Measure Lifetime Derived from a Field Study of Age at Replacement

David Robison, MicroGrid
David Cohan
Bruce True, Portland General Electric

The traditional engineering technique for estimating the expected lifetime of a measure is a survivorship study. For this type of study, a cohort of units is followed through time to derive the median age at the time of failure. Unfortunately, this methodology does not lend itself well to field study of long lived measures—such as water heaters, HVAC units, etc.—because a complete study would require many years to obtain results. Longitudinal studies of cohort groups are usually only able to track survivorship for a few years and then must extrapolate the failure curve to the remaining years. This overlooks the potential for accelerated failure toward the end of the lifetime.

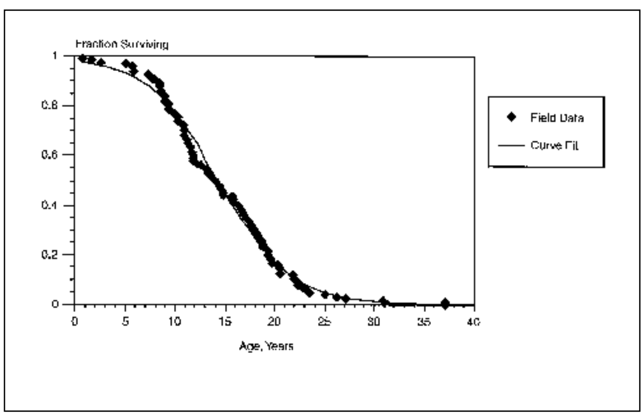
With our alternative approach—the age distribution approach—field data are collected in a cross-sectional study to provide the median age at time of replacement. This provides equivalent information to a survivorship study if care is taken to recognize the underlying assumptions regarding uniform technology as well as population size and stability.

Our field study collected age at time of replacement for a sample of residential water heaters and used the distribution of ages to estimate the mean expected lifetime. Adjustment for population changes was applied during the analysis. Uniformity of the measures was assessed through a survey of manufacturers. The technique is widely applicable to investigating the lifetimes of long lived energy efficient appliances and could help to improve the data used for program planning.

INTRODUCTION

Engineers usually estimate product lifetimes by compiling a survivorship curve such as the one shown in Figure 1 (Nelson 1982). The median expected lifetime is usually considered the average lifetime for computing cost-effectiveness (Short 1986). In other words, it is thought that half the units will survive for longer periods than the median expected lifetime and half will survive for shorter periods. Survivorship curves have been estimated for solar and heat pump water heaters (Robison 1987), tank wraps, and other conservation measures (Bordner et al. 1993). The lifetime estimates for water heaters have been based on industry opinion rather than field data (McMahon and Xi 1994). Since appliance lifetimes may approach 20 years, long-term studies are not practical. Studies have been limited to data collected over about 5–10 years. This paper proposes an alternative methodology for estimating product lifetimes and applies the methodology to estimate the lifetime of residential water heaters.

Figure 1. *Survivorship Curve*

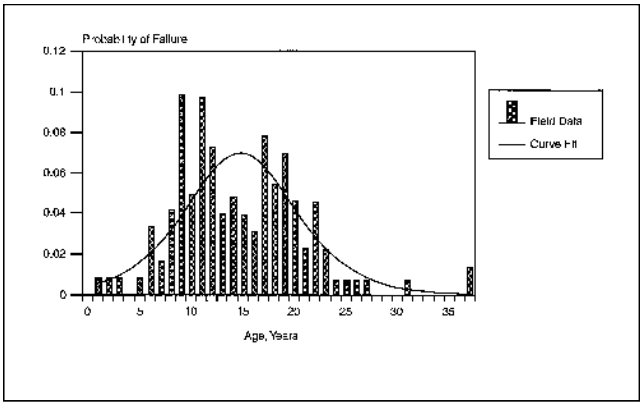


METHODOLOGY

Figure 1 presents a survivorship curve as the cumulative probability of a unit surviving over an extended period of time. Figure 2 shows that the same information can be presented as the probability of failure in a specific year. In survivorship analysis, this distribution is called the Hazard Function. The median point of the distribution in Figure 2 is the same as the mean time to failure in Figure 1.

The distribution in Figure 2 does not have to be based on a longitudinal study. A similar probability distribution of age at time of death can be collected looking at a sample of units replaced over a short time. This characteristic pro-

Figure 2. Probability of Failure in a Given Year (The Hazard Function)



vides a practical technique to develop survivorship information without requiring a long-term study.

The application of cross-sectional results as representative of longitudinal results involves at least two implicit assumptions:

(1) The technology involved is assumed to be uniform over the time period studied.

(2) The relative proportions and size of the underlying population are assumed to be uniform.

In our study, we carefully considered both assumptions to ensure that the extrapolation of cross-sectional results is valid. We verified the first assumption—that the technology of water heaters has not changed in ways that would affect lifetime—by surveying manufacturers and asking their opinions. The consensus was that the manufacture of water heater tanks has not changed significantly over the last several decades. This assumption might not hold true in the future if, for example, the standard technology changes to plastic tanks.

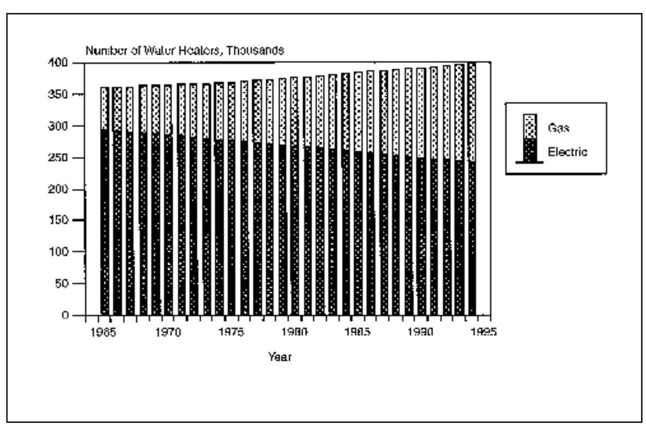
For the second assumption, we first note that the cross-sectional data take the failure rate of an age group as representative of that particular age cohort. For example, suppose that we gather information showing that 5 of 100 water heaters were replaced at year 10 of their life. The probability of failure is then 5% of the fraction of the water heaters that were installed ten years ago. Of course, we do not know how many water heaters were installed ten years ago. However, we only need to know the relative number of installations to estimate the probability of failure. That is, if we assume that the installation rate has been a constant function of the total stock of water heaters and that our sample of 100 units is representative over the entire period considered, we can still assume that the 5% failure rate is appropriate for year 10.

With this approach, one implicitly assumes that the underlying population has remained stable. However, with water heaters we know that the customer population has grown: there are more customers with newer homes or younger installations. Therefore, when examining the water heater ages at the time of failure, young ages need to be made less frequent and older ages need to be made more frequent. That is, if we have five units replaced at year5 and five units replaced at year 10, the relative frequencies are not the same. The five units replaced at year 5 represent a smaller fraction than the five units replaced at year 10 because there were more water heaters installed at that point in time. Adjusting for the change in the underlying population shifts the age distribution curve outward, and the mean age at the time of failure is extended a couple of years.

Although a population adjustment might sound complicated, in practice it is not difficult because the distribution median is relatively robust compared with changes in the underlying population. In other words, the median result is not highly sensitive to assumptions regarding population change. Thus, it is unnecessary to know the exact population change to apply an approximated adjustment.

To demonstrate how a population adjustment is carried out, Figure 3 shows the estimated relative numbers of gas and electric water heaters in the Portland General Electric service territory over the last 30 years. These data were provided by Portland General Electric planning staff. Although the total number of residents has grown slowly, the number of electric water heaters has declined, and the number of gas water heaters has increased dramatically. The individual failure probabilities in any specific year are weighted by the ratio of the population in that year to the mean population.

Figure 3. Market Share of Gas and Electric Water Heaters Over the Last 30 Years



The adjustment factors derived from the population estimate are shown in Figure 4.

RESULTS

We discussed the results under two headings: collection of field data and data processing.

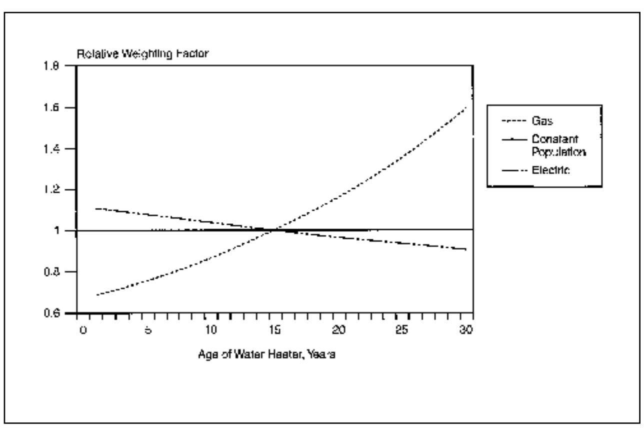
Collection of Field Data

For this study, we arranged for plumbers to log the model and serial number of old water heaters that were being replaced. Over a period of about four weeks, we collected information on 305 water heaters. We then attempted to identify the date of manufacture of the replaced units from their serial numbers. The date of manufacture provided a reasonable proxy for the date of installation. Although manufacturers had no information on how long units are retained in inventory, they indicated during a telephone survey that units are typically installed within three months of their manufacture.

The date of manufacture is encoded in the serial number. By contacting the manufacturers, we were able to decipher the code and definitively calculate the age of 147 of the 305 water heaters. We were also able to make reasonable assumptions about the age of some additional units. We ended up with a total of 164 observations on electric water heaters and 39 observations on gas water heaters.

Several factors complicated the identification of the model numbers. The water heater industry has gone through a long period of mergers and consolidation. Several of the major brands available ten or more years ago are no longer manufactured, and we were not able to interpret these manufacturing codes. Sensitivity analysis indicates that these ‘‘orphan’’ brands do not bias the results. Although the orphan brands were older than the current brands, they were also cheaper and less durable. Thus, there is no indication that the lifetimes of the orphan brands differ from the median lifetime. In some cases, the plumbing contractors were not able to read the numbers. With some recently manufactured units, the industry changed from stamped metal tags to tags printed with a dot matrix printer, and the new tags were often difficult to read. About half of the serial numbers that were not useful came from the orphan brands. The remaining serial numbers were either not readable or were misread in the field.

Figure 4. *Weighting Factors Derived from Population Estimates*



Of the useful observations, 7 of 164 electric water heaters were being replaced even though they had not yet failed totally. In these cases, the participants expected their water heaters to fail soon, so elected to replace them even though they were still working. These observations were included in the analysis because a small percentage of water heaters are typically installed before their total failure (e.g., during remodeling). Thus, there was no rationale to exclude these observations, and their inclusion did not change the average age.

We believe that the serial numbers provide a conservative estimate of age. The older the tank, the more likely that the manufacturer had gone out of business, and therefore we were unable to interpret the manufacturer’s codes. Thus, many older tanks were likely excluded from the analysis.

Data Processing

Under data processing, we consider four subjects: population adjustment, curve fitting, hypothesis test, and survivorship results.

Population Adjustment. Electric water heaters needed a small population adjustment, and gas water heaters needed a relatively large population adjustment. The adjustment factor was the ratio of the underlying population size to the constant population size. For convenience, the average population was taken as a constant, and the age frequencies were weighted by the ratio:

$$w_t = \frac{Population_t}{Population_k}$$

where

w_t	=	the adjustment factor for each time period t
$Population_t$	=	the underlying population at any point in time
$Population_k$	=	the constant population, in this case the average population during the time period

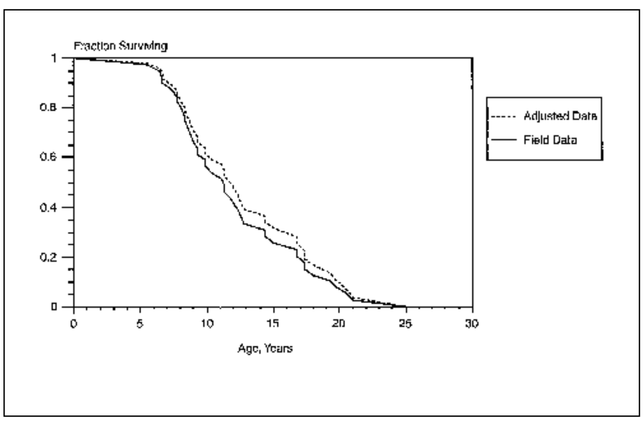
For a growing population, the failure probability observations related to older units are weighted more, and the failure probability observations related to younger units are weighted less. This approach corrects for skewness due to demographic trends. The adjustment factors used for this analysis are shown in Figure 4. In the case of electric water heaters, the population adjustment was not significant because the changes in the number of water heaters were relatively small compared with the large population base. However, the same was not true for gas water heaters.

Figure 5 shows the effect of the adjustment on the survivorship curve for the gas water heaters. The survivorship curve shifts slightly toward longer lifetimes. For this analysis, the adjustments shown in Figure 5 were applied to the field data before applying the logistic curve fit. This procedure differs from survivorship analysis in common statistical packages, which assume that each lifetime observation counts as an single, integer event and which do not allow for population weighting.

Curve Fitting. Both Figures 1 and 2 were based on field data collected for electric water heaters during this study and were designed to demonstrate variations in real world data. When sample sizes are relatively small, random errors affect the observed number of failures within any one year (see Figure 2). However, the survivorship curve is a useful way to present results because it smoothes variations in the observed field data (see Figure 1).

The survivorship curve typically follows a logistic distribution which lends itself well to analysis. A logistic distribution is defined by the odds ratio (i.e., the ratio of the survival probability to the failure probability). For any point in time, the odds ratio can be derived as:

Figure 5. *Illustration of Population Adjustment for Gas Water Heaters*



$$odds\ ratio = \frac{survival\ probability}{(1 - survival\ probability)}$$

With a logistic distribution, the odds ratio log is a linear function of time. Thus, a linear regression model provides an appropriate curve-fitting line of the form:

$$Ln\ (odds\ ratio) = \alpha + \beta + age + \epsilon$$

To illustrate, the logistic linear regression is shown in Figure 6. This linear regression provides the curve fit shown in Figures 1 and 2. One advantage of a logistic curve is that it provides an easy way to compute the median. The median is the point where half the population is surviving. The odds ratio of the median is .5/(1.5) or one. The log of the median is zero. Thus, the median occurs at a time equal to the regression intercept divided by the negative of the regression slope.

The confidence limits are slightly more complicated to compute. The confidence interval for estimated log odds ratio is given by:

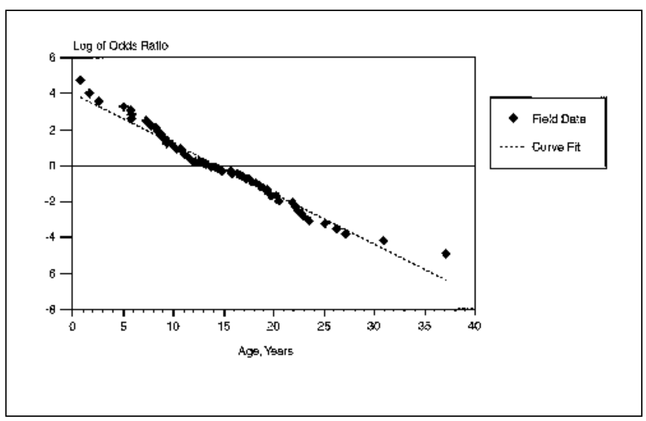
$$confidence\ limit = Z * SE(\alpha/\beta)$$

where

Z	=	standard deviation score (1.645 for a 90% confidence limit)
$SE(a/b)$	=	standard error of the estimated median

Unfortunately $SE(a/b)$ has no convenient closed analytical form. It cannot be calculated algebraically; instead, it must be determined through complex matrix calculations. To avoid this complexity, we substituted a procedure based on the standard error of the estimated function, S_F, which is evaluated for specific points on the time axis and is given by:

Figure 6. *Logistic Linear Regression*



$$S_F = \sqrt{s^2 * \left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}\right]}$$

where:

S_F = standard error of the estimated function
s = standard error of the estimate from the regression
N = number of observations
x_0 = the specific value of the independent variable being examined
$\bar{x}$ = the mean value of the independent variable in the regression
x_i = the independent variable for i = 1 through N

It is easiest to iteratively solve for those values of the independent time variable that provide a 50% survival probability including the upper bound on the log and a 50% survival probability including the lower bound. For the data sets in this study, the difference between the upper and the lower confidence limits was very small (well below the precision of the data measurements). Thus, a single plus/minus value was reported for the confidence limits.

Hypothesis Test. The first task of the analysis was to determine whether the two types of water heaters had different survivorship curves. For this hypothesis test, we applied a t-test of category variables. Our observations were pooled into a common regression with dummy variables to represent the different fuel type categories. This regression showed that the two types of water heaters had different survivorship curves. The difference in the curves was significant at the 99% probability level. Thus, we conclude that gas and electric water heaters follow two different survivorship curves.

Survivorship Results. Survivorship curves were computed for the two types of water heaters, as shown in Figure 7. The regression results are summarized in Table 1.

Figure 7. *Survivorship Curves for Gas and Electric Water Heaters*

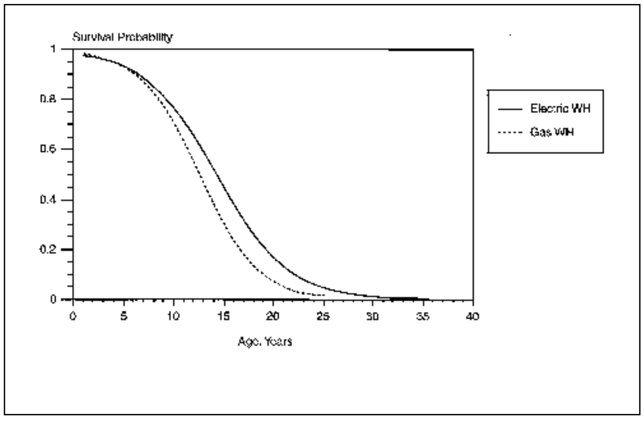


Table 1. *Summary of Measure Life Study*

Water Heater Type	n	Median Lifetime	90% Confidence Range
Electric	164	14.3	±0.6
Gas	39	12.6	±2.1

CONCLUSIONS

A cross-sectional study of age at time of replacement can be used to estimate product lifetime. However, one must be aware of the assumptions implicit in this method, namely, that the technology has remained stable and that the underlying population has remained stable. If the population has been changing, it is possible to adjust the data for those changes.

The recommendations that emerge from this study are:

- Plan to have a completed sample size of about 100 for adequate confidence limits on the results.
- Confirm with manufacturers how they code their serial numbers. Our vendor logs would have been more complete had we provided specific instructions to describe the digits needed to identify a valid serial number.

ACKNOWLEDGMENTS

This study was supported by Portland General Electric.

REFERENCES

Bordner, Robert, et al. 1993. ‘‘Ten Years Later: Are the Measures Still There and Producing Savings?’’ *Energy Program Evaluation: Uses, Methods, and Results*. Chicago: National Energy Program Evaluation Conference, CONF-910807, August.

McMahon, James, and Xiaolin Xi. 1994. ‘‘Uncertainty Analysis of Life-Cycle Cost: Residential Electric Heat Pump Water Heaters.’’ *Proceedings: ACEEE 1994 Summer Study on Energy Efficiency in Buildings*, pp. 8.131–8.139. Washington, D.C.: American Council for an Energy-Efficient Economy.

Nelson, Wayne, 1982. *Applied Life Data Analysis*. New York: Wiley & Sons.

Robison, David. 1987. *Cost and Performance of Solar and Heat Pump Water Heaters in the Pacific Northwest*. Bonneville Power Administration, June 4.

Short, Walter. 1986. *A Method for Including Operation and Maintenance Costs in the Economic Analysis of Active Solar Water Heating Systems*. Golden, Colo.: Solar Energy Research Institute, SERI/TR-253-2616, April.