Survival Analysis of SCE/CPUC CFL Lab Study

Brett Close

June 4, 2015

Submitted to Miriam Fischlein
Southern California Edison

0 Executive Summary

Compact fluorescent lamps have provided a large portion of the savings delivered through California's energy efficiency programs for over a decade. CFLs, as they are commonly known, owe this privileged place within the portfolios of California's Investor-Owned Utilities (IOUs) primarily to three factors: first, their large improvement in efficiency compared to the incandescent base technology; second, to their relatively low price premium over the base technology and compared to other efficient technologies; and third, their vast improvement in life span compared to the base technology. This report focuses on the third characteristic: the life of the CFL.

The report details the methods and results of an analysis of a CFL Laboratory Study conducted on behalf of Southern California Edison and the California Public Utilities Commission Energy Division to test the life times of CFLs with different characteristics under different usage profiles. The purpose is to understand how those lamp characteristics and usage profiles affect lamp life, and produce estimates of lamp life for the lamps provided by SCE through its programs. This analysis is necessary because medians taken directly from the Laboratory Study results do not provide a meaningful measure of the life of program lamps. They do, on the other hand, provide a powerful set of information from which careful modeling can provide consistent, efficient, and valid estimates of lamp technical life.

The analysis models the conditional median life of lamps using a Weibull Accelerated Failure Time model with an instrumental variables approach. In the first stage, it uses lamp characteristics that directly affect lamp life and ones that affect lamp life only through their influence on cycling robustness as instruments in a linear survival model of lamp cycles. Then, in a second stage it uses the predicted number of cycles and the direct covariates in a Weibull model to explain lamp life. Marginal effects of characteristics are calculated using finite differences approximations. Due to the nonlinear and two-stage model, confidence intervals for marginal effects and lamp lives are based on a Wild Bootstrap.

Various measures indicate the model fits the data well. Particularly, the overall median predication for the sample matches with an error of 1.7%. The median of a synthetic sample survival curve is within 10%, and provides a conservative estimate. In both cases, the true values appear to be within the confidence bounds of the model.

The model is then used to make estimates of the median lamp life for lamps provided by SCE through its programs. These estimates are based on lamp characteristics and program volumes reported by SCE and an average cycle time from the 2005 KEMA metering study.

In general, estimates are larger for these predictions than for the values from the laboratory study because the sample for the study was quite different from the program lamp population. Specifically, there was a shift away from standard lamps into specialty lamps, which tend to have a longer life.

Through this intensive modeling process, the significant challenges of the nonlinearity of the response and the endogeneity of lamp life and cycles are successfully met in a way that provides a consistent and efficient estimate of lamp life.

Recommendations from this research are as follows:

- SCE should adopt the "mean median" lamp hours for the lamp types from Table 5, reproduced as Table 1, as planning estimates for lamp EULs.
- SCE should use the model results to estimate lamp lives for the program populations for newer populations, and use those values for planning purposes.
- The Energy Division should undertake an updated lighting metering study to gather newer data on average cycle times and average hours of use, and how they correlate with lamp conditions.

Table 1: Recommended Lamp Life by Lamp Category

Lamp Category	Recommended EUL
Basic Spiral	4,047
Specialty Shape	6,300
Specialty Controls	4,414
High Wattage	9,171

Results reproduced from Table 5

Contents

1 Introduction

Compact fluorescent lamps have provided a large portion of the savings delivered through California's energy efficiency programs for over a decade. CFLs, as they are commonly known, owe this privileged place within the portfolios of California's Investor-Owned Utilities (IOUs) primarily to three factors: first, their large improvement in efficiency compared to the incandescent base technology; second, to their relatively low price premium over the base technology and compared to other efficient technologies; and third, their vast improvement in life span compared to the base technology. This report focuses on the third characteristic: the life of the CFL. CFLs tend to be rated between 6,000 and 12,000 hours of use, approximately 6-12 times as long as incandescent lamps, with 10,000 hours being the most common. But empirical and anecdotal evidence suggests that the life of a lamp depends heavily on how it is used, and may be significantly lower than rated life.

This report details the survival analysis conducted on behalf of Southern California Edison (SCE) of data from the joint CFL Laboratory Study funded by the California Public Utilities Commission Energy Division and SCE. Why is this analysis necessary? Why not just rely on the empirical sample median from the laboratory study? The laboratory study was a very important data collection activity, but there is no reason to expect that a median life from that study directly represents a relevant median life for use by program administrators or regulators because the sample of lamps did not directly reflect the mixture of lamps included in utility programs, and the usage profiles did not directly represent the usage profiles of lamps used by IOU customers. Why can't the median from the laboratory sample simply be scaled by some factor? A cursory look at the data from this study as well as previous studies [1][2][3] clearly indicates that cycle time and the number of cycles experienced by a CFL have a non-linear effect on the technical life. As such, any method that aggregates the independent variables on the front end will yield a biased result. The median of a non-representative sample is not a consistent estimator of the median of the population. Furthermore, any analysis that relies on a median as an input will be inefficient because it throws out the majority of the information; a method that relies on a median from a non-symmetric distribution will also introduce bias. Biasedness, inconsistency, and inefficiency are generally the three least desirable characteristics for an estimator to have.

The report begins with a brief overview of the history of the study and a description of the relevant technical characteristics of CFLs. Section 2 describes the basic statistical theory of survival analysis and the methodologies implemented in this study. Section 3 describes the data used and descriptive results from the laboratory study. Section 4 presents results from the analysis in terms of the marginal effects of lamp characteristics and model fit to the data. Section 5 provides predictions of lamp life for lamps promoted through the SCE Upstream Lighting Program. That is, results in Section 3 are observational in nature, and are not based on modeling; results from Section 4 show the results of modeling on the laboratory sample; and results in Section 5 use the model to provide projections for lamps outside the laboratory sample. Section 6 provides conclu-

[1] Ji, Yunfen, Robert Davis, and Weihong Chen. "An Investigation of the Effect of Operating Cucles on the Life of Compact Fluorescent Lamps". 1998 IESNA Annual Conference, Paper No. 35.

[2] National Lighting Product Information Program (NLPIP). "Specifier Reports: Screwbase Compact Fluorescent Lamp Products". Volume 7, Number 1, June 1999.

[3] US EPA. "Durability Testing for ENERGY STAR® Residential Light Fixtures". April 30, 2003. Submitted by Lighting Research Center, Rensselaer Polytechnic Institute.

sions and Section 7 provides recommendations from the research.

1.1 History of the Study

SCE has been working with the CPUC Energy Division and its consultants on a study of CFL switching effects on technical survival since 2009, and data collection continued through 2014. Previous studies of CFL survival based on usage characteristics suffered from two primary faults. First, the samples used were small in the total number of lamps, the number of lamps in each category, and the range of usage conditions tested. The analysis of the data from these studies relied primarily on taking sample medians for each group. Additionally, lighting technology has changed significantly over the last 20 years, and drawing conclusions about current products based on older data is unlikely to be valid. The study addresses the limitation of the previous work in three ways, including 1) collection and use of data from recent CFL population promoted by the statewide IOU programs, 2) using a large and diverse sample of lamps subjected to a wide range of operating characteristics, and 3) implementing improved data analysis models given the improved level of data collected.

1.2 Brief Technical Overview of CFLs

Compact Fluorescent Lamps emit light by exciting mercury gas with an electrical current. This is the source of both their improved efficiency and increased durability compared to incandescent lamps. Incandescent lamps operate by passing current through a metal filament to heat it sufficiently to glow with enough brightness to provide illumination. This means that the majority of energy is used to produce radiation outside the visible spectrum (which is why incandescent lamps are so much hotter than CFLs), and that an incandescent lamp fails when the filament breaks. CFLs on the other hand produce a narrow spectrum of light, which is then absorbed by phosphors in the coating on the lamp's tube and re-emitted as a spectrum designed to have some set of desirable qualities.

In order to create the current flow through the mercury gas, CFLs need to provide sufficient current and voltage. This leads to the two primary failure modes for CFLs. The first is that the electrodes in the tubes can fail, the second is that the ballast can fail. Because common CFLs have the tube integrated with the ballast, the failure of either leads to the failure of the whole lamp, in contrast to linear fluorescents in which the lamp and ballast tend to be separate.

2 Methodology

The fundamental question when considering the effective useful life of an energy efficient measure in a California energy efficiency program is what is the conditional median survival time of the measure. We are interested in the median because CPUC policy dictates the median as the definition of the life of a measure, but on a deeper level this makes sense because the distribution of measure lives is often non-symmetric and with a long tail. That means that a small number of individual measures may last much longer, making the estimation of the mean difficult in practice because the mean relies on the value of every member of a population (so data have to be collected until all the measures have failed). The median, on the other hand, is the time at which half the lamps have failed. So, in principle, the median can be easier to estimate because it does not require waiting until all the lamps have failed.

We care about the conditional median because we want to know how measures with different characteristics perform. These characteristics include the technical characteristics of any equipment involved (e.g. the wattage and other characteristics of the lamp), as well as the usage characteristics (e.g. where, how, and how often it is used). The conditional median is more difficult to estimate than the unconditional median because it requires more modeling, but it allows us to project the conditional median of measures outside the sample in a way that would be impossible for the unconditional sample median. That is, based on our sample and modeling we could predict the median life of a measure that was not included in the sample of data, either in terms of lamp characteristics or usage characteristics, but there is no reason to expect that the median life of a lamp outside the sample would be the same as a lamp in the sample because the sample is not representative of real-world conditions. That is, the conditional median provides external validity, whereas the unconditional median lacks external validity.

The final piece of the fundamental goal is the survival time. Data about lamp failure have a peculiar structure (discussed more in-depth below) that requires special analytical techniques. The traditional analytical approach for questions of "time to event" (i.e. the time until a lamp fails) is survival analysis. Survival analysis is a set of tools for duration data that incorporates both time-to-failure information as well as censoring time for study participants that do not fail (or have not yet failed).

Thus, the analysis is based on modeling of lamp life and then using the model to estimate the median for relevant measure characteristics. To be clear, this estimate will not produce the true effective useful life of a lamp. The true EUL is based not solely on the technical life of the lamp, but on how long the savings of the measure persist. That is, it is the time until the lamp fails, or is removed because it has become too dim, the tenants of the space want a newer technology, it is broken by an errant football thrown in the living room, or ceases to provide energy savings for any other reason. As such, the median technical lamp life provides an upper bound that may reasonably approximate the true effective useful life for most circumstances.

2.1 Overview of Survival Analysis

Survival analysis is a well-developed set of analytical tools for dealing with time-to-event data based on the special structure of this type of data. This subsection will describe the nature of the data and then explain the basic characteristics of techniques involved in survival analysis.

2.1.1 Time-to-Event Data

Consider a CFL burning in a socket. In principle it could fail at any time. Thus it is useful to think of the lamp facing a hazard of failure at any point during its operation, but once it has failed it cannot fail in the future because it is removed from operation. The hazard it faces at any moment is the likelihood of it failing at that moment, conditional on the fact that it has survived until that time. Notably, the hazard for a lamp type at a given time is different from the probability of failure for a lamp type at that time. If we think of this issue in terms of the distribution of failure times for a population of similar lamps, we can rely on the cumulative distribution function, $F(t)$[4] as the chance that the lamp will fail at a time less than equal to t. That is,

$$F(t) \equiv Pr(T_i \leq t) = \int_0^{T_i} f(t)dt, \tag{1}$$

where T_i is the failure time and $f(t)$ is the probability density function of the failure time. Because every lamp fails only once, the survival function, $S(t)$ is

$$S(t) = 1 - F(t), \tag{2}$$

the likelihood of failing after time t. Then the hazard function, $h(t)$ is,

$$h(t) \equiv Pr(T_i = t | T_i \geq t) = \frac{Pr(T_i = t) Pr(T_i \geq t | T_i = t)}{Pr(T_i \geq t)} = \frac{Pr(T_i = t | T_i \geq t)}{Pr(T_i \geq t)} = \frac{f(t)}{S(t)}, \tag{3}$$

where the first equality after the equivalence is from Bayes' Theorem, and the second one is because $Pr(T_i \geq t | T_i = t) = 1$ by definition. Next, note that by differentiating $S(t)$ in Equation 2 with respect to t we find,

$$\frac{\partial S}{\partial t} = -f(t), \tag{4}$$

because the partial derivative of the cumulative distribution function is the probability density function. So given any of the four functions, $F(t)$, $f(t)$, $S(t)$, or $h(t)$ we can find any of the others.

An additional wrinkle in the data is that we often do not have the failure time for all members of the sample. Specifically, for any lamp that has not yet failed, we don't know its failure time, only that has survived at least until now. That is, we know its failure time is something greater than its current time, in this case called its censoring time. We then say that its failure time is censored, meaning that we believe it has a failure time, and we do not know what it is. Thus for the sample we have either a failure time or a censoring time.

2.1.2 Survival Analysis Methods

There are three categories of methods in survival analysis: non-parametric, semi-parametric, and parametric. The Kaplan-Meyer method, as non-parametric survival analysis is known, looks at the information for each study participant and constructs a survival curve and non-parametric survival

[4] In what follows, t is used to represent the time variable and T_i represents the failure or censoring time observed for an individual lamp.

function, where these are based on the observed failure times in the sample. The advantage of this method is that it does not rely on any assumptions about the distribution of the effects. The disadvantage is that it cannot incorporate any covariates as part of the analysis, except as sub-setting characteristics. As such, it is descriptive, but not useful for extrapolating directly to the population.

The primary semi-parametric methodology is the Cox Proportional Hazard Model, also known as the Cox Partial Likelihood Model. It incorporates a non-parametric baseline hazard condition with a parametric assumption about how the covariates (characteristics of the lamps and the test conditions) proportionally scale the underlying baseline hazard. It has the major advantages of requiring few assumptions about the distribution of failure times, and providing estimates of the marginal effects of the covariates on lamp life. The Cox Proportional Hazard Model is the most commonly used in the social sciences.

The final group is parametric models, most of which are in a class known as Accelerated Failure-Time (AFT) Models. These models rely on a regression-like framework to model the failure time of lamps based on covariates. They rely on the assumption that all lamps have the same basic survival function, but that covariates have the affect of accelerating or decelerating the effective time. That is, $S(t)$ (and thus $F(t)$, $f(t)$, and $h(t)$) has the same shape for each lamp, but the time scaling varies based on lamp characteristics and usage. This allows for specific estimation of the marginal effects of the covariates and predictions out of the sample. While it requires assumptions about the form of the distribution, fortunately there are very flexible distributions that make this requirement relatively innocuous compared with other assumptions required in evaluation of energy efficiency programs. And, as we'll see in Section 4, the Weibull model appears to fit the data very well.

Unfortunately, while lack of a parametric assumption for the hazard function in the Cox Proportional Hazard Model is desirable, the model relies heavily on the proportional hazard assumption as the model parameters are identified by comparing the characteristics of lamps remaining in the risk set at each time. For this study, a test of the proportional hazard assumption indicated it is very unlikely to hold for some variables. While the Weibull model used in the analysis enforces a proportional hazards assumption, identification comes from modeling the failure time directly, rather than conditioning on the proportional hazard. Additionally, accelerated failure time models are better suited to making out of sample predictions and more fully reflect information from censored members of the sample. Because the Weibull distribution appears to fit the data well, this parametric structure is a justified tradeoff compared to the proportional hazards assumption.

2.1.3 Accelerated Failure Time Models

Let's return to the basic framework developed in Section 2.1.1. For each lamp we have an associated failure time or censoring time, T_i. Then, based on whether the lamp time is a failure or a censoring time we can associate a new variable C_i defined as

$$C_i = \begin{cases} 1 & \text{if lamp i has failed} \\ 0 & \text{if lamp i has been censored} \end{cases} . \tag{5}$$

Then, assuming all lamp's failure times are independent (i.e. a lamp going out or staying on doesn't effect whether any other lamp fails or stays on), we can define the likelihood for the whole sample as,

$$\mathcal{L} = \prod_{i=1}^{N} [f(T_i)^{C_i}] * [S(T_i)^{1-C_i}], \tag{6}$$

where N is the total number of lamps, such that for each lamp we rely on one contribution to the likelihood because every lamp has either $C_i = 1$ and $1 - C_i = 0$ or $1 - C_i = 1$ and $C_i = 0$. Lamps that failed enter through the distribution directly and censored lamps enter through the survival function. Because the likelihood is monotonic, it is maximized by the same value that maximizes its log. So we define the log-likelihood, L as,

$$L = ln(\mathcal{L}) = \sum_{i=1}^{N} \big(C_i * ln(f(T_i)) + (1 - C_i) * ln(S(T_i))\big). \tag{7}$$

Then, given a flexible parametric specification for $f(T_i)$ (and thus $S(T_i)$), we can estimate the effects of the covariates on lamp survival by maximizing the log likelihood, and thus the likelihood of observing the sample that we did in fact observe.

In an accelerated failure time model, the form for the hazard rate is given for each lamp, and varies among lamps with different characteristics. One of the most common practices is to specify a Weibull distribution for the life of the lamp, leading to a hazard of the form,

$$h(t|X_i) = exp(X_i'\beta)\alpha t^{\alpha-1}, \tag{8}$$

where X_i is a vector of covariates (including a constant term), α is the shape parameter of the distribution and $exp(X'\beta)$ is the scale parameter. As we saw above, given a hazard rate, we can determine the density function and survival function and thus calculate the log likelihood and estimate of β by maximum likelihood.[5]

Note that, as was mentioned previously, the effect of the covariates is to change the hazard rate between lamps, keep it constant for lamps with the same characteristics at a given time, and let it vary over time. Two reasons for this specification are its relative flexibility, and that it is positive for all values of X_i and β (as is necessary for survival times). Since the exponential model is nested in the Weibull model (with $\alpha = 1$), the restriction of the model to the exponential model can be tested directly. Other distributional assumptions, such as log-logistic and log-normal, are used at times, but the Weibull model was selected here for its flexibility and the fact that it allows the hazard to rise over time, as we expect for lamps that undergo stress during operation. It also appears to fit the data in this study quite well on average.

2.1.4 Coefficients, Marginal Effects and Standard Errors

As is common in non-linear models (and different from linear models), the marginal effect of a change in a covariate is not equal to the coefficient parameter. Instead, the coefficient parameter

[5] The model can also be thought of in a regression-like setting with a model of the form, $\alpha ln(T_i) = -X_i'\beta + u_i$, where u_i is distributed type one extreme value. But due to censoring, this still must be estimated by maximum likelihood.

reflects the hazard ratio for two lamps that differ only in one characteristic. That is, for two lamps i and j at time t that differ only in characteristic k, the hazard ratio is,

$$HR_{i,j} = \frac{h_i(t)}{h_j(t)} = \frac{exp(X_i'\beta)\alpha t^{\alpha-1}}{exp(X_j'\beta)\alpha t^{\alpha-1}} = \frac{exp(X_{ik}'\beta)exp(X_{i,-k}'\beta)}{exp(X_{jk}'\beta)exp(X_{j,-k}'\beta)} = \frac{exp(X_{ik}'\beta)}{exp(X_{jk}'\beta)} = exp((X_{ik} - X_{jk})'\beta). \tag{9}$$

In the two-stage model, the interpretation of the coefficients becomes a bit more difficult due to the dual effect covariates can have directly on the outcome and through the endogenous regressor as instruments. Because of this issue, the marginal effects in the model are calculated numerically, as discussed below.

Additionally, in a standard maximum likelihood model, the covariance matrix from which standard errors are calculated is just the inverse of the information matrix. In this case, because the instrument is a calculated rather than fixed covariate, this method does not work. Instead, I estimate confidence intervals from a bootstrap without calculating standard errors, as discussed below.

2.2 Model Estimation

The survival analysis in this report relied on an intensive process of model development. This began with a general investigation of functional forms, continued with an enhanced 2-step model, followed by a model selection process. Once the final model was selected confidence intervals were estimated using bootstrapping and median lives were estimated.

As discussed above, the Weibull distribution is the most common distribution assumed for survival analysis models. Other commonly used distributions include the normal, log-normal, logistic, and log-logistic distributions. Results for general models including a selection of the lamp characteristic were compared and the Weibull model had a better fit to the data in terms of log likelihood (for the exponential distribution, which is a nested model of the Weibull) and in terms of the fit of the model for the normal, log-normal, logistic, and log-logistic models. Model convergence with the generalized gamma distribution, a more flexible generalization of the Weibull distribution, was very challenging due to the increased complexity of the distribution and likelihood function, leading to a lack of confidence that results for the generalized gamma model were true maximum likelihood results. For these reasons, the Weibull was retained as the distribution for the modeling.

With the distribution selected, the focus of modeling moved to selecting the best lamp characteristics to include. Certain covariates did not show up as statistically significant in any of the models tested and were not considered for a final model. For the remainder, models were compared, as discussed below in Section 2.2.2.

2.2.1 Endogeneity

A fundamental challenge of the modeling exercise is to use the information coming out of the laboratory study as effectively as possible in a way that is possible to apply to lamps outside the sample. This is because the number of cycles a lamp survives is a powerful predictor of its life, but the number of cycles a lamp will survive is not observable for individual lamps that aren't tested. In a fundamental way, lamp life and lamp cycles are jointly determined, and lamp cycles are endogenous to the estimation of lamp life. Simply leaving lamp cycles out of the model leads to models with significantly less explanatory power, so a method to include that information is necessary. This endogeneity makes sense when one considers how CFLs operate and how they fail. CFLs face hazard of failure both from operation and from start up: the current running through the electrodes and the tube degrades the lamp, as does the sharp voltage spike and rapid heating that occur when a lamp is first lit. It is this second source of degradation that leads lamps with different cycle times to have such vastly different life times.

Luckily, there are well established methods for dealing with endogeneity, primarily the method of instrumental variables. This method is a two-step method in which the endogenous variable is modeled in the first step as a function of the other explanatory variables from the main model and additional variables that help explain the endogenous variable but the not the variable of interest, and then the fitted value for the endogenous variable is used in the main model in place of the observed value. That is, the number of cycles is modeled first with a larger selection of variable, and then a predicted value for the number of cycles, which we can think of as cycling robustness, is used in the model for lamp life. The additional variables in the first stage, known as instrumental variables, are necessary for identifying the model based on variation in the data, rather than the parametric form of the model.

While this technique is most commonly used in least squares models, it is valid in a larger class of models that includes the maximum likelihood models used in survival analysis. For example, the Heckman Selection Model, a very common method in social sciences that won James Heckman the Nobel Prize in Economics, relies on a maximum likelihood estimate in the first step. The Heckman Selection Model was developed specifically to address this problem of endogeneity. What we need for the two-step model to be consistent is that the both the first and second steps are consistent themselves, which is true in this case. That is, this is a perfectly common and legitimate procedure.

The two-step model follows a common structural modeling approach with a modification. The number of cycles is modeled as a flexible polynomial series function of the explanatory variables and instruments in the sense that the model is treated as a series approximation of the true model. Because of the censoring of the data, a pure linear least squares model would not be a consistent estimator of the number of cycles, so a normal (i.e. Gaussian) survival model is used to estimate the first stage. Then predicted values from the first stage are used in the Weibull model of the lamp life.

2.2.2 Model Selection

The final model selection was a question of which covariates to include in the second stage model. As long as the first stage contains relevant explanatory variables and instruments, there is no need to limit the number of included regressors as efficiency of the model parameters is not a concern, nor was overfitting; all that matters is removing extraneous variation. The final cycling model has an R^2 value of approximately 0.56, which is quite high considering the amount of variation amongst lamps with the same characteristics. Because of the design of the study, there are approximately five lamps of the same type (that is, with all the same lamp characteristics) run under the same cycle. So the best possible model could only predict the mean of a group, not the value of an individual lamp. By taking lamp type averages the R^2 of best possible hypothetical would be 0.71. So the cycling model has an effective R^2 of 0.79, that is, it explains 79% of the variation of the best possible model compared to the most simple model using the overall sample mean.

Model selection was based on both the two-step model, and a base model of just one step, where the observed values of the lamp cycles were used in place of the predicted values, but the form of the equation otherwise followed the second step of the two-step model. This was in order to consider both external and internal validity. The final second stage model was selected among three competing models, one, Model 1, which included all variables that were at least marginally significant in a single stage model with actual cycles, one, Model 2, that included variables that were significant near the 95% level, and one, Model 3, that only included variables that were significant above the 99% level. The final model was selected based on parsimony (including the fewest possible number of regressors to achieve the same fit), model fit to the data, and external validity, with the most emphasis put on the last. This was assessed using hold out samples, also known as k-fold cross-validation.

The method was a follows. The sample was randomly partitioned into 4 groups (i.e. $k = 4$ here). Then each of the models was estimated on each subset of the sample with one group missing. Predicted values were then calculated for each observation based on the model that excluded it. Six measures were then calculated for each of the models: the average prediction error, the average absolute value of the prediction error, and the square root of the average of the square of the prediction error (i.e. the RMS prediction error), each for the whole sample and for the portion of the sample that had actually failed. The average prediction error is the least important, but indicates how skewed the distribution of errors is (because these are nonlinear models, prediction errors are not mean-zero even in the full model). The average absolute value treats all errors equally and will favor a model that predicts medians well.[6] As the object of interest in this study is a median, this is desirable. The RMS prediction error puts more emphasis on very large prediction errors, which are undesirable. These were calculated on the whole sample and the subsample that had failed because the predictions would not be expected to be correct for the lamps that did not fail.

Each of these measure were calculated for the single-step version of the model, as well as the two-step version of the model. Results are shown in Table 2.
While Model 2 slightly outperforms Model 1 in some measures for the base model, Model 1 is optimal in the two-step model by every measure. For this reason, Model 1, the model with the most complete set of explanatory variables in the second stage, was chosen.

2.2.3 Confidence Intervals

A significant drawback of the two-step model is that the typical asymptotic standard errors that are easy to calculate are no longer valid. This is because asymptotic standard errors require that

[6] While ordinary least squares will yield a model that fits the mean of the data, a least absolute deviations model will fit the median of the data.

Table 2: Model Cross-Validation Results

	Base			Two-Step		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Mean Residual	-3,194	-3,182	-3,198	-1,919	-1,943	-1,960
Mean Absolute Residual	3,485	3,475	3,491	3,601	3,621	3,635
RMS Residual	13,838	13,733	13,820	6,318	6,342	6,526
Mean Residual (Failed)	-614	-616	-618	-2,243	-2,264	-2,279
Mean Absolute Residual (Failed)	942	945	948	3,221	3,237	3,245
RMS Residual (Failed)	3,091	3,088	3,132	5,460	5,487	5,610

the sample be independent. With a two-step model, the calculated value for the cycles is no longer independent as the predicted value is dependent on the sample.

One way to address this is through an adjustment in the information matrix used to calculate the variance matrix. This is difficult and tedious. A better method, employed here, is to estimate confidence intervals from the bootstrap, rather than from asymptotic standard errors.

The bootstrap relies on resampling to develop a consistent estimate of the distribution of relevant parameters of a model. While the most common bootstrap techniques resample from the empirical distribution function of the data, this is inefficient in a setting with an experimental design where random assignment creates independence of the error term and covariates. The best and most common solution is to use what is known as the Wild Bootstrap.

The Wild Bootstrap maintains the same experimental design matrix but resamples from the residuals in a way that preserves the first two moments of the distribution of the residuals. This maintains the key characteristics of the error structure that make the experimental design so desirable, while still allowing for an investigation of the distribution of relevant parameters. This procedure is again carried out in two steps. For each bootstrap replication, each observation is randomly assigned either a negative or positive shock to its residual from the cycling and the survival models[7], and the bootstrap values of the cycles and life of each lamp are then calculated as the predicted value plus the altered residual. The cycling model is then evaluated based on the original design matrix and the bootstrap cycling value. This bootstrap cycling model is then used to predict the cycling value that is used in the survival model with the original design matrix and the bootstrap survival time. This process is repeated a total of 9,999 times, and then confidence interval for the parameter estimates can be estimated from the distribution of bootstrap values of the parameters.[8]

[7]The residual from the cycling model is not the difference between the actual and the predicted values, which is not mean zero, but a modified value based on the form of the survival model.

[8]An additional advantage of the bootstrap is that it allows us to adjust the original parameter estimates in addition to calculating confidence intervals. Although the survival analysis models are unbiased, there will be deviations within any sample from the true value. The mean values of the parameters from the bootstrap distribution can be used to correct for small-sample deviation from the true model parameters.

2.2.4 Estimation of Median Life

Although the model is formulated to estimate a conditional mean, estimation of conditional medians (or any quantile of the distribution) in a Weibull model is very simple. In general, for something whose expected life is distributed Weibull with mean life μ and scale factor σ, time at which $q\%$ have failed is,

$$T(\frac{q}{100}) = \mu(-ln(1 - \frac{q}{100}))^{\frac{1}{\sigma}}. \tag{10}$$

Then the median life is just,

$$Median = \mu(-ln(.5))^{\frac{1}{\sigma}}. \tag{11}$$

3 Data

This section describes the data used for this analysis. This first part discusses the preparation and cleaning of the data. The second presents some key qualitative findings that come directly from the data.

3.1 Description of Laboratory Experiment

The experiment was designed to test the dependence of lamp life on lamp characteristics and operating conditions. This design was developed collaboratively between SCE and the CPUC's Data Management and Quality Control (DMQC) group. A designed sample of lamps were to be procured and then operated under specified conditions until failure. A subsample of lamps would also be tested periodically to determine how the true (as opposed to rated) wattage, color rendering index, color correlated temperature, power factor, and lumen output changed over time.

The sample was designed to represent the diversity of lamps available through the IOU upstream lighting programs and through the market at the time the study began in 2010. It was not, on the other hand, designed to be representative of the program distribution of lamps. Instead, it was designed to cover the relevant types of lamps offered through the programs or otherwise available, but with the distribution designed to be able to detect effects, rather than to allow for direct extrapolation to the population. It also included non-Energy Star lamps, which are not available through the programs, to test for difference in durability between Energy Star and non-Energy Star lamps.

The design of the sample to detect effects relevant to IOU program lamps required a number of characteristics. First, lamps were selected to have a wide variety of characteristics, and combinations of characteristics representative of IOU programs, but not to represent the mix of lamps themselves. This increases the ability of statistical methods to model the effects of these characteristics and combinations of characteristics. Second, we wanted a large enough sample to be able to achieve reasonable confidence and precision, and to be able to have multiple lamps of each type with each of the assigned treatments in order to be able to observe the variability among lamps of the same type. Similarly, the cycling regimes were designed to detect effects, rather than represent the population. The times selected were fixed 2-, 5-, 15-, 30-, 45-, 60-, 90-, 180-, and 720-minute cycles and a variable cycle with an average on-time of 30 minutes. The distribution for the 30-minute variable cycle followed the distribution of average on-times for the 2005 CFL Metering Study.[9] Therefore, both specialty lamps, and extreme cycle times were over-sampled relative to the population. This makes good sense given the problems with direct extrapolation discussed above.

For more information about the design of the experiment, consult the report from the CPUC DMQC.[10]

3.2 Data Preparation

The data from the lab experiment were delivered in a cleaned and usable form, but not the necessary form for this analysis. The data include individual lamp-level data on time of failure if failed, lamp characteristics, including type (spiral, reflector, globe, A-lamp), control (standard, dimmable, three-way), Energy Star status, wattage, lumens, rated life, and experiment usage characteristics (base up, average cycle time, high or medium power for 3-way lamps).

The dependent variable, failure/censoring time, was created by setting it to the failure time if the lamp had failed or to the total operation time for the cycle on which the lamp was operating. Likewise, the failure variable was set to 1 if it had failed, and 0 if it had not. This allows the R statistical software, in which the analysis was conducted, to create a formatted data set appropriate for survival analysis.

The categorical variables were each converted into dichotomous variables to indicate the presence of the specific attribute. For lamp control, dummies were created for dimmable, 3-way high, and 3-way medium to indicate whether the lamp was dimmable, 3-way and operated as a regular lamp, or 3-way and operated on the middle power setting. Dummies were also created for reflector, globe, and A-lamp types. Standard control and spiral were not designated to avoid perfect collinearity in the data matrix.

As the data were periodically updated over the course of the study, the data sets were reconciled between each version and differences, such as failure time changing for an already failed lamp, were investigated to ensure data quality.

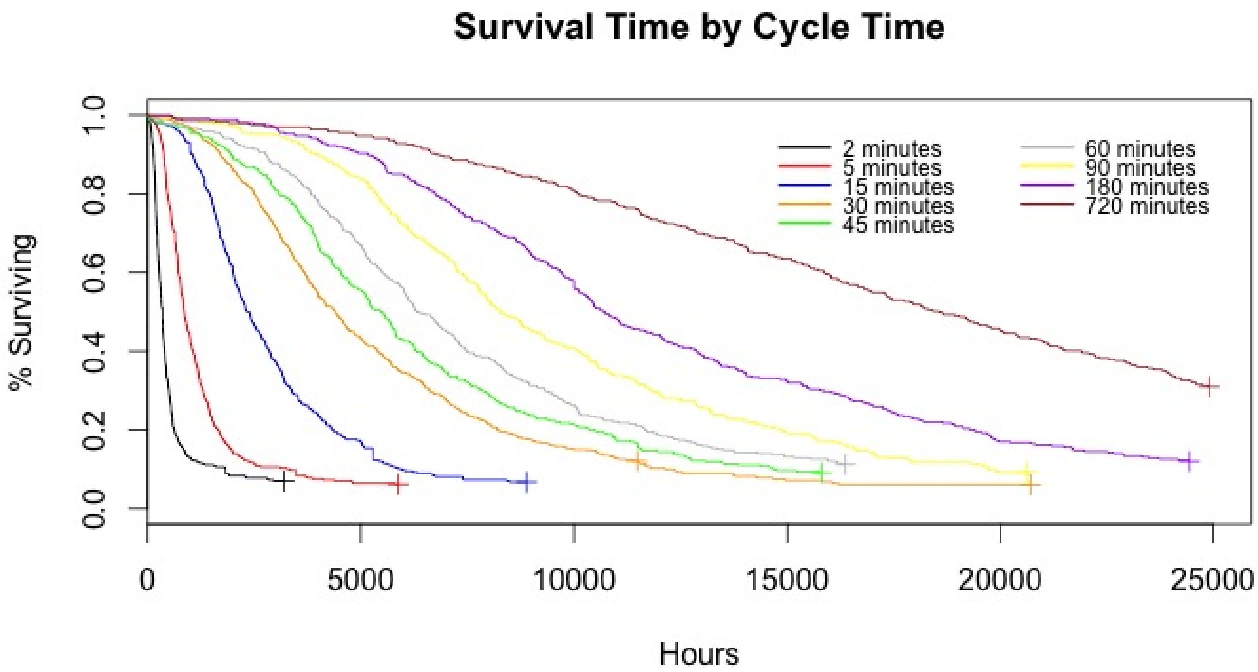
3.3 Description of Laboratory Results

This section describes some of the qualitative results with respect to lamp life that came out of the laboratory study. Figures in this section present the empirically observed failure times for lamps

[9] "CFL Metering Study". Kema-Xenergy. February 2, 2005. Available at: http://calmac.org/publications/2005_Res_CFL_Metering_Study_Final_Report.pdf

[10] "CFL Laboratory Testing Report Preliminary Results from a CFL Switching Cycle and Photometric Laboratory Study". James J. Hirsch and Associates, and Erik Page & Associates, Inc. Submitted to the California Public Utilities Commission Energy Division, May 21, 2012.

Figure 1:



in the study. That is, there is no modeling involved with the survival curves presented here. The curves show the percent of the sample surviving as a function of the run hours. Therefore, a curve that lies above another curve at a point has more lamps surviving at that point, and a curve that lies above another curve at all points beyond zero represents lamps that lasted unambiguously longer conditional on the sample selected and its operating conditions but not the other lamp characteristics.[11] Kaplan-Meyer confidence bounds are included only for two-way comparisons in the interest of readability.

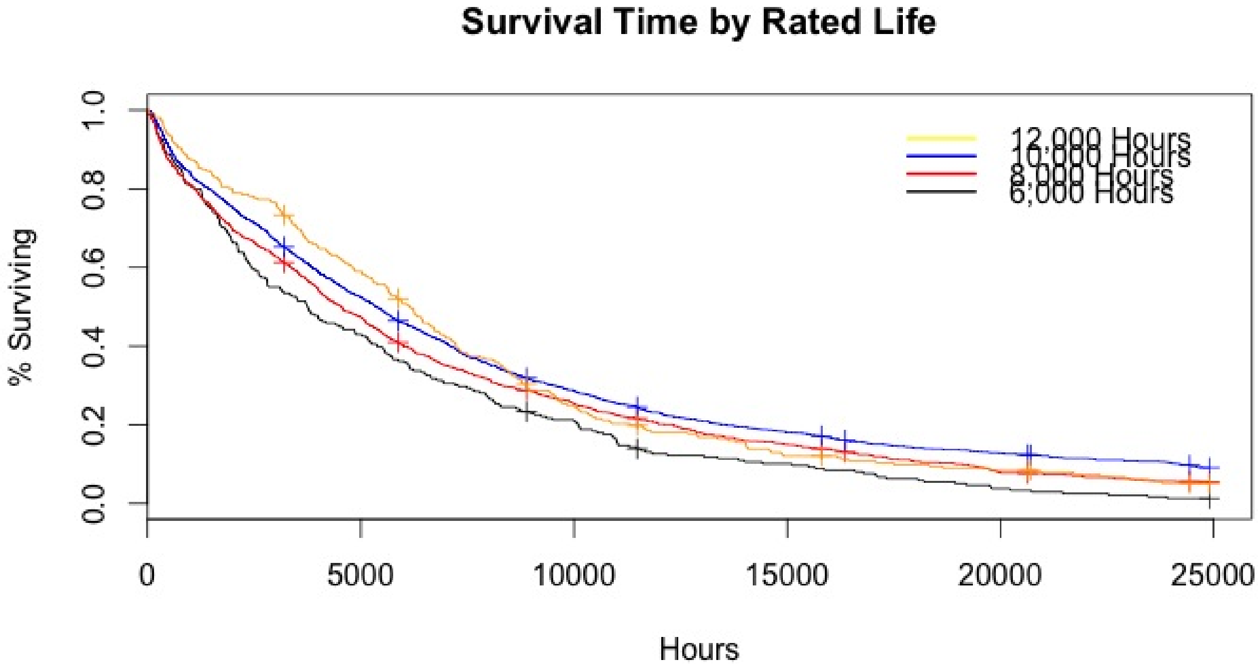
As expected, the cycle time of the lamps shows a very clear effect on the life of the lamp, as shown in Figure 1. Cycle time has a strong effect on lamp life throughout the distribution of the sample, as well as at all values of the cycle time. That is, for example, the very first short-cycle lamps died before the very first long-cycle lamps, and there are significant increases in lamp life going from 2 minutes to 5 minutes, as well as from 180 minutes to 720 minutes.

The effects of rated life on lamp life are less clear, as shown in Figure 2. While there is a clear and expected relationship between lamps rated at 6,000 hours, 8,000 hours, and 10,000 hours, lamps rated at 12,000 hours initially perform better than the others, but then decline somewhat drastically and settle close to the value for 8,000 hour rated lamps in the tail of the distribution. As this occurs after the median life of the lamp, this does not necessarily indicate problems with the lamp testing.

Perhaps the more surprising result come from a comparison and ENERGY STAR and non-

[11]That is, just because the lamps of one category in this study lasted unambiguously longer than another group, does not mean that all lamps of this category would last unambiguously longer than the other.

Figure 2:



ENERGY STAR certified lamps, shown in Figure 3. Non-ENERGY STAR lamps actually had longer lives throughout most of the distribution. Considering the pointwise 90% confidence bands for the Kaplan-Meyer survival functions indicates that, while the sample size for non-ENERGY STAR lamps was much smaller, this difference appears to be real.

Another notable result is the difference between standard and specialty lamps. Standard lamps are spiral lamps with regular controls and wattage less than 30W; specialty lamps have a different shape, different controls, and/or higher wattage. The results are shown in Figure 4. Once again, specialty lamps appear to last longer than standard lamps at all parts of the distribution, and the difference appears to be statistically significant in most of the distribution.

One major surprise was that the lamp orientation, base up vs. base down, did not show up as significant in the various specifications tested. There has been a lot of anecdotal evidence and discussion that many CFLs die prematurely because of being used in a base up orientation. There is not evidence from this study to support that idea. The Kaplan-Meyer survival curves for lamps in these two groups are shown in Figure 5. Based on the commonly perceived impact of base orientation, this result is quite striking. These two groups are nearly indistinguishable at all points in the distribution, and nowhere are they statistically distinguishable. Recall that the Kaplan-Meyer survival curve is not a modeled curve, it is drawn directly from the data. A test of the hypothesis that the hazard rate is equal at all times throughout the distribution fails to reject the null even at the 80% confidence level.

Figure 3:

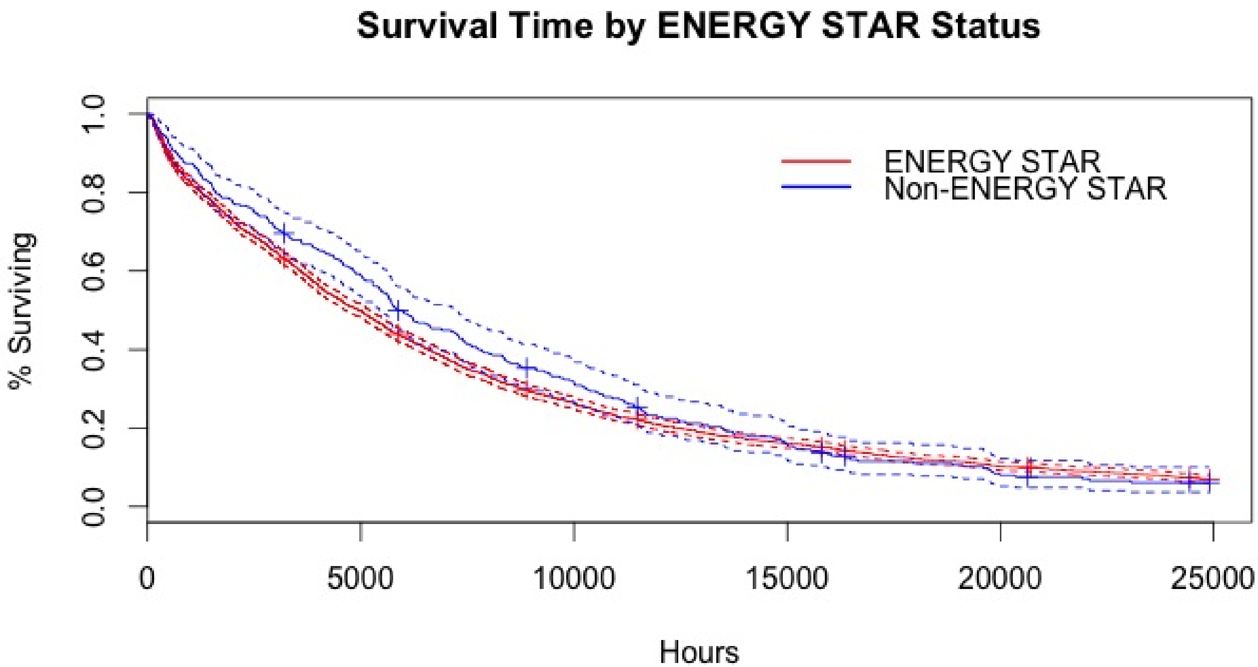


Figure 4:

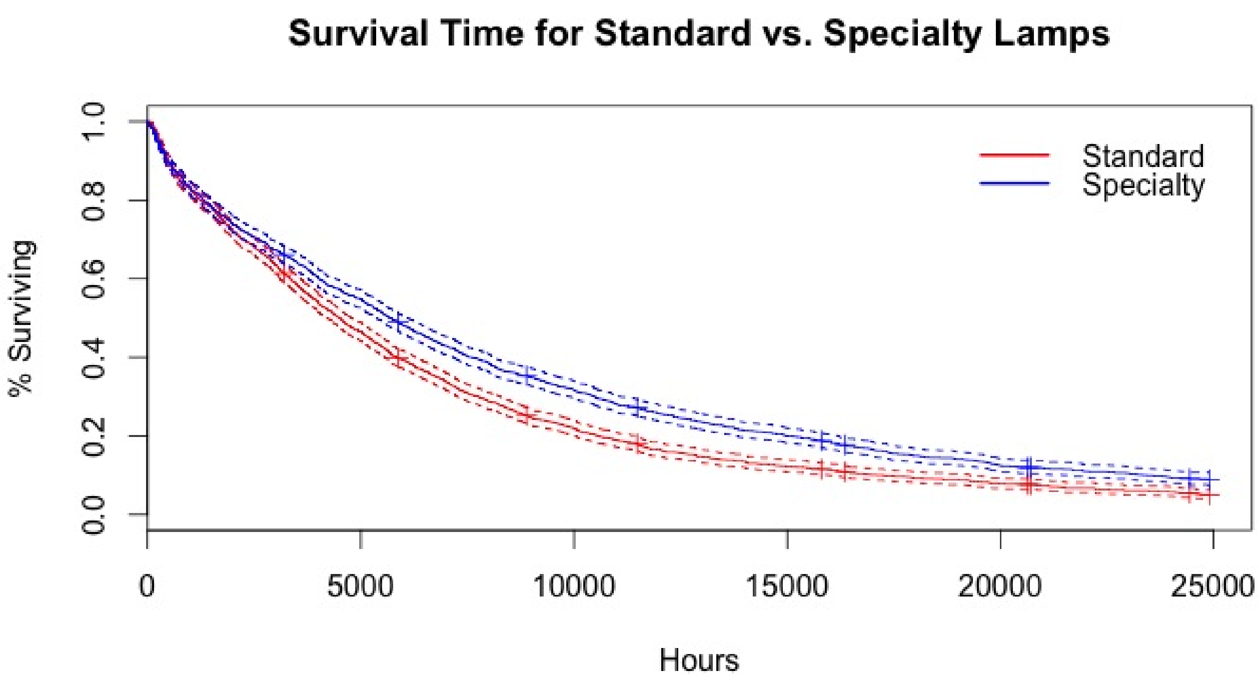
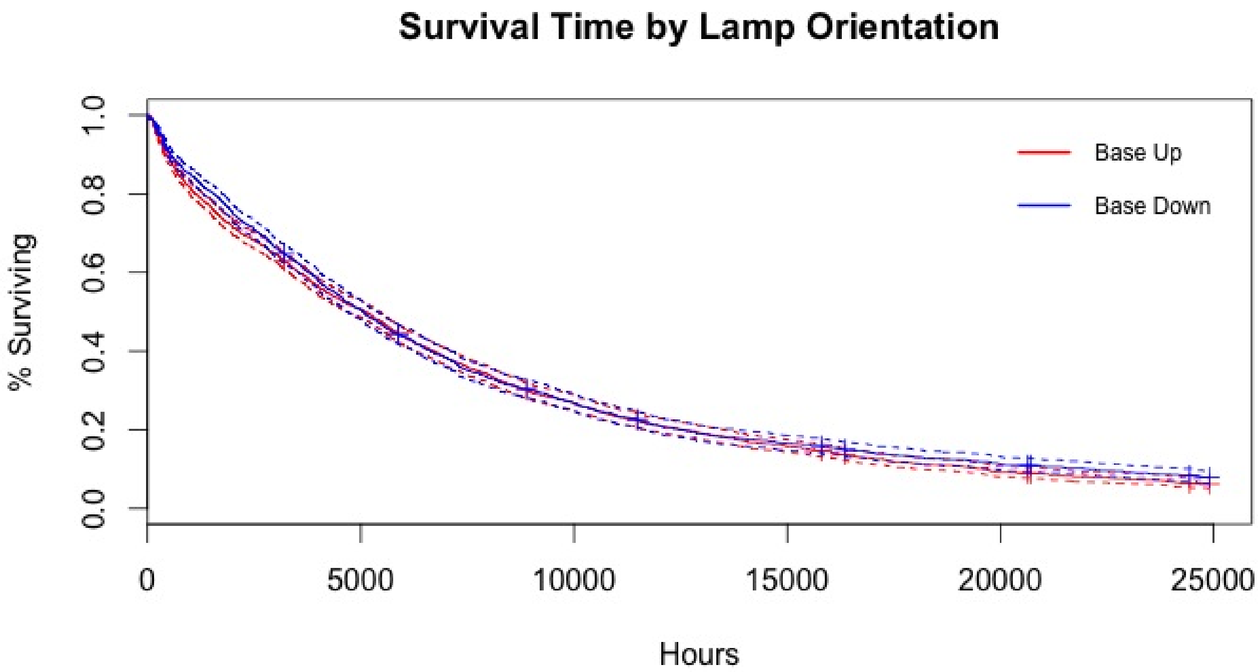


Figure 5:



4 Results

This section discusses the results of the survival analysis and model predictions within the sample. The first subsection discusses the marginal effects implied by the model. The second subsection presents results on how well the model predictions fit the sample data for various important lamp categories, showing results based on various averages for lamp shape, control type, and cycle time. The final subsection presents the model fits for the entire sample, considering various ways to estimate that. All the analyses in this section deal with lamps observed in the sample.

4.1 Marginal and Incremental Effects

Marginal and incremental effects indicate the change in the average lamp life that the model predicts would occur if the characteristics of a lamp changed. In a linear model, the marginal and incremental effects are simply the coefficients of the model, and are by construction constant across all ranges of the lamp characteristics. Two significant implications of the nonlinear model used here are that the model coefficients are not directly interpretable or meaningful without further calculation, and that marginal effects must be calculated and vary based on lamp characteristics. For this reason, I do not present model coefficients. Instead, I present marginal and incremental effects.

Additionally, because this is a nonlinear model with an exponential form, it contains all interactive effects between included variables. To see why, consider a simplified model, with just two characteristics and no intercept, of the form,

$$\mathbb{E}[T_i|x_1, x_2] = exp(ax_{1i} + bx_{2i}). \tag{12}$$

Then the change in expected life in going from x_1 to x_1' is $exp(ax_1' + bx_2) - exp(ax_1 + bx_2)$, which is different for different values of x_2. For this reason, the marginal effect of an interaction between variables is even less well defined that the marginal effect for a single variable. All the interactive terms were significantly different from zero with 90% confidence in the first-stage cycling model.

Both marginal effects (those for continuous variables) and incremental effects (those for categorical variables) in a nonlinear model can be calculated analytically by taking the partial derivative of the expected value of the dependent variable. But in this two-step model, the partial derivative is somewhat burdensome to calculate for computational reasons due to the fact that variables can impact the outcome directly in the second stage, as well as indirectly through the predicted cycle time in the first stage. Instead, I estimate the marginal effect by taking a finite differences approximation, that is, calculating the predicted value at two values of the target variable while holding all other constant, and then dividing the difference in the predictions by the difference in the input. Incremental effects are calculated by taking the difference in the predicted values.

As mentioned above, marginal effects for one variable depend on the values of the other variables in the model. Following the standard practice for nonlinear models, I've held the other values at their mean value. Such an "average" lamp does not exist, and includes non-sensical values for the categorical variables, such as roughly 0.07 A-lamp and 0.92 ENERGY STAR certified. The goal is to calculate values that are indicative of the change in lamp life for lamps given a change in their characteristics, while understanding that the actual predicted change will be different for lamps with different sets of characteristics. For lamp type and control, I used the average values for lamps with those characteristics. For example, the estimate for A-lamps is based on the predictions for an A-lamp with average wattage, lumens, and rated life but no other specialty characteristics compared to a non-A-lamp with the same wattage, lumens, and rated life and no other specialty characteristics. For the value of the cycling time, I selected the variable cycle, as that is closest to the average conditions found in the actual usage. Results are shown in Table 3. Confidence limits are based on the results of the bootstrapping procedure described above.

As expected, the rated life of a lamp has a statistically significant positive effect on its life. For an average lamp, increasing the rated lamp by one hour would increase the life by about one and a half hours, holding all other characteristics constant. Similarly, increasing the wattage by one watt, while holding all other characteristics *including the lumens* constant, increases the life of an average lamp by about 725 hours. Increasing lumens by one lumen, on the other hand, again keeping all other characteristics *including wattage* constant, decreases the average life by about seven and a half hours. A-lamps, reflectors, and 3-way lamps do not appear to have a statistically significantly different life compared to other similar lamps. Globes, candles, and dimmable lamps appear to have significantly longer lives. ENERGY STAR-certified lamps appear to live statistically significantly shorter, although the effect is not large. This likely indicates that part of the effect in Figure 3 is due to other lamp characteristics correlated with whether a lamp is ENERGY STAR certified or not. The result for dimmables is quite surprising. While dimmables lived significantly longer in the sample than non-dimmables, the difference was not nearly as great as the incremental effect indicate, an increase of roughly $1,700$ hours in life compared to other similar lamps. The large

Table 3: Marginal and Incremental Effects for an "Average" Lamp

	Lower Confidence Limit	Marginal/Incremental Effect	Upper Confidence Limit
Rated Life	1.00	1.47	1.93
Wattage	493.53	727.83	952.65
Lumens	-10.52	-7.57	-4.43
A-Lamp	-615.61	498.45	1,976.34
Globe	1,920.57	3,544.15	5,051.22
Candle	2,356.17	5,323.82	7,998.00
ENERGY STAR	-1,488.45	-690.85	208.66
Reflector	-199.32	619.35	1,601.43
Dimmable	1,216.67	1,755.17	2,316.61
3 Way	−2,703.40	-819.00	920.55

90% confidence interval based on Bootstrap with 9,999 replications

value is likely due to the interactive effects between dimmables and other characteristics being particularly strong in the average range.

4.2 Model Fit

The present subsection provides results on the model fit for various groups of average lamps. Subsubsections cover lamp shape, controls, and cycling times. The model fit considered is the fit between the nonparametric Kaplan-Meyer survival function observed in the data, and the model prediction of the survival function. That is, it compares the percent surviving in the sample at any time with the model's prediction of the percentiles of the survival function. As with the marginal and incremental effects of the previous section, these results are based on average lamps, rather than the full sample of relevant lamps. The reason for this is that while for large groups of lamps, we see a good representation for the survival function, this is not possible to do at the lamp group category. That is, with 251 A-lamps we can get a sense of the survival function, but with only five lamps of a given type, it is impossible to know which percentiles to predict.

4.2.1 Lamp Shape

Figures 6 through 8 show the fit results for various lamp shapes. Candles were omitted as the sample size of 100 was so small that the confidence bounds on the Kaplan-Meyer estimate are too large to make a comparison very meaningful. For this and the following subsubsection, cycling regimes are restricted to the 30 minute variable regime.

In each case, the fit is quite good, with the predicted survival curve for the average lamp being within the 90% confidence bounds well over 90% of the time.

4.2.2 Controls and ENERGY STAR-Certification

Figures 9 and 10 show the results for 3-way and dimming controls.

Figure 6:

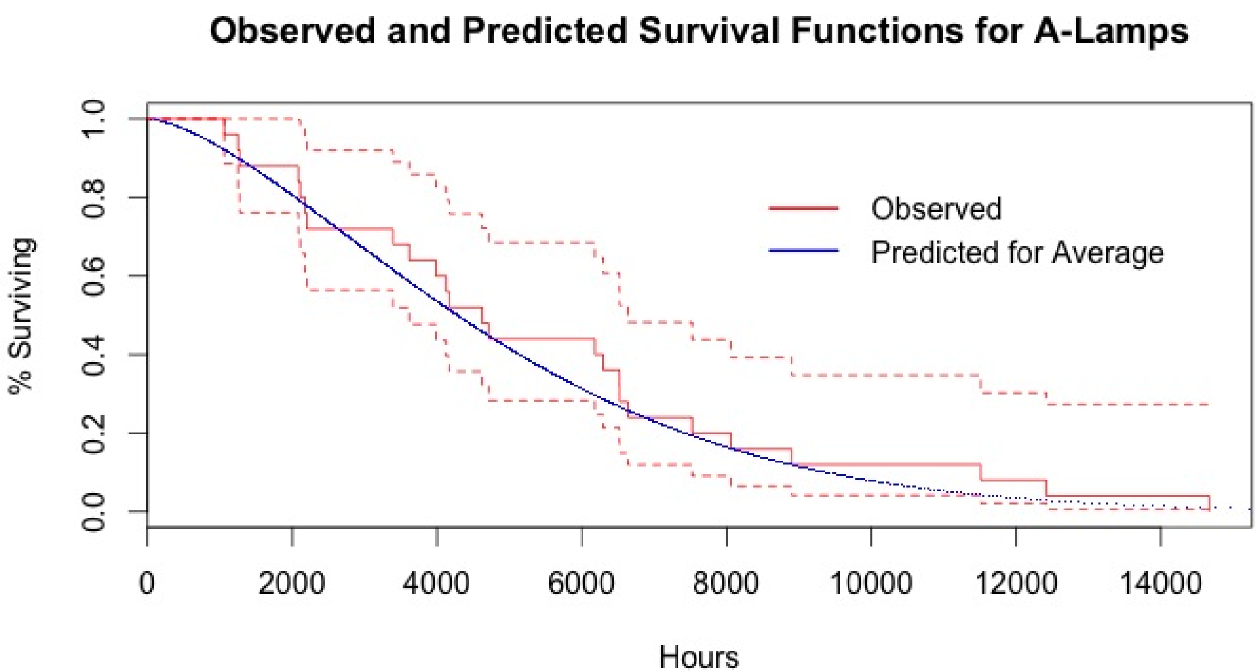


Figure 7:

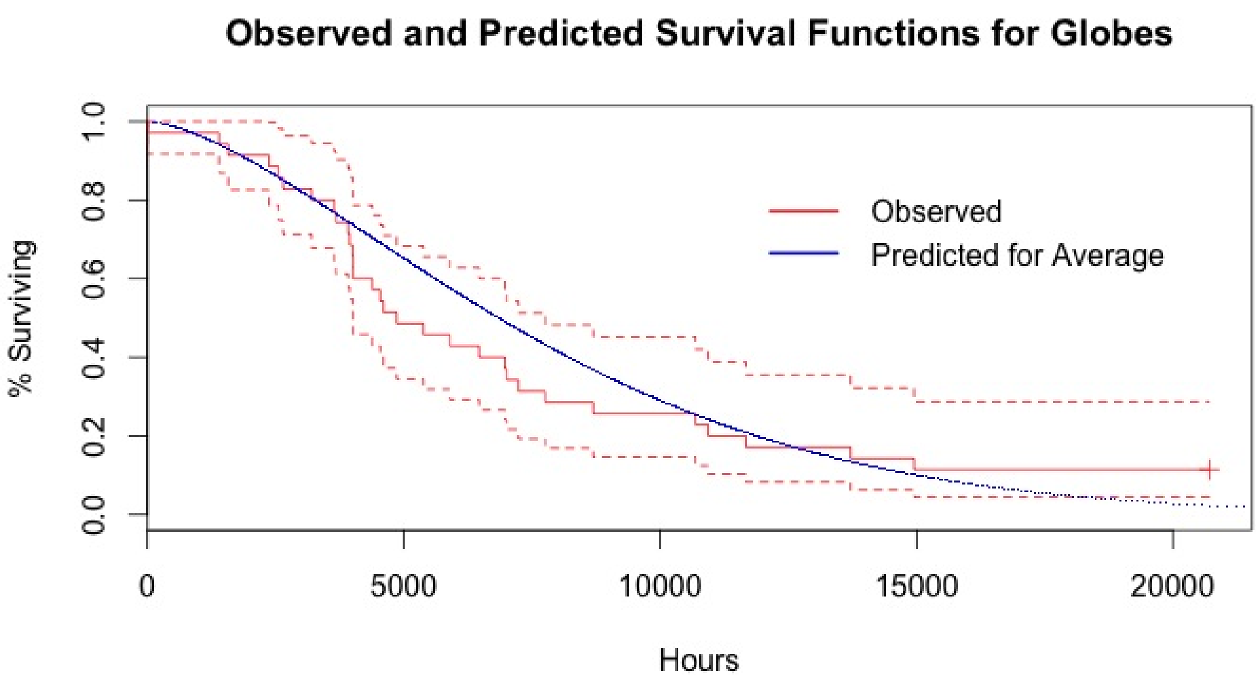


Figure 8:

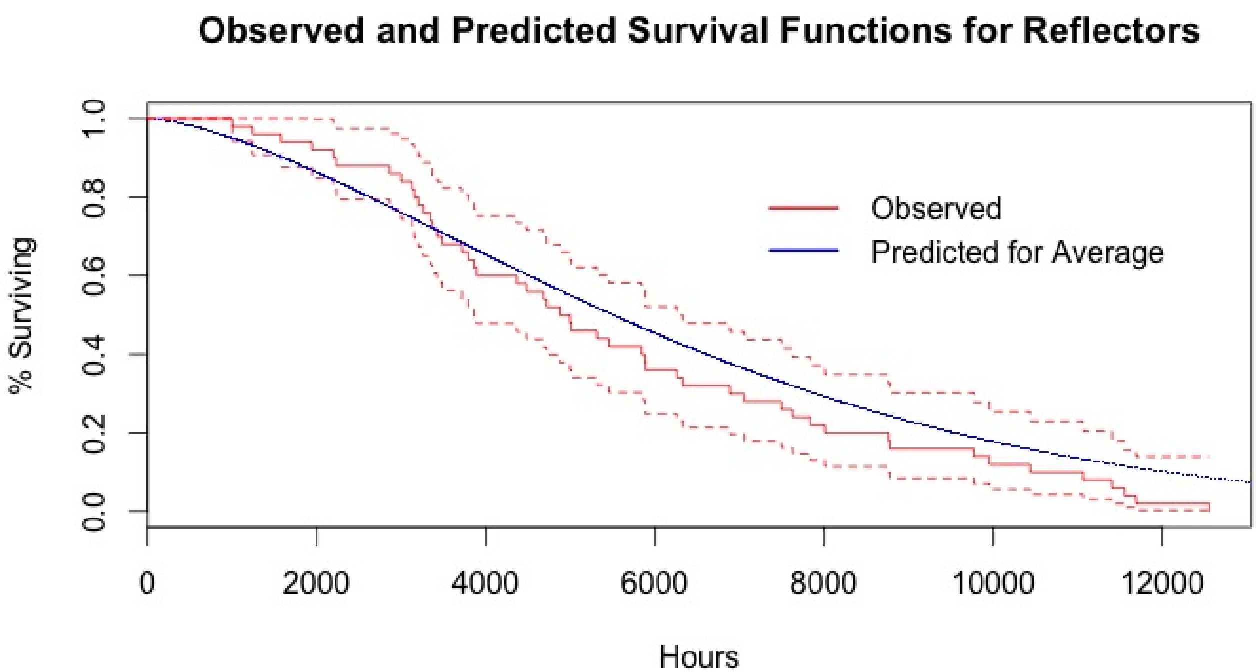


Figure 9:

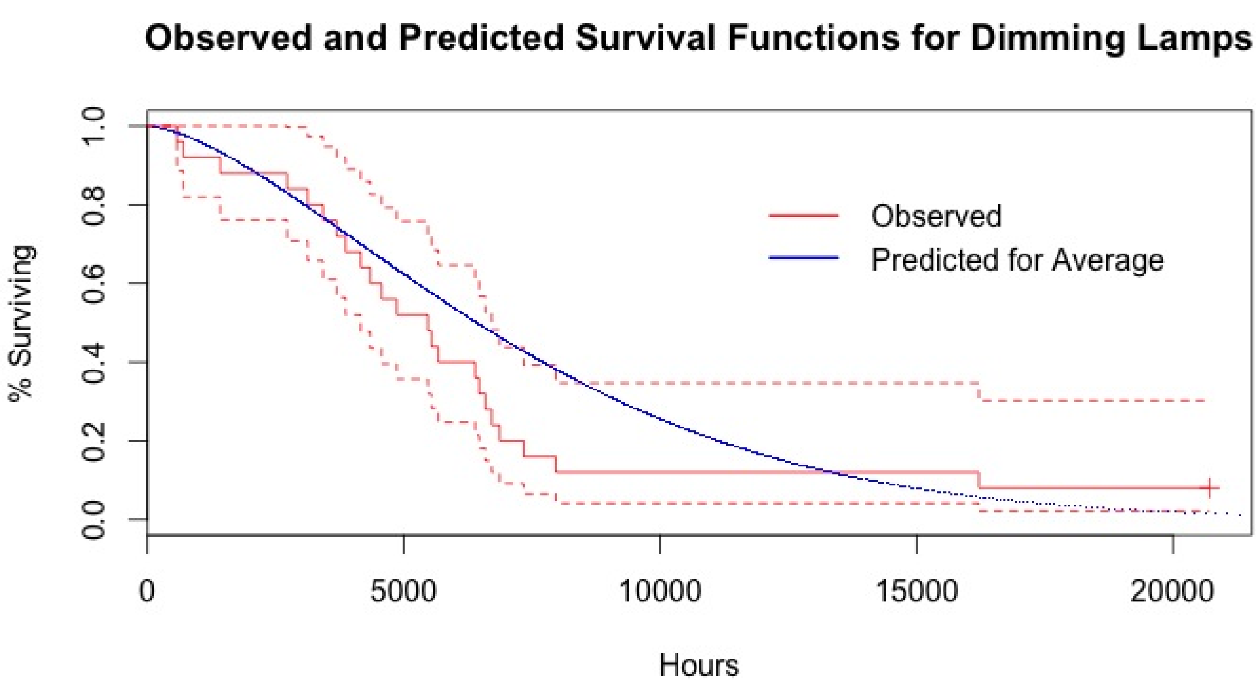
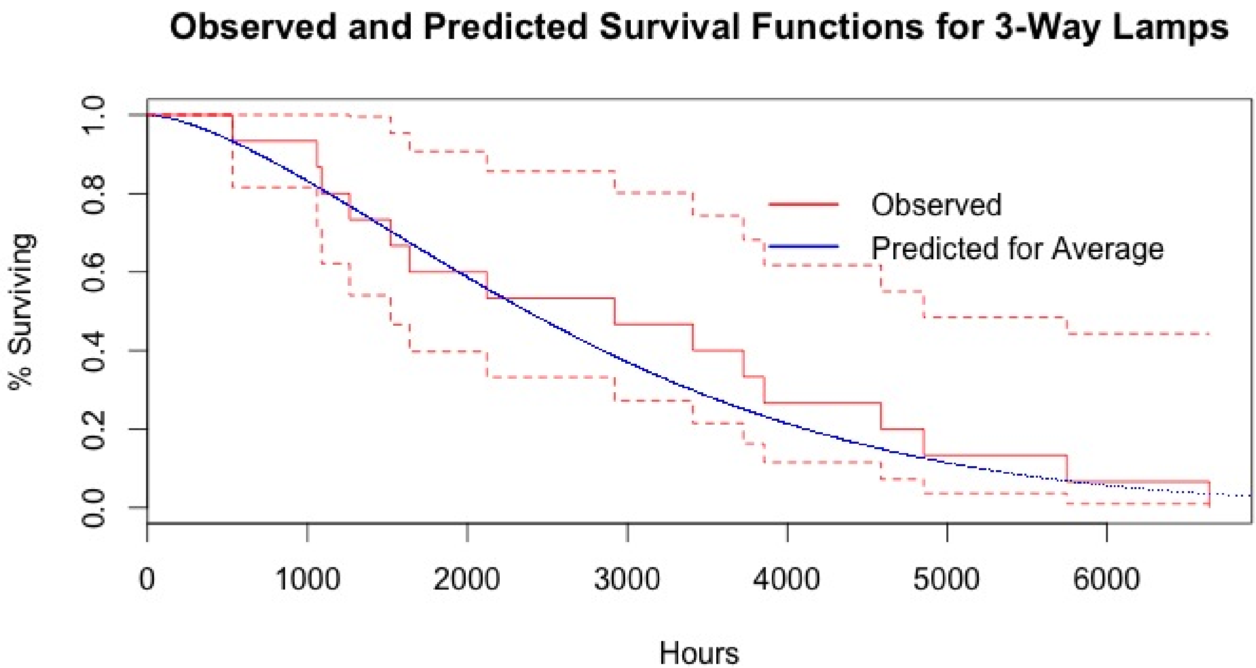


Figure 10:



Again, the fit is quite good and within the confidence bounds well over 90% of the time. The model appears to have had trouble fitting the relatively steep drop and long tail of the dimming lamps. It should be noted that the result for dimming lamps is based on the lamp operating in the dimming model and the result for the 3-way lamp is for the lamps running in the medium setting. No significant effect was found for 3-way lamps tested in the high setting. As such, these estimates likely overstate the true life of these lamps as they are unlikely to be operated below full capacity all the time.

Figure 11 shows the results for ENERGY STAR certified lamps. The fit is not as tight in this case, as to be expected for this larger and more diverse group of lamps. That is, the simplification of an "average" lamp is not nearly as valid here, weakening the fit.

4.2.3 Cycle Time

Figures 12 through 15 show the fits for 15 minute, 30 minute fixed and variable, and 45 minute cycles. As with the ENERGY STAR lamps in Figure 11, these estimates compare predictions for an "average" lamp that covers all the different types of lamps in the sample to the actual results for all the different types of lamps in the sample, so the estimates would not be expected to fit as well as for the narrower categories.

Nonetheless, the predictions fit reasonably well and match the overall shape reasonably well. While the predictions for the 15 and 30 minute lamps tend to over-predict survival, the 45 minute lamps are under-predicted.

Figure 11:

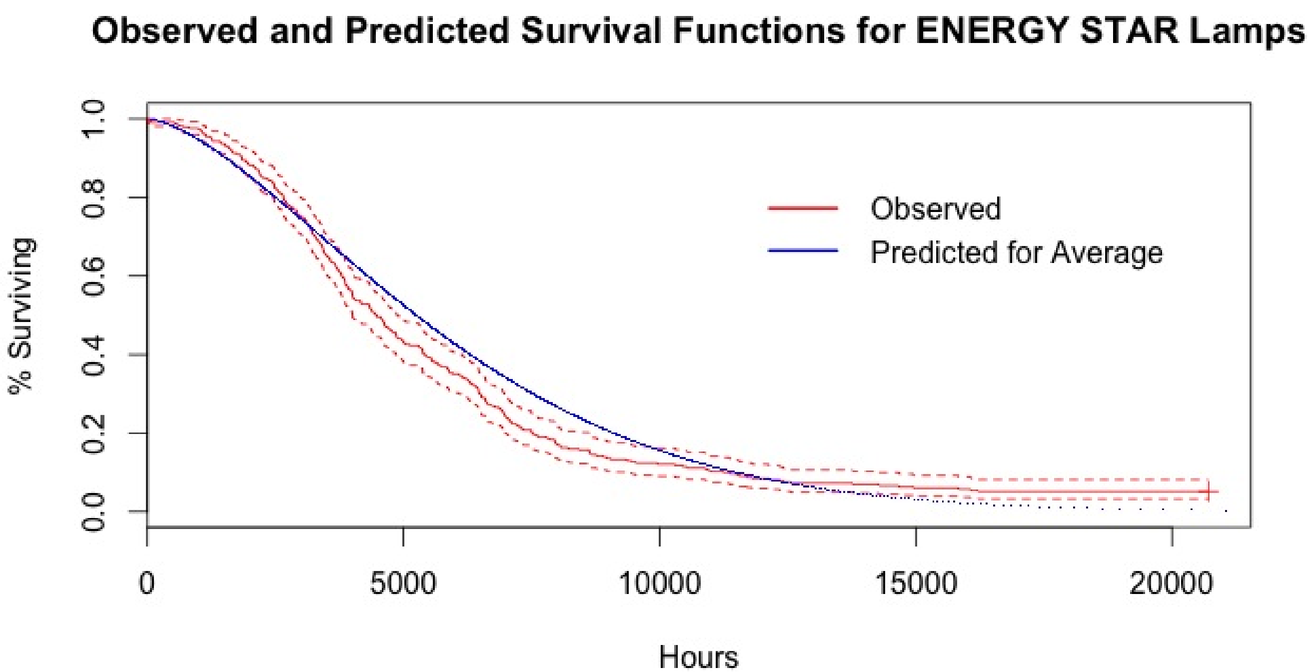


Figure 12:

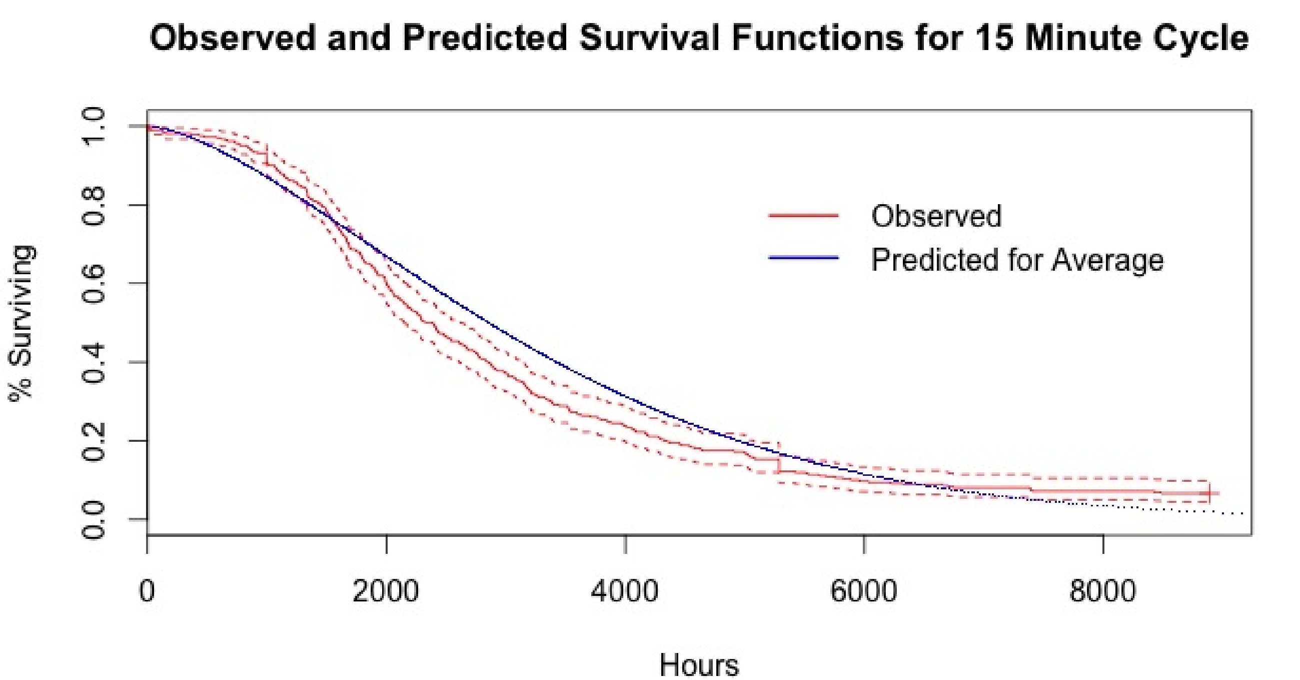


Figure 13:

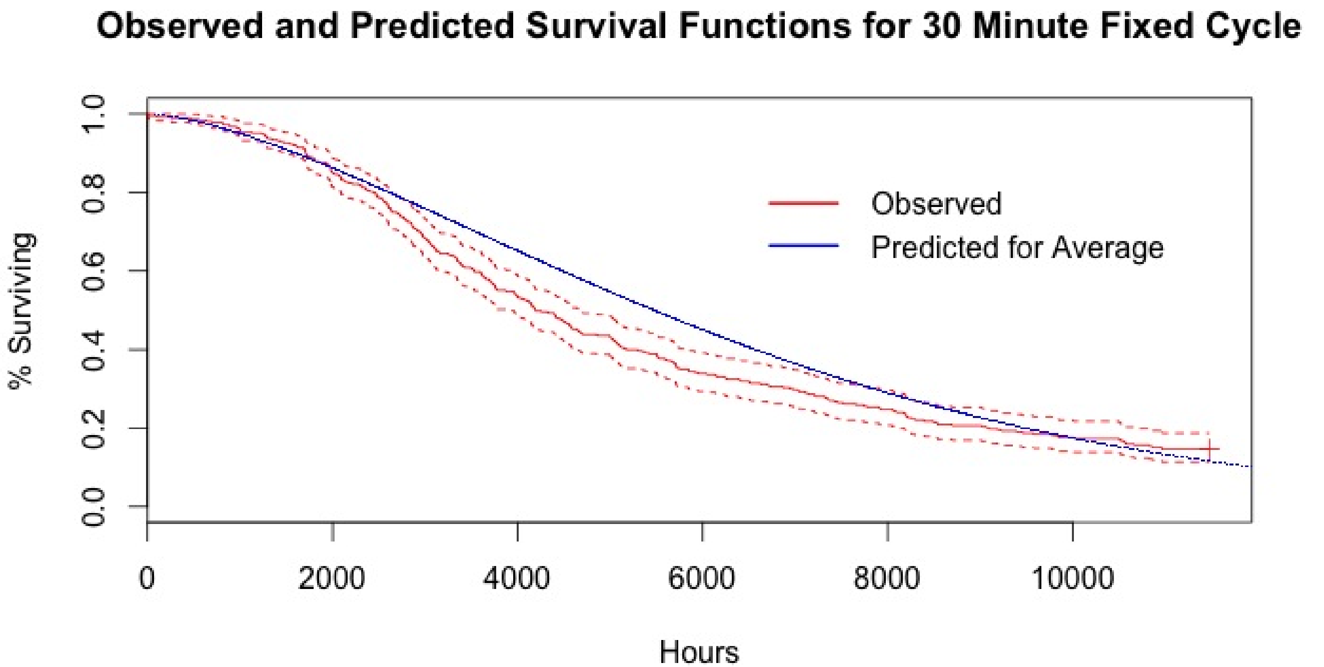


Figure 14:

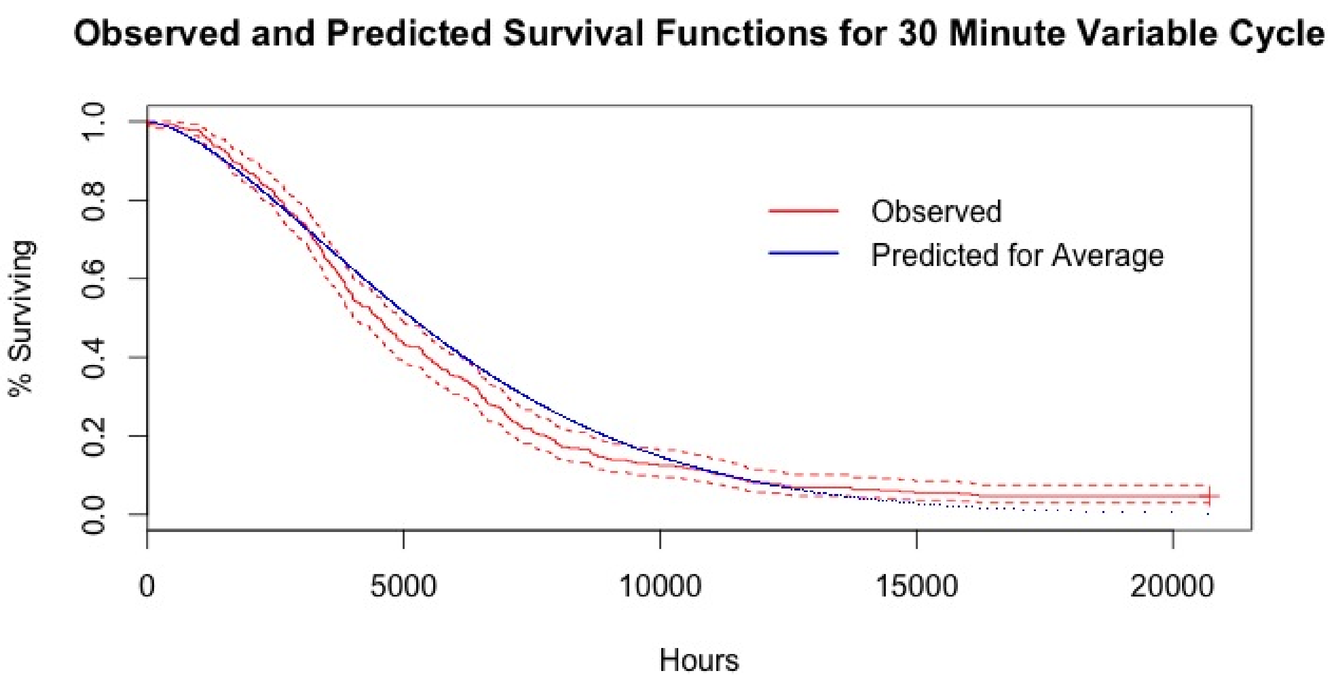
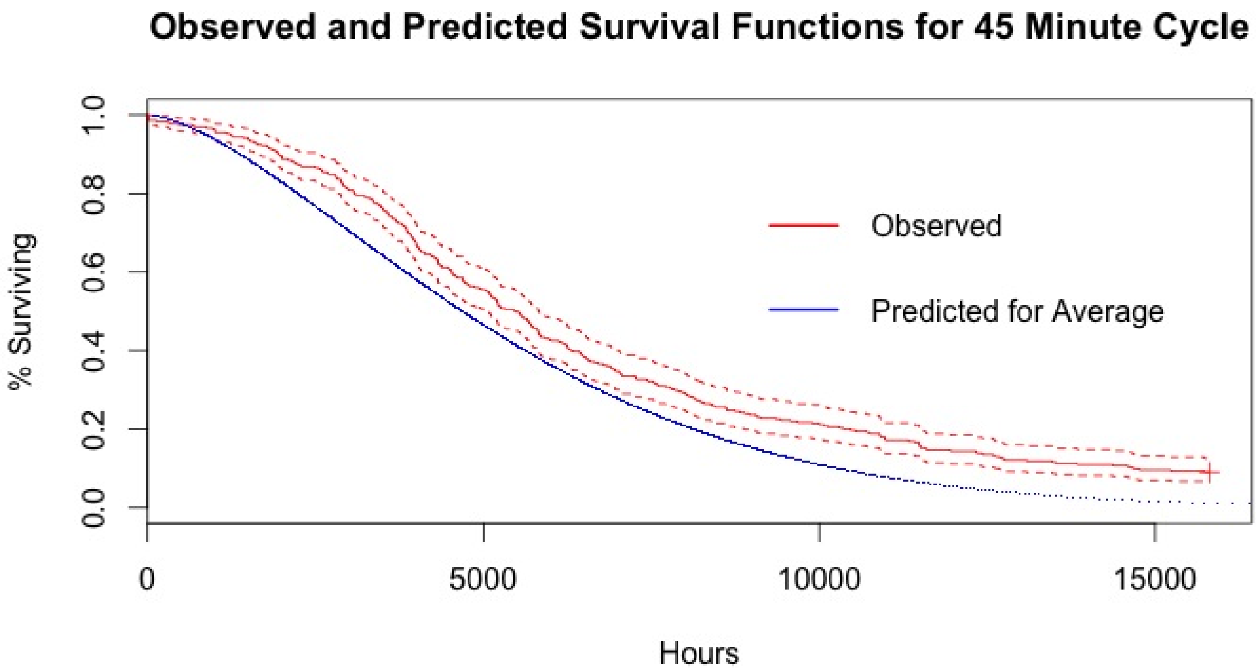


Figure 15:



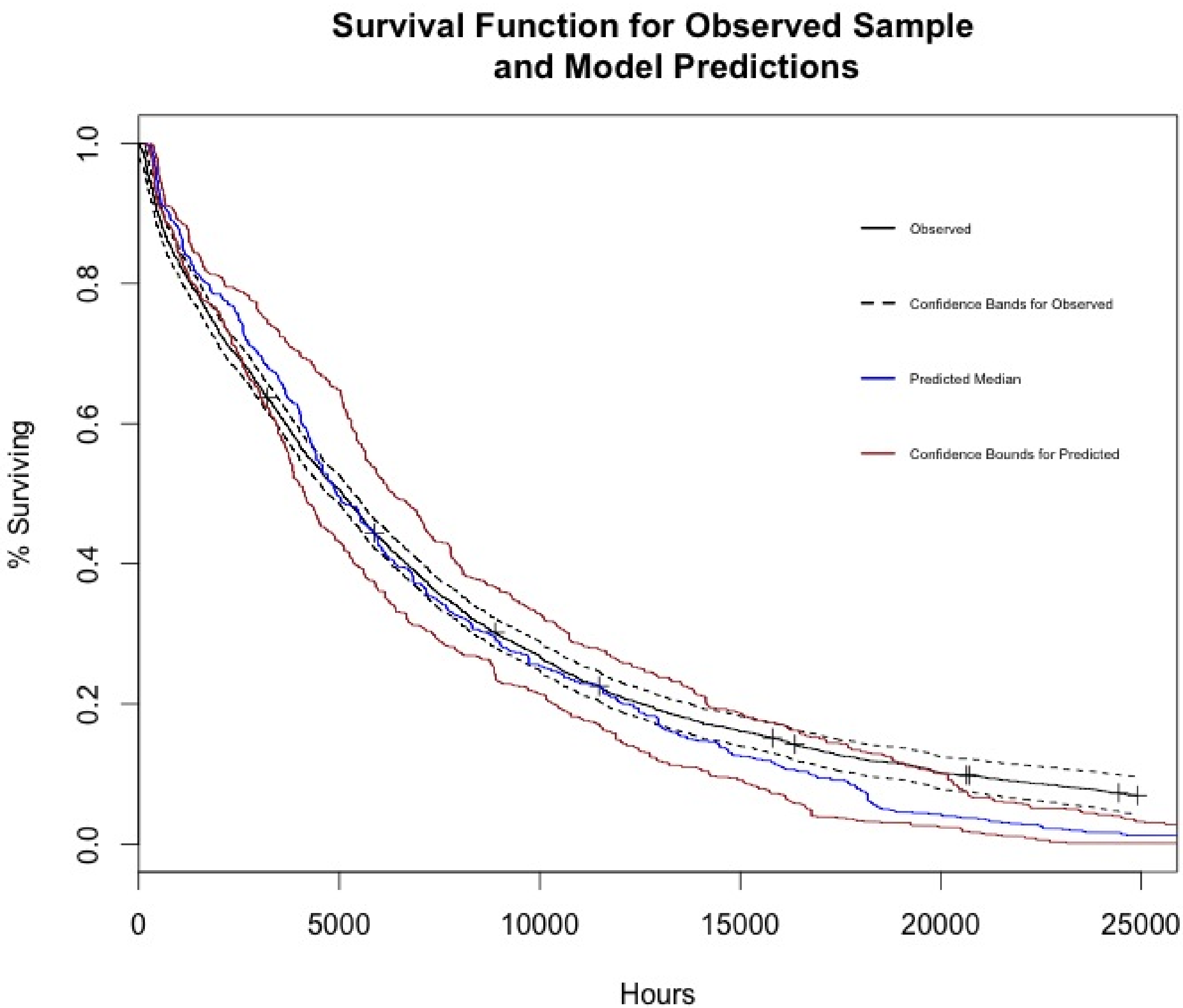
4.3 Full Sample Model Predictions

The previous subsection presented estimates of the survival function for "average" lamps of various categories and compared to the actual lamps in those categories. This subsection considers the fit of the full sample. As discussed previously, the challenge is determining what percentile to predict for each lamp. An additional challenge is calculating confidence intervals. Because of these challenges, I present two results for the full sample.

First, the predicted distribution of the median. That is, for each lamp, I estimated the predicted median life of the lamp and used that as an estimate of the survival function. This is in fact, not quite the same as the survival function, though, as it aggregates the distribution for each lamp at the very beginning, and so misses the full distribution, especially any effects from long tails. As such, it will over-estimate the survival function for low values, as it treats lamps that are more likely to be from the low percentiles of the distribution as median lamps, and under-estimate the tail of the distribution, treating high percentile lamps again as medians. The results are shown in Figure 16. The expected features for the early and late failing lamps are clear, but the predicted median function tracks the empirical survival function remarkably well from about 55% to about 20%. The confidence intervals for the predicted values are point-wise confidence bounds for the median drawn from the bootstrap results. That is, they represent the median lamp life at each percentile for the 5th and 95th percentile distributions of possible lamp lives from the bootstrap.

Perhaps the most striking feature of this analysis is the fit at the median. The estimated median of the median distribution is 4920 hours, just 89 hours off from the true sample median of 5009 hours. **That is an error of 1.7%.**

Figure 16:



Survival Function for Observed Sample and Model Predictions
Observed
Confidence Bands for Observed
Predicted Median
Confidence Bounds for Predicted
% Surviving
Hours
1.0
0.8
0.6
0.4
0.2
0.0
0
5000
10000
15000
20000
25000

The second analysis is an estimate of the full survival function of the sample. This was constructed by calculating the predicted value for each lamp for each percentile of its distributions. That is, it represents a estimate of the survival function of the sample. The major drawback is that bootstrap confidence bounds are not practical to calculate for this due to the number of computations required. Instead, I rely on the invalid, but easily calculable asymptotic confidence bounds. These are meant only to give a rough sense of the relative uncertainty of the estimates and are almost certain to underestimate the true confidence bounds. Results are shown in Figure 17. The fit is almost exact beyond the 35 percent surviving, and somewhat conservative before that. The Kaplan-Meyer survival function estimate does appear to be within the pointwise 90% confidence bounds approximately 90% of the time and would almost certainly be with the appropriate (but unavailable) bootstrap confidence bounds. The fit at the median of the distribution is not quite as tight as the previous estimate. The estimate of 4550 underestimates the true value, but the true value is well within the understated confidence bounds.

5 Model Predictions

This final analytical section presents the model predictions for the lamps offered by SCE through its upstream lighting program in 2010-2012. That is, unlike the previous section that matched model predictions to lamps that were actually observed in the sample, the results of this section are predictions out of the sample to lamps that were promoted through the SCE portfolio, but not directly represented in the laboratory study. The estimates are all based on the model developed and analyzed in the previous sections, cycling times from the KEMA (2005)[12], and lamp characteristics are based on data from SCE for the population of lamps supported through the upstream lighting program. Two types of predictions are reported. First, predictions for all lamps by room type; second, predictions by lamp type based on the average cycling time.

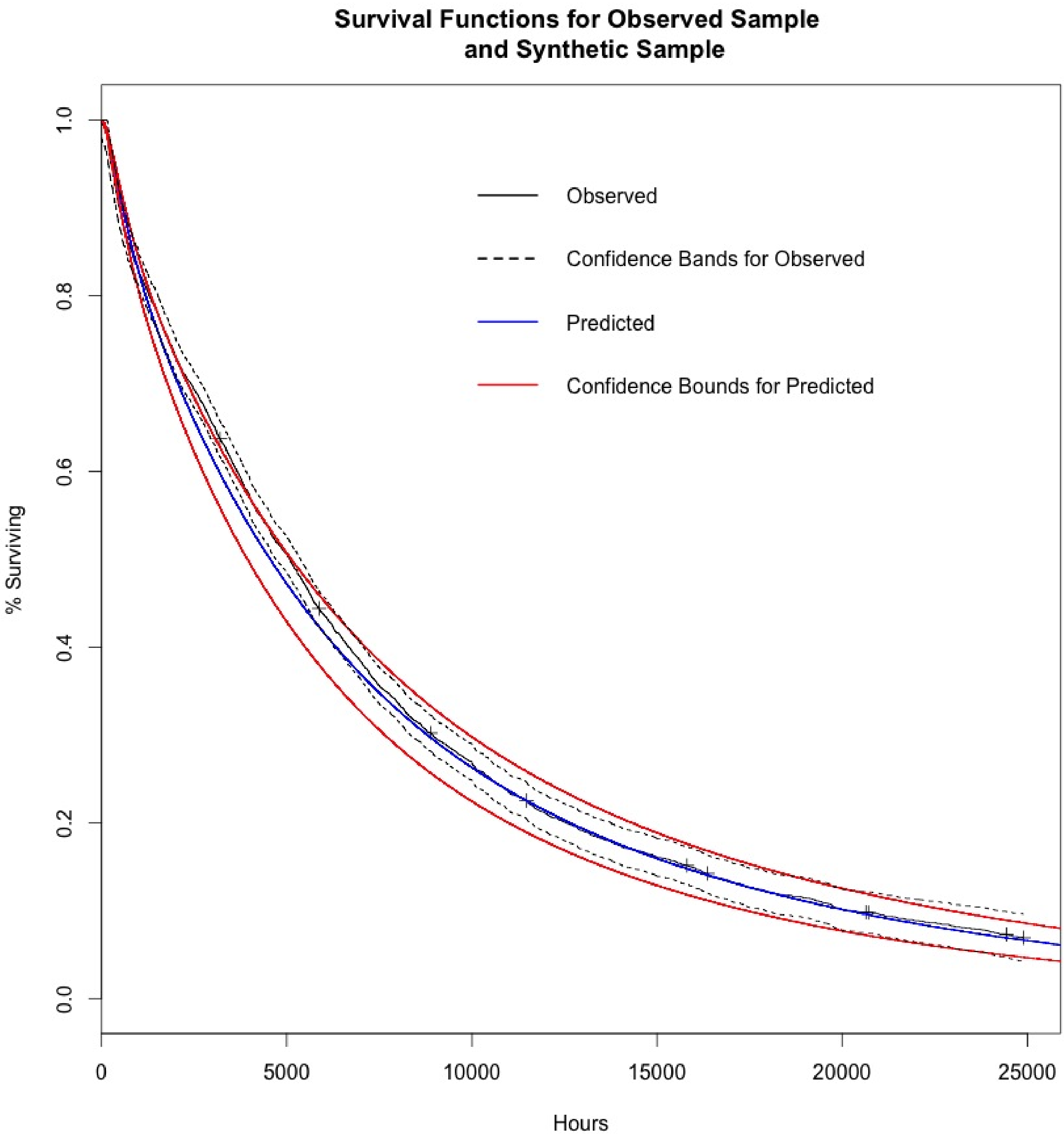
Although more recent lighting studies have measured lighting usage in California, none of the more recent studies have produced estimates of average cycle times that are reliable. As the model is based on treating the cycling times used in the laboratory study as categorical variables, it isn't possible to directly estimate lamp life for intermediate cycling times. Instead, I have used a linear interpolation between the nearest study cycling times. All values are weighted values, using the number of lamps of each type in the population in the weighting. As before, the confidence bounds are based on the bootstrap. The median is the value of interest and the mean is presented only for reference.

In general, estimates are larger for these predictions than for the values from the laboratory study because the sample for the study was quite different from the program lamp population. Specifically, there was a shift away from standard lamps into specialty lamps, which tend to have a longer life.

The columns labeled "Mean Median" in Tables 4 and 5 provide the weighted mean estimate of

[12] "CFL Metering Study". Kema-Xenergy. February 2, 2005. Available at: http://calmac.org/publications/2005_Res_CFL_Metering_Study_Final_Report.pdf

Figure 17:



Survival Functions for Observed Sample
and Synthetic Sample
Observed
Confidence Bands for Observed
Predicted
Confidence Bounds for Predicted
% Surviving
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000
25000
Hours

the median life of each lamp. The columns labeled "Median Mean" provide the weighted median value of the mean life of the lamp. In this case, "Mean Median" is the more appropriate measure as the program supported a very large number of lamps (over 2.5 million), but a relatively small number of distinct lamp models (48).

5.1 Lamp Life by Room

Results in this subsection are for all lamps in various rooms based on the average cycle time for that room. That is, it is not an estimate specifically for the types of lamps that tend to be used in any given room, but for the program population, using the cycling time associated with a specific room. An improved estimate could be made if the distribution of lamp types by room were available, but unfortunately I did not have access to data that would provide that distribution. The results are shown in Table 4.

Table 4: Lamp Life by Room

Room	Ave. Cycle Time	Mean	Median Lower Bound	Mean Median	Median Mean	Median Upper Bound
All Lamps	32	9,241	5,201	7,273	7,640	12,529
Bedroom	29	9,290	5,820	7,311	7,629	9,600
Bathroom	11	9,598	4,472	7,553	4,319	16,280
Family Room	51	10,794	5,806	8,495	11,018	17,124
Garage	41	9,053	5,647	7,125	7,363	9,400
Hallway	27	8,816	5,476	6,938	7,097	9,199
Kitchen	38	8,384	4,467	6,598	7,440	14,234
Living Room	68	19,984	11,549	15,728	20,964	23,512
Laundry Room	16	6,210	3,684	4,887	4,610	6,904
Other	26	8,579	5,307	6,752	6,831	8,996

Results based on room cycle times from KEMA (2005)
Values based on all lamps, not room-specific
90% Confidence Interval Based on Bootstrap with 9,999 Replications

5.2 Lamp Life by Lamp Type

A more meaningful breakdown is the lamp life by lamp type. For each of the estimates, the predictions are based on the average cycling time for all lamps from KEMA (2005).[13] Again, that means the estimate is not based on any differential usage between lamp types. There is quite a marked difference between basic spiral lamps and specialty lamps. While this is true for each category of specialty lamps, it is particularly true for high-wattage lamps, i.e. those with wattage greater than 30 watts. Thus, the lamp selection of the program compared to the lamp selection for the study sample has driven the increase in the median life. As discussed previously, the estimates for

[13] Ibid.

Table 5: Lamp Life by Lamp Category

Lamp Category	Mean	Median Lower Bound	Mean Median	Median Mean	Median Upper Bound
All	9,241	5,201	7,273	7,640	12,529
Basic Spiral	5,142	3,417	4,047	4,891	5,155
All Specialty	9,555	5,337	7,520	7,802	13,093
Specialty Shape	8,005	4,715	6,300	6,774	10,490
Specialty Controls	5,609	3,365	4,414	5,568	6,060
High Wattage	11,653	6,192	9,171	12,586	16,630

Results based on overall average cycle time of 32 minutes from KEMA (2005)
90% Confidence Interval Based on Bootstrap with 9,999 Replications

specialty controls may overstate the improvement compared to standard lamps.

6 Conclusion

The goal of this study was to estimate the technical life of CFLs based on data from a laboratory study. Numerous challenges presented themselves: the highly nonlinear response of CFL life to various characteristics, especially cycling time; how to project out of sample when the sample was designed for statistical power rather than as a representative sample; how to use the study results efficiently for predicting out of the sample when a key characteristic (robustness to cycling) is not directly observable; how to deal with significant variation in lamp life between lamps with the same characteristics and usage profiles; how to estimate confidence intervals; and how to deal with estimation of medians, among others.

These questions and challenges were addressed through the means of a two-step survival analysis and Wild bootstrapping. The process allowed me to use highly useful cycling data to form a better model while still making predictions about non-study lamps. It allowed me to maintain the powerful experimental design characteristics while estimating valid confidence intervals. It allowed me to reflect information effectively about both the lamps that failed in the study and the ones that did not fail. And it facilitated easy estimation of median values.

My recommended values for estimates of technical lamp life from the SCE program are the values found in the “Mean Median” column. They represent the expected value of the median lamp life, which is the closest estimate to the desired value of the true median of the population. A direct estimate of the median of the full distribution would, while technically possible, be practically unfeasible due to the computational intensity of the process.

7 Recommendations

Recommendations from this research are as follows:

- SCE should adopt the "mean median" lamp hours for the lamp types from Table 5, reproduced as Table 6, as planning estimates for lamp EULs.
- SCE should use the model results to estimate lamp lives for the program populations for newer populations, and use those values for planning purposes.
- The Energy Division should undertake an updated lighting metering study to gather newer data on average cycle times and hours of use, and how they correlate with lamp conditions.

Table 6: Recommended Lamp Life by Lamp Category

Lamp Category	Recommended EUL
Basic Spiral	4,047
Specialty Shape	6,300
Specialty Controls	4,414
High Wattage	9,171

Results reproduced from Table 5

A Appendix

No formal comments or questions were received regarding this study following the public presentation on April 7, 2015. However, a substantive concern was raised during the presentation regarding the influence of lamp wattage on predicted lamp life. The concern was that there were a small number of high wattage lamp models in the study in comparison to the prevalence of high wattage lamps in the program population considered. This Appendix addresses that concern and assess the robustness of the model with regard to wattage. First, I will clarify the concern, and then I will present three robustness checks.

The specific concern presented was that there was only one high wattage lamp model in the sample and the results for the high wattage lamps were due to that one lamp model. There are two important clarifications to make. The sample actually contained three lamp models with wattage greater than or equal to 30 watts, as well another at 29 watts, two at 27 watts and six at 26 watts. That is, there were three lamp models in the high wattage category, as well as nine others with wattages near the high wattage range.

Additionally, the idea that the projected life of lamps in the high wattage category is determined only by the sample for the high wattage category is mistaken. The effect of wattage estimated by the model is due to the gradient of lamp survival time across the full range of lamp wattage while controlling for other lamp characteristics. Said another way, the effect of wattage on lamp life in the model is determined by differences between the survival times of lamps across all values of the wattage. A model that was built on cell averages would have this problem, but the purpose of this study was specifically to avoid that problem. That being said, the external validity of the model does rely on having sufficient variation in lamp characteristics in the sample.

A first robustness check is to include higher order terms (squares and cubes) for wattage in the model. While the base model used in this study includes a quadratic term in the cycling model, it is only first order in the survival time model. Adding higher order terms reduces the clarity of interpretation of effects, but it controls those effects more flexibly, allowing for more complicated patterns of influence and approximating other types of non-linearities.

To investigate this effect, I introduced second- and third-order terms in the survival time model and a third-order term in the cycling model. While this did reduce the point estimates of the predicted lamp lives, a Wald test of the hypothesis that the higher-wattage model was different from the base model used in the study was highly insignificant (with p-values above 0.99). Although higher-order terms do provide additional explanatory power, the overall model is not significantly different and runs the risk of over-fitting the sample.

A second robustness check is to evaluate a "placebo" model of the effect. The fundamental concern is that, although there may be an important impact of increased wattage over some ranges of values, this effect does not apply throughout the entire range of wattages, notably the upper end. If this is the case, then replacing a lamp in the upper end of the wattage range with a lamp that is identical except for a higher wattage, would have no effect on the survival outcome. But, it would attenuate the effect of wattage in the model.

To test this effect, I compared the outcome in the base model to a model in which wattage of all lamps with wattage greater than 25 was increased by five. A t-test of the difference in the wattage coefficients was insignificant, and a Wald test of the equality of the overall models was highly insignificant.

A third and final test is to implement a weighted model in which the weights of the lamps in the sample are scaled to be in proportion to the lamps in the population. This will put more emphasis on specialty lamps in the sample, at the expense of basic spiral lamps that are much less prevalent in the program population considered, with the goal of having the model reflect the population more directly.

A test of the equality of these models was significant, but the weighted model actually increased the predicted lamp life for lamps in the population. For this reason, and because the model is less general, I don't recommend using these results, favoring instead the results presented in the main body of the study.