Incorporating real-world data into measure lifetime estimates: How long does energy efficient equipment really last on-site?

Rose Woods
Skumatz Economic Research Associates, Inc. (SERA)
USA

Lisa A. Skumatz, Ph.D.
Skumatz Economic Research Associates, Inc. (SERA)
USA
skumatz@serainc.com

Keywords

measure lifetimes, EULs, measure retention, savings lifetimes, energy efficient measures, estimated program savings, technical degradation

Abstract

In work for a large west coast agency, the authors conducted a detailed assessment of more than 100 measure retention / lifetime studies to identify the real-world lifetimes for numerous energy saving measures used in: 1) Residential, 2) Low income, 3) Commercial / industrial, 4) agricultural, and 5) military buildings. Estimated useful lifetimes (measure lifetimes), in conjunction with energy savings estimates, are the key elements in computing energy savings for a program or intervention. Many of these EULs had not been updated for more than a decade.

The study provides quantitative information on: 1) updated measure lifetimes for scores of commonly-installed program measures; 2) information on measures for which reliable lifetime information is not currently available; 3) information on technical degradation studies, and 4) best practices for measure lifetime analyses.

For each study, the authors conducted an exhaustive review of program measures, sampling methodologies and approach; field work; data validation; and analysis steps to determine whether reliable lifetime information could be gleaned from the study. The paper summarizes these results and provides information on updated / recommended EUL estimates for energy efficiency equipment. The implications related to gaps in available EUL information – especially as it relates to key equipment – and reliable estimates of EUL values for planning for future programs are presented.

Background on Measure Lifetimes / EULs

Measure lifetimes, or "expected useful lifetimes" (EULs) represent the median number of years that an energy efficiency measure can be expected to remain in place and operational in the field. This figure represents a critical component in the computation of total energy savings that can be attributed to the measures installed under a program – the shorter the assumed lifetime, the smaller the savings. Energy savings calculations are used in both program planning and program benefit cost analysis. The calculation of gross[1] program energy savings, in simple terms, is the multiple of the number of measures installed under the program times the kilowatt hour savings per measure annually, times the number of years the measure can be expected to perform (the expected useful life) times any technical degradation factors relevant (a factor representing the decay in the performance of the measure over time).[2] The

1. Gross savings are the saving installed under the program, using either estimated, deemed, or other savings for the measures. The concept of net savings brings in another critical link, the attributable savings, or those savings that are due to the program, above and beyond what would have happened had the program not been in place. This involves a number of refinements in the computations. For example, a detailed impact evaluation or metering work may be needed to verify the gross savings actually delivered by the measures installed. Then, a net to gross (NTG) analysis is needed (see Skumatz, et.al. 2004 for a detailed discussion of techniques for this analysis). NTG work uses survey techniques, difference of differences, or other methods to identify 1) "free riders", or those participants that would have installed the measure even without the influence of (or incentive from) the program (decreases savings attributable to the program), and 2) spillover, or additional savings the program did not "count" or incentivize that its influences caused in the marketplace indirectly. Both participants and non-participants can contribute to spillover, which tends to increase the attributable savings. The combination of these two effects represents the NTG ratio, the percentage applied to program-installed gross savings that are attributable to the program. While an important and very interesting topic, this NTG topic is auxiliary to the paper's focus on measure lifetimes.

2. For dollar (or Euro) savings over a life, a discount rate may be applied; for energy savings in kilowatt hours, they would not be.

longer the lifetime – all else held constant – the higher the attributable energy savings for the program. Hence, this figure is an important component of computations for planning and benefit-cost purposes. [3]

In particular, the lifetime estimates are also inputs into the calculations of energy savings that underlie award of shareholder returns for investor-owned utilities (IOUs) in California – the return that utilities and their shareholders get as incentives for for investing in energy efficiency programs. For the California computations, the *a priori* lifetimes and associated validation study schedule and methods are prescribed by the State regulatory agencies.

In several projects, the authors were tasked with reviewing more than 100 EUL / validation studies that were performed as part of the evaluation process "protocols" required by the State.[4] We reviewed the studies to:

- Assess conformance with prescribed protocols, and review the methods, quality, and the justifiability of results. The analysis allowed us to identify methodological problems with past studies and develop "best practices" for the conduct of EUL studies.
- Assemble the EUL results for the various measure types and identify measures that require updates to the assumed measure lifetimes. The results of this review were used to update the EULs for a large database used extensively in planning and evaluation protocols.
- Identify measures for which insufficient retention study information is available.

The studies that were analyzed covered measures for a wide variety of programs in the residential, low income, commercial, industrial, agricultural, and military sectors. For each study, the consultant team conducted an exhaustive review of program measures, sampling methodologies and approach; field work; data validation; and analysis steps. The results have implications for others conducting retention analyses, as well as those considering or revising protocols or standards related to these studies.

For each study, the consultant team conducted an exhaustive review of:

- Program information, share of program savings covered by the analysis, measures included, and other topics related to justification and context for the studies
- Sampling methodologies, sample quality and justification, quality of field work – including data collection approach, treatment of sample, quality of program records and field work practices; and
- Data validation and verification, treatment of sample attrition and sample, statistical approach, consideration of alternative models and treatments, and the justifiability of the results reported.

We reviewed retention studies and scored them based on: 1) conformance with prescribed protocols for retention analysis; 2) sample characteristics, size, data collection procedures; 3) modeling approach, estimation method and consideration of alternative models; and 4) results and implications

ASSESSMENT OF MEASURE LIFETIME STUDY "BEST PRACTICES"

The review of the many retention studies showed that the majority of the studies attempted to follow the guidelines and suggestions from the protocols, and the studies were generally able to provide reasonable and useful EUL estimates. However, there was significant variation in the quality and thoroughness of the studies. As a consequence, we were able to summarize several common mistakes in the studies [4].

SMALL SAMPLE SIZE.

The most common problem with the studies that we evaluated was an insufficient sample size. In some cases, a small sample size was the result of an inadequate data collection effort, and therefore easily avoidable. However, other studies worked from poor population lists (usually obtained from program tracking data). In such cases, sample size complications were far beyond the control of the research team.

Suggestion: Any possible effort should be taken to ensure a sufficient sample size. Inadequate samples can lead to several insurmountable analytical problems, from large confidence intervals to models that do not converge. Utilities need to maintain high quality lists – with a view toward evaluation and not just rebate or program invoicing applications. In addition, sample sizes need to vary based on the expected failure rate; those items with long lifetimes will likely need extremely large sample sizes to detect failures.[5]

FAILURE TO TEST OTHER MODELS

Another common mistake was the failure to test several models using different statistical distributions when estimating survival functions. Often, this occurred when a research team tested different functions for one measure, then applied that function

3. A reviewer notes several other issues associated with measure lifetimes and savings computations. They note that although a longer lifetime always provides more energy savings, the increment of those savings relative to the marketplace is decreasing as new, much more efficient, equipment is entering the market. Thus, if the increment of savings delivered by the program is decreasing, so they argue the exact lifetime, discounted by this factor, becomes relatively less important (especially for a long-lived measure). A discussion of factors related to baseline are found in Sebold, et. al. 2001. However, the point is that the measure has a long life and if the program had not induced purchase of an efficient measure, an inefficient measure would be in place over that lifetime. If the program reached the person at the point they were considering replacement, then the savings over the full lifetime are relevant. If the program led to early replacement, then the program may only properly be able to take credit for the full savings up to the point when the measure would have been replaced, and then only for an increment over the period after that. The reviewer also note that another methodological issue is the relationship between the program set up and the methods applied. Short run programs have much more difficulty in providing reliable estimates of total energy savings than long run programs, because in long run programs there is more time to check actual lifetimes, even without statistical extrapolation methods.

4. The study also examined "technical degradation" studies, which examined the pattern of degradation of incremental savings performance by measures over time.

5. For example, if multiple failures are desired to provide a reasonable chance for a model "fit" say 3 years after installation, and the lifetime of the measure is 15 years (median), and a "normal" curve is assumed for failures with a 5 year standard deviation, preliminary computations by the authors suggest it may take surveys (phone or on-site) of 450 sites to detect 2 failures. Different numbers of observations would be needed for different assumptions about measure lifetime, distribution, variance, and years after program installation, and the sample sizes are very sensitive to (unknown) standard deviations. The California Public Utilities protocols provide sample size and accuracy guidelines.

to the rest of the measures covered by the study. Because different measures act differently, the same model assumptions will not always be justifiable from one type of equipment to the next. This caveat is especially important when parametric models are being used. Although failure rates may accelerate with time for both refrigerators and air conditioners, they may not accelerate in the same way.

Suggestion: Statistical programs make it fairly easy to test log-logistic, weibull, gamma, and other distributions. Testing alternatives allows the researcher to identify the best "fit" for the failures over time, and improve the chance of identifying appropriate EULs. Failure rates for different types of measures, and for measures of different lifetimes may be expected to accelerate in different patterns, and the research should account for that.

AMBIGUOUS FAILURE DATES

There is often a tendency for inspections and surveys to fall short in their attempts to obtain approximate failure dates. Even if the exact date of failure is unknown, any additional information regarding when the measure in question failed can be used to narrow the censoring interval. Follow-up questions, such as year of failure, season or month of failure may produce more accurate responses than simply asking whether the measure is still in place at the time of the interview. Accurate failure date responses are easier to obtain when the measure being studied is more noticeable. Even the maintenance supervisor for a large and busy building is likely to know approximately when an energy management control system stopped working.

Suggestion: The best remedy for large failure date intervals when measures are small and numerous (such as light bulbs) is more frequent surveys – though this course of action can be expensive. However, if the measure is common or responsible for a large share of savings (and potentially a large share of earnings claims) the extra investment may be well justified. If dates cannot be recalled, follow-up questions that identify the season and year of failure or removal are essential. In addition, utility records must also clearly note installation date (which did not always happen).

POOR DOCUMENTATION

Reports need to clearly document the methods, procedures, and analyses conducted, and their justification. The biggest problem that we encountered in the review was documentation inadequate to determine exactly what procedures had been followed, hypotheses tested, modeling applied, coding adjustments made, weighting schemes used, etc. Some reports had included formulae that were not relevant to the models estimated. The documentation step is frequently overlooked but extremely important.

Suggestion: Simply put, the documentation must be sufficient to facilitate both (1) thorough understanding of the methods used to conduct the study (and justify methodological decisions), and (2) the conclusions drawn from the study. Regardless of whether a potential reader is reviewing the study for accuracy, to assess shareholder earnings claims, or simply trying to gain insights from its conclusions, documentation is critical.

FAILURE TO EXAMINE RESULTS IN CONTEXT

Very few studies looked "outside themselves". There are now many retention studies that have been conducted across the nation for a large number of measures (including previous studies for the same program in many cases).

Suggestion: Discussion of results is improved if results are reviewed and compared to other studies – earlier studies of the same program, or for programs in other locations – to identify similarities, patterns, and differences, and provide a context or benchmark for the findings.

"CONVERGENCES" ON ESTIMATED EULS

Many of the retention studies assessed programs that included more than one measure. As a result, the extensive review of over 100 studies conducted in this project examined EUL results for 301 energy-efficiency measures. The quality review allows us to "score" each study in terms of its quality; data, methods, and analysis quality all entered into the assessment. Studies that received a score of "A" to "C" (in a range of high of "A" to a low of "F") based on how well they conformed to "best practices" (above) were considered reliable sources of updated information about median EUL estimates. Figure 1 presents the EUL estimates for 20 saving measures that could be confirmed, justified, or revised based on the studies that we reviewed. The study provided information on 43[6] common measure types and variations, and the results computed by the authors were incorporated into the large California statewide DEER database[7] on incremental savings, incremental costs, and lifetimes.[8]

The analysis also pointed out several "gaps" in EUL analyses. Many of these gaps are being addressed by upcoming statewide research in California.

- CFLs and lighting measures are more complex to measure than other energy efficiency equipment because the lifetimes vary dramatically based on operating hours. As such, the retention studies / results are beginning to be revised to be stated in terms of operating hours, and detailed operating hour studies are being conducted.
- Many of the measures studied had only limited retention studies, or were only examined by retention studies that did not score well in our analysis. This is important because many of these measures are responsible for significant savings in programs across the State of California. Measures of this type include: air compressor equipment (9 equipment types, mostly industrial); air conditioning (8 types, including all sectors); cooking measures (2 types, commercial); controls and heating (9 measures, mostly commercial/industrial or agricultural); lighting (16 measures, many sectors); motors and pumps (25 measures, all sectors); process

6. The results covered more than 20 measures; some of the results were for specific measures within a type (ranges of efficiency for air conditioners, refrigerator variations). Some of the measures had very few studies, had only relatively poorly conducted studies, or had other issues that did not support conclusions regarding updated lifetimes.

7. DEER stands for the Database for Energy Efficient Resources, and links can be found under energy.ca.gov/deer/.

8. (1) Measures and results are presented without distinguishing between end use or sector. The end uses covered included: air conditioning; HVAC; clothes washers; EMS; lighting; motors, drives and pumps; process; and refrigeration. (2) The sectors covered included: commercial, industrial, residential, residential new construction and agricultural.

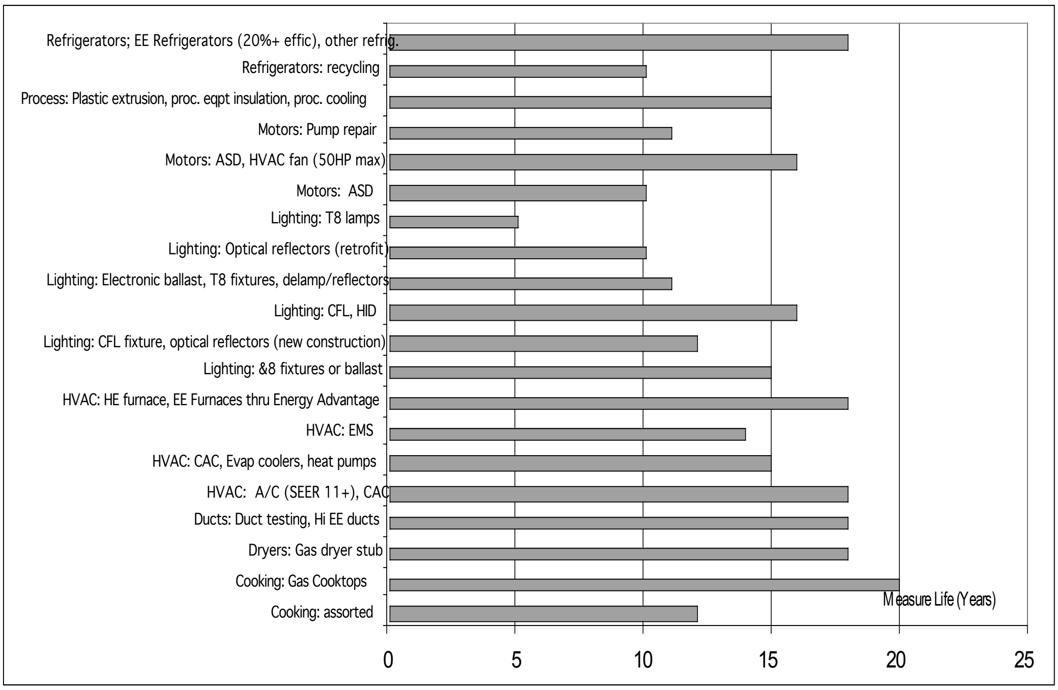

Figure 1: Summary of Final Lifetime Estimates of Energy Efficient Equipment – by End Use Type

related equipment (12 specific industrial measures); refrigeration (several); shell measures (including insulation and glass); and several other types of measures.

- The study found no retention studies that addressed measure lifetimes for a number of other measures included in the statewide DEER database. These include: air conditioning equipment (more than 16 specific measure types in all sectors); dryers, washers, and water-related measures (about 25 measures), cooking measures (5 types), controls and heating (17 measures), lighting (12 measure types), motors and pumps (6 types), refrigerators (dozens of measures), shell measures (16 measures) and several other types.[9]

Summary and Implications on Measure Lifetime Results

Measure lifetimes are a critical component of the computation of energy savings deriving from energy efficiency programs. Assumptions about these lifetimes – in combination with the per-measure savings – drive the benefits side of the benefit/cost ratios associated with programs and measures. This study provided a thorough and practical review of the more than 100 retention studies that had been conducted in California, and 1) identified strengths and weaknesses of studies, and developed "best practices" recommendations for this type of study; 2) used the results to provide updated measure lifetimes for key measures; and 3) identified those measures for which inadequate retention information is available, indicating additional EUL research is needed.

The research provided an opportunity to examine "best practices" in retention studies, and the paper lists itemized suggestions on sample sizes, modeling approach, data collection, documentation, and the importance of reviewing the results in context – both over time and compared to results from other programs including similar measures.

Most importantly, the results were used to provide updated EULs (provided in Figure 1), which are a critical component in the overall evaluation of realized energy savings.[10] The study also provided information useful in helping prioritize future research on measure lives. Those measures with inadequate studies to support reliable EULs included CFLs (research is underway), air compressors, many motors and pumps, process equipment, and a number of other measures used in energy efficiency programs in the US.

References

Skumatz, Lisa A. and John Gardner, "Revised / Updated EULs Based on Retention and Persistence Studies Results", Prepared for Southern California Edison by SERA, Superior, CO, July, 2005.

Skumatz, Lisa A., Daniel Violette, and Rose A. Woods, 2004. "Successful Techniques for Identifying, Measuring, and Attributing Causality in Efficiency and Transformation Programs", Proceedings of the American Council for an Energy Efficient Economy Summer Study on Buildings, Asilomar, CA.

Sebold, Fred, Alan Fields, Lisa A. Skumatz, Shel Feldman, Miriam Goldberg, Kenneth Keating, and Jane Peters, 2001. "A Framework for Planning and Assessing Publicly-Funded Energy Efficiency", Report prepared for Pacific Gas and Electric Company, San Francisco, CA.

9. Of course, the data in Figure 1 are based only on the limited number of retention studies available per measure. More information is provided in the full study, referenced as Skumatz and Gardner, 2005.

10. Although, as noted in previous footnotes, not the only component (technical degradation, NTG, baseline, etc.).