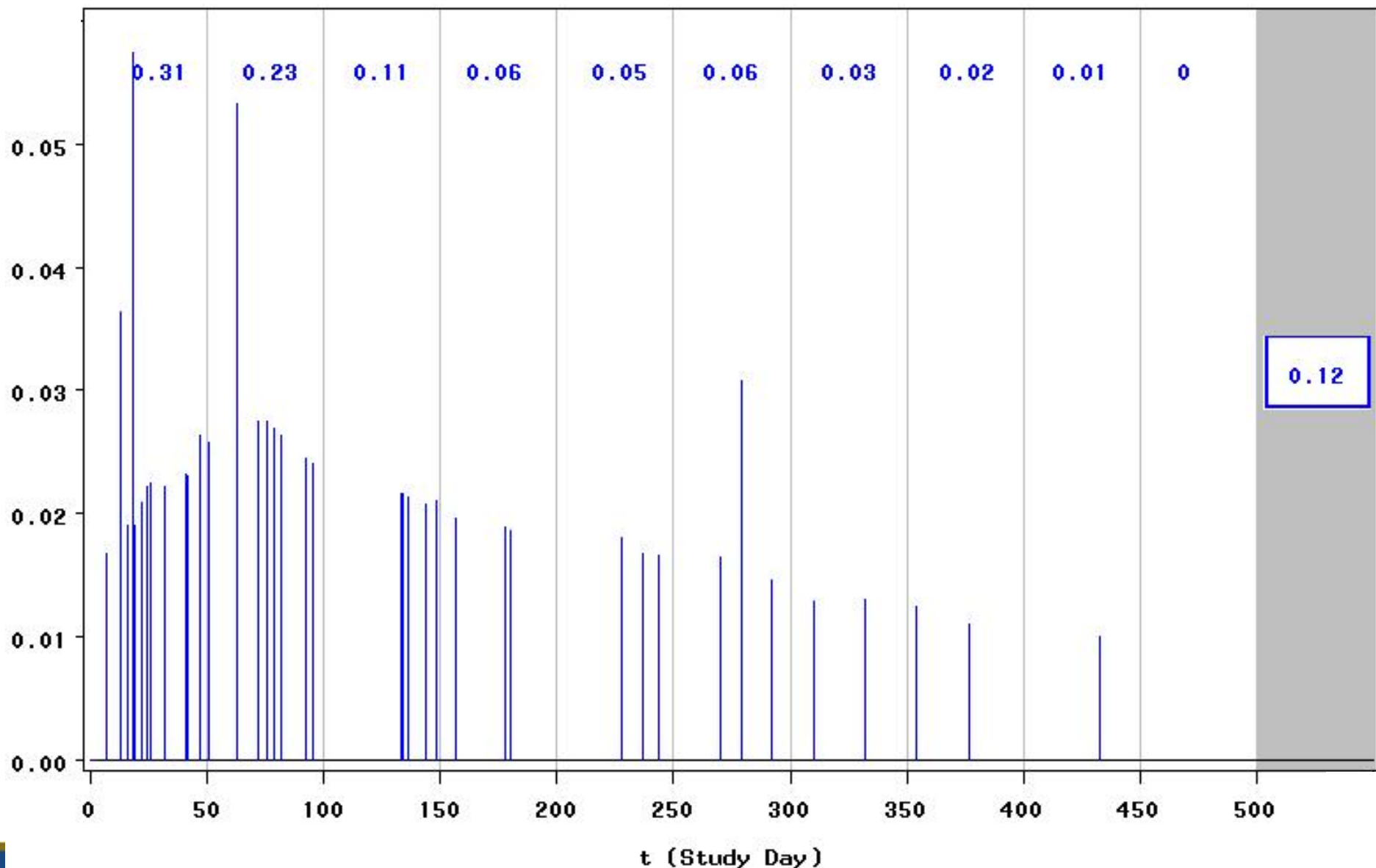
Density Function of T Given Prematurely Censored at day=4

$\Pr(T=t \mid C1=4)$ for $\alpha=-1.0$, $\tau=730$, $\text{trt}=2$



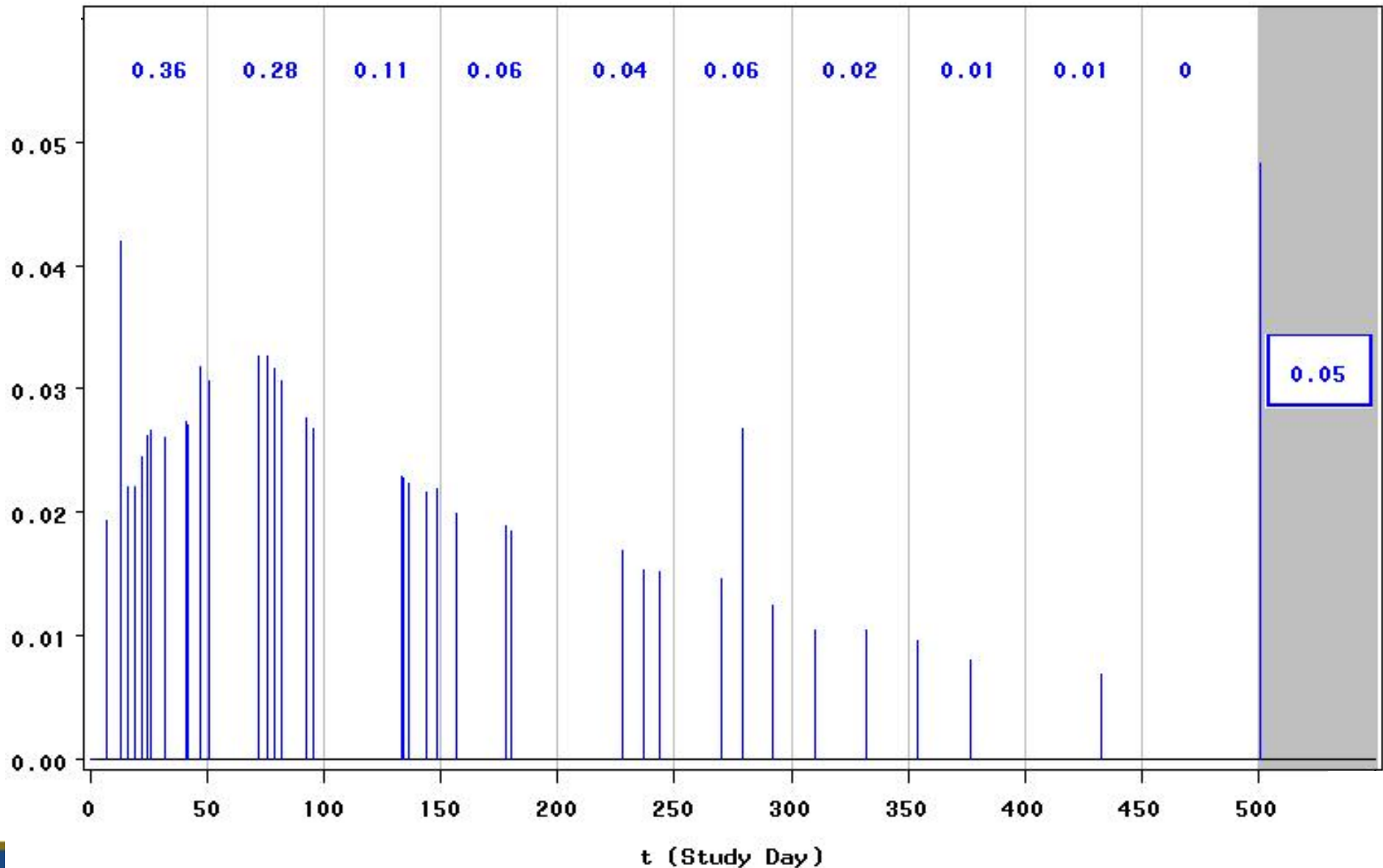
Density Function of T Given Prematurely Censored at day=4

$\Pr(T=t \mid C1=4)$ for $\alpha=-1.5$, $\tau=730$, $\text{trt}=2$



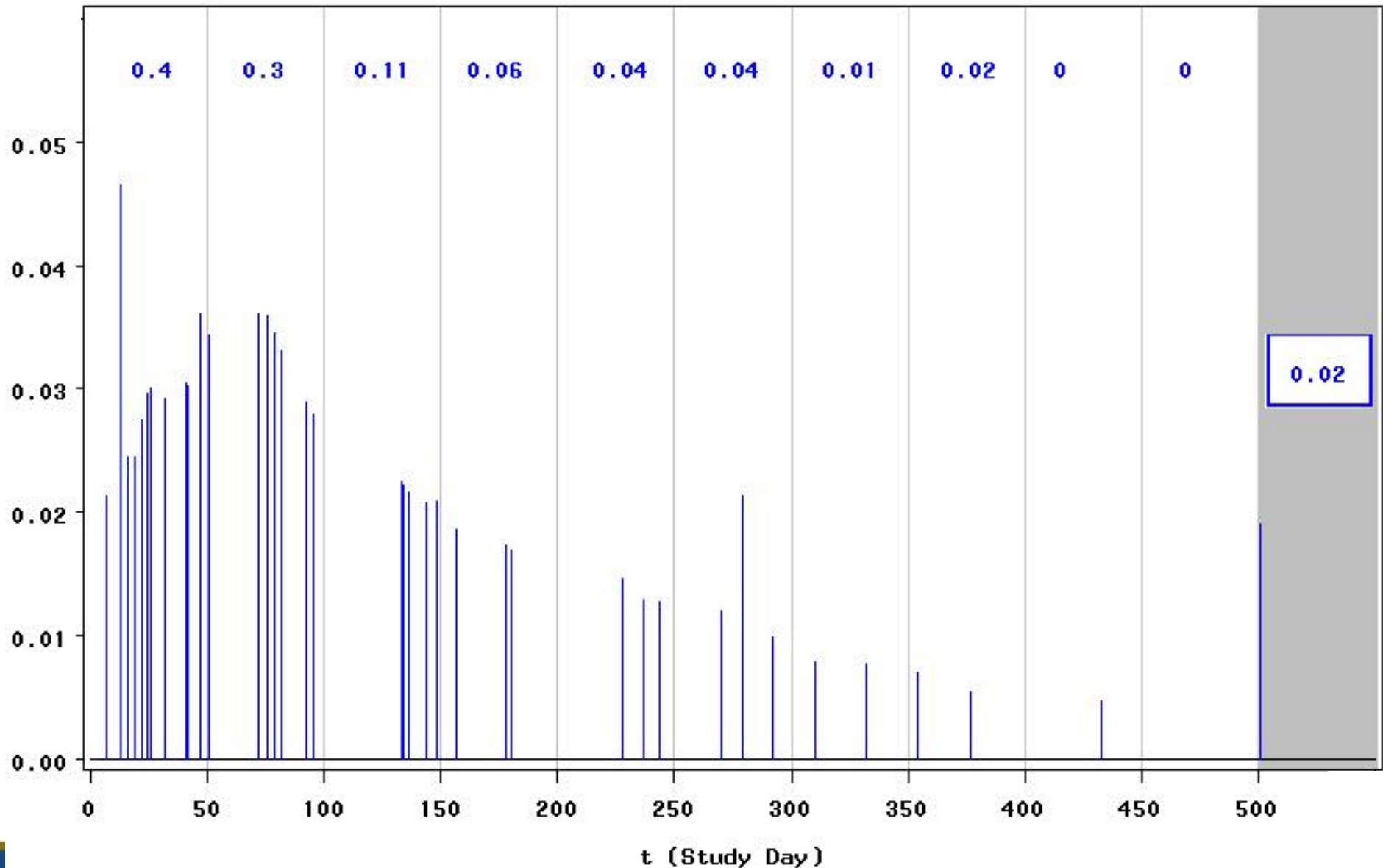
Density Function of T Given Prematurely Censored at day=4

$\Pr(T=t \mid C1=4)$ for $\alpha=-2.0$, $\tau=730$, $\text{trt}=2$



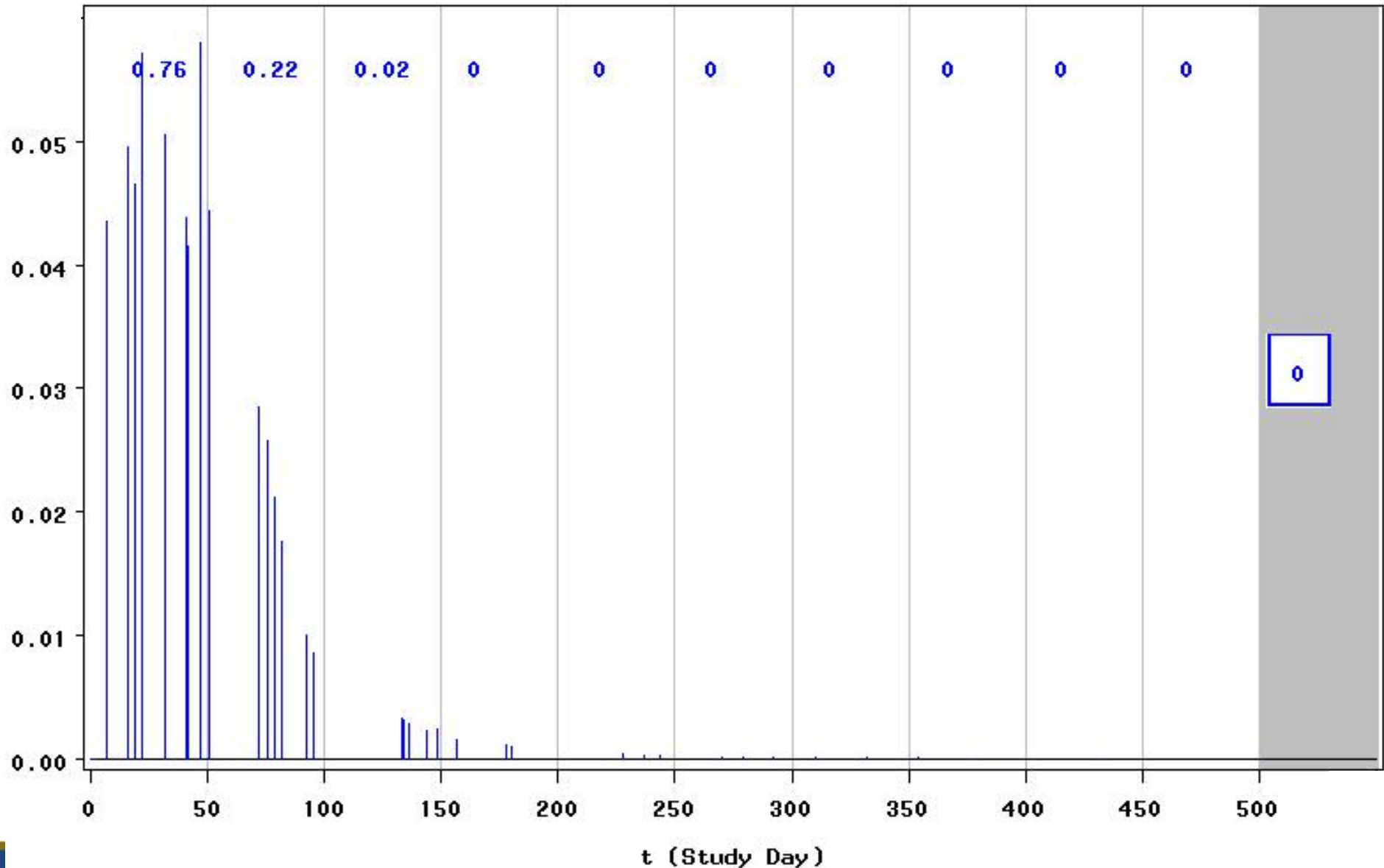
Density Function of T Given Prematurely Censored at day=4

$\Pr(T=t \mid C1=4)$ for $\alpha=-2.5$, $\tau=730$, $\text{trt}=2$



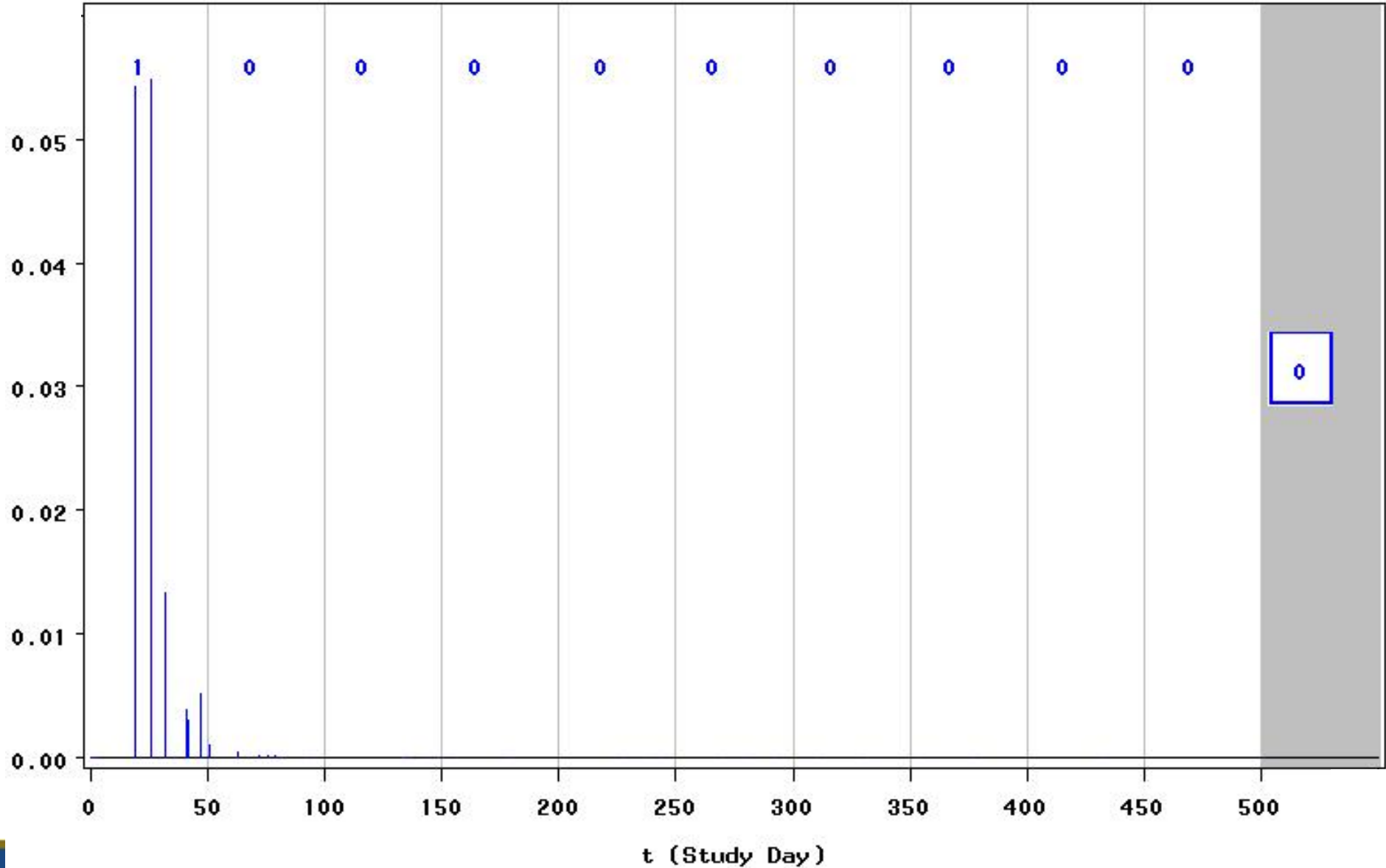
Density Function of T Given Prematurely Censored at day=4

$\Pr(T=t \mid C1=4)$ for $\alpha=-10$, $\tau=730$, $\text{trt}=2$



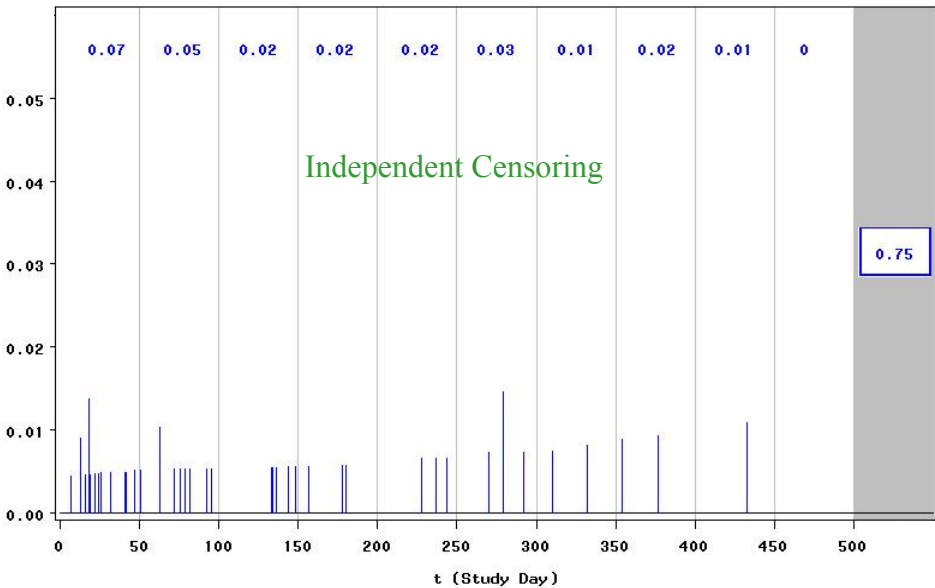
Density Function of T Given Prematurely Censored at day=4

$\Pr(T=t \mid C1=4)$ for $\alpha=-50, \tau=730, \text{trt}=2$

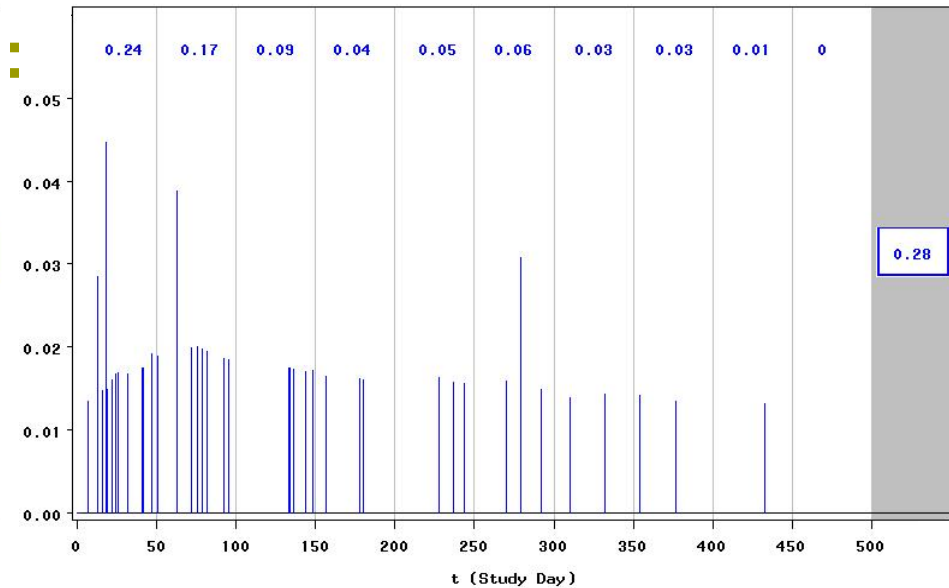


Pr(T=t | High Dose, C₁=4 days)

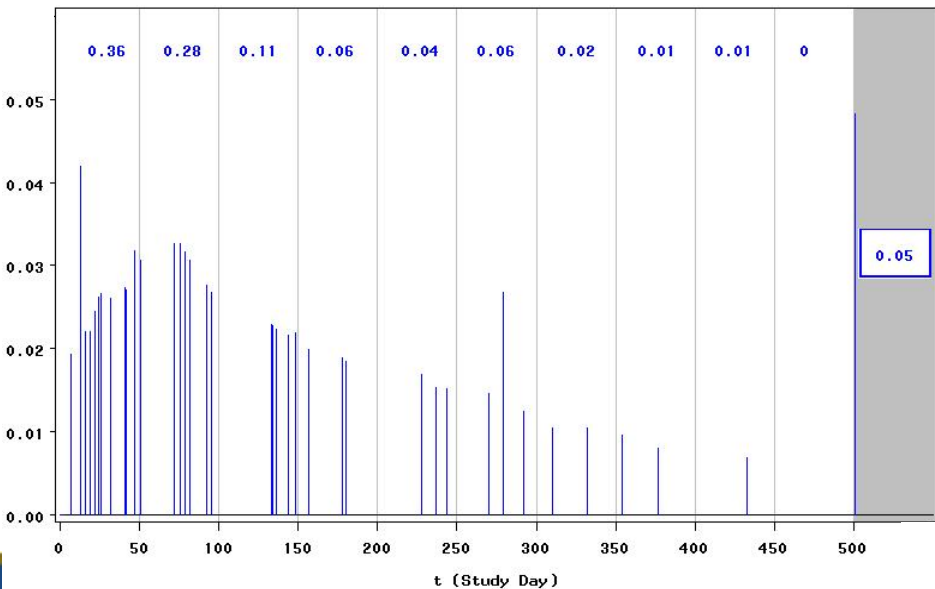
$\alpha = 0, \tau = 730$



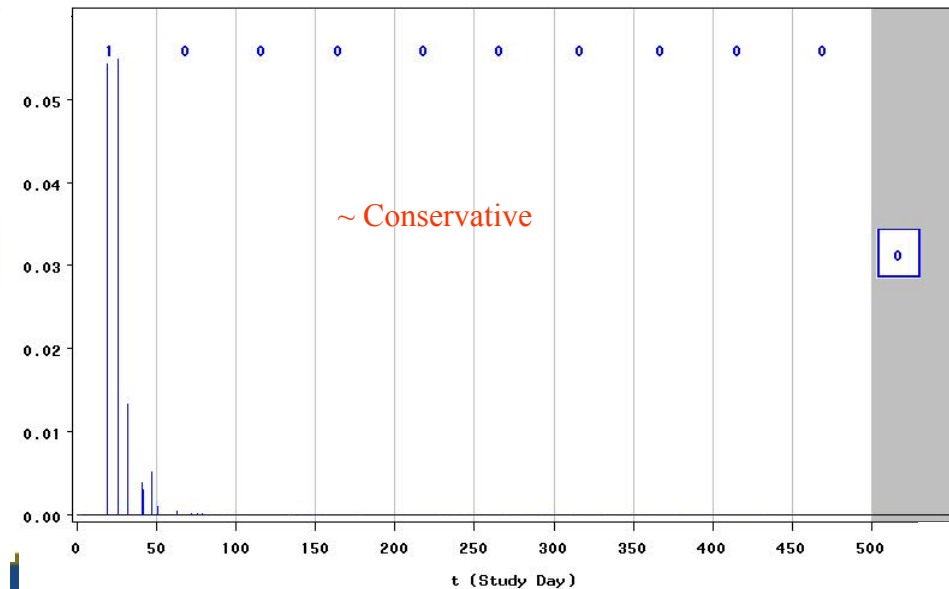
$\alpha = -1, \tau = 730$



$\alpha = -2, \tau = 730$



$\alpha = -50, \tau = 730$



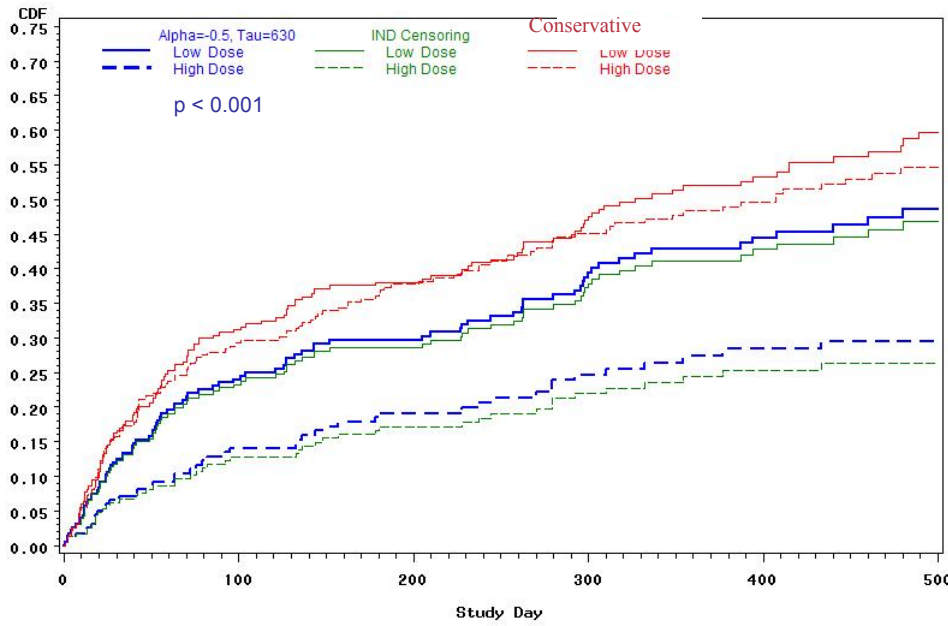
Sensitivity to Variations in Tau

Table 5: P-values for Comparison Between Treatments
Based on the Censoring Bias Function Method

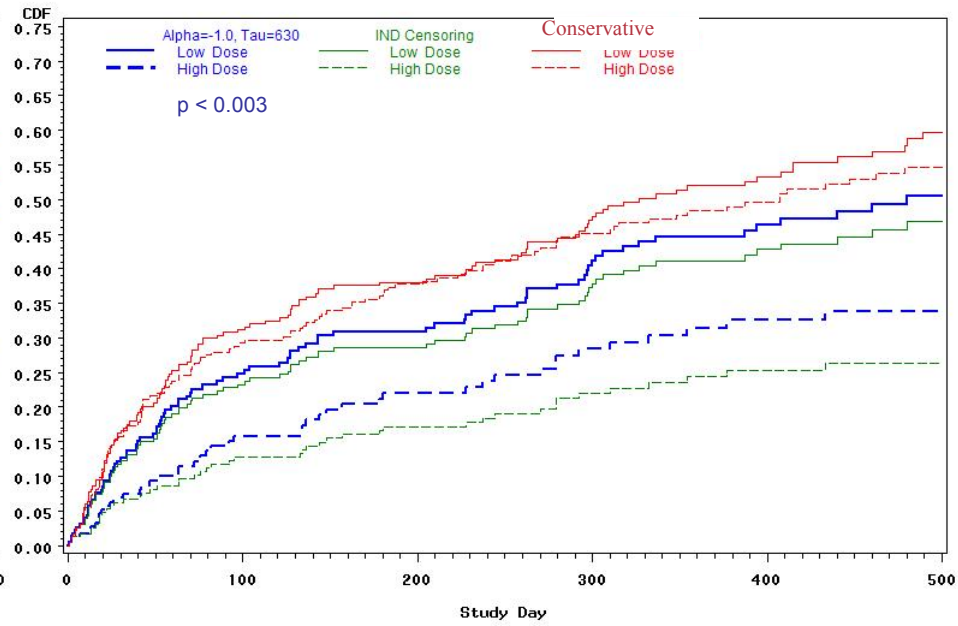
	α					
	0	-0.5	-1	-1.5	-2	-2.5
$\tau = 630$	<0.001	<0.001	<0.003	0.009	0.028	0.053
$\tau = 730$	<0.001	<0.001	<0.004	0.018	0.043	0.079
$\tau = 830$	<0.001	<0.001	<0.006	0.027	0.065	0.102



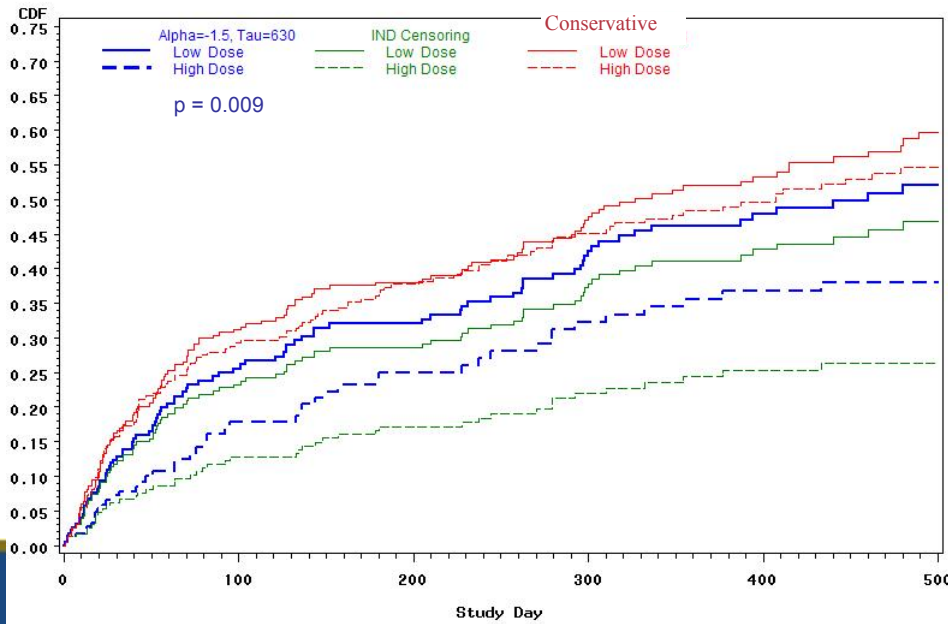
Epilepsy Event CDFs: Alpha=-0.5, Tau=630



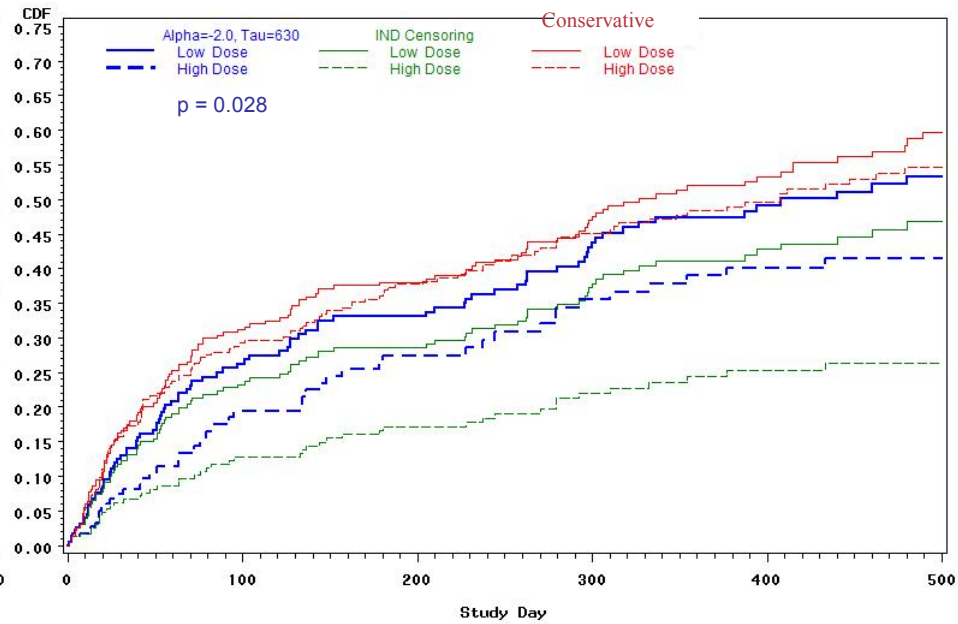
Epilepsy Event CDFs: Alpha=-1.0, Tau=630



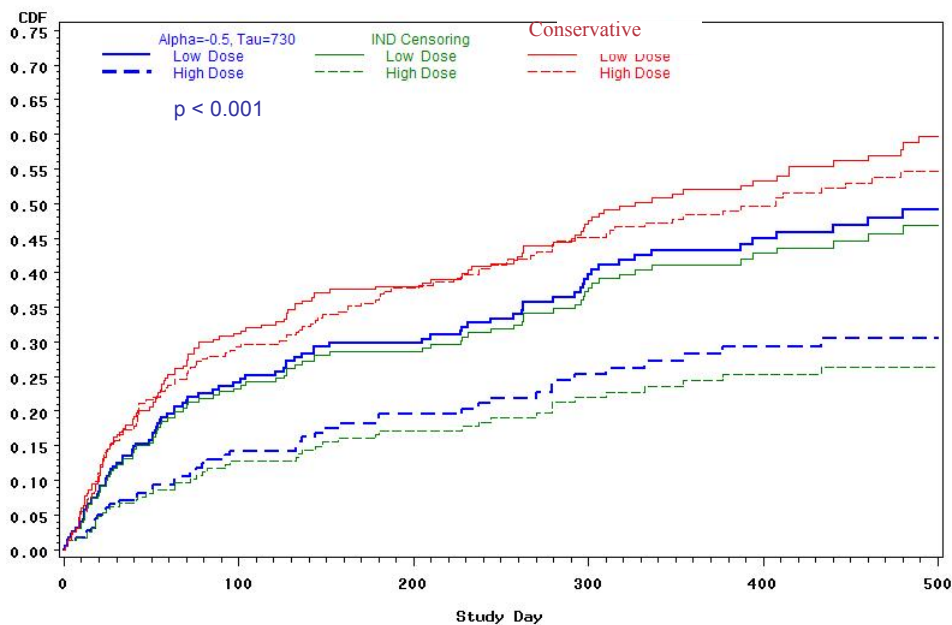
Epilepsy Event CDFs: Alpha=-1.5, Tau=630



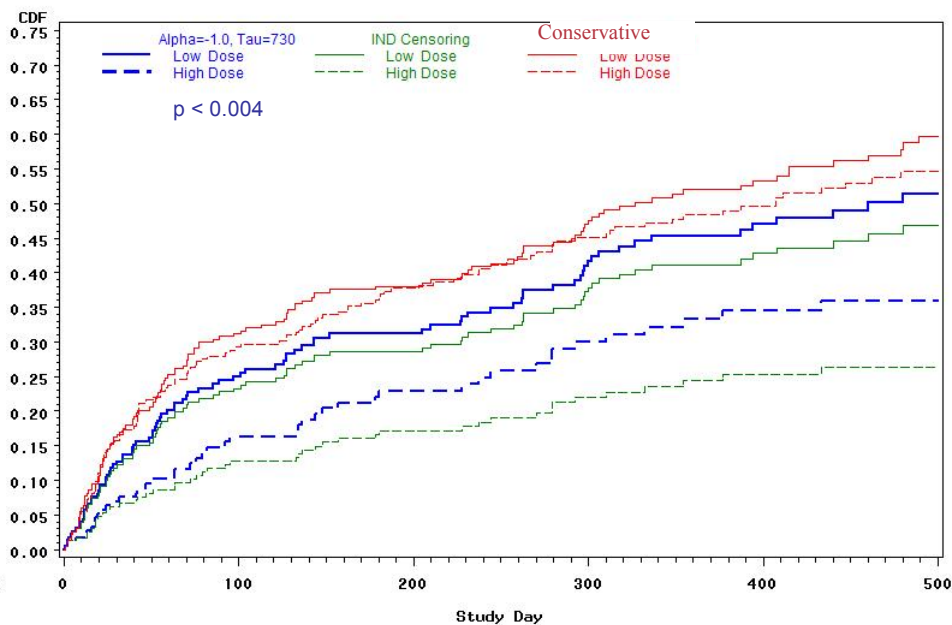
Epilepsy Event CDFs: Alpha=-2.0, Tau=630



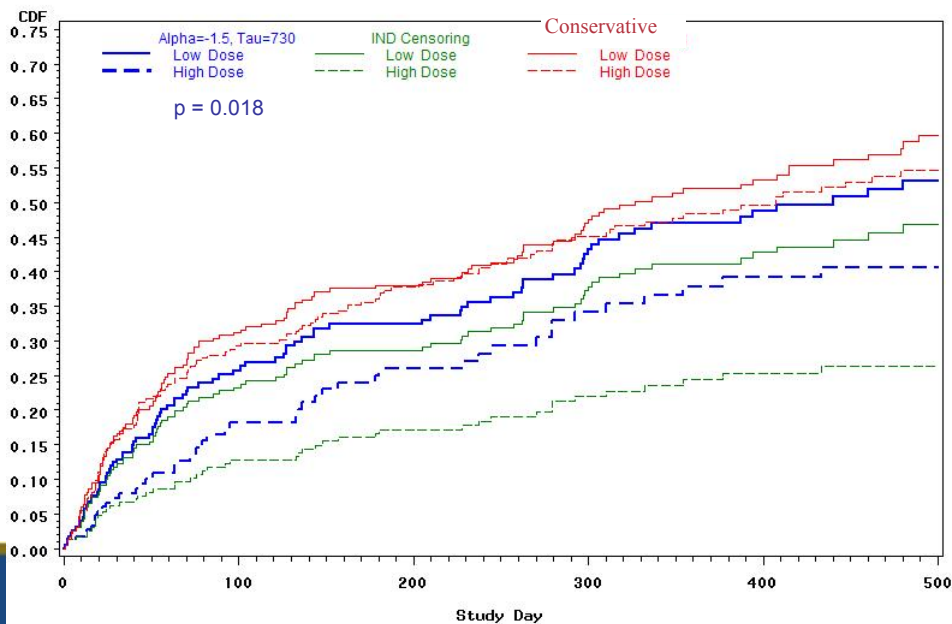
Epilepsy Event CDFs: Alpha=-0.5, Tau=730



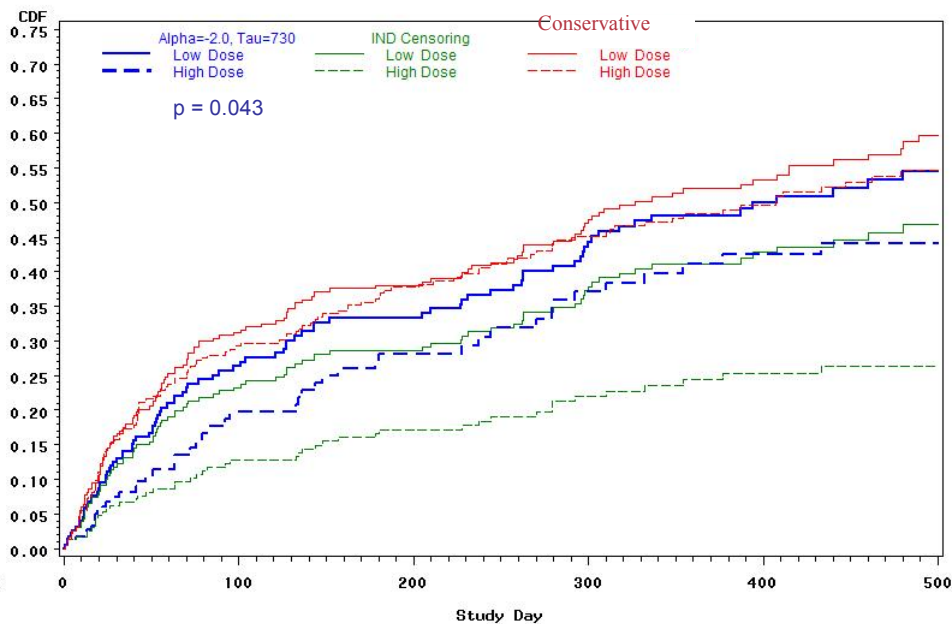
Epilepsy Event CDFs: Alpha=-1.0, Tau=730



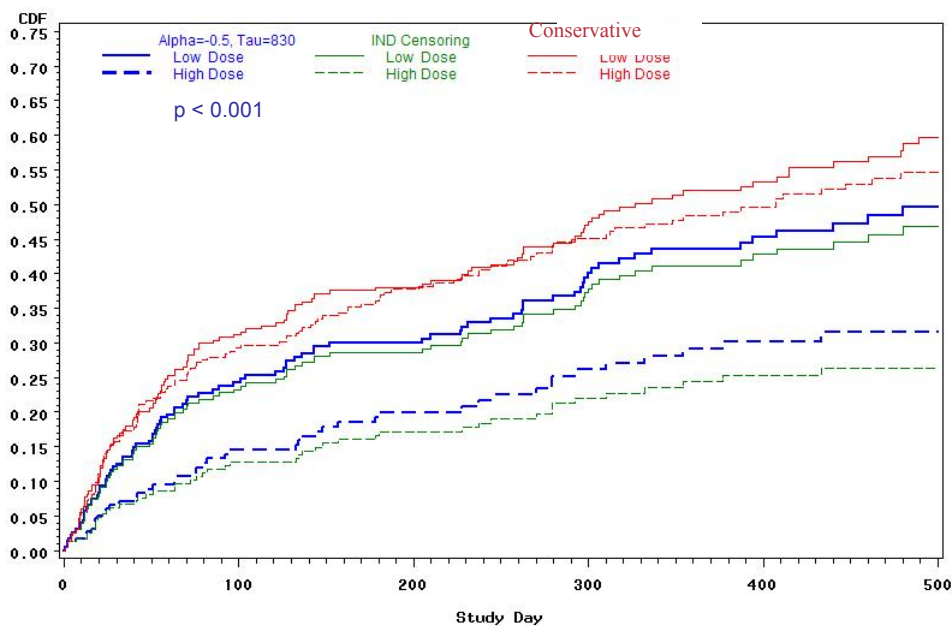
Epilepsy Event CDFs: Alpha=-1.5, Tau=730



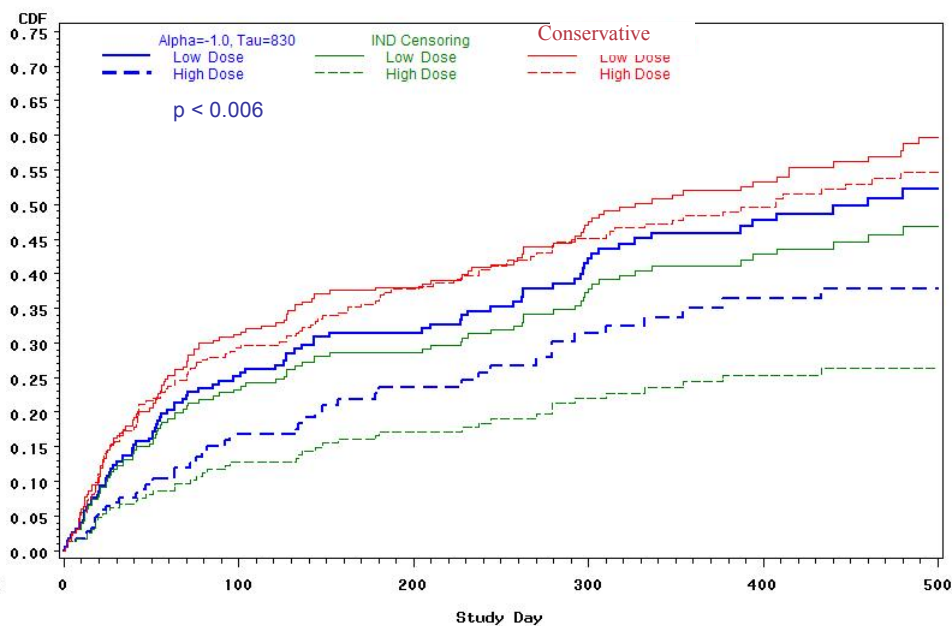
Epilepsy Event CDFs: Alpha=-2.0, Tau=730



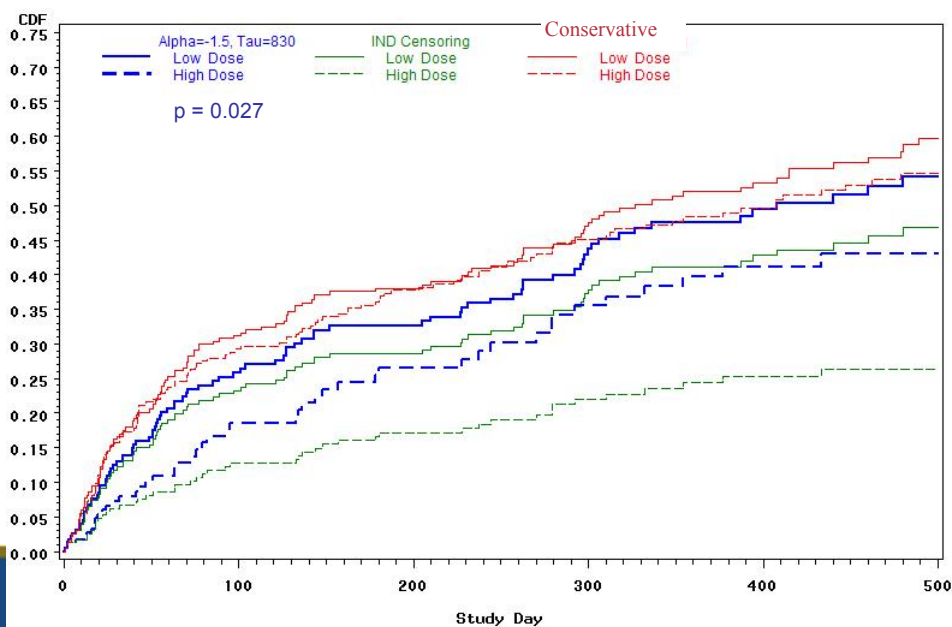
Epilepsy Event CDFs: Alpha=-0.5, Tau=830



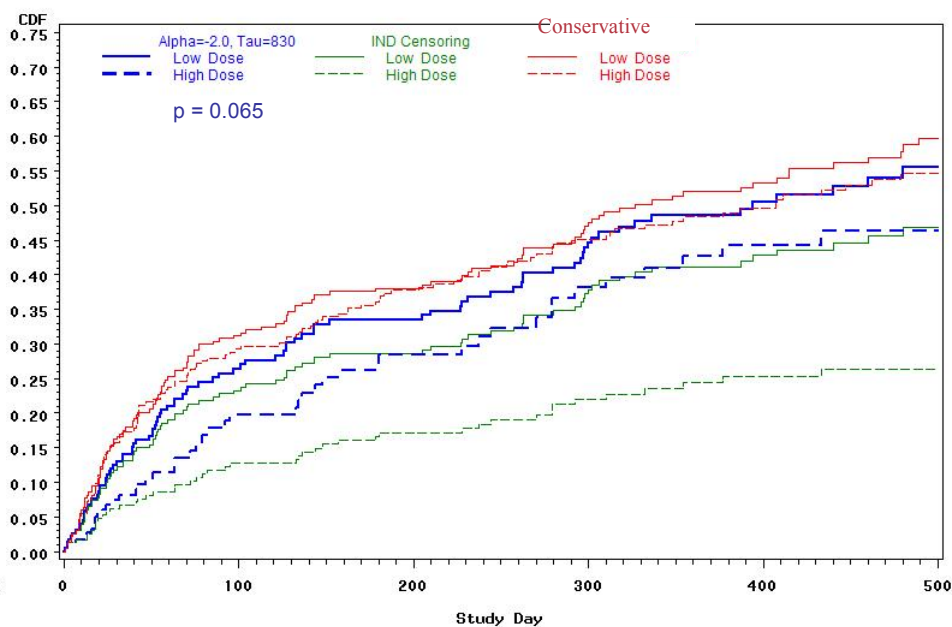
Epilepsy Event CDFs: Alpha=-1.0, Tau=830



Epilepsy Event CDFs: Alpha=-1.5, Tau=830



Epilepsy Event CDFs: Alpha=-2.0, Tau=830



Conclusion

- ❑ Results appear to be robust to deviations from non-informative censoring.
- ❑ But, is it
 - ❑ scientifically convincing?
 - ❑ convincing for regulatory purposes?



Estimation of Sensitivity Analysis Parameters: Topiramate Study 105



Topiramate Study 105

- ❑ In this study, there were 140 subjects who were treated with topiramate and had follow-up data after stopping their double-blind medication.
- ❑ Almost all these subjects stopped their medication prematurely due to adverse events, lack of efficacy, or other reasons.



Topriamate Study105

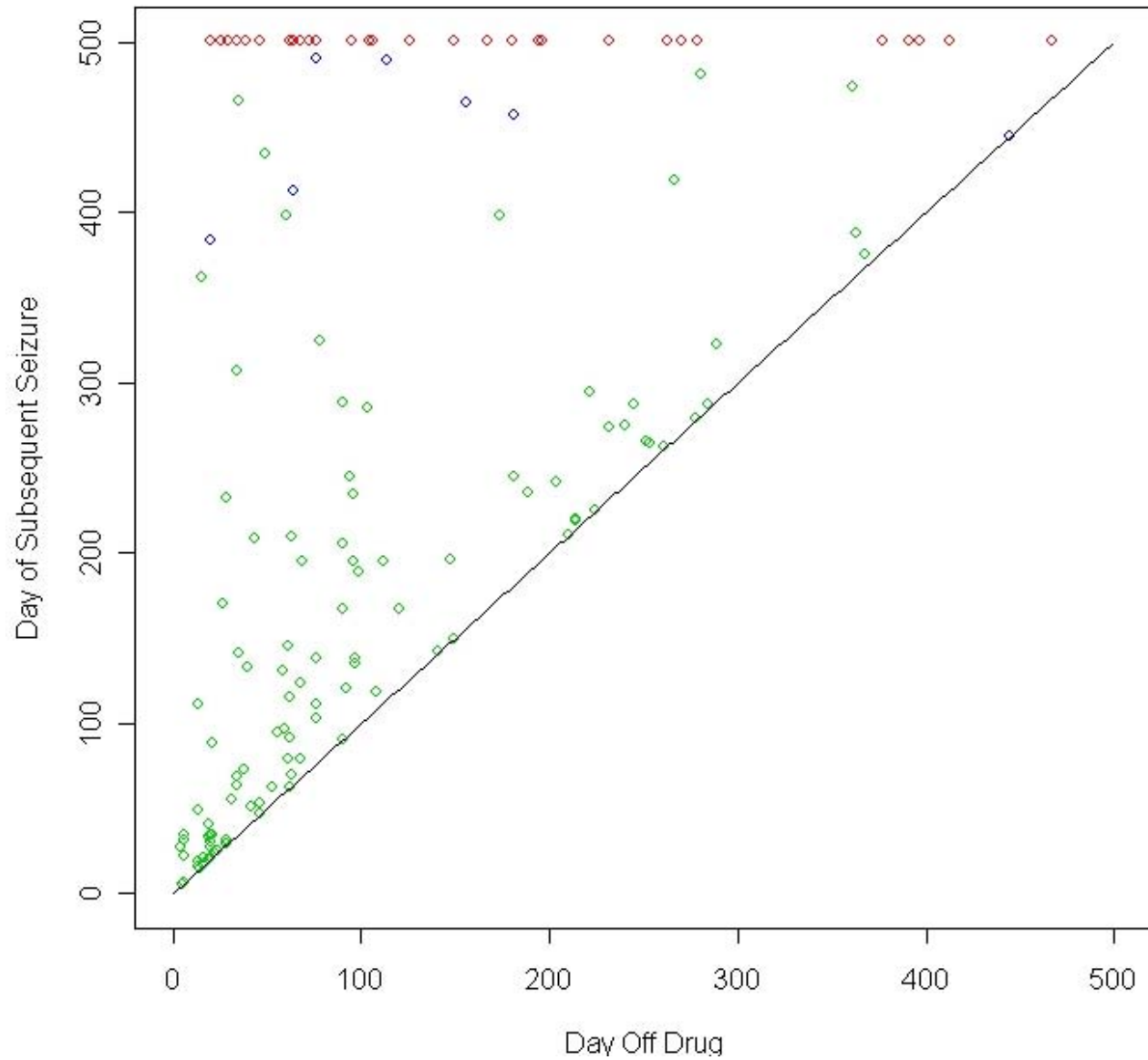
- ❑ Subjects in Study 105 who would have been considered as withdrawals in Study 106 form a population with more frequent seizures than the premature withdrawals in Study 106.
- ❑ One would reasonably expect that the time to seizure after stopping topiramate is shorter in Study 105 as compared to Study 106.
- ❑ Thus, the dependence between treatment discontinuation and subsequent seizure should be greater than the unknown truth in Study 106.



Analysis

- We used Day 500 as the analysis day.
- For subjects who prematurely withdrew prior to the minimum of time of next seizure and 500 days, we conservatively assumed that their subsequent seizure time was on the day of premature withdrawal.





Notation

C_1^\dagger = day of treatment discontinuation.

T^\dagger = minimum of day of next seizure and $500 + \epsilon$

C_2^\dagger = minimum of day of administrative censoring and $500 + \epsilon$

$X^\dagger = \min(C_2^\dagger, T^\dagger)$

$\Delta^\dagger = I(T^\dagger \leq C^\dagger)$

The observed data for an individual is $(C_1^\dagger, X^\dagger, \Delta^\dagger)$.



Model

We specify a hazard model for C_1^\dagger given T^\dagger of the following form:

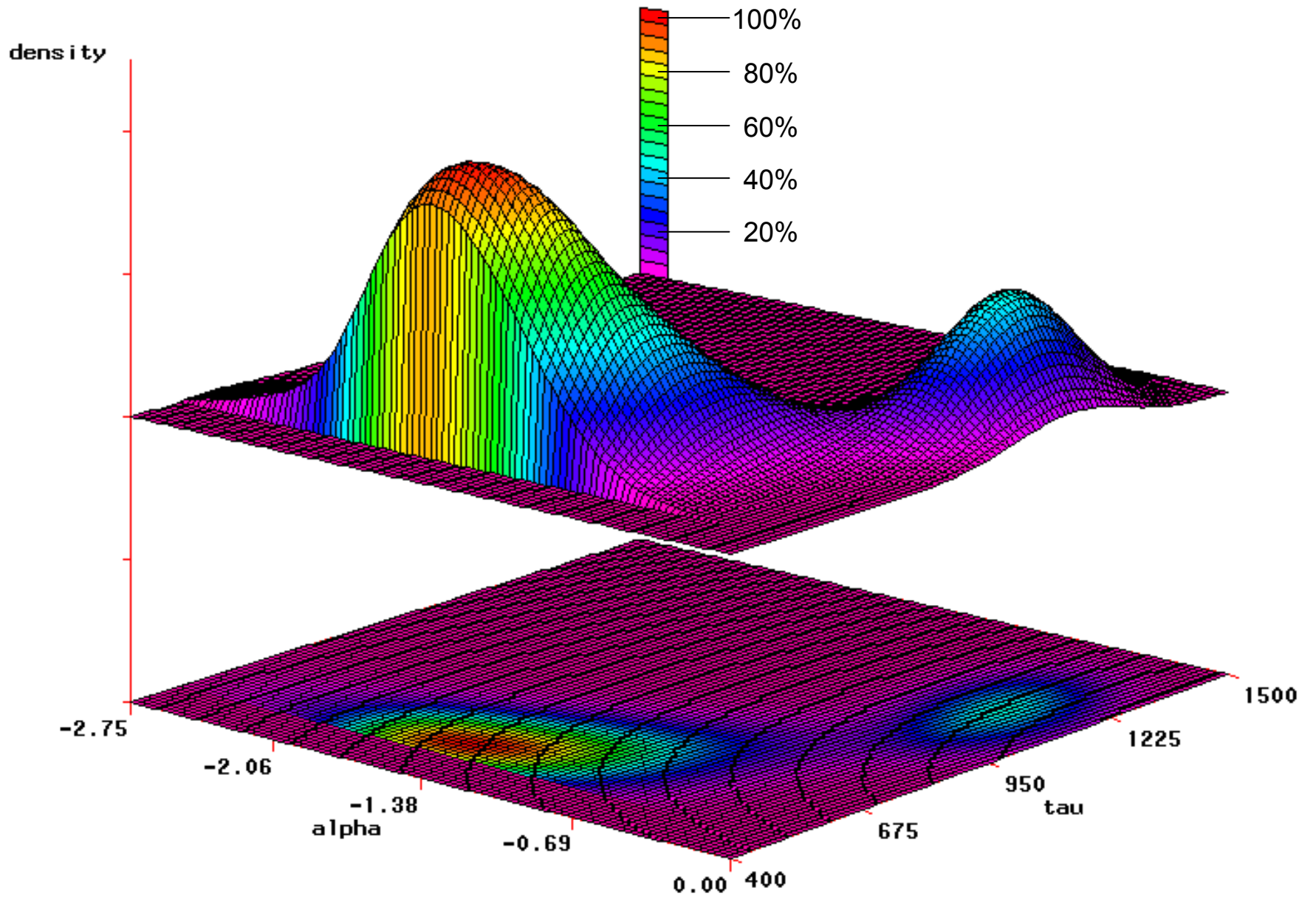
$$\lambda_{C_1^\dagger}(t|T^\dagger) = \lambda_0(t) \exp\{q(t, T^\dagger; \beta^\dagger)\} \quad t < T^\dagger$$

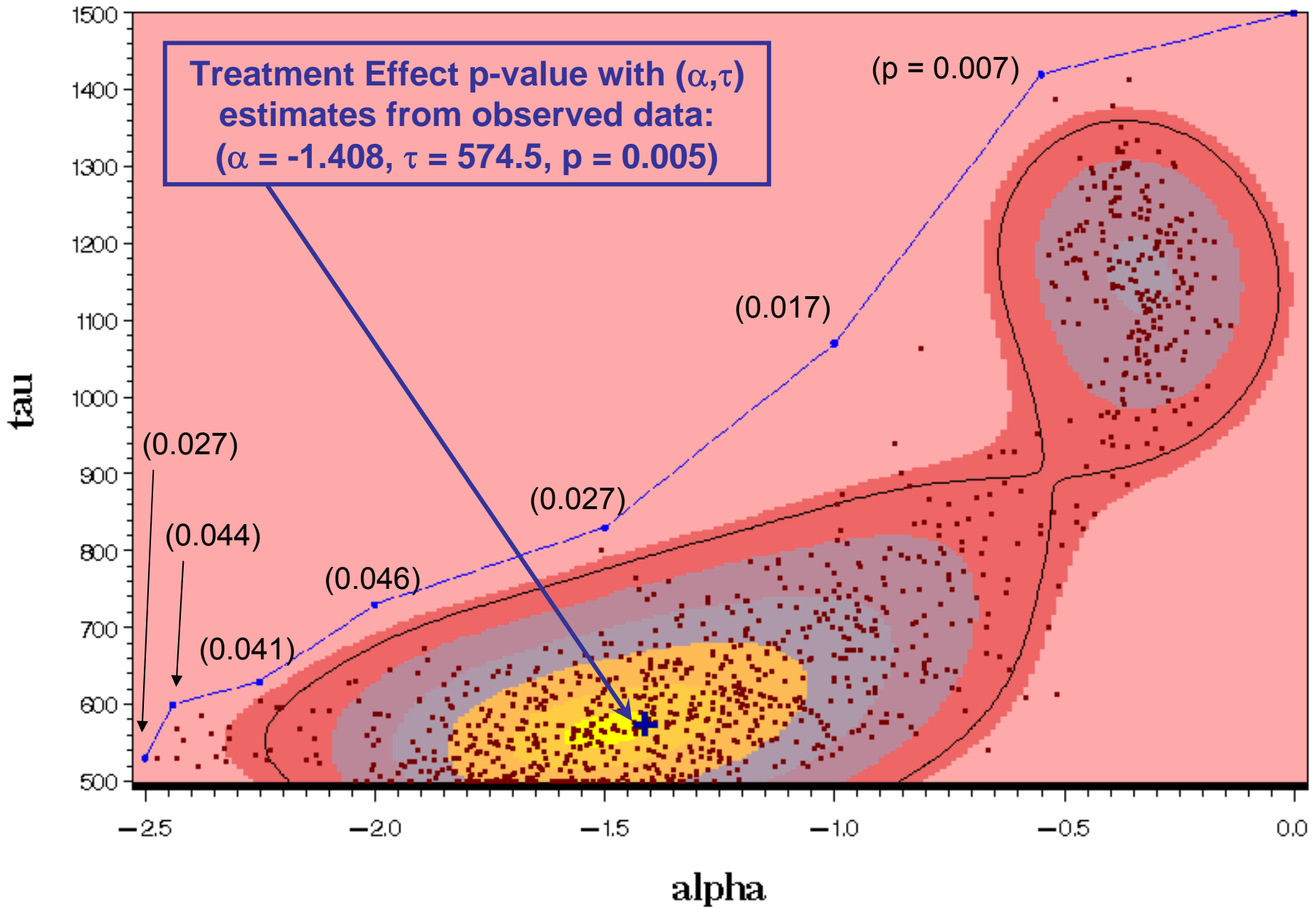
where $q(t, T^\dagger; \beta)$ is defined as above.



$$\text{Alpha} = -1.408$$
$$\text{Tau} = 574.5$$







Conclusion

- The estimate of alpha and tau and the associated confidence region is highly conservative.
- This result demonstrates that the original primary analysis of Toprimate Study106 is robust to potential informative censoring.



Lessons Learned

- ❑ Collect outcome data on subjects even after treatment termination.
- ❑ Collect factors that are prognostic for premature censoring and outcome.
- ❑ Make use of other data sources.



Clinical Trial Registration - ICMJE

- ❑ “In return for the altruism and trust that makes clinical research possible, the research enterprise has an obligation to conduct research ethically and to report it **honestly**.”
- ❑ “Registration is only part of the means to an end; that end is **full transparency** with respect to the performance and reporting of clinical trials.”
- ❑ In my view, the evaluation of the robustness of trial conclusions is an integral part of honest and transparent reporting.



From Theory to Practice

- ❑ We have well-developed sensitivity analysis methodologies for analyzing time-to-event and longitudinal data.
- ❑ We have a team to work collaboratively with pharmaceutical companies and the FDA to implement this methodology.
- ❑ We are partnering with software development experts to develop a suite of software tools.

