

MEASURING PERSISTENCE: A LITERATURE REVIEW FOCUSING ON METHODOLOGICAL ISSUES

Amy K. Wolfe
Marilyn A. Brown
David Trumble

March 1995

Prepared for the
Office of Energy Efficiency and Renewable Energy
U.S. Department of Energy

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
Managed by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
for the
U.S. DEPARTMENT OF ENERGY
Under Contract No. DE-AC05-84OR21400

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED *WV*

MASTER

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

TABLE OF CONTENTS

LIST OF TABLES & FIGURES.....	v
EXECUTIVE SUMMARY.....	VI
LIST OF ACRONYMS.....	xii
1. BACKGROUND AND PURPOSE.....	1.1
2. ISSUES FUNDAMENTAL TO PERSISTENCE AND ITS MEASUREMENT.....	2.1
2.1 DEFINITIONS AND SCALES OF ANALYSIS.....	2.1
2.2 PURPOSE OF STUDIES.....	2.9
3. RESEARCH DESIGN AND DATA COLLECTION ISSUES.....	3.1
3.1 ASSUMPTIONS AND CONTROL GROUP OPTIONS.....	3.1
3.2 RETROSPECTIVE VERSUS PROSPECTIVE DESIGNS.....	3.4
3.3 FREQUENCY AND TIMING OF DATA COLLECTION.....	3.5
3.4 SAMPLING AND ATTRITION.....	3.6
3.5 DATA COLLECTION METHODS.....	3.8
3.5.1 Telephone and Mail Surveys.....	3.9
3.5.2 On-Site Inspections.....	3.11
3.5.3 Billing Data.....	3.13
3.5.4 Metering/monitoring.....	3.16
4. SECTOR AND END USE.....	3.17
4. DATA ANALYSIS ISSUES.....	4.1
4.1 EQUIPMENT SURVEY/ENGINEERING STUDIES.....	4.2
4.1.1 Assessing a Measure's Physical Persistence.....	4.3
4.1.2 Engineering Estimates.....	4.5
4.1.3 Savings Projections.....	4.5
4.1.4 Summary.....	4.7
4.2 CONSUMPTION EVALUATION STUDIES.....	4.7
4.2.1 Univariate Analysis of Weather-Adjusted Consumption.....	4.7
4.2.2 Multivariate Analysis of Weather-Adjusted Consumption.....	4.10
4.2.3 Projecting Savings Estimates.....	4.14
4.2.4 Summary.....	4.16

5. SUMMARY OF FINDINGS	5.1
6. DISCUSSION AND CONCLUSIONS.....	6.1
7. GLOSSARY	7.1
8. REFERENCES.....	8.1

LIST OF TABLES

Table 2.1 Definitions of Persistence	2.2
Table 2.2 Purposes for Studying The Persistence of Savings.....	2.10
Table 3.1 Examples of Control Group Options.....	3.2
Table 3.2 Example From California Protocol.....	3.7
Table 4.1 LILCO's Persistence Study for Selected Lighting Measures.....	4.6
Table 4.2 Average Net Savings by Utility	4.10
Table 4.3 Summary of Wisconsin Gas Company Median Polish Results.....	4.14
Table 4.4 Historic Costs And Net Savings Of Bonneville's Weatherization Program.....	4.16

LIST OF FIGURES

Fig. 2.1 Persistence of Savings from a Single DSM Measure: The Measure Scale	2.6
Fig. 2.2 Persistence of Savings from a Package of DSM Measures: The Cohort Scale	2.7
Fig. 2.3 Persistence of Savings Over the Lifetime of a DSM Program: The Program Scale	2.9

EXECUTIVE SUMMARY

This literature review was conducted as part of a larger project to produce a handbook on the measurement of persistence. The past decade has marked the development of the concept of persistence and a growing recognition that the long-term impacts of demand-side management (DSM) programs warrant careful assessment. Although increasing attention has been paid to the topic of persistence, no clear consensus has emerged either about its definition or about the methods most appropriate for its measurement and analysis. This project strives to fill that gap by reviewing the goals, terminology, and methods of past persistence studies. It was conducted from the perspective of a utility that seeks to acquire demand-side resources and is interested in their long-term durability; it was not conducted from the perspective of the individual consumer.

Over 30 persistence studies, articles, and protocols were examined for this report. The review begins by discussing the underpinnings of persistence studies: namely, the definitions of persistence and the purposes of persistence studies. Then, it describes issues relevant to both the collection and analysis of data on the persistence of energy and demand savings. Findings from persistence studies also are summarized. Throughout the review, four studies are used repeatedly to illustrate different methodological and analytical approaches to persistence so that readers can track the data collection, data analysis, and findings of a set of comprehensive studies that represent alternative approaches.

DEFINITIONS, PURPOSES, AND SCALES OF ANALYSIS

The literature review revealed that the term "persistence" is defined in numerous different and sometimes conflicting ways. Most often, persistence refers to the long-term temporal pattern of energy savings and load reductions from DSM investments. This is the definition that we emphasize. However, persistence also is often defined as the operational life of an energy conservation measure, or the degree of retention of measures. Other definitions refer to the time-line of energy savings relative to expectations or the degradation of energy savings over time.

Persistence studies are conducted for many different purposes. Although planning and evaluation are the two overarching purposes, these broad goals can be refined further to distinguish the following purposes: resource planning, load forecasting, annual program planning, program evaluation, lost revenue recovery, shareholder incentive calculations, and the determination of performance contracting payments. The methods used to satisfy each of these purposes varies. For instance, univariate or multivariate regression analysis of billing data may be sufficient for determining shareholder incentives and performance contract payments. On the other hand, equipment surveys and on-site inspections, in conjunction with multivariate analysis, may be necessary for program evaluation because they can offer insights into the causes of any degradation or increase in savings over time. Another kind of goal for persistence studies is to discover methods for improving programs by changing procedures that appear to lead to premature measure removal

or by altering the program to delete specific types of equipment that tend to be removed or become dysfunctional.

The array of purposes for persistence studies and different approaches taken in such studies, is influenced by the diversity of factors that can impair the persistence of measures and energy savings. These factors include the technical life of a measure, which is influenced by technical degradation under optimal conditions; market factors such as remodeling and occupant turnover; and operational life, which can be affected by how the measure actually is installed, sized, operated, and maintained.

To enhance the understanding of persistence, three scales of analysis were delineated. Each of the scales, described below, encompasses different time frames and scopes of investigation:

- **measure** – used in studies that predict the pattern of year-to-year savings from the installation of a particular DSM measure by a single cohort of participants in a DSM program;
- **cohort** – used in studies that predict the pattern of year-to-year savings from the installation of a collection of DSM measures by a single cohort of participants in a DSM program; and
- **program** – used in studies that predict the pattern of net program savings from packages of measures for all years (across multiple cohorts) of program participants in a DSM program.

The *measure* scale addresses the energy savings associated with a particular measure during its lifetime and usually focuses on a single cohort of participants. Studies at this scale, which constitute a significant proportion of persistence work to date, frequently seek to determine the technological and behavioral reasons why particular measures achieve certain levels of savings over time. In contrast, studies falling within the *cohort* scale tend to emphasize program persistence and not the savings associated with particular measures. The *program* scale is the least common, perhaps because it requires the longest time frame and, possibly, the largest financial investment to track net savings over multiple participant cohorts. Many factors can influence the trajectory of a program's energy savings, such as changing program features and operations, changing baseline efficiencies, and the effects of participant self-selection on savings over time.

THE DESIGN OF PERSISTENCE STUDIES

The various definitions of persistence, goals of studies, and the scales of analysis summarized above affect the selection of methods for collecting and analyzing persistence data, as do other factors such as project budgets and regulatory influences. Control group decisions provide an illustration of this point. If the focus is on the persistence of gross energy savings from installed energy conservation measures, control groups are unnecessary. Control groups, however, are required to ascertain net savings over time because it is critical in these cases to distinguish program effects from other effects. The options identified in the literature review include control groups

consisting of non-participants and control groups consisting of later participants. Establishing and maintaining adequate control groups is particularly challenging for persistence studies, for the following reasons:

- the need for control group members to remain unaffected by a program over an extended period;
- the cost of selecting an appropriate control group and obtaining data from such a group;
- changes in programs over time;
- market transformation;
- sample attrition due to mobility and other changes to the end user;
- changes in the composition of participants over time (e.g., from a group of people or buildings with great potential for saving energy to those with significantly lower potential savings); and
- changing economic conditions including energy prices that can influence patterns of energy consumption.

Persistence studies can adopt a retrospective or prospective research design. The vast majority of studies are retrospective, using data from past participants and selecting analytical methods to interpret existing data. Another possibility is prospective studies, which use data from current and future participants and which allow desired analyses to determine what data are collected.

DATA COLLECTION METHODS

Four primary data collection methods are used in persistence studies. These methods are telephone and mail surveys, on-site inspections, billing histories, and metering or monitoring. Metering refers to the collection of whole-building or end-use energy consumption data and monitoring denotes the collection of relevant, non-energy consumption data (e.g., temperature, humidity, and duty cycle). The kinds and quality of data the four main data collection methods can provide vary, as do their costs. Method selection entails trade-offs between the costs of obtaining high-quality data and the potential benefits to the utility.

Telephone and mail surveys are often used to study measure persistence. They also may be used for eliciting background information about customers, their attitudes, needs, and behaviors to shape studies, to acquire information to help explain or confirm the findings obtained via other methods, or to elicit the customers' degree of satisfaction over time with particular energy conservation measures. Although they are relatively inexpensive to administer and can be cost effective, the main limitations of these methods are that they may produce flawed data because of their reliance on the expertise, knowledge, and biases of respondents.

On-site inspections may be the method most appropriate for evaluating and verifying the physical persistence of energy-conservation measures, the factors influencing the lack of such persistence, and the operating performance of energy conservation measures. They have been used primarily to determine measure retention rather than measure performance. If measure retention is the persistence goal, on-site inspections offer several advantages. First, inspections may be the only means of acquiring reliable information about whether or not measures remain in place and operational, and about where measures are located. Second, inspections are conducive to conducting face-to-face surveys with on-site personnel. Finally, high-quality measure retention information obtained through on-site inspections is useful in conditional demand studies conducted for the purpose of estimating the persistence of savings.

However, site inspections tend to be costly, although their costs depend upon the amount of information collected and the degree of data precision sought. Beyond their high costs, disadvantages of using on-site inspections include their reliance on the expertise of on-site auditors; consistency among auditors; access limitations both to equipment and to some locations within facilities; their limitations in documenting historical and behavioral characteristics; and the possibility of a large non-response bias.

Information collected from telephone or mail surveys, and from on-site inspections, may serve to increase the accuracy of engineering estimates by providing data on the numbers of measures in place and (for on-site inspections) operating as anticipated. Surveys (mail, telephone, or on-site) also help to explain why measures may not achieve the manufacturers' estimates of savings by ascertaining information about behavior. Particularly for on-site surveys, DSM professionals should evaluate the extent to which field work is necessary to provide information that may increase the accuracy of estimates, given that field work tends to be costly.

Billing data typically are used to evaluate the net energy (electric and/or gas) savings attributable to energy conservation programs rather than particular energy conservation measures. They are amenable, therefore, to cohort- and program-scale analyses. In this context, the main advantages of billing analyses are that they use large amounts of readily available data; that they provide information at an aggregate level; and that they tend to be cost-effective. Billing analyses generally rely on assumptions about the physical presence of installed energy conservation measures; involve the use of comparison groups; and include some form of weather normalization so as to determine the persistence of net energy savings. A limitation of billing analyses is sample attrition over time, due to data quality problems that often reduce the size of the sample available for analysis and other factors that reduce the sample size (e.g., changes in participant and non-participant populations). Also, billing analyses generally are inadequate for distinguishing the savings achieved from individual energy conservation measures.

Metering or monitoring may be used to provide detailed information about the energy and load effects of specific measures at specific premises, over time. They also are a good means for acquiring data about non-energy consumption characteristics, such as those related to weather or the

times and manner of equipment use, that are important for distinguishing program or equipment effects from other effects and for understanding why certain levels of energy savings are achieved. Such insights may lead to increased persistence through changes in program design or delivery. Nevertheless, these measurements can produce biased estimates of energy consumption because of the limited duration of many metering projects and the inability of short-term metering to characterize changes over time. It also is difficult to generalize to a larger population when small samples are used for metering or monitoring. Further, metering and monitoring tend to be equipment- and cost-intensive; they probably constitute the most expensive of the four categories of data collection methods.

DATA ANALYSIS METHODS

Methods used to estimate the persistence of savings range from engineering methods based on equipment surveys, to univariate and multivariate analysis of energy consumption data.

Equipment survey/engineering methods estimate gross savings and, although they focus on individual measures, often aggregate measures when calculating savings. These methods may be used for different persistence study goals, such as program evaluation or determining shareholder payments. The most comprehensive of these methods considers both survival energy savings (which is the gross energy savings from measures installed by the program) and stimulated energy savings (which reflects any savings due to installations of additional high-efficiency measures by participants or non-participants as the result of the program but without further program incentives). These kinds of surveys usually are used to collect data only once or twice from any single cohort of participants.

Univariate analyses of energy consumption, frequently conducted for the purpose of determining shareholder payments, typically rely solely on weather normalization to analyze the savings from energy conservation programs. Once weather-normalization data adjustments are made, analyses simply highlight any energy consumption differences.

In addition to adjusting for weather, multivariate analyses of the energy savings from DSM programs strive to include additional factors. These factors include electricity prices, pre-participation consumption, sample selection bias, building square feet, the use of alternative energy sources such as wood, and long-run average heating degree days. Multivariate analyses of this type can test hypotheses about the causes of savings persistence, enhancement, or degradation, which can lead to improvements in program design. Therefore, these methods more often are used for the purpose of program evaluation than to determine shareholder payments.

A major difference in the use of uni- and multivariate analyses is in their relative dependence on control groups in determining net energy savings. Both types of approaches are highly dependent on control groups, but multivariate analyses also can make statistical adjustments for differences between control and participant groups. However, because multivariate analyses require more data than do univariate analyses, they are more costly to perform.

FINDINGS OF PERSISTENCE STUDIES

Results of persistence studies, taken as a group, are inconclusive and in some cases inconsistent. Measure persistence studies generally have found that the *ex-post* estimates of year-to-year savings associated with measures were less than the *ex-ante* engineering estimates, although this is not always the case. For instance, when stimulated energy savings (i.e., market transformation) is considered, *ex-post* estimates have exceeded *ex-ante* estimates. Further, significant variability has been observed in the persistence of savings associated with different measures and different sectors.

Persistence studies that focus on the cohort scale, which generally have evaluated residential DSM programs, also have produced varying results. The results of these studies range from a decline in savings to a net increase in savings over time within cohorts of participants. These studies often fail to report confidence intervals or levels of precision, so it is not possible to determine whether the differences are significant.

The one example of an analysis of persistence at the program scale indicated that savings can decrease more precipitously across successive years of participants than across the years following program participation for a single cohort of participants. The importance of this finding underscores the need to conduct persistence studies that analyze the full lifecycle of program impacts.

In part because of the inconclusiveness of findings from persistence studies, one recommendation is to use methods that will provide a basis for statistically strong results. Along these lines, it is important to report levels of precision and variability along with the point estimates of savings over time. At the same time, costs must be considered. Increasingly, utilities are looking for relatively inexpensive and straightforward methods for evaluating their DSM programs. Finally, no single method best meets the variety of users' information needs while remaining within their budgetary constraints.

Beyond these methodological considerations, the future may present a different type of challenge to the conduct of persistence studies. We do not know how utilities will address persistence as they adapt to a competitive world. On the one hand, accurate information about persistence or how to improve persistence may enhance a utility's competitive edge. On the other hand, expensive persistence studies may be an unaffordable luxury. Competition also may result in multiple utilities cooperating to conduct persistence studies.

LIST OF ACRONYMS

ANOVA	Analysis of variance
CCIG	California Conservation Inventory Group
DNAC	Change in normalized annual consumption
DSM	Demand-side management
ESCO	Energy service company
HE	High efficiency
HRCP	Hood River Conservation Project
HVAC	Heating, ventilation, and cooling
IRP	Integrated resource planning
LILCO	Long Island Lighting Company
NAC	Normalized annual consumption
NEPSCO	New England Power Service Company
ORNL	Oak Ridge National Laboratory
PRISM	Princeton Scorekeeping Method
SRC	Synergic Resources Corporation

ACKNOWLEDGEMENTS

This literature review is part of a larger project sponsored by the U.S. Department of Energy and the Electric Power Research Institute to produce a handbook on the measurement of persistence. Secretarial support was provided at various stages of this project by Sherry Surdham, Connie Dagley, and Dan Bell. Ed Lapsa and Dan Bell produced the graphics for this document. Valuable comments on earlier drafts of this literature review were provided by Les Baxter, Eric Hirst, and Ed Lapsa (ORNL), Rob Bordner and Ron Moe (Synergic Resources Corporation), John F. Nelson (Wisconsin Gas Company), Kim Johnson Oswald (Strategic Utility Market Services), Lee Parker (Applied Energy Group, Inc.), Jane S. Peters (Barakat & Chamberlin, Inc.), and Ed Vine (Lawrence Berkeley Laboratory). The comments provided by these individuals led to significant improvements in this final report.

1. BACKGROUND AND PURPOSE

The durability or persistence of savings from demand-side management (DSM) investments is an issue of importance to many stakeholders, including: DSM program managers and evaluators, resource planners, load forecasters, DSM forecasters, utility regulators, public interest groups, and customers. Inaccurate assumptions about the persistence of savings can bias estimates of program benefits. These biased estimates may in turn lead to suboptimal program design decisions and to cost-ineffective DSM investments. In addition, they may cause utilities to receive inappropriate net lost revenue adjustments, or energy service companies (ESCOs) may receive performance-based payments that do not reflect actual DSM benefits. Ultimately, utilities may invest in unnecessary supply options or fail to plan for necessary future energy resources if forecasted DSM impacts are inaccurate.

The purpose of this literature review is to provide a framework for developing a handbook on persistence methods. To this end, the review discusses the concept of persistence, describes the methods used to measure it, and briefly summarizes the findings of past persistence studies. The review focuses on measurement issues because it is believed that the use of improved and more consistent methods will help stakeholders make better-informed decisions about DSM programs and future resource requirements. This review was conducted from the perspective of a utility that seeks to acquire DSM resources and is interested in their long-term durability; it was not conducted from the perspective of the individual consumer. It also focuses on net, rather than gross, savings of energy conservation programs.

The past decade has produced conflicting information about the operational lifetimes of individual DSM measures, the trajectory of savings from DSM installations, and the ability of programs to generate sustained savings over extended periods of operation. Some of this confusion has resulted from the use of different operational definitions of persistence and the application of different measurement methods.

Currently, there is no consensus either on how best to measure persistence or on how frequently it should be measured (e.g., Braithwait *et al.* 1994). There are several possible reasons for this lack of agreement. One reason may be the relative newness of the persistence concept and of serious attempts to measure it. Another reason is that persistence has been defined and analyzed at many different scales. Three different scales are described below, each covering a unique time frame and level of DSM investment:

- **measure** – this scale is used in studies that predict the trajectory of year-to-year savings from the installation of a particular DSM measure by a single cohort of participants in a DSM program;
- **cohort** – this scale is used in studies that predict the trajectory of year-to-year savings from the installation of a collection of DSM measures by a single cohort of participants in a DSM program; and

- **program** – this scale is used in studies that predict the trajectory of net program savings from packages of measures for all years (multiple cohorts) of program participants in a DSM program.

The *measure* scale is the least comprehensive and the one that is used most often in persistence studies. At the other extreme, the *program* scale is the most comprehensive and the one that is examined least often. Different measurement methods tend to be used for each of these scales of analysis, although methods are not entirely consistent even within each scale.

Differences in methodology also occur because persistence is measured for many different purposes. In particular, there are two overarching reasons why persistence studies would be undertaken: to serve as a basis for DSM or integrated resource planning (IRP) or to evaluate (e.g., determine what savings are achieved and why) or validate (i.e., assess performance relative to a specified goal) the savings from DSM measures or programs. Throughout this review we identify instances where different scales of analysis and different study purposes require different measurement methods.

This review is organized into eight chapters. Chapter 2 discusses the variety of definitions of persistence and the various reasons for measuring it. Chapter 3 discusses research design and measurement issues, including the use of control groups, units of measure, sampling, sample attrition, and data collection methods. Chapter 4 discusses alternative data analysis approaches for equipment surveys or engineering studies and univariate and multivariate energy consumption analyses used to estimate current and future savings. The review ends with a brief summary of persistence study findings (Chapter 5), a concluding discussion (Chapter 6), a glossary (Chapter 7) and a list of references (Chapter 8).

2. ISSUES FUNDAMENTAL TO PERSISTENCE AND ITS MEASUREMENT

Two conclusions are clear from a review of the literature; namely, that persistence is an issue of growing concern and that the term "persistence" has been used in many different ways. The combination of these two observations creates conditions for considerable ambiguity and miscommunication among stakeholders. It also complicates the task of developing a methods handbook. Therefore, this section seeks to identify the multiple dimensions of persistence so as to lay the foundation for later discussions of current methods and for the handbook itself.

2.1 DEFINITIONS AND SCALES OF ANALYSIS

Persistence in this report is defined as the long-term temporal pattern of energy savings and load reductions from DSM investments. The savings of concern may involve a single energy conservation measure or a package of measures installed in a DSM program.

Questions about the durability of DSM savings first were raised seriously about a decade ago (e.g., Hirst, White, and Goeltz 1984, and 1985). Since then, persistence has been defined in many different ways. At one extreme, the term "persistence" has been referred to as "subsequent-year savings" and "long-term savings," emphasizing the positive; at the other extreme, terms such as "degradation" or "decay" have been used, emphasizing the negative (e.g., Jeppesen and King 1993; Violette *et al.* 1991; White and Brown 1992). An insightful analogy is provided by Braithwait *et al.* (1994), who liken persistence studies to economic studies of the depreciation of investment goods.

Table 2.1 illustrates the range of definitions of persistence used to date. For instance, persistence is often defined as the operational life of an energy conservation measure or the degree of retention of measures. Alternatively, definitions have referred to persistence as the time-line of energy savings relative to expectations.

Table 2.1 also identifies in bold-face type the four persistence studies that are used repeatedly throughout this literature review to illustrate measurement options. They were chosen because they span a wide array of different scales and sectors, and because they provide detailed descriptions of alternative data collection and analysis methods.

Conceptually and in practice, persistence is considered at many different scales, as illustrated in Table 2.1. These scales add another layer of complexity to the discourse about persistence, for a number of reasons described below.

First, studies at each of the scales would focus on different factors to explain the persistence of energy savings. The *measure* scale may focus on the technological and behavioral factors that influence the operation and use of energy conservation measures. For *cohorts*, the emphasis is on large-scale factors that influence energy consumption (e.g., weather), on the consumption of energy,

Table 2.1 Definitions of Persistence^a

Authors	Definition	Motivation	Focus	Sector	End-Use
		Planning (P) Eval. (E)	Program (P) Measure (M)	Res. (R) Comm. (C) Ind. (I)	Multiple or Specific
Braithwait, Maniaci, Blyer, and Attenberger 1994	"The term persistence refers to the extent that the energy savings and the energy efficiency measures that are promoted by utility DSM programs are maintained over time. A distinction is usually made between the persistence of energy savings (both total and net) and the persistence of the energy efficiency measures...themselves..." (p. 3). Distinguish "three distinct but related concepts or elements of persistence": measure retention (effective measure life); measure performance; and persistence of net energy savings.	E	P, M	multiple	multiple
Brown, Berry, Balzer, and Faby 1993	Discuss savings over lifetime of weatherization measures, which is assumed to be 20 years; an appendix discusses weighted estimates of measure lifetime; no discussion or definition of "persistence" per se.	E	P, M	R	multiple
Brown and White 1992	No definition. Report discusses the decay in program energy savings over time, by multiple cohorts of participants.	E	P	R	multiple
Coates 1992	No explicit definition; paper deals with 1st-, 2nd-, and 3rd-year energy savings of a program.	E	P	C	multiple
Delaware Office of the Public Advocate and Delmarva Power & Light Company 1993	"Persistence measures whether energy savings continue as expected over a conservation technology's useful life" (p. 29).	E	P	multiple	multiple
Eto, Vine, Shown, Sonnenblick, and Payne 1994	No explicit definition. Distinguish short- medium- and long-term persistence. Also consider economic life of savings.	E	P	C	lighting
Guyant, Hopkins, and Reid 1989	Focused on "longevity," long-term benefits, enduring high levels of efficiency	E	M	R	oil heat systems
Hickman and Brandis 1992	Looked at "measure retention," focusing mainly on premature measure removal.	E	P, M	C	multiple
Hirst, White, and Goeltz 1985	No definition; looked at durability of energy savings one-, two-, and three-years after program. May be "first study to closely examine actual (measured) energy savings due to a conservation program over such a long time span" (p. 2).	E	P	R	multiple
Hirst and Sabo 1991	No definition of "persistence," but define: "lifetime of the programmatic energy effects" as "the median length of time (in years) that a DSM program produces energy effects; it is a function of both the lifetimes of the measures installed and operating practices" (p. 70).	E	P, M	n/a	n/a
Jacobson, Miller, Granda, Conant, Wright, and Landsberg, 1992	No explicit definition; compared engineering estimates with measured savings.	E	P	multiple	multiple
Jeppesen and King 1993	Use definition from 1990 <i>Report on (California) Statewide Collaborative Process</i> (p. A-27 of that report): "Persistence (and decay) refers to any decline in energy-saving effectiveness that may take place over a measure's useful life" (p. II-2).	E	M, but look at measures as part of programs	C	multiple

^a Bold-faced items are those documents that are highlighted as examples throughout the remainder of this report.

Table 2.1 Definitions of Persistence (cont.)

Authors	Definition	Motivation	Focus	Sector	End-Use
		Planning (P) Eval. (E)	Program (P) Measure (M)	Res. (R) Comm. (C) Ind. (I)	Multiple or Specific
Keating 1991	"Defined at the utility level, persistence is equivalent to the long-term temporal changes in net program impacts; defined at the societal level, it is equivalent to the long-term temporal changes in total impacts. Note that the persistence of both total and net impacts includes technical and operating characteristics" (pp. 89-90). Also discusses two dimension of measure persistence: lifetimes of DSM measures installed and ways measures operated.	P, E	P, M	multiple	multiple
Massachusetts Department of Public Utilities 1993	No overt mention of persistence, but tables include lifetime savings—their present value of measures and the MWh (kW-Year) savings of programs	E	P, M	multiple	multiple
Miller, Blake, Dagher, Schutte, and Wright 1992	No definition; looked at short-term persistence of measures.	E	M	small C/I	lighting
Minnesota DSM Evaluation Consortium Working Document 1993	Persistence is the "[r]ate of energy-savings effectiveness that takes place over a measure's useful life" (p. 3).	E	P	n/a	n/a
Misuriello and Hopkins 1992	No definition; discuss "the availability of DSM savings over the life of the measure" (p 20)	E	M	multiple	multiple
Narum, Pigg, and Schlegel 1992	"There are two main ways in which persistence can be defined, and, in turn, evaluated: 1) at the measure-specific level (measure retention, measure lifetimes, and measure performance considerations); and 2) at the program level (total and net impacts)" (p. 2).	E	P	R*	multiple
New Jersey Board of Regulatory Commissioners 1993	States that utilities must specify procedures to allow for, or to verify, persistence, but provides no definition of persistence.	E	M	multiple	multiple
Pacific Gas & Electric <i>et al.</i> 1993	A persistence study is defined as "A study to assess changes in net program load impacts over time" (p. A-8).	E (verification)	P*	multiple	multiple
Parker 1993	Defined in terms of the technical persistence of operational life (the percent of originally rebated measures installed and operating, or replaced with other equipment at the end of its operating life) and the persistence of efficient technology (the actual inventory of efficient equipment installed and operating, whether the equipment was installed directly by the program, replaced in kind, or added by the customer without a rebate) (see pp. 28-29).	E	M	C&I	multiple
Pennsylvania Public Utility Commission 1992	Discuss persistence of net savings as long-term effects of program measure. "This is referred to as the 'snap-back', 'rebound', or 'take-back' effect" (p. 16), which can be caused by behavioral responses or equipment degradation.	E	P, M	multiple	multiple
PSCW ca. 1992	Persistence is "the extent to which achieved first-year savings last as long as anticipated" (p. 2)	E	M, P	n/a	n/a
Robinson 1992	No definition. In part this study reports on an effort to help assure the persistence of savings via direct bulb installation.	E	P, M	R	lighting

* Unlike most studies reported in this table, this study focused on gas-utility programs rather than on electric-utility programs.

Table 2.1 Definitions of Persistence (cont.)

Authors	Definition	Motivation	Focus	Sector	End-Use
		Planning (P) Eval. (E)	Program (P) Measure (M)	Res. (R) Comm. (C) Ind. (I)	Multiple or Specific
U.S. EPA 1993a	"Subsequent-year energy savings: The energy savings occurring in the annual periods following the first year" (p. 21).	E (verification)	M	multiple (utility focus)	multiple
U.S. EPA 1993b	No definition. May claim savings throughout useful life of measure(s). Useful life is the median number of years that a measure produces energy savings.	E (verification)	M, P	multiple	multiple
Vine 1992	Measure persistence: the number of years a measure is installed and its level of efficiency. Program persistence: continuation of participants' energy savings over time. Look at gross and net savings. [No explicit definition given; above definitions based on Vine's recommendations for persistence assessments]	P, E	P, M	multiple	multiple
Violette, D., M. Ozog, M. Keneipp, F. Stern, and P. Hanser 1991	"Refers to any decline in energy-savings effectiveness that may take place over a conservation measure's life. This is a function of both consumer behavior and equipment degradation" (p. C-6).	E	M	multiple	multiple
White and Brown 1990	Report focuses on an impact evaluation three years after weatherization, looking at measures installed and actual electricity saved. No mention of term "persistence."	E	P	R	multiple
White and Brown 1992	No definition; looked at consistency or decay of energy savings over time.	E	P	R	multiple
White, Stovall, and Tonn 1992	No definition, but interested in "durability or persistence of program benefits, especially load savings..." (p. 25) one-, two-, and three-years after weatherization.	E	P	R	multiple

and on participant/non-participant differences in energy actions. *Program*-scale persistence can encompass the factors relevant to the other scales, and adds the element of cross-cohort comparison to distinguish year-by-year, cohort-by-cohort effects.

Second, the scales extend the distinction between the two dimensions of persistence typically mentioned in the literature, program and measure persistence (e.g., Braithwait *et al.* 1994; Keating 1991;¹ Narum, Pigg, and Schlegel 1992; Vine 1992). "Program persistence" has been used to refer to the durability of net energy savings (those savings attributable directly to the program) of energy conservation programs. The three scales proposed in this review disaggregate the conception of program persistence reported in the literature into cohort-year analyses and analyses over the lifetime of programs. "Measure persistence" has consisted of the long-term gross energy savings associated with specific energy conservation measures. Distinctions between program and measure persistence sometimes have been fuzzy in practice. For example, program persistence may be gauged simply by summing or combining the persistence of separate energy conservation measures. In other cases, because a relatively small group of similar measures constitute a program (e.g., some lighting

¹ Keating distinguishes the persistence of program measures from what he calls "the overall definition of persistence" (p. 89) that can be viewed from either a utility or societal perspective. The utility perspective focuses on the long-term temporal changes in net program impacts" (p. 89) and the societal perspective focuses on "the long-term temporal changes in total impacts" (p. 90). However, in this text, the studies that fall under the "total-impact perspective" are those that are concerned with measure lifetime and the studies that he labels the "net-impact perspective" are equivalent to those labeled in this literature review and by Vine (1992), e.g., as program persistence.

programs), the persistence of the program may be determined by focusing on one energy conservation measure (e.g., high-performance light bulbs in Robinson 1992).

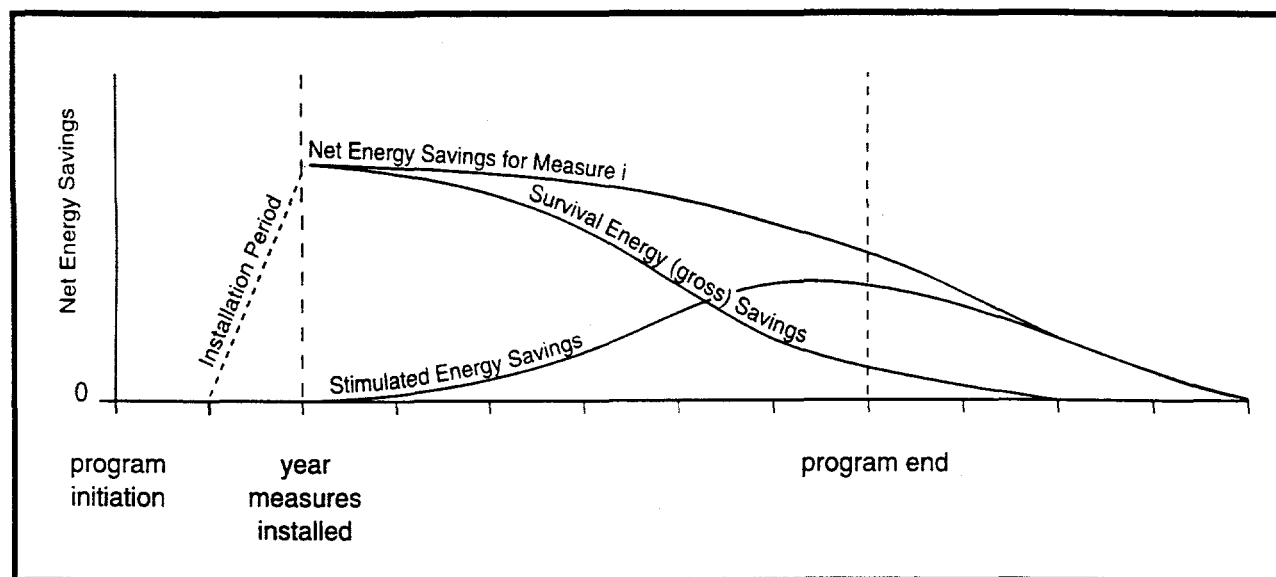
Third, the scales tend to distinguish between net and gross savings; this distinction has tremendous implications for the methods for studying persistence and for analytical results. Because net savings are those savings attributable to the program in question, evaluating the persistence of those savings requires methodological rigor (thus, relatively high financial costs). In particular, they require control groups, so that it is possible to detect the differences between what happened because of the program and what would have happened in its absence. As will be discussed later in this report, establishing control groups for persistence studies is made particularly difficult because elements of the program, participants' energy conservation needs, the pool of appropriate comparison groups (e.g., non-participants who can be "matched" with participant groups), and economic conditions (e.g., the price of electricity) change over time. These kinds of shifts make it difficult to track groups, measures, and savings over time in ways that allow evaluators to state with confidence that the effects they see are attributable to energy conservation measures or programs. In contrast, ascertaining gross energy savings generally requires different and simpler methods from studies of net savings, ranging from engineering estimates to on-site surveys to submetering. Results from gross energy savings studies, however, are of limited value in evaluating the persistence of a program's energy savings. Measure-life studies tend to fall into the category of gross savings. While cohort-scale studies generally aim to ascertain net savings, the control-group constraints sometimes are so limiting that only estimates of gross savings result. Program-scale analyses aim to provide estimates of the persistence of net energy savings.

Measure Persistence. The measure scale focuses on the energy savings associated with a particular measure during its lifetime for a particular cohort year of participants. As noted earlier, studies of the persistence of energy conservation measures are more prevalent than studies of program persistence.

The lifetime savings for an energy conservation measure for a single cohort year of participants is illustrated in Fig. 2.1. The figure shows the net energy savings for a single measure as a function of two kinds of energy savings. First, net energy savings is influenced by "survival energy savings," which is the gross energy savings from measures installed by the program. The second influence is "stimulated energy savings," which reflects possible additional savings due to actions taken both by participants and non-participants, such as installing or replacing more measures as the result of the program but without further program incentives.² The stimulated energy savings recognize a program's ability to transform markets for DSM measures. These "market transformation" effects are sometimes called market progression, surge, or free drivership, depending on whether they involve participants or non-participants. In general, they refer to a program's ability

² The literature considers "stimulated energy savings" to mean the addition of measures by program participants. We are expanding and enhancing that definition to include the installation of measures by both participants and non-participants.

to accelerate market trends toward increasing energy efficiency, above and beyond the investments that are caused by the program's incentive payments. Market transformation is particularly relevant to evaluations of the net savings of energy conservation programs, which is the primary focus of this review.



**Fig. 2.1 Persistence of Savings from a Single DSM Measure:
The Measure Scale**

Both survival and stimulated energy savings depend upon measure life and measure performance or efficiency over that life. Measure life is "the time during which the measure is installed and is accruing energy efficiency benefits" (Jeppesen and Rudman 1993, p. 521). It is a function of the measure's technical life, market factors (e.g., remodeling, occupancy or residence changes, behavior), and operational life factors (e.g., installation and sizing, operation, maintenance). In contrast, measure efficiency or performance is "the actual energy efficiency performance of the measure" (Jeppesen and Rudman 1993, p. 521). It is a function of the technical degradation of the measure (including wear-related factors) and behavioral factors (such as operation and maintenance activities).

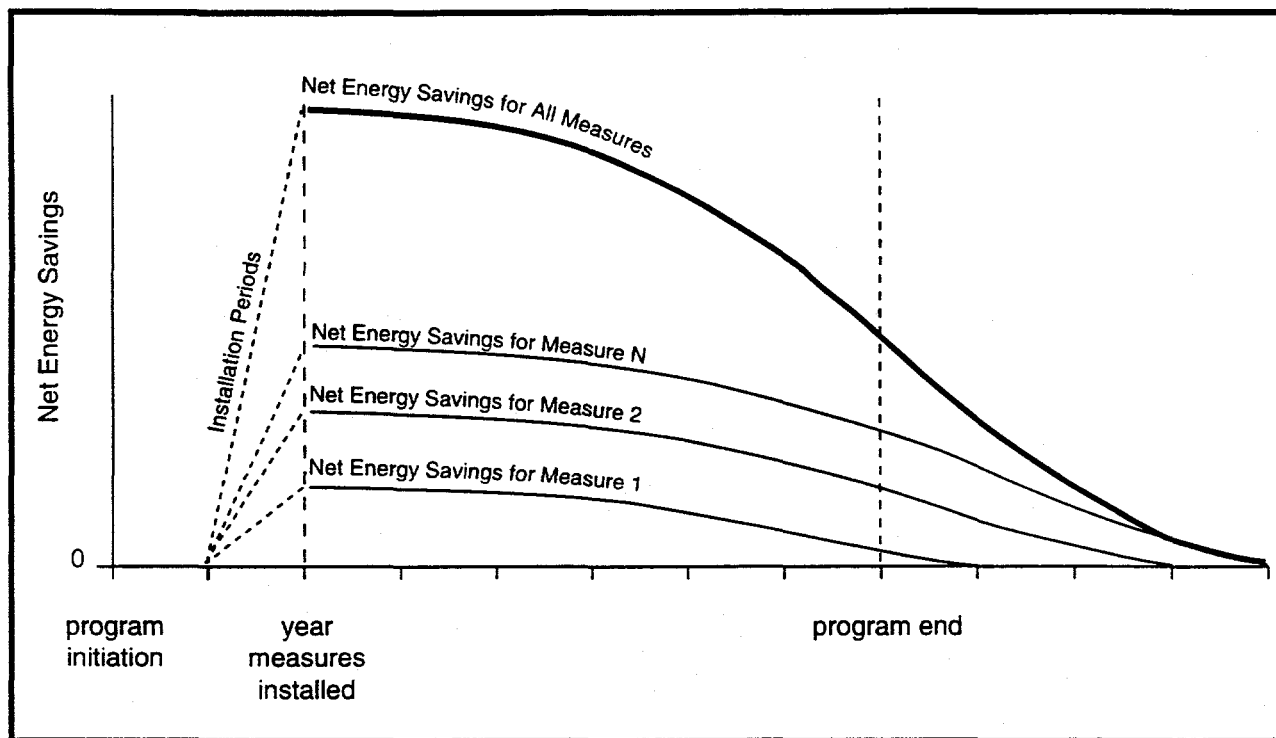
The following is a more complete list of the many factors that can reduce measure life and impair measure performance:³

- improper energy conservation measure installation;
- inadequate maintenance of installed measures;

³ The literature on persistence contains varied and sometimes conflicting use of terms. For instance, the LILCO study (Parker 1993) distinguishes "measure lifetime" (or "operating life") from "measure life" (or "technical life"). Measure lifetime, which refers to the field longevity of energy conservation measures, is comprised of a combination of technical endurance (or degradation) and behavior. The concept of measure life, as used in the LILCO study, refers to engineering or manufacturers' projections and does not include behavioral elements.

- technological problems with installed measures (e.g., bad lightbulbs);
- premature removal of the energy conservation measure;
- replacement of energy conservation measures with less efficient measures;
- patterns of energy conservation measure usage that adversely affect measure life;
- surge effects (i.e., "the tendency for some consumers to react to the savings realized by the initial conservation measure installation by expanding their adoption to other conservation measures" [Jeppesen and King 1993, p. II-6;]⁴ and
- building renovation or turnover (Braithwait *et al.* 1994; Parker 1993).

Cohort Persistence. This scale specifies the lifetime energy savings for all measures installed in a particular cohort year. The scale is illustrated in Fig. 2.2.



**Fig. 2.2 Persistence of Savings from a Package of DSM Measures:
The Cohort Scale**

Many studies of program persistence are performed at the cohort scale. Some program persistence studies concentrate on the consistency of savings over time, by comparing later-year savings with first-year savings (Vine 1992). However, others (e.g., White and Brown 1990; Narum,

⁴ Surge effects refer to behaviors adopted by individual program participants; free-driver effects refer to the behavior of non-participants (Jeppesen and King 1993).

Pigg, and Schlegel 1992) evaluate savings among successive cohorts of program participants. Where these evaluations of multiple cohorts have focused on the persistence of savings across cohorts as well as on the year-to-year savings of individual cohorts (as in Brown and White 1992), they are considered to address the "program" scale.

Persistence at the cohort scale usually is assessed at an aggregate level, without attempting to estimate the energy savings from individual measures. An example of this aggregate approach is provided by White and Brown (1990) in their evaluation of the Bonneville Power Administration's Residential Weatherization Program. They calculated net electricity savings from billing records (weather-normalized) and did not estimate the impacts of different weatherization measures.

Authors of these kinds of program studies tend to take one of two approaches to the factors contributing to a degradation in energy savings over time. One approach is to discuss only generally the kinds of factors that affect persistence, perhaps distinguishing factors that influence measure lifetime and measure performance from those that affect the market.

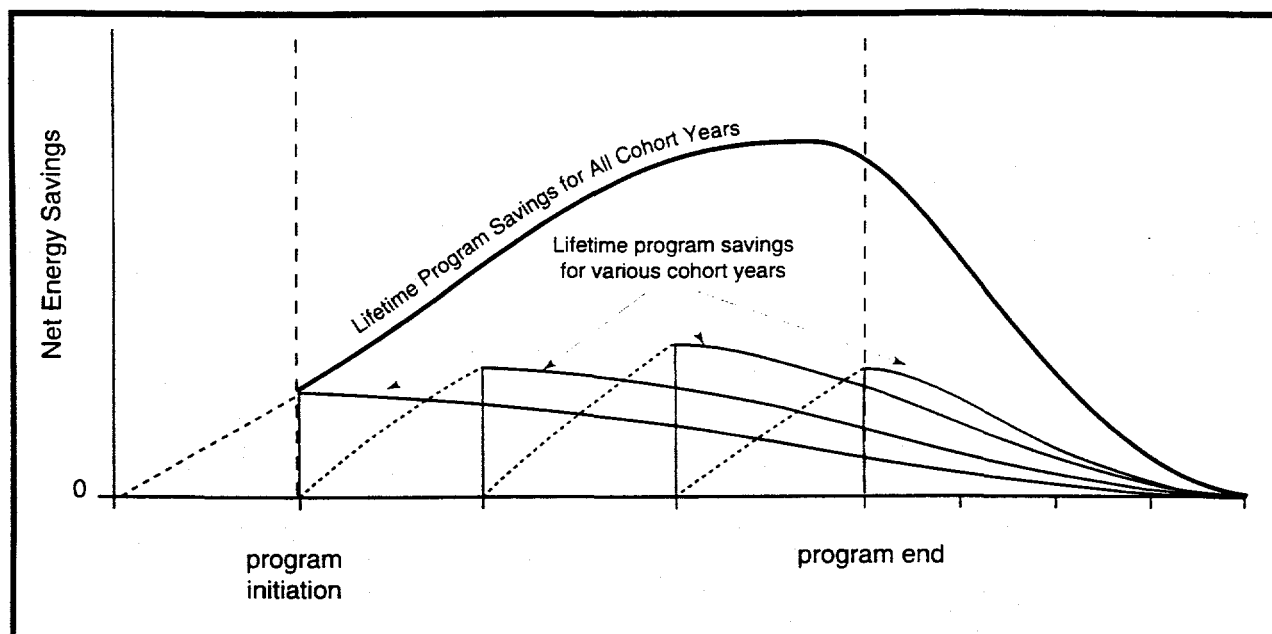
Another approach is to investigate specifically the contribution of one or more of these factors to program persistence. This approach may overlap considerably with studies that focus on energy conservation measures. Further, the two approaches may call for different kinds of data analysis; descriptive statistics may be adequate for the former, while multivariate inferential statistics may be more appropriate in the latter case.

Program Persistence. This scale refers to the net savings for all measures installed throughout a program's lifecycle. It is represented in Fig. 2.3. Program assessment may be the most important persistence question for forecasting and planning.

The program scale has received little attention in the persistence literature, perhaps because relatively few programs have gone through their entire lifecycles and because of the significant cost associated with tracking net savings over multiple cohorts of program participants. This scale also includes net savings that may extend beyond the operational life of a program. Even though a particular program's activities may be completed, the measures installed in earlier years still may be operating or performing as "high-efficiency" measures. In addition, replacement energy savings (see Fig. 2.1) still may be accruing as the market for DSM measures is transformed.

As with the cohort scale, evaluators can assess specific factors that might affect energy savings over the lifecycle of a program, or they may simply postulate possible causes by analyzing the aggregate patterns. Along with the many factors that affect measure lifetime and measure performance, there are additional factors to consider at this scale that may cause changes in the nature of participants and the DSM measures they adopt. These factors include any changes in local energy prices that may influence the demand for DSM measures, changing economic conditions that might affect the ability of a utility's customers to purchase DSM measures, changes in efficiency standards (that may influence baseline measures), changing demographic characteristics, program ramp-up, and the effects of participant self-selection on savings over time. In the case of Brown and

White (1992), for instance, the Residential Weatherization Program attracted very different participants over its ten-year lifecycle. During its first few years of operation, the program attracted intensive energy users with great potential for cost-effective energy savings. During its later years, the program attracted participants with lower utility bills and less saving potential.



**Fig. 2.3 Persistence of Savings Over the Lifetime of a DSM Program:
The Program Scale**

2.2 PURPOSE OF STUDIES

There are two overarching reasons why persistence studies would be undertaken, to serve as a basis for DSM or integrated resource planning (IRP) or to evaluate or verify the savings from DSM measures or programs. These two general rationales for studying persistence can be refined further, as in Table 2.2. Note that this table reflects a utility perspective and not a customer perspective. Further, the table does not address another possible goal of persistence studies, to discover methods for improving program design and delivery. Program design improvements could involve altering procedures that lead to premature measure removal; program delivery could be improved by avoiding equipment brands that tend to be removed (Oswald 1995).

The methods required to satisfy planning and evaluation — particularly validation — activities are somewhat different from one another. Estimates of predicted measure or program persistence are key for planning purposes, while actual performance over time is central to evaluation. Planning and evaluation may be intertwined in that (a) the accuracy of prediction may be improved by evaluation results (though new data or modeling errors, or new variables may contribute to forecasting errors), (b) the process of evaluating often entails making estimates, particularly since many conservation programs have not been in place long enough to allow truly long-term persistence

studies (e.g., Eto *et al.* 1994; Keating 1991; Vine 1992), and (c) evaluations can contribute to an understanding of why predictions and forecasts differ (Baxter 1995). Planners also may benefit from acquiring information from evaluators to help them design programs that maximize persistence (e.g., Braithwait *et al.* 1994; Keating 1991; Vine 1992). Nevertheless, the distinction between planning and evaluation is important for persistence in cases when the information or activities needed for each endeavor differs.

Table 2.2 Purposes for Studying the Persistence of Savings

1. DSM Planning or IRP:
• resource planning (i.e., to assess the most cost-effective means of meeting future energy and load requirements), • load forecasting (i.e., to estimate future load requirements, based on alternative DSM investments with varying levels and trajectories of savings), and • annual program planning (i.e., to allocate resources to next year's portfolio of DSM programs).
2. Program Evaluation:
• program evaluation (i.e., to assess the cost-effectiveness of a program based on measured savings over time and to understand factors influencing program impacts), • lost revenue recovery (i.e., to compensate utilities for the loss of revenues from lower electricity sales due to DSM programs), • shareholder incentives calculations (i.e., to determine utility incentive payments that reflect the magnitude and longevity of DSM savings), and • performance contracting payments (i.e., to determine payments to ESCOs that reflect the magnitude and longevity of DSM savings).

Compare, for example, utilities' needs when planning their resource acquisition relative to their needs when validating conservation program or measure effectiveness. For a specific energy conservation measure, planners could use an equation that combines engineering estimates of technical measure life with foreseeable behavioral factors such as maintenance activities or building renovations to project the measure's performance over its life.

Evaluation generally requires evidence of actual measure performance (with performance being a function of technology and behavior), particularly when public utility commissions specifically state that they do not want utilities to rely on engineering estimates. For example, the Massachusetts Department of Public Utilities issued an order that would allow utilities to earn financial incentives for its conservation and load management programs so long as the utility can measure program savings and not rely on engineering estimates (Jacobson *et al.* 1992). The billing

analyses and site visits that typically are associated with evaluation exercises easily could be absent from planning activities. The results of billing analyses and site visits could provide useful and much-desired information for planners, but are not generally considered necessary for planning.

While planning and evaluation activities are integral to utilities, attention to persistence currently is not. The general awareness that engineering estimates of energy conservation measures typically overstate both energy savings and measure durability seems to have led to a number of studies evaluating actual savings and durability. Table 2.1 shows that investigations focus predominantly on evaluation. While several documents state that persistence is important to planning activities or to provide information that could be incorporated by planners (e.g., Braithwait *et al.* 1994; Keating 1991; Vine 1992), they do not provide measurement guidelines. A key conclusion from this review of persistence is that the methods handbook should recognize the different purposes for studying persistence and should discuss methods appropriate for each purpose.

3. RESEARCH DESIGN AND DATA COLLECTION ISSUES

A variety of methods are used to collect data on persistence, including surveys, billing data analyses, on-site inspections, and metering and monitoring. Multiple methods often are used in single studies; these methods also may be combined with engineering estimates to produce an evaluation. Other methods for acquiring information to determine persistence include using engineering assumptions and various parameters from previous studies (Braithwait *et al.* 1994; Jeppesen and King 1993). The choice of method is influenced by the study objective(s).

This section describes the main data collection methods currently in use. First, however, some fundamental research design and data collection issues are discussed. These issues are relevant to virtually all persistence studies, though their particulars may vary according to the type of study undertaken, its purpose, etc.

3.1 ASSUMPTIONS AND CONTROL GROUP OPTIONS

The assumptions underlying studies may strongly influence both research design and results. Stated in the extreme, there would be no perceived need for persistence studies if the past assumptions guiding the use of engineering estimates in determining measure life and energy savings remained unquestioned today. Although persistence studies are a recent phenomenon, it is clear that results based solely on engineering estimates may differ considerably from results based on "field" performance (e.g., Braithwait *et al.* 1994; Eto *et al.* 1994; Hirst, White, and Goeltz 1985; Jeppesen and King 1993; Keating 1991; Skumatz *et al.* 1991).

This section focuses on the working assumptions that researchers make, specifically on the assumptions (or choices) that influence research design and data collection. There are two caveats to this discussion. First, the phrase "working assumptions" is used to emphasize that these assumptions deliberately may be used to facilitate the completion of a project, regardless of whether they are accurate. For example, Brown *et al.* (1993) assumed a 20-year lifetime of all installed weatherization measures for the purpose of evaluating savings over that period of time though they specifically stated that some weatherization measures have a much shorter expected lifetime and discussed estimated measure lifetimes in an appendix. Second, financial constraints may force some utilities to make assumptions (e.g., about the physical presence of measures at sites or about the accuracy of survey results in determining the physical presence of measures) instead of measuring persistence.

Assumptions and choices about control groups are extremely important in studies that aim to estimate net energy saving, because of the necessity of attributing savings to a particular program (e.g., Keating 1991; Vine 1992). This attribution depends, in part, on the determination of how program participants would have behaved in the absence of a program. Free drivership (non-participants adopting program measures either consciously due to the program or because of changes in the market [Saxonis 1991]) is among the confounding variables in distinguishing the participant study group from a control group. "The potential implications of this difficulty are that efforts to

measure net program savings will be costly, and may provide results that are not sufficiently robust to significantly improve the existing uncertainty" (Braithwait *et al.* 1994, p. 16). In addition, the long timeframe of persistence studies complicates the size and composition of control or comparison groups (see *Attrition* in Section 3.4).

Various control group options have been used in previous persistence studies. Table 3.1 describes the approaches used by the four persistence studies highlighted by this literature review. These options mirror those available in non-persistence studies. However, the time dimension of persistence studies challenges the integrity of control groups.

Table 3.1 Examples of Control Group Options

- **no controls:** Long Island Lighting Company (Parker 1993) — gross savings estimated, based on counts of installed equipment in commercial facilities
- **no controls:** Hood River Conservation Project (White, Stovall, and Tonn 1992) — estimated gross savings, based on differences between pre- and post-weatherization load consumption, normalized for long-term weather conditions and by comparing consumption during peak days
- **control group of non-participants:** Bonneville's Residential Weatherization Assistance Program (Brown and White 1992) — estimated net savings, based on comparisons of participants and non-participants; the non-participant sample was weighted to match the geographic distribution of participating households
- **control group of later participants:** Wisconsin low-income residential weatherization program (Narum, Pigg, and Schlegel 1992) — estimated net savings, based on year-to-year comparisons of early and later program participants, where later program participants are the control group for earlier program participants

When comparing participants with non-participants, it is typical to specify that non-participants be matched in some way with participants to provide a reasonable indication of what program participants might have done (or how buildings or conventional equipment might have performed) in the absence of a program. Therefore, billing-data analyses may use eligible non-participants as comparison groups to enable researchers to distinguish program-related changes in energy consumption from non-program-related factors (e.g., White and Brown 1990). Using eligible non-participants has been thought to control for free-ridership effects (participating in programs even though program-sponsored energy conservation measures would have been adopted in the absence of the program) by providing a basis for estimating what participants might have done in the absence of a program (Keating 1991). In some cases, however, this assumption may not be valid. Participation in voluntary programs may distinguish participants from non-participants in ways that influence their energy use, undermining the notion "that non-participants behaved the way participants would have behaved without the program" (Train 1994, p. 426).

Narum, Pigg, and Schlegel (1992) used a variation of the above themes in their study of the persistence of energy savings in low-income Wisconsin residences. They used as their comparison

group "a similar group of untreated buildings (composed of later participants), with the assumption that the untreated buildings' energy usage patterns is [sic] a reasonable representation of what the treated buildings would have done had they not participated in the program" (p. 3).

In some cases, no control groups are used and comparisons are made pre- and post-program implementation. This method is known as the time-series approach (Violette 1991). The rationale given for a study of the persistence of load reductions due to the Hood River Conservation Project's residential retrofit demonstration project was that the use of a control group "would have interfered with the maximum possible penetration goal of the [Hood River Conservation Project]" (White, Stovall, and Tonn 1992, p. 1). There were so many participants in the program that the pool of non-participants was too small for meaningful comparisons. The main advantage of the time-series approach is that dwelling and demographic factors are unlikely to change over a few-year timeframe. However, the approach cannot account for such factors as market progression, and typically does not address the impacts of weather (the White, Stovall, and Tonn 1992 study is an exception) or electricity prices (Violette 1991).

Studies focusing on measure performance also entail comparisons, but in these cases the comparisons are with conventional equipment or measures. The LILCO commercial/industrial persistence study (Parker 1993) is an example in which these kinds of comparisons are made. Gross energy savings were estimated on the basis of assumed energy savings associated with the equipment that was in place. As Keating (1991, p. 94) states, "Measuring persistence without a comparison group implicitly assumes that standard equipment experiences no deterioration in performance over time."

Control group considerations are applicable to any evaluation that seeks to determine net energy savings. Persistence studies pose particular challenges to establishing and maintaining adequate control groups, for the following reasons:

- the need for control group members to remain unaffected by a program for an extended period;
- the cost of selecting a control group and obtaining data from such a group;
- changes in programs over time;
- market transformation;
- sample attrition over time;
- changes in the composition of participant cohort groups over time (with different potentials for energy savings); and
- changing economic conditions that can influence patterns of energy consumption.

3.2 RETROSPECTIVE VERSUS PROSPECTIVE DESIGNS

By whatever definition, persistence deals with time. There are four related aspects of time that have particular relevance to research design and measurement. These aspects are the timespan addressed by a study, whether study designs are retrospective or prospective, the frequency with which persistence studies ought to be performed, and the timing of persistence studies relative to measure or program implementation.

Retrospective studies look at past events and use existing information to estimate persistence. The majority of published studies are retrospective (e.g., Brown and White 1992; Hirst, White, and Goeltz 1985; Jacobson *et al.* 1993; Narum, Pigg, and Schlegel 1992; Parker 1993; Skumatz *et al.* 1991; White and Brown 1990; White, Stovall, and Tonn 1992). In contrast, prospective studies are designed to estimate persistence by following participants over time. The research design is established at the beginning of, or early in, a program's life and data are collected in subsequent years. There are advantages and disadvantages to both kinds of study. Retrospective studies, for instance, are limited by the quality, kinds, and completeness of data available: the vagaries of recollection; and fallible reports of current behaviors. The major advantage of retrospective studies is the relative accessibility of certain kinds of data such as maintenance records, test measure life estimates, building permits and utility service records, and billing data. Prospective studies' main advantages are their ability to provide better quality, more detailed, and more timely data for utilities than can retrospective studies. Because prospective studies are designed before data are collected, they may be better able to collect the data most appropriate to answer research questions.¹ Although sampling issues are important for retrospective and prospective studies, prospective studies may provide researchers with more options. For instance, researchers may choose to follow measures that may be moved from one site to another (Vine 1992). Retrospective studies do not track the location of removed measures, presumably due to the difficulty, time, and expense involved in re-creating those migration patterns.

The California Conservation Inventory Group (CCIG) measure life study (Jeppesen and King 1993; Jeppesen and Rudman 1993) explored options for, and components of, prospective research designs to evaluate measure life for a group of extant program participants that would be followed longitudinally in future years. CCIG's attention to prospective research design issues is rare in the literature currently available. Although the extensive discussion cannot be summarized here adequately, an attempt is made to sketch many of the salient points.

Prospective study goals, according to the CCIG study, would be to produce reliable and defensible estimates of DSM measure energy savings over time; the estimates would contribute to DSM evaluations, forecasting, and shareholder earnings. In part because of the expense involved in

¹ Braithwait *et al.* (1994) point out that prospective studies eventually require information about past programs, much like retrospective studies. However, because plans for collecting such data already will be in place, the chances of collecting accurate and appropriate data may be better with prospective than with retrospective studies.

conducting long-term studies, CCIG members compiled a list of the 10 highest priority measures for both commercial and residential sectors. These measures were judged likely to be important to future DSM efforts and their effective measure lives were deemed relatively uncertain. Study authors envisioned a five-year study and recommended using single- (univariate) or multiple-parameter (generally, multivariate) survival functions to estimate a measure's overall mean effective life or the differential effects of causes of failure on a measure's overall mean effective life, respectively. In addition to the typical data requirements for measure persistence studies (measure presence or cause for removal, installation and sizing, remodeling, changes in occupancy or residency, measure operation, measure maintenance, additional failure data, and information on customer behavior), the study adds population characteristics for residences, businesses, and industries as well as tracking data on measure migration. Several data collection methods are suggested: (a) on-site inspection to check measure replacement, measure performance, and to provide consistent data during the course of study; (b) supplemental telephone or mail surveys to follow-up site inspections; and (c) monitoring energy and patterns of equipment use to gauge the operation of appropriate measures. No comparison groups are recommended because the entire focus is on the effective measure life of high-efficiency measures. The authors do not suggest trying to replace those customers who withdraw from the study because replacement is likely to be cost-ineffective and because they prefer direct observation to respondent recollection. Nevertheless, the authors suggest incorporating an oversampling strategy in the research design and to promote the retention of an adequate sample size (with the sampling unit typically consisting of a location at which measures are installed rather than facility occupants).

3.3 FREQUENCY AND TIMING OF DATA COLLECTION

The timespan addressed by a study influences the selection of research methods. Because a number of measures and programs whose expected durability is considerable have only relatively recently been in effect, insufficient time has elapsed to measure actual long-term persistence. As Eto *et al.* (1994) point out, the persistence literature currently consists primarily of short-term studies (first few years after installation), not medium- or long-term studies. This situation translates into evaluation studies that combine measurement with estimation. The total period of study (or, of anticipated program or measure lifetimes) also may constitute an important factor when considering how much to invest in that study. The financial costs of purchasing, installing, and operating monitoring equipment, for instance, may be viewed quite differently for a two-year effort than for a twenty-year effort.

There apparently is no consensus about the frequency with which persistence studies should be conducted. Protocols such as those adopted by the California Public Utilities Commission may prescribe study frequencies (Pacific Gas & Electric *et. al* 1993). As an example, these protocols (p. 17) indicate that Pacific Gas & Electric's, San Diego Gas & Electric's, and Southern California

Edison's new construction programs have a 10-year measurement period during which there are four scheduled persistence studies. The persistence studies are as follows: (a) fourth year, load impact study ("an analysis of the net program impacts in the designated load impact year" [p. 17]); (b) fourth and ninth years, retention studies (collecting "data on the fraction of measures or practice[s] (sic) remaining in a given year that will be used to provide a revised estimate of its effective useful life" [p.17]); and (c) fourth year, performance study ("a time series analysis of the relative change in the performance/efficiency of high efficiency equipment or high performance shell measures over time" [p.17]). These studies are required for the utilities' earning claims. Table 3.2 displays the frequency of required persistence studies.

The author of the LILCO persistence study (Parker 1993) suggests performing measure persistence studies every two or three years. Reasons for this recommendation are to avoid customer irritation with annual surveys focusing on the same item, to allow time for economic conditions to change, and to avoid allocating substantial financial resources to efforts whose results likely will be only marginally different.

Braithwait *et al.* (1994) discuss the timing and frequency of persistence studies in terms of the tradeoffs between the costs incurred and the value of the information obtained.² So, for instance, a short-term persistence study would provide valuable information in cases where impacts are quite uncertain. Persistence studies could be initiated somewhat later when there is good information about installed measures and their effects are predicted to be stable for several years. Another alternative is to time persistence studies so as to be most appropriate for the types of equipment installed.

3.4 SAMPLING AND ATTRITION

Sampling choices should be driven by the study goals, the desired units of measure, the duration of the study, data availability, budget, time, and personnel constraints, etc. For persistence studies, changes that occur over time also influence sampling. These changes may involve such items as program modifications, evolving participation in multi-year conservation programs (new program participants over time, people or organizations dropping out of programs), energy-related behaviors (surge effects, takeback, market progression, maintenance and replacement habits, etc.), and other activities (e.g., building vacancies, turnover, renovation, or demolition; people relocating). These dynamic conditions make sample identification and sample retention particularly challenging.

Attrition. Sample attrition is a fact of life in energy savings studies. However, its effects are amplified in persistence studies because increasing numbers of participants drop out over time. A variety of factors contribute to sample attrition during the course of data collection. Some of these factors have the effect of reducing the sampling frame (the total set from which a sample can be drawn) rather than of causing the attrition of an already-selected sample.

² The LILCO study (Parker 1993) makes a similar point, but in terms of economics rather than in terms of the timing of persistence studies.

Table 3.2 Example From California Protocol

Persistence Studies Required For An Earnings Claim For Pacific Gas & Electric, San Diego Gas & Electric, and Southern California Edison

Program	Required Persistence Studies	
	Measurement Period	Measurement Schedule
Residential, Weatherization Retrofit Incentives, and Appliance Efficiency Incentives—Space Conditioning	10 years	1. 4th year load impact study 2. 4th and 9th year retention 3. 4th year performance
Residential and Appliance Efficiency Incentives—Lighting and Refrigeration	To be determined by statewide studies	To be determined by statewide studies
Commercial Energy Efficiency Incentives	10 years	1. 4th year load impact study 2. 4th and 9th year retention 3. 4th year performance
Industrial and Agricultural Energy Efficiency Incentives	7 years	1. 3rd year load impact study 2. 3rd and 6th year retention 3. 3rd year performance
New Construction	10 years	1. 4th year load impact study 2. 4th and 9th year retention 3. 4th year performance
Miscellaneous	10 years	1. 1st, 4th and 9th years retention
Residential Direct Assistance	10 years	1. 1st and 9th year retention
Residential Energy Management Services	7 years	1. 4th year practice retention 2. 6th year practice retention
Commercial and Industrial Energy Management Services	10 years	1. 4th year practice retention 2. 9th year practice retention
Agricultural Energy Management Services	7 years	1. 4th year practice retention 2. 6th year practice retention

Source: Pacific Gas & Electric *et al.* 1993

Among the factors that may contribute to sample attrition are the following:

- incomplete billing or consumption history, if the study is a billing analysis (e.g., Keating 1991; Narum, Pigg, and Schlegel 1992; White and Brown 1990);

- participation in other energy conservation programs, either before or after program in question, if such participation affects related equipment (e.g., Narum, Pigg, and Schlegel 1992);
- earlier participation in the same energy conservation program (e.g., Narum, Pigg, and Schlegel 1992);
- non-participants electing to participate in the program (e.g., Keating 1991);
- insufficient or anomalous data that prevent researchers from assigning buildings, customers, or energy conservation measures to appropriate groups (e.g., Narum, Pigg, and Schlegel 1992);
- building vacancies, remodeling, turnover, or demolition (e.g., Jacobson *et al.* 1993; Parker 1993); and
- participants or energy conservation measures moving (e.g., Jeppesen and King 1993).

Responses to sample attrition fall into three broad categories. First, the research design deliberately can include initial oversampling in anticipation that there will be sample attrition. This fix can be simple, as long as the attrition occurs without bias across sample subgroups. Second, attempts can be made to replace the losses. Replacements may have to be identified to match the composition of the cases lost, if attrition bias is found to exist. Finally, researchers simply can perform analyses with fewer data points. If this approach is taken, analysts should test to determine whether or not attrition bias exists, and then apply weightings to correct for any bias present. An example of this weighting is provided by Brown and White (1992), who corrected for differential rates of attrition among the homes served by the eight utilities that participated in Bonneville's data gathering project.

3.5 DATA COLLECTION METHODS

Four principal methods are used to acquire information about persistence—telephone and mail surveys, billing analyses, on-site inspections, and metering or monitoring. The kinds and quality of data these methods can provide vary, as do their costs. In some cases, regulators provide guidance about the methods that should be used. For instance, the California PUC protocol states that for a load impact study beyond the first load impact year, the “study should make every effort to include an analysis of the billing data from both the comparison and participant groups that were used in the first year impact study (even though some sample attrition is expected)” (Pacific Gas & Electric *et al.* 1993, p. 18). The choice of methods also is influenced by the goals of persistence studies.

Several authors are sensitive to the relative financial costs of implementing data collection methods (e.g., Parker 1993; SRC 1992; Violette *et al.* 1991). However, Braithwait *et al.* (1994) go further to frame method selection in the context not only of the value to the utility of the information likely to be obtained, but also in the context of the risks to the utility of uncertainty about the persistence of energy savings. These authors categorize such risks in terms of (a) the value of DSM

as a resource, (b) the prudence of utilities' expenditures and the size of their incentive payments, and (c) the costs associated with not improving programs. So, for example, where planned resource acquisition is limited, persistence information for resource acquisition has relatively little value. The risks of uncertainty with regard to utilities' expenditures and incentive payments vary according to a combination of the following three factors, according to these authors: the direction of bias (e.g., overestimates versus conservative estimates) and uncertainty of persistence information used in DSM planning and evaluation; the magnitude and nature of participants' rebate payments (e.g., large rebate on energy conservation measures paid by utilities versus primary financial burden on participants); and the characteristics of utilities' incentive payment agreements (e.g., based on estimates versus evaluations or verification).

Clearly, choices about which method(s) to adopt must take into account the trade-offs between data quality and costs of obtaining those data relative to the potential benefit to the utility. Nevertheless, it is important to recognize that "in nearly all cases, energy savings resulting from the installation of [energy efficiency measures] can only be *estimated* by combinations of engineering analysis and econometric analysis of changes in energy consumption, while controlling for other factors. Furthermore, attributing the appropriate portion of those savings to the program (i.e., net savings) requires inferences from analysis of customer surveys" (Braithwait *et al.* 1994, p. 12).

3.5.1 Telephone and Mail Surveys

Telephone and mail surveys are basic tools used to conduct many persistence studies, particularly those whose aim is to study measure persistence or total impacts (Braithwait *et al.* 1994; Keating 1991). As Vine (1992) states, they are proven methods that are relatively inexpensive to administer, cost-effective, accurate for certain kinds of data, and can be used to elicit a variety of data. Their limitations are in their relative lack of depth and their reliance on self-reported data, which can be erroneous or biased. Vine (1992) supports the use of such surveys in the following cases (clearly evaluation-oriented, not planning-oriented):

- (1) certain kinds of measures—e.g., residential measures, inexpensive and low risk measures, and measures with good program tracking data, (2) particular programs—programs that are one-measure oriented (non-customized), (3) presence of equipment—rather than the condition or efficacy, and (4) attitudinal and behavioral questions—e.g., customer's experience and satisfaction with the measure. (p. 1078)

Keating (1991) also suggests using surveys to identify free riders and free drivers, although he recognizes that surveys may not be particularly good tools for eliciting motivations for adopting energy-efficiency measures or behaviors.

There is some evidence that data about the third item in Vine's list, presence of equipment, may not be ascertained accurately from telephone or mail surveys. Results of the LILCO persistence study of Commercial Audit and Dollars and Sense Programs (Parker 1993; Velcenbach and Parker 1993; Parker 1994) indicate that mail or telephone surveys may not provide accurate information about the presence of measures, instead suggesting a walk-through of the entire facility (including

store-rooms). Nevertheless, telephone and mail surveys are popular tools for tracking and verifying the presence and proper operation of energy efficiency measures (e.g., Jeppesen and King 1993).

***Example case: California Conservation Inventory
Group measure life study (Jeppesen and King 1993)***

- **focus:** effective measure life
- **goal:** part of larger effort; this part tested “the feasibility of using a *retrospective* research approach to the study of effective measure life” (p. III-1)
- **sector:** residential and commercial/industrial
- **target population:** energy efficiency program participants (4 utilities)
- **comparison/control group:** none (non-statistical sampling; study focused on methods)
- **data sources:** utility records on customers and measures installed; telephone survey; on-site inspections
- **time span of study:** study conducted ca. 1992; retrospective portion focused on four time spans — 1976-1980, 1981-1985, 1986-1989, 1990-1991
- **sample size and attrition:** residential — 100 completed interviews (“sample points”), available sample of 1,149; commercial — 56 completed interviews (“sample points”), available sample of 120.

The overall study sought “to obtain recommendations on future research and methods for estimating effective measure life” (p. I-2). Study components consisted of a literature review, a feasibility study of the effectiveness of retrospective measure life research, and recommended research designs for future measure life studies. Telephone surveys were part of the retrospective facet of the research, “designed to be an evaluation of the research processes required for a retrospective research approach” (p. III-1). Three particular items were investigated: sampling, focusing on data availability and quality; the ability of program data to identify both customers and the measures installed; and the adequacy of retrospective data collection methods such as telephone- and on-site surveys. Secondly, although non-statistical sampling methods were employed, researchers collected substantive data from participants on selected, high-priority conservation measures.

The authors suggested a two-step sampling process in which program data were scanned before drawing the sample. They also discovered that respondents were more likely to be willing to provide information over the telephone than to agree to on-site surveys.

These tools may be used in conjunction with other methods such as site inspections (e.g., Skumatz *et al.* 1991). Telephone surveys may be sufficient for residential sector measures or when, in the future, data on programs improve. However, such surveys may be inadequate as the sole source of information about commercial and industrial programs. In such cases, an effective combination may be to collect data from a small on-site inspection sample and a larger, overlapping telephone survey sample (Braithwait *et al.* 1994).

Another use of telephone interviews is to acquire background information that helps shape subsequent research. As an example, Skumatz *et al.* (1992) used telephone surveys in the first phase of research, specifically “to obtain a better understanding of commercial renovation, remodeling, and turnover rates, and the resultant effects, if any, upon energy-using equipment lifetimes” (p. I-3). Telephone interviews were conducted with a total of 106 knowledgeable respondents such as

“contractors, architects, property management firms, government agencies, large private firms, and chain operators” (p. I-3). These interviews elicited information about the types of building changes (renovation, hard-, and soft-remodels) that had occurred; business segment turnover; percentage of buildings retaining same business segment after turnover; building changes typically accompanying turnover; the disposition of equipment removed during renovation or remodel; factors influencing equipment replacement decisions; energy-efficient equipment availability; and experiences regarding equipment operation and maintenance.

3.5.2 On-Site Inspections

On-site inspections are most appropriate to evaluate and verify the physical persistence (presence) of energy conservation measures and the factors influencing the lack of such persistence. The information they may provide about physical persistence is more detailed and accurate than information provided by other methods. Inspections (and metering and monitoring) also may be the only means of ascertaining information about measure performance (Braithwait *et al.* 1994). Jacobson *et al.* (1993) suggest that on-site inspections of non-lighting measures may be particularly fruitful since inspectors may determine where along a continuum of performance those measures may fall. Lighting measure performance, in contrast, may be judged dichotomously — as either working or not working. The amount of information collected and the degree of data precision affect the duration of site visits and their costs; information about measure performance therefore takes more time and costs more money than simply verifying the presence of measures (Braithwait *et al.* 1994). These practical considerations may explain why on-site inspections have been used primarily to determine measure retention rather than measure performance.

A particularly well-developed set of inspection forms was developed for LILCO (see example case). Based in part on a review of several persistence studies, the researchers concentrated on the status of rebated, recommended, or in-place equipment. A lengthy data collection form was used. This form contained, as an example, a site verification sheet to provide a mechanism for collecting data in an organized and replicable way, and for developing tables based on a variety of scenarios (all equipment inspected, equipment prematurely replaced — and the reasons for replacement, equipment not identifiable as rebate equipment, etc.). The form included the following categories: code, device/measure, number rebated or percentage installed, number or percentage verified, status (still operating, never installed, etc.), disposition (where equipment is located, including “unknown”), replacement (if replaced, or with what kind of equipment and rebate, if any), and comments. The “replacement” category allowed researchers to quantify free drivers.

Inspectors were provided with a set of assumptions to help them fill out the site verification form. The format of these assumptions was if-then scenarios that described, as ‘ifs’, rebate documentation, on-site inspection findings, and qualifiers (e.g., customer contact cannot provide adequate information). The assumptions constituted the ‘thens’. Apparently, auditors sometimes found it difficult to match the measures inspected with the measures that had been recommended for

rebate. (Jacobson *et al.* 1993 and Jeppesen and Rudman [1993] reported a similar findings for commercial lighting measures.)

***Example case: Long Island Lighting Company
(Parker 1993)***

- **focus:** physical measure persistence
- **goal:** "to estimate the lost savings due to premature replacement" (p. 8) for Dollars & Sense and Audit programs
- **sector:** commercial and industrial
- **target population:** customers receiving rebates or for whom audits were completed in 1988, 1989, and 1990
- **comparison/control group:** none
- **data sources:** utility facility, rebate, audit, energy savings, and account turnover data; on-site inspection data; and face-to-face interviews with facility managers
- **time span of study:** study conducted in 1992; records used from 1987 through 1991
- **sample size and attrition:** over 600 site surveys; no attrition figures

Research design consisted of two major parts, preliminary database work and inspections. Preliminary database compilation and manipulation was performed for the following purposes: to determine the sample frame; to define what information was sought; and to establish how to obtain the desired information. Three databases ultimately were created — Audit, Rebate, and Overlap. The Audit database, consisting of customers who received audits and recommendations, included information about the facility and the audit. Facility information contained facility name and address, contact person, telephone number, and peak demand. Audit information comprised audit number, date, recommended measures, reported percent of installation, kW saved, annual energy savings, and data from the original audit (XENCAP audit) describing the base case, energy-efficiency measure, and quantities and locations of equipment. The Rebate database also included the rebate amount, both kW and dollars and the quantities of rebated equipment, in terms of installed tonnage (e.g., central air conditioners and chillers), BTU output capacity (e.g., room air conditioners), or kW (e.g., non-electric cooling, thermal energy storage, custom rebates). As its title indicates, the Overlap database consisted of those facilities that received both audits and rebates.

Information about account turnovers was added to each of the databases. Identifiers were placed on facilities in the databases that experienced account turnovers from 1987 through 1992. (The utility tracks account turnover data.) Account turnovers indicate meter installation in new or existing buildings, changes in meter or billing names or filing for bankruptcy protection. Account turnovers are not necessarily equivalent to ownership or tenant changes.

The second aspect of the LILCO study consisted primarily of on-site inspections. Researchers tried "to inspect 100% of the facility and rebated or audited measures" (p. 14). When equipment was not accessible (e.g., ballasts), the goal was to inspect about 10% of the quantity rebated through a random selection process that relied on the facility manager's cooperation. When only visual inspection was possible, researchers had to rely on information from the facility's staff.

In addition to inspections, in-person interviews were scheduled with facility managers, particularly the people responsible for the facility's energy use. Ideally, these interviews were conducted at the time of the site inspection.

Vine (1992, p. 1078) listed the following advantages of on-site inspections:

- verifying independently (objectively) measure installation and the quality of installation;
- determining whether measures were applied appropriately;
- determining whether measures are sized appropriately and whether they are operating;
- identifying the energy systems affected by measures;
- determining the maintenance level of a particular measure (e.g., its cleanliness);
- examining measures' efficiency levels and their performance degradation or failure; and
- ascertaining measure-specific information (model numbers, temperatures, and lighting levels).

Braithwait *et al.* (1994) added two items to these advantages. First, on-site inspections have the capability of providing information about opportunities for additional energy conservation measures. Second, and as the LILCO study discussed (Parker 1993), they allow for face-to-face interviews with people who can explain observed conditions.

Disadvantages include varying levels of expertise among on-site auditors; limited access to equipment and measures (e.g., insulation, ballasts); limited or no access to certain locations within a facility (e.g., laboratories, operating rooms, buildings containing hazardous waste, clean rooms, locations where security clearances are required); the inability of inspection alone to document historical and behavioral characteristics; the possibility of a large non-response bias; and high costs relative to billing analyses and telephone or mail surveys (Parker 1993; Vine 1992). Vine promotes the use of on-site inspections for "high priority" measures, such as: "measures with complicated installation (e.g., HVAC systems and energy-intensive industrial process measures); measures with high savings, high costs, or high risk; measures needing high maintenance; and measures with poor program tracking data" (p. 1078).

3.5.3 Billing Data

The use of billing or consumption data is associated primarily with program persistence studies that aim to evaluate net impacts. Studies of electricity savings from Bonneville Power Administration's Weatherization Program are prominent examples of reliance on billing-data analyses (e.g., Hirst, White, and Goeltz 1985; White and Brown 1990).

Sample attrition due to lost or incomplete billing records, participants moving, or non-participants electing to participate can be a serious problem for billing analyses. Attrition rates can be very high. Other problems with billing data methods may be the confounding effects of free-ridership, free-drivership, and self-selection bias — a systematic difference between control and participant groups in which participants' generally greater awareness of conservation issues would influence them to use less energy and to adopt more energy conservation measures than control groups, even in the absence of a DSM program (Violette *et al.* 1991).

Keating (1991) cautions that billing analyses may be unsuitable for large commercial and industrial facilities because of the absence of appropriate comparison groups. Therefore, it is possible that only an estimate of savings can be obtained.

Beyond these items, it is particularly difficult to distinguish research design and measurement elements from data analysis elements for billing analyses. A considerable amount of data processing is necessary, for instance, to determine which observations are usable for analysis (e.g., Violette *et al.* 1991). While these data processing procedures arguably are more appropriate to discuss in terms of data analysis than data collection, they are discussed here because they also are inherent to research design.

In an early evaluation of Bonneville's Residential Weatherization Program, White and Brown (1990) used a number of screens that, together with weather normalization procedures, reduced their participant sample from 513 to 252 and their non-participant sample from 1,339 to 688. (Another element that was important in reducing the sample size was that the screening criteria had to apply for three years, so that the authors could determine third year post-retrofit savings.) Screens that reduced the number of participants' worksheets obtained (1,502) to the number included in the study sample (513) included (1) removal of households that were not single-family, non-low-income customers; (2) removal of households that previously participated in other programs or that previously had been excluded from Bonneville's residential weatherization program samples; and (3) removal of households that were weatherized under the 1986 Long-Term Residential Weatherization Program either early or late in the year. Other screens were the following: (1) removal of households that moved during the study period; (2) removal of households with anomalous billing records; and (3) removal of households that were unsuitable for the weather-normalization model, PRISM. (See chapter 4 for a more detailed discussion of weather normalization). PRISM analyses removed households for the following reasons: (1) less than the minimum of two years' worth of billing history; (2) under 3,000 kWh whole-house electricity use in any single year; (3) PRISM electric space heating estimates of zero or less; (4) failure to have 240 consecutive days of billing history in each year; (5) households whose energy-use patterns were not in alignment with typical all-electric households' energy use patterns.

The later Bonneville study (Brown and White 1992) used similar screening procedures as the 1990 study. However, the latter study also employed less restrictive sample selection criteria, resulting in expanded sample sizes. The 1992 study included households with low covariances of normalized annual consumption (NAC), meaning that such households' consumption of electricity did not change significantly with heating degree days. Using households with low NAC covariances increased sample sizes, particularly for non-participants. In 1988, the participant sample grew from 324 to 356 and the non-participant sample grew from 1,009 to 1,170. The situation in 1989 was similar in that the participant sample increased from 383 to 433 and the non-participant sample increased from 1,238 to 1,466.

***Example case: Bonneville's Residential Weatherization Program
(Brown and White 1992)***

- **focus:** program savings
- **goal:** "to provide an impact evaluation of Bonneville's 1988 and 1989 long-term Residential Weatherization Program" (p. 1.1)
- **sector:** residential, single-family homes
- **target population:** utilities (6 in 1988 and 9 in 1989) and participating households
- **comparison/control group:** control group — eligible non-participating households; comparisons also made pre- and post-installation
- **data sources:** billing records, National Oceanic and Atmospheric Administration
- **time span of study:** 1 year before through 2 years after installation
- **sample size and attrition:** For 1988, participants — originally 356, ultimately 315; non-participants — originally 1,170, ultimately 1,084. For 1989, participants — 433; non-participants — 1,466.

This study represents one of a series of studies evaluating Bonneville Power Administration's long-term Residential Weatherization Program. The focus of this study was on single-family, non-low-income households. Data were collected from two primary sources: participating utilities provided billing records and the National Oceanic and Atmospheric Administration provided weather data (average daily temperature) for the same time period for which billing data were obtained. The pattern of electricity consumption of a sample of participants was compared with that of a sample of non-participants.

***Example case: Wisconsin's Low-Income Weatherization Assistance Program
(Narum, Pigg, and Schlegel 1992)***

- **focus:** program savings
- **goal:** to determine the persistence of energy savings in large groups of customers at a program-wide level
- **sector:** residential; low-income, gas-heated buildings (single- and multi-unit buildings)
- **target population:** low-income, gas-heated buildings participating in Utility Weatherization Assistance Program
- **comparison/control group:** control — untreated buildings (later participants); compared treated and untreated buildings year-to-year
- **data sources:** 2 utilities' weatherization program databases, fuel consumption information
- **time span of study:** up to 8 years post-installation
- **sample size and attrition:** for Wisconsin Gas Company — 9,956 buildings in weatherization database, consumption data matched for 7,259 buildings, final sample 5,129; Madison Gas & Electric Company — 2,969 buildings in weatherization database, consumption data matched for 2,270 buildings, final sample 1,553 buildings

This study is similar to the evaluation of the Bonneville Power Administration's long-term residential weatherization program in its focus on overall program savings from packages of weatherization measures. The primary methodological difference between the two studies is in the use of control groups. The Bonneville Power Administration study used two kinds of comparisons — pre- and post-participants as well as participants versus non-participants. The Wisconsin study, in contrast, compared participants and non-participants in any particular year with the preceding year such that energy use was assessed from year to year.

3.5.4 Metering/monitoring

Metering or monitoring may be used to provide detailed, direct information about load effects or about measure performance (Braithwait *et al.* 1994; White, Stovall, and Tonn 1992). Metering entails the use of meters to collect whole-building or facility energy consumption data. Monitoring refers to the collection of non-energy consumption data such as temperature, humidity, and duty cycle at particular premises (Pacific Gas & Electric *et al.* 1993). Although they may be the best available methods for actually measuring energy consumption, results still may have to be stated as estimates for four reasons (Braithwait *et al.* 1994; Vine 1995). First, because direct information is available only for the duration of the metering or monitoring, analysts must make assumptions when generalizing the results from a short period of time (e.g., a few weeks) to an entire year. Second, unless there is long-term metering or monitoring, those methods do not provide information about changes over time. Third, generalizing the results of metering and monitoring a small segment of the population to the larger population requires assumptions to be made. Fourth, when considering net savings, analysts still must consider what portion of the measured savings to attribute to the program.

Example case: Hood River Conservation Project (HRCP)
(White, Stovall, and Tonn 1992)

- **focus:** persistence of program benefits, particularly load savings
- **goal:** to determine the persistence of typical and peak load savings of retrofit demonstration project three years after measures installed
- **sector:** residential
- **target population:** electrically heated houses participating in project
- **comparison/control group:** no control group; comparisons made pre- and post-weatherization
- **data sources:** metering data collected every 15 minutes; three weather stations
- **time span of study:** 1 year before through 3 years after installation
- **sample size and attrition:** originally 320 homes; ultimately 220 homes

The HRCP was intended "to install as many cost-effective retrofit measures in as many electrically heated homes as possible in the community of Hood River, Oregon" (p. 1). This study reported findings from a sample of 320 homes whose end-use loads were submetered for one year before the measures were installed and for three years after installation. It used a time-series approach to focus on "the electric load profiles, load savings, persistence of savings, and fuel switching for [those homes]" (p. 4). Every 15 minutes, submetered data were collected on interior temperature and on electric space heating, electric water heating, and wood fuel space heating end-use loads. Extensive weather data also were collected at 15-minute intervals from three weather stations; these data were averaged. These field data were collected by Pacific Power & Light Company and transmitted to Oak Ridge National Laboratory for compilation into a database. In addition to these data, the database contained audit and weatherization information as well as data collected during four occupant surveys of the submetered houses. The database was submitted to rigorous quality assurance measures.

The original study sample consisted of 320 participating homes, each of which was to be submetered. These homes were selected to represent a cross-section of the community. Sample attrition over a four-year timespan resulted in 220 homes being available for analysis.

3.6 SECTOR AND END USE

This section discusses the relationship between customer sectors (residential, commercial, and industrial) or energy end uses (e.g., lighting and space heating) and persistence. In this context, the main question is, what influence does, or should, sector and end use have on the methods used to evaluate persistence? With regard to sector, forms of the four major categories of methods (surveys, billing analyses, on-site inspections, and metering and monitoring) have been used for residential and commercial/industrial sectors. The literature occasionally addresses the role of sector on the selection of study methods. For example, billing-data methods may be inappropriate for large commercial and industrial sectors because of the difficulty in designating control groups (Keating 1991). Keating also indicated that sample attrition and attrition bias may be particularly strong in the commercial sector.

Sector may influence how measure persistence studies are conducted. Some researchers recognize that evaluating the physical persistence of measures in commercial and industrial facilities, particularly via site visits, may simultaneously provide opportunities to promote measure persistence (e.g., Keating 1991; Parker 1993). Inspectors can check to see if the equipment is in place, running properly, etc. as well as provide facility managers with appropriate educational information.

Table 2.1 shows that, among the studies that specify a sector, all sectors are represented. Most commercial or commercial/industrial sector studies focus on measure persistence (e.g., Parker 1993), although some studies emphasize whole-building energy savings. In contrast, the residential sector studies tend to be less measure- or end-use-specific, instead focusing on whole-building energy savings. A number of residential-sector persistence studies evaluated weatherization programs (e.g., Brown and White 1992; Hirst, White, and Goeltz 1985; Narum, Pigg, and Schlegel 1992; White and Brown 1990 and 1992).

Like sector, end use seems to play a limited role with regard to the methods for ascertaining persistence. It is not surprising that persistence-related studies tend to focus on particular end uses more often when they emphasize measure life or lifetimes than when their goals are to evaluate programs (exceptions include end-use-specific programs). Billing and consumption data typically are not used to evaluate measure lifetimes. Therefore, to the extent that studies specifying end use aim to evaluate physical measure persistence and not net energy savings, such studies tend to use methods other than billing analysis.

The practical problems posed by verifying the existence, proper operation, and energy savings of measures associated with different end uses also may influence the selection of methods and the degree of confidence placed in the results obtained through those methods. As an example, it is quite difficult to verify commercial lighting measures. Survey respondents may not provide accurate information about the installation and operation of those devices, perhaps because of their reliance on assumptions about the facility in question instead of on an intimate knowledge of the particulars of lighting in that facility. On-site inspections therefore may provide information superior to that obtained through telephone or mail surveys. However, field inspectors report their difficulties

in determining whether the commercial lighting measures observed during on-site inspections actually are the program-supported measures (e.g., Jacobson *et al.* 1993; Jeppesen and Rudman [1993]; Parker 1993). The methods handbook will grapple with the tradeoffs among methods in terms of their financial costs, time, personnel, and data reliability.

4. DATA ANALYSIS ISSUES

There is considerable overlap between the issues relevant to research design and implementation and the issues relating particularly to data analysis. Since data analysis methods are often planned in conjunction with the research design, many of the issues are identical. For example, data analysis methods are influenced by study objectives; different methods are suitable for different study goals.

All the studies considered in this review of persistence studies may be identified either as Equipment Survey/Engineering or Consumption Evaluation studies. The Equipment Survey/Engineering studies obtain gross savings estimates by combining survey data on installed and operating equipment with engineering estimates of their unit savings. Equipment Survey/Engineering survey studies tend to focus either on individual measures or packages of measures. To date, they have not been employed to address program life-cycle issues by examining cross-cohort differences. Such an effort would require multiple surveys to track different cohort groups over the installed measure's life. In contrast, Consumption Evaluation studies tend to use pre- and post-retrofit consumption data to estimate a net savings effect. They typically are based on utility bills, although submetering is occasionally used; thus, they generally are limited to whole-building savings effects, which are associated with a package-of-measures scale of analysis. Consumption Evaluation studies also have been the approach used in the few program life-cycle studies completed to date.

A major difference between these two types of studies is the way in which persistence estimates typically are benchmarked. Equipment Survey/Engineering studies usually are benchmarked by pre-retrofit audit information. On the other hand, Consumption Evaluation studies generally use a pre-retrofit year of consumption data to benchmark savings.

Both types of studies can be used to examine program lifecycle performance in terms of persistence within and across cohorts of participants. Equipment Survey/Engineering studies are usually based on an initial audit or equipment survey and at least one follow-up survey to assess persistence. Although Consumption Evaluation studies covering multiple cohort groups from several post-retrofit years can estimate both within- and across-cohort effects, few studies have considered across-cohort effects. In some cases, different cohort groups are assumed to be the same to have a sufficient number of observations with several years of post-retrofit experience.

The issue of gross versus net persistence also is a defining feature of the two types of studies. Equipment Survey/Engineering studies often do not include any explicit net calculations to account for what individuals would have done in the absence of the program. However, these studies implicitly may make net adjustments by the way in which equipment replacement is handled. Consumption Evaluation studies typically include a non-participant group, which may be either a matched "control" group or a group of later participants.

This section focuses on the two basic data analysis issues of estimating program effects and projecting program savings. The estimation issue is the problem of determining the mean and

variance of the parameters denoting program savings. The projection issue is the problem of how those parameter estimates are used to determine overall or future program savings.¹ The section is divided into two parts: the first deals with Equipment Survey/Engineering studies and the second with Consumption Evaluation studies.

4.1 EQUIPMENT SURVEY/ENGINEERING STUDIES

Studies based on Equipment Survey/Engineering estimates can be an economical alternative to those based on consumption histories, depending on the level of survey and analysis effort. At a minimum, a telephone survey of "representative" cases with a simple tabulation of results is likely to be relatively inexpensive. On the other hand, an in-depth site survey covering a large number of participants is likely to be one of the most expensive options. The resulting saving estimates are, of course, highly dependent on the underlying assumptions of a measure's life and unit savings, which in turn depend on assumptions regarding usage and the performance of alternative measures.

Algorithms used to provide engineering estimates span a range of complexity. Some account for interactions across end uses, while others do not. Some address time-of-use issues and can estimate demand reductions, while others cannot. Engineering estimates also may be calibrated to known measurements from billing or metering data, using statistically adjusted engineering methods. For example, in auditing window film retrofits for the New England Power Service Company, the HEC Inc. used a building energy simulation program, "TRACE 600" by the Trane company (Fleming Group 1994; HEC Inc. 1993 and 1994). This model used field-verified data instead of assumed benchmark estimates, which were used in the initial energy audit. Although Equipment Survey/Engineering studies may use statistical methods, they are distinguished by their use of engineering assumptions as inputs.

The LILCO commercial/industrial persistence study provides a good example of this type of study (Parker 1993; Velcenbach and Parker 1993). The portion of this study that presented findings in terms of energy savings may be categorized as using an engineering approach calibrated to audit information obtained through on-site inspections. The same measures of carryforward were used to provide estimates in terms of both energy and unit counts. On-site calculations were conducted by auditors completing the persistence data collection forms following guidelines described in Chapter 3. This information was used to produce percentage estimates of carryforward which then were used to project cohort-year program savings.

The Jacobson *et al.* (1993) study also relied on on-site inspections to determine measure persistence in the commercial and industrial sector. Similar to the LILCO study, these authors used both engineering calculations and measure counts to estimate energy savings.

¹ For example, due to data attrition, a subset of program participants may be examined to estimate average household savings. In this case, the estimation problem is a single parameter one. The program savings then may be projected by applying the average household savings for the estimation subset to the entire population of program participants. The

The literature reviewed in this report may under-represent the amount of work utilities are doing to assess the persistence of energy efficiency measures. For example, Niagara Mohawk Power Corporation's (1994) *Annual Evaluation Report: 1993 Demand-Side Management Program* refers to a persistence study of a few energy efficiency measures. Neither these kinds of documents nor filings or hearings before Public Utility Commissions were reviewed for this report, however.

4.1.1 Assessing a Measure's Physical Persistence

Perhaps the most comprehensive measures of physical persistence were developed in the LILCO study (see sidebar). Its four measures take into account a variety of factors including the normal replacement of high-efficiency equipment for similar types of equipment and the premature replacement of high-efficiency equipment for less efficient equipment. It also differentiates between high-efficiency equipment installed with or without a rebate.

Persistence Statistics Used in the LILCO Study

$$\text{Technical Persistence of Measures' Operational Life} = 1 - \left[\frac{\text{Premature Replacement}}{\text{Number Inspected}} \right]$$

$$\begin{aligned} \text{Persistence of Efficient Technology} = 1 - & \left[\frac{\text{Additional HE Equip.}}{\text{Number Inspected}} \right] - \\ & \left[\frac{(\text{Premature Replacement}) + (\text{Normal Replacement by Non - HE Equip.})}{\text{Number Inspected}} \right] \end{aligned}$$

Where:

Number Inspected = Program measures surveyed less those that could not be identified or inspected or excluded for other survey reasons.

Premature Replacement = High-efficiency equipment prematurely replaced by less efficient equipment, which reduces the persistence of energy savings by reducing operational life.

Normal Replacement by Non-HE Equipment = Number of measures replaced after normal operating life, but not with high-efficiency (HE) equipment.

Additional HE Equipment = Additional high-efficiency equipment obtained without a rebate.

Two units of measure are used, percentages and kWh.

Source: Parker (1993)

In the final analysis, the LILCO study only used the carryforward equation to project program savings. By definition, carryforward cannot exceed 100%. The definition also means that

variance estimate may be used to test whether program savings are significant for a given confidence level and may also be used to determine overall program confidence levels.

the utility takes credit for the savings associated with a particular piece of equipment even after the equipment has worn out and been replaced, as long as two conditions are met. First, the replacement equipment must be of the same type. Second, the utility does not take credit for additional high-efficiency equipment obtained without a rebate (i.e., resulting from market progression). Thus, the resulting savings estimate must be interpreted as an estimate of gross savings.

Carryforward Statistics Used in the LILCO Study*

$$\text{Carryforward} = \left[\frac{(\text{Installed \& Operating}) + (\text{Normal Replacement With No Rebate})}{\text{Number Inspected}} \right]$$

$$\text{Carryforward Including Additional Equipment} = \text{Carryforward} + \chi \left[\frac{\text{Additional HE Equip.}}{\text{Number Inspected}} \right]$$

Where:

Number Inspected = Program measures surveyed less those that could not be identified or inspected or excluded for other survey reasons.

Installed & Operating = Number of measures still installed and operating.

Normal Replacement With No Rebate = Number of measures replaced after normal operating life with high-efficiency equipment, without a rebate.

Additional HE Equipment = Additional high-efficiency equipment obtained without a rebate.

χ = Proportion of additional high-efficiency measures obtained without a rebate which survey indicates is due to the program.

Two units of measure are used, percentages and kWh.

* Carryforward is an adjustment to cumulative savings that consists of estimated annual energy savings after one year of a measure's life (Parker 1993).

Source: Parker (1993)

The Jacobson *et al.* (1993) study used energy audit data on installed measures to form an engineering estimate of "actual savings." These savings estimates then were compared to the initial pre-retrofit engineering estimates with the resulting percentage change denoted as an estimate of net persistence. No attempt was made to verify pre-retrofit conditions, confirm retrofit wattages, or evaluate program retrofit decisions. This study identified both instances of retrofit savings being greater than expected as well as less than expected. The net findings in some cases are greater than 100%.

4.1.2 Engineering Estimates

The LILCO study used estimates of expected operating lifetimes to project total energy savings. These estimates were based on manufacturer information, studies, and expert opinion, and were used in equations that drew from information gathered on-site.

The post-retrofit engineering savings estimates of the Jacobson *et al.* (1993) study were based on observed equipment counts as well as revised engineering estimates of the associated savings. For measures such as lighting, savings were determined by equipment counts for which few changes in engineering estimates of savings were made. In contrast, for measures such as window film, there was substantial revision of the associated engineering estimated savings with almost all measures still in place. Inputs to the engineering estimates like location, cooling practices, and released shading coefficient differed from those used in the pre-retrofit audit.

4.1.3 Savings Projections

The LILCO study estimated persistence rates for individual measures for each cohort group. However, these individual measure-estimates of persistence were aggregated to a class of measures to calculate annual gross savings by program and cohort group. For example, a category of efficient lighting was used to calculate program savings for individual cohort years. Table 4.1 presents the persistence estimates (i.e., the estimates of "carryforward") for individual lighting measures and the overall category average used to calculate program savings. The carryforward percentages range from a low of 45% for 60-watt lamps installed in 1989, to 100% for several measures in each of the three years of analysis. The "Carryforward Including Additional Equipment," which is not reported in this table, exceeded 100% for four of the measures (including high and low-wattage high-intensity discharge lamps, reflectors, and high-efficiency ballasts) in at least one of the three years, but these "market transformation" effects were not part of the final savings estimates. Table 4.1 also shows the expected life for these lighting measures, which varies from two to twelve years. The Jacobson *et al.* (1993) study used a stratified sampling approach in surveying sites to ensure a representative sample. Average "proposed" savings (pre-retrofit audit estimates) and average "actual" savings at the time of the site audit for each stratum then were used to calculate population-weighted persistence based on both the results for each of the n sites and for each cohort year of the program.

$$\text{Population Weighted Persistence} = \frac{\sum_{i=1}^n (\# \text{ of Sites})_i (\text{Average Actual Savings})_i}{\sum_{i=1}^n (\# \text{ of Sites})_i (\text{Average Proposed Savings})_i}$$

The overall population of retrofit sites is assumed to have the same mix of measures as the survey sample. Since the proposed mix of measures is known for this population of sites, it would have been more representative to have examined the ratio of actual to proposed savings for individual

measures. "Actual" program savings then could be projected by multiplying the number of population measures by the measure-specific ratio of actual-to-proposed.

For both of the above studies, the resulting persistence estimates for different cohort groups are based on different numbers of post-retrofit years. It therefore is not possible to determine any within-cohort or across-cohort persistence of savings without a second survey.

Table 4.1 LILCO's Persistence Study Results For Selected Lighting Measures

	Expected Life (in years)	kWh Savings and Percent Carryforward		
		1988 Cohort	1989 Cohort	1990 Cohort
High Efficiency Ballasts	4	1,969,981 97%	2,298,686 95%	1,977,902 99%
Fluorescent Current Limiters	12.5	736,415 99%	1,798,473 100%	1,318,178 93%
Fluorescent Fixtures	10	2,942,916 100%	30,162,373 99%	6,837,883 100%
High Intensity Discharge $\leq$ 200 Watts	2	190,787 91%	1,244,165 96%	21,283 100%
High Intensity Discharge $\geq$ 200 Watts	4	2,942,916 100%	6,369,245 98%	812,651 98%
Reflectors	12.5	1,251,276 100%	15,539,819 100%	9,860,480 100%
Compact Fluorescent Lamps	2	899,977 56%	2,073,328 85%	1,543,749 93%
34-watt, 4' High Efficiency Lamps	4	2,023,344 63%	3,255,790 74%	1,444,247 72%
60-watt, 8' High Efficiency Lamps	2.4	684,926 65%	325,998 45%	640,499 73%
Totals: Annual kWh Savings Percent Carryforward		14,349,211 86%	65,191,380 96%	32,925,710 96%

Source: Parker 1993

4.1.4 Summary

Persistence estimates from the two Engineering studies discussed in this section are gross estimates. In both cases, within- and across-cohort persistence effects are not addressed, since each study is based on a single survey. In addition, even though these studies focus on individual measures, the persistence estimates used in calculating savings are based on aggregated categories of measures. As a result, it is difficult to assess lifetime program performances from these estimates of persistence.

4.2 CONSUMPTION EVALUATION STUDIES

Almost all of the persistence studies that use measured consumption data from billing analysis or metering, are based on pre- and post-consumption measurements. A central feature of the variety of data analysis methodologies that have been applied is weather normalization. Since weather is such an important factor, some studies have focused exclusively on this one factor. Typically, the Princeton Scorekeeping Method (PRISM) (e.g., Hirst, White, and Goeltz 1985; Keating 1991; Narum, Pigg, and Schlegel 1992; White and Brown 1990) is used to estimate normalized annual consumption (NAC) for a building's pre- and post-consumption years. A notable exception is the White, Stovall, and Tonn (1992) study, which is based on a comparison of matched weather days. Weather-normalization techniques, described in the evaluation literature, are not presented in detail in this review.

In modeling consumption changes at the building level, the studies in this class exhibit a range of model structure. Both the White, Stovall, and Tonn (1992) and the Brown and White (1990) studies assume that consumption changes are primarily a function of weather and program effects. Once consumption is weather-adjusted, the analysis essentially is univariate. In contrast, the Narum, Pigg, and Schlegel (1992) study assumes that additional factors, such as year-specific effects, also are determinants of consumption changes. The resulting multivariate model is estimated with ordinary least squares, as well as with more sophisticated methods.

4.2.1 Univariate Analysis of Weather-Adjusted Consumption

The studies discussed in this subsection are based on simple comparisons of weather-adjusted consumption; other factors, such as changes in the price of electricity or the local economy, were not considered explicitly. However, the studies that include a control group implicitly are accounting for these other factors (assuming that the control group reasonably represents the treatment group).

White, Stovall, and Tonn (1992) examined 28 winter days for each year of a four-year study period, which included one pre-retrofit year and three post-retrofit years. Similar winter days for each year were selected for direct comparison to weather normalize the load data. The matched days that were selected covered a typical range of conditions, from extremely cold to mild winter days, with similar humidity and barometric pressure. Thus, this study examined 112 days of a possible 1,460. All changes in loads among years are attributed to the retrofit program.

White, Stovall, and Tonn (1992) developed their load profiles by averaging hourly data across households. Load savings and cumulative load savings were calculated as:

$$\text{load savings} = \text{load}_{(\text{base year})} - \text{load}_{(\text{next succeeding year})}$$

$$\text{cumulative load savings} = \text{load}_{(\text{base, or pre-program year})} - \text{load}_{(\text{current year})}$$

The final results of this study show a high degree of variability. The whole-house electricity load on weekdays for all households at the peak hour of 8:00 AM decreased 15% in the first post-retrofit year, decreased another 11% in the second, and increased 13% in third year. Space heating electricity loads demonstrated a similar pattern with decreases of 30% and another 22% in the first and second years, respectively, and a 27% increase over the second-year figures in the third year. Similar results were obtained for weekends, with loads increasing between the second and third year. Water heating electricity loads also were erratic, with a 3% decline the first year, a 3% increase the second year, and a 10% increase in the third year.

These results are difficult to explain solely as a product of the retrofit program. In general, we expect maximum retrofit savings to occur upon installation and that these savings would decline over time. If year two results are ignored, one might conclude gross persistence is roughly 100%. Alternatively, if full retrofit benefits were not achieved until year two, either because all retrofits were not completed before the first retrofit year or some other reason, then one could conclude that persistence is considerably less than 100%. The large variations of the second and third year indicate the need for additional study, such as an analysis of a larger proportion of the data or considering other time-dependent factors such as electricity price.

The most common univariate approach is to weather-normalize annual consumption (NAC) with a PRISM-type model and compute mean and median changes in pre- vs. post-NAC. A whole series of studies of this type have been conducted to monitor program performance in the ten-year history of the Bonneville Power Administration's residential weatherization program, 1980-1989. These studies all employ a heating-only version of the PRISM model to estimate household NAC, which is computed as:

$$\text{NAC} = 365a + bH_o(t)$$

where,

365a = the fixed amount of base load electricity consumed by a household in one year,

$H_o(t)$ = the heating-degree days (base t) in a typical year, so that

$bH_o(t)$ = the proportional amount of heating fuel relative to the outdoor temperature, adjusted for long-term outdoor temperatures.

The three parameters of the model, a, b, t, usually are estimated with twelve or so monthly bills. The change in household NAC from the pre-retrofit year to one or more post-retrofit years, DNAC, is then the focus of investigation.

The early Bonneville studies of this type (Hirst, White, and Goeltz 1984; Hirst, White, and Goeltz 1985; and Goeltz, Hirst, and Trumble 1986) compute the average DNAC for both a

participant group and a non-participant group. Net savings simply are the difference in average change between the two groups.

$$\text{Average Net Savings per Household} = \text{Avg}[\text{DNAC(P)}] - \text{Avg}[\text{DNAC(C)}]$$

where P and C denote the participant and control groups, respectively.

Using this approach to examine a 1981 cohort group, Hirst, White, and Goeltz (1984) concluded that net savings increased by 10% (3,370 and 3,750 kWh/year) between the first and second post-retrofit years. However, with nine fewer households for the same cohort group over the same period, Hirst, White, and Goeltz (1985) concluded that net savings was unchanged at 3,800 kWh/year. Additionally, second to third year post-retrofit net savings were concluded to have decreased by 11%. Goeltz, Hirst, and Trumble (1986) examined both a 1982 and 1983 cohort group using the same approach. The results indicate that the 1982 cohort group had net savings decline of 20% (4,100 to 3,300 kWh/year) and 24% (3,300 to 2,500 kWh/year) over a three year post-retrofit period for each successive year. The 1983 cohort group showed a similar decline of 15% over the first two post-retrofit years. Because savings variances estimates for this method were not reported in any of these studies, it is not possible to determine whether there is a statistically significant change in net savings, that is, whether persistence is 100%.

The later Bonneville studies (Schweitzer, Brown, and White 1989; White and Brown 1990; and Brown and White 1992) employ utility weights in calculating overall net savings. This weighting approach is used to adjust for differences in participation rates across utilities, which vary from 80% to 1%. At the utility level, the same procedure as described above is employed to calculate utility net savings.

Utility-level results reported by these studies (see Table 4.2) show a wide range of within cohort net savings persistence estimates. The results for the 1986 cohort group in this table provide a comparison of the overlapping studies of Schweitzer, Brown, and White (1989) and White and Brown (1990). The latter study used slightly fewer observations than the former, due to data attrition, to obtain third-year post-retrofit savings estimates. For example, in the case of Tacoma, this difference in sample size reduced the 21% increase in net savings for post-retrofit year two to only a 3% increase. Overall, this table indicates the high degree of variability in persistence estimates and highlights the sensitivity of this analytical approach to small changes in sample size. Persistence estimates over a two-year period vary from a high of 21% to a low of -66%.

Schweitzer, Brown, and White (1989) conclude that net cohort savings decreased from 11.8% to 10.6% over the first two post-retrofit years. Similarly, White and Brown conclude that net savings for the same cohort group over the same time period decreased from 12.6% to 8.7%, with third-year net savings increasing slightly to 8.9%. Brown and White (1992) conclude that total net cohort savings for a later group decreased from 9.3% in the first year to 8.5% in the second. None of these studies consider whether this within-cohort decrease is statistically significant, nor do they present variance estimates for net savings. However, the Brown and White (1992) study does report the

standard error of gross savings, which suggests that the difference between 9.3% and 8.5% is not significant statistically.²

Table 4.2 Average Net Savings By Utility

	# of Partici- pants	# of Non- Partic.	Post Retrofit Savings				
			Year 1 (kWh)	Year 2 (kWh)	Percent Change Year 1 & 2	Year 3 (kWh)	Percent Change Year 2 & 3
EUGENE							
'86 Cohorts (1)	46	158	3.672	2.368	-36%		
'86 Cohorts (2)	41	149	4.338	2.451	-43%	3.093	26%
'88 Cohorts (3)	97	142	2.640	2.435	-8%		
TACOMA							
'86 Cohorts (1)	50	224	4.037	4.886	21%		
'86 Cohorts (2)	44	201	2.608	2.698	3%	2.160	-20%
'88 Cohorts (3)	76	323	1.341	971	-28%		
RICHLAND							
'86 Cohorts (1)	42	116	2.105	715	-66%		
'86 Cohorts (2)	31	96	2.781	1.132	-59%	2.245	98%
'88 Cohorts (3)	50	75	2.603	3.162	21%		

Sources: (1) Schweitzer, Brown, and White (1989); (2) White and Brown (1990); and (3) Brown and White (1991).

4.2.2 Multivariate Analysis of Weather-Adjusted Consumption

Two of the first studies to apply multivariate analysis to the problem of estimating the persistence of savings were Hirst, White, and Goeltz (1984) and Kushler, Witte, and Crandall (1984). These studies employed a two-stage approach. The first stage applied the PRISM model to estimate NAC for each household over a study period, which included one pre-retrofit year and two or more post-retrofit years. The second stage of the analysis developed a cross-section/time-series model to estimate net household savings while controlling for factors such as the price of electricity, pre-participation consumption, and sample selection bias. This two-stage approach had been used earlier in several different studies to examine a single year of post-retrofit consumption.

The Hirst, White, and Goeltz (1984) study examined two years of post-retrofit consumption. After first applying the PRISM model to estimate NAC, the resulting cross-section/time-series data set of 1,200 observations, collected over three years for 237 participants and 163 non-participants, was used to estimate a one-way random effects model. This model can be expressed as:

² The standard error of net savings can be approximated by assuming the covariance of participant and comparison gross savings is zero, in which case the difference in net savings, 8.5% versus 9.3% is not significant.

$$\begin{aligned} \text{NAC}_{i,t} = & C + a_1 \text{Income}_i + a_2 \text{Household Members}_i + a_3 \text{Sqft}_i + a_4 \text{Wood}_i + A_5 * \text{Electricity} \\ & \text{Price}_{i,t} + a_5 \text{Primary Electric}_i + a_6 \text{Participant}_i + a_7 \text{First Retrofit Year} + \\ & a_8 \text{Second Retrofit Year} + a_9 \text{Sample Selection Correction} + u_i + e_{i,t} \end{aligned}$$

where u_i and $e_{i,t}$ are independently distributed with zero means and positive variances σ_u^2 and σ_e^2 . This error specification allows for the within-household correlation of the model residuals.

The significantly positive estimate reported for coefficient a_6 indicates that higher consumption households are participating in the program in the first program year relative to later years. Many subsequent studies report similar findings. Savings persistence is reflected in the differences between retrofit coefficients, a_7 and a_8 . The estimation results indicate that net savings increased by roughly 2%-3% over a two year post-retrofit period, but it is not clear whether this difference is significant statistically.

A similar evaluation of the Michigan Residential Conservation Service program (Kushler, Witte, and Crandall 1984) also collected and analyzed two years of post-audit energy consumption for gas-heated homes. The results showed a slight increase in net annual savings from the first to the second post-audit year of about 6%.

The statistically significant sample selection term in the above Hirst, White, and Goeltz (1984) model indicates that sample selection bias is a problem. The auxiliary model upon which this correction term is based indicates that households with higher heating operating and capital costs were more likely to participate in the program for that cohort year. Correcting for sample selection yields roughly 6% higher estimates of net savings for both retrofit years. This sample selection bias correction procedure was not used in the Hirst, White, and Goeltz (1985) follow-up study of the same 1981 cohort group, which included a third post-retrofit year.

The above multivariate approach was improved in the study by Goeltz, Hirst, and Trumble (1986). Through a series of diagnostic tests, both the model and error specification were expanded and refined. The resulting model specification was expanded to include interaction and non-linear factors. Significant coefficients were estimated for long-run heating degree days (based on PRISM reference temperature estimates), interactions with electricity price, building square feet, and wood use. Squared terms for building square feet, wood use, and long-run heating degree days all were found to be significant statistically. The variance specification also was expanded to include factors such as income, building square feet, wood use, as well as the estimated expected NAC. The estimated net savings results showed some variation among first-, second-, and third-year savings. However, these differences were not significant statistically. This study also included estimated models of NAC changes yielding similar coefficient estimates. The conclusion was that net household savings were reasonably constant over the three-year post-retrofit evaluation period.

A comparison of these multivariate result with the NAC results described in the previous section on univariate approaches indicates that generally the estimates from the two approaches are fairly close. In almost all cases, the difference in estimated net savings is less than 6%. The one

exception is Goeltz, Hirst, and Trumble (1986) where the two approaches differ by 25% for the 1984/85 heating season. In this case, the NAC estimates indicated a 21% decline in net savings, whereas the multivariate approach indicated a 9% increase in net savings. One possible reason for this difference is the delayed effect of sharp electricity price increases over the period between 1981 and 1984.

The White and Brown (1990) study also used ordinary least squares to estimate NAC changes as a function of pre-retrofit NAC, total weatherization costs, long-run average heating degree days, and the location of the household (east or west of the Cascade Mountains), as well as a number of specific weatherization measures. More detailed information that was used in prior studies was not available for this study. This modeling effort did not account for heteroskedastic (i.e., non-constant household error variances) model error, yielding large variance estimates for the model parameters. The results were therefore of limited value for hypothesis testing. Since these estimates explained very little variance in gross energy savings (less than 10%), they were not reported.

Similarly, the Narum, Pigg, and Schlegel (1992) study of the persistence of savings from weatherizing low-income Wisconsin residences performed a multivariate analysis of NAC estimates. Like the previously discussed studies, Narum, Pigg, and Schlegel depended on billing data and employed their version of a PRISM model. In contrast to these previous studies, the authors focused on low-income residences (rather than non-low-income residences) that could include multifamily buildings (instead of only single-family residences). Since buildings were the unit of analysis for this study, household consumption was not clearly represented. Another difference between the Wisconsin and Bonneville studies is that the former investigated gas energy consumption and the latter looked at electric energy consumption.

The model specification employed by Narum, Pigg, and Schlegel (1992) was a two-way additive model that distinguished two kinds of effects, those gas consumption changes that affect all buildings (whether weatherized or not) and those changes that result from weatherization. The following equation was used to estimate year-to-year changes in energy use for the duration of the program for each of the two utilities:

$$DNAC_{i,t} = a_0 TREND_t + a_1 PRE_{i,t} + a_2 WEATHERIZATION_{i,t} + a_3 POST I + a_4 POST II_{i,t} + a_5 POST III_{i,t} + a_6 Y86:87_t + a_7 Y87:88_t + a_8 Y88:89_t + a_9 Y90:91_t + e_t$$

Where:

TREND represents change in NAC that is common to all buildings in all years

PRE represents change in NAC that occurred among unweatherized buildings

WEATHERIZATION represents change in NAC going into and coming out of the year of weatherization

POST I represents change in NAC between the first and third years after weatherization

POST II represents change in NAC between the third and fifth years after weatherization

POST III represents change in NAC between the fifth and eighth years after weatherization

Y86-87 represents change in NAC common to all buildings from 1986-1987

Y87-88 represents change in NAC common to all buildings from 1987-1988

Y88-89 represents change in NAC common to all buildings from 1988-1989

Y90-91 represents change in NAC common to all buildings from 1990-1991.

This specification is fairly restrictive in that energy-savings effects are assumed to be a constant for all buildings, both single- and multi-unit, and for a mixture of shell and heating measures that change over time, with heating system replacements becoming increasingly common. Results for building subsets based on number of units, ownership, and type of energy measure appear to yield significantly different estimates. A model explaining overall program results should incorporate these known factors.

This model does not estimate any cohort-specific effects. The primary reason for this absence is that late participants define the benchmark by which net savings are estimated. The coefficients of the POST variables are key parameters modeling net savings persistence and are not specific to a particular cohort group. If these coefficients are not significantly different from zero then the hypothesis of 100% persistence cannot be rejected for all cohort groups.

Two different estimation methods were used to estimate the coefficient of this model. Standard ordinary least squares produced mean effects while an ANOVA variant, the median polish, produced median effects. Since median data are less influenced by extreme data values than means, the authors relied primarily on the median polish technique in deriving their findings. However, coefficients of interest for **persistence** effects generally were insignificant statistically. Variance estimates from first-stage PRISM estimates could have been used to obtain more efficient estimates with a generalized least squares estimation method. The resulting variance could alter the conclusion regarding insignificant persistence effects. This estimation method in effect would down-weight high-variance households, which was a primary reason for using the median polish method.

An example of the estimation results for this study are presented in Table 4.3. Model results for the overall Wisconsin Gas program are in the upper portion of this table with result for shell and heating measures provided separately below. The initial savings due to weatherization are summarized by the column denoted Wx. A comparison of this effect for the individual and combined types of measures indicates that there are significant differences, with both types of measures yielding the highest savings, followed by heating measures and then shell measures. The POST coefficients are labeled 1, 2, and 3 in this table. The only significant POST coefficient in this table is that labeled 3 for heating measures with a median value 75.9, indicating a decrease in net savings.

Table 4.3 Summary of Wisconsin Gas Company Median Polish Results

		Weatherization Effects		Post-Retrofit Periods		
		Pre	Wx	1-3 Years	3-5 Years	5-8 Years
PROGRAM						
Model Results		15.9	-155.6	-4.6	0.0	6.9
Therms	Mean	16.1	-155.5	-4.5	0.4	6.1
	St. Dev.	5.2	6.8	3.8	2.6	5.5
	Low CI	7.4	-166.8	-10.9	-4.3	-2.1
	Hi CI	24.4	-144.5	1.8	4.3	15.9
	Count	3,690	4,040	3,382	1,844	760
SHELL MEASURES						
Model Results		3.9	-121.4	-4.4	0.0	2.9
Therms	Mean	8.2	-120.1	-2.5	1.6	3.8
	St. Dev.	7.4	7.0	3.5	3.9	5.6
	Low CI	-8.3	-133.0	-10.2	-6.4	-6.3
	Hi CI	16.1	-109.9	1.4	6.4	12.1
	Count	1,354	1,853	1,986	1,333	646
HEATING MEASURES						
Model Results		18.6	-142.9	0.0	-4.6	75.9
Therms	Mean	12.8	-146.2	-1.5	-6.1	66.3
	St. Dev.	12.1	19.4	10.3	12.0	43.0
	Low CI	-1.4	-174.8	-16.9	-24.3	5.1
	Hi CI	38.5	-110.9	16.9	15.1	146.6
	Count	401	334	292	124	14
BOTH MEASURES						
Model Results		23.4	-294.7	-14.2	0.0	19.9
Therms	Mean	25.8	-293.4	-11.1	0.4	24.5
	St. Dev.	13.9	17.5	10.0	5.3	18.1
	Low CI	0.4	-323.5	-30.7	-8.7	-9.9
	Hi CI	46.3	-265.9	2.2	8.7	49.7
	Count	1,040	1,356	1,038	384	99

Source: Narum, Pigg, and Schlegel (1992)

Overall, the estimation result indicates that net savings increased over the eight-year post-retrofit period by as much as 30% for the Wisconsin Gas Company and 62% for Madison Gas & Electric. However, these results also are not significant statistically.

4.2.3 Projecting Savings Estimates

The results of these energy consumption studies are used a variety of different ways to extrapolate to a larger scale. White, Stovall, and Tonn (1992), for instance, assumed that their Hood River sample was representative of the Pacific Power service area. They then determined kW impacts

by multiplying the kW per house for their sample by the "number of customers with permanently installed electric space heating equipment" (p. 7).

On the other hand, White and Brown (1990) weighted the gross utility savings estimates of both participant and control groups in calculating net cohort program savings. Participant and control weights were used to reflect the proportionate representation for each utility cohort group relative to the total population for all the utilities of the program. The utility weights for participants WP_i were the number of households retrofitted by utility (i) relative to the number households weatherized by all program utilities for a particular cohort year. The utility weights for control group households, WC_i , were the number of eligible households for a particular utility relative to the number of eligible households for all program utilities. Net program cohort savings therefore was expressed as:

$$\text{Net program cohort savings} = \sum_{i=1}^n \{WP_i * \text{Avg}(i)[\text{DNAC}(P)] - WC_i * \text{Avg}(i)[\text{DNAC}(C)]\}$$

For example, when Brown and White (1992) examined first-year savings, a total of 356 households had sufficient data to be considered. Of this total, 97 participants (28%) were from one utility for which a total of 3,971 households were retrofitted. Since the total number of households retrofitted by all program utilities was 4,898, this particular utility had a participant weighting factor of 81%. Thus, less than a third of the participant observations were given a majority of weight in computing net program savings. The utility control weight for the same utility was only 11%, with the largest weight of 56% allocated to another utility. The program net savings estimate therefore was determined largely by two utilities, the participant savings of Eugene, Oregon and the control group savings of Seattle, Washington. This net estimate raises the question of whether Seattle households adequately represented Eugene households.

The early Bonneville studies (Hirst, White, and Goeltz 1984 and 1985; and Goeltz, Hirst, and Trumble 1986) focused on unweighted net household savings. Net within-cohort savings for these studies were found to increase, remain constant, or decrease over the first two retrofit years. However, it appears that all these change in net savings are not statistically significant. Similarly, the later Bonneville studies, which all used a weighted program net savings estimate, uniformly concluded that net savings decreases over a two year post-retrofit period. Only the last study in this series provided any estimate of savings variance, from which it appears that net savings have not changed significantly.

Brown and White (1992) also provided a summary of across-cohort performance for the Bonneville residential program (Table 4.4). Since the program itself changed over time in terms of the measures provided and in terms of how it was administered a simple net kWh/year would not be very informative. Thus, a measure of the cost of net savings is presented in Table 4.4 for different cohort groups for the first post-retrofit year. This table indicates that the cost of net savings increased, particularly for the 1989 cohort group. A statistical test of this observation would of

course be desirable to conclude whether across-cohort persistence is significantly different from 100%.

Table 4.4 Historic Costs And Net Savings Of Bonneville's Weatherization Programs

Program Year	First Year Net Savings	Regional Levelized Costs (1989-\$)		Bonneville Levelized Costs (1989-\$)	
		Per Unit	Mills/kWh	Per Unit	Mills/kWh
Pilot (1980-82)	3,840	3,220	40	(Loan Prg.)	<40
Interim (1982-83)	4,200	2,540	29	1,900	21
RWP 1985	2,610	2,580	47	1,780	32
RWP 1986	3,060	3,130	48	1,750	27
RWP 1988	2,180	2,310	50	1,530	33
RWP 1989	1,330	2,320	83	1,470	52

Source: Brown and White (1992)

4.2.4 Summary

Both the univariate and multivariate approaches strongly depend on a control group to estimate net savings. To the extent that the multivariate model structure "explains" cross sectional variation in consumption, the net savings benchmark is adjusted for differences between treatment and control group factors of consumption. However, univariate approaches, such as NAC estimate tabulations, depend exclusively on having a matching control group in estimating net savings. As discussed earlier, the NAC results appear to be highly sensitive to small changes in sample size. Such control group shortcomings should not matter as much over short evaluation periods as over longer time periods, or, more generally, during periods when key factors like electricity prices are relatively stable.

A related advantage of the multivariate approach is the ability to test hypotheses regarding the persistence of net savings, which can lead to improvements in program design. Within the framework of generalized least squares, a wide variety of hypothesis tests may be conducted relatively easily. The NAC approach which yields estimates of the variance of net savings or weighted net savings estimates typically are used. Hypothesis tests regarding changes in net savings therefore are not easily performed.

The main drawback of the multivariate approach is the cost of acquiring household information beyond the initial audit information and the monthly billing data. The early Bonneville studies used one or more surveys to collect information such as number of household members and wood consumption. However, new multivariate methods recently have been applied to the problem

of estimating net savings using only initial audit and monthly billing information (Trumble and MacDonald 1994). These methods easily could be extended to the problem of estimating the persistence of net savings.

The two different types of studies, Equipment Survey/Engineering and Consumption Analysis, are directed at two different notions of persistence, primarily because of their differences in how savings and persistence are benchmarked. The Equipment Survey/Engineering studies, reviewed above, provide both an assessment of whether proposed savings are being realized and useful information about possible problems. Repeated surveys of this type would yield persistence estimates more similar to those obtained by Consumption-based studies, which are benchmarked by first year net savings estimates. In this case, both types of studies could examine both within- and across-cohort effects.

Due to differences in definition, conclusions about persistence from the two Equipment Survey/Engineering studies reviewed here are difficult to obtain. The measure used by LILCO in computing program savings by definition cannot exceed 100%.³ In contrast, the measure employed by NEPSCO allows actual "savings" to exceed proposed savings. Reported category results in several cases are greater than 100%.

The Consumption-based studies generally indicate that net cohort savings are either roughly constant or slightly decrease over the first two to three years post-retrofit. The multivariate results reviewed here do not indicate any statistically significant changes in net cohort savings. However, the more prevalent NAC results do suggest some decline in net savings. Unfortunately, it is not clear whether these within-cohort changes are statistically significant. Regarding across-cohort persistence, only the Brown and White (1992) study really addressed this issue. For the Bonneville Residential Weatherization Program both net kWh measures as well as cost-adjusted measures indicate a decline in net savings across cohort groups. Whether this decline is statistically significant and what its underlying causes are is not clear.

³ Results for the "technical persistence of operational life" cannot exceed 100%; the "persistence of efficient technology" does not have this constraint.

5. SUMMARY OF FINDINGS

Just as the goals, research designs, and methods of persistence studies have varied, so have their results. This chapter briefly summarizes the results of persistence studies, focusing on the four example studies that have been discussed throughout this literature review. Because the review did not focus either on the customers' perspective or on methods for improving persistence, no such findings are presented. The review is organized by scale of analysis, focusing first on measure-scale studies then on cohort-scale studies, and finally on program-scale studies. As noted earlier, persistence studies to date have been conducted primarily at the first two scales of analysis.

Overall, the results of persistence studies at all three scales are inconclusive and sometimes contradictory. Measure-scale persistence of savings for different measures and sectors shows considerable variability. Further, while *ex-post* estimates of year-to-year savings associated with measures generally are less than *ex-ante* engineering estimates, the opposite has resulted when stimulated energy savings (i.e., market transformation) is considered. Results of cohort-scale studies, typically focused on the residential sector, also vary. Both declines in savings and net increases in savings have been reported within cohorts of participants evaluated over time. Another reason why cohort-scale studies are inconclusive is that it is impossible to determine the statistical significance of observed differences because confidence intervals and levels of precision frequently are not reported. Because only one program-scale analysis was identified, it is difficult to generalize that result to the set of program lifecycle impacts.

Virtually all of the persistence studies so far conducted have been retrospective in design. The research conducted by Jeppesen and King (1993) points to some of the limitations of retrospective analyses (see example case, below).

Ex-ante vs. ex-post findings. Studies of the persistence of energy conservation measures generally have found that the savings associated with measures, when calculated on the basis of inspections or consumption data, differed from *ex ante* engineering estimates. Typically, *ex ante* engineering estimates overstate actual savings, for many of the reasons discussed earlier, including their failure to account adequately for human interventions such as premature measure removal and building renovations or remodeling.

Variation in explanations of findings — measure scale. Studies to date also indicate significant variability in the explanations for persistence of savings associated with different measures and different sectors. The LILCO study (Parker 1993) found, for instance, that commercial and industrial lighting measures with the shortest anticipated operating life were among the least persistent measures installed through their programs because they did not remain in place and in operation as long as other high-efficiency measures. However, it found a relatively small premature removal rate for lighting measures in these commercial and industrial sectors. In contrast, Vine (1992) and others have concluded that residential lighting measures (in particular compact fluorescent bulbs) are

among the least persistent measures (along with low-flow showerheads and door weatherstripping), largely due to high premature removal rates.

Example case: California Conservation Inventory Group effective measure life study (Jeppesen and King 1993)

Selected Findings

- “no studies were identified that actually tracked measure performance decline over time in customer installations” (p. II-7)
- most efforts to gauge commercial effective measure life have focused on difference among commercial sectors rather than on the measures themselves; few studies have targeted program participants and energy efficiency measures directly
- “energy savings estimates from billing analysis cannot be interpreted to reveal the influences of measure life, behavior, or technological degradation of savings” (p. II-8)
- the retrospective analysis feasibility assessment indicated that:
 - there were wide variations according to measure in both data identification and sampling feasibility;
 - data on commercial measures were particularly difficult to access;
 - residential records tended to be adequate but commercial records frequently were inadequate;
 - there was a much higher completion rate for residential than for commercial telephone surveys; and
 - only a limited number of measures are appropriate for retrospective analysis, due largely to the adequacy and reliability of data over time.

Despite these variations, general estimates of levels of measure persistence have been offered. For instance, Braithwait *et al.* (1994) concluded that persistence rates range from 92 to 100% for commercial energy conservation measures, when estimated one- to five-years after installation.

Variation in findings—cohort scale. Studies focusing on the cohort-scale persistence of energy savings also have produced varying results. Most of these studies center on residential energy these findings (see example boxes, below).

Narum, Pigg, and Schlegel (1992) found an overall increase in net energy savings within cohort groups from the first through eighth years after weatherization. However, the utilities and types of customers examined differed in certain respects. Specifically, savings persistence was greater among gas accounts for multi-unit buildings than for single-unit buildings for the Wisconsin Gas

Company; no such difference was apparent in the results for Madison Gas & Electric. The same type of variability across sectors of customers was found in the studies of measure life reported earlier.

Example case: Long Island Lighting Company (Parker 1993)

Selected Findings

- least persistent lighting measures were bulbs with the shortest expected operating life; these bulbs had the lowest kWh/unit impact and high unit turnover
- relatively small premature removal rate for all lighting measures

Summary findings for the lighting component of the Dollars & Sense Program for Non-Free Riders and Rebated Equipment

Year	Technical Persistence of Operational Life		Persistence of Efficient Technology		Carryforward* Adjustment (with added equipment)		Carryforward* Adjustment (without added equipment)	
	kWh	Percent	kWh	Percent	kWh	Percent	kWh	Percent
1988	16,461,255	99%	16,284,537	97%	14,910,635	89%	14,349,211	86%
1989	66,638,160	98%	67,501,706	100%	66,015,616	97%	65,191,380	96%
1990	33,583,824	98%	34,194,415	100%	33,458,049	98%	32,925,710	96%

* carryforward is "the estimate of annual energy savings following year one of a measure's operational life" (p.37)

The results of the White, Stovall, and Tonn (1992) study indicated an overall pattern of persistent load savings three years after installation for a cohort group. However, the load savings measured in each of three successive years fluctuated considerably (see example box). It generally is believed that within-cohort group persistence declines gradually, but such a trend cannot be confirmed statistically from this study or from Narum, Pigg, and Schlegel (1992).

Program-scale findings. The one example of an analysis of persistence across a program's lifecycle indicated that savings can decrease more precipitously across successive cohort years of participants than across the years following program participation for individual cohorts. Brown and White (1992) describe a general pattern of decline in net energy savings, both within and across cohort groups. Across cohorts, persistence appears to decrease dramatically over time (see Brown and White example case), although no statistical tests of significance are conducted to confirm such a trend. In addition, the factors influencing this decline could only be hypothesized. They include a range of factors such as self-selection bias, changes in the retrofit program and how it was administered, and changing economic conditions such as rising energy prices that have motivated

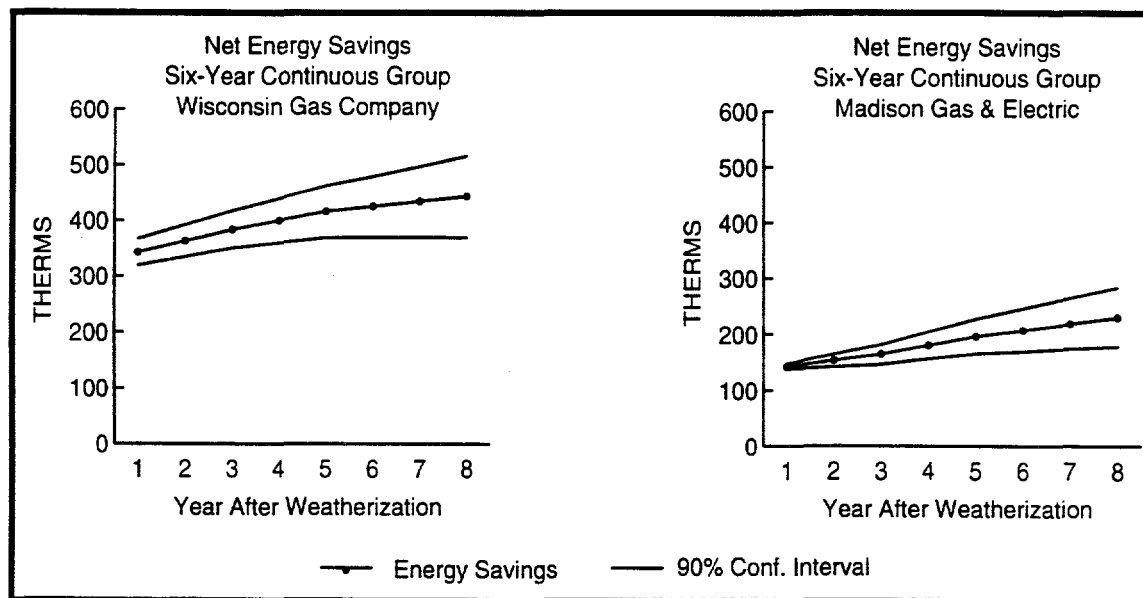
households to invest in retrofit measures outside of the program's activities. These factors have caused the program to move from an initial focus on high electricity users to more recent cohorts of participants who did not consume as much electricity prior to participation and who therefore did not offer the same high level of savings potential.

**Example case: Wisconsin's Low-Income Weatherization Assistance Program
(Narum, Pigg, and Schlegel 1992)**

Selected Findings

- 1st-year energy savings: 19% (343 ± 25 therms) for Wisconsin Gas Company; 13% (142 ± 30 therms) for Madison Gas & Electric Company
- overall, net savings increased
- persistence varied among subgroups within programs

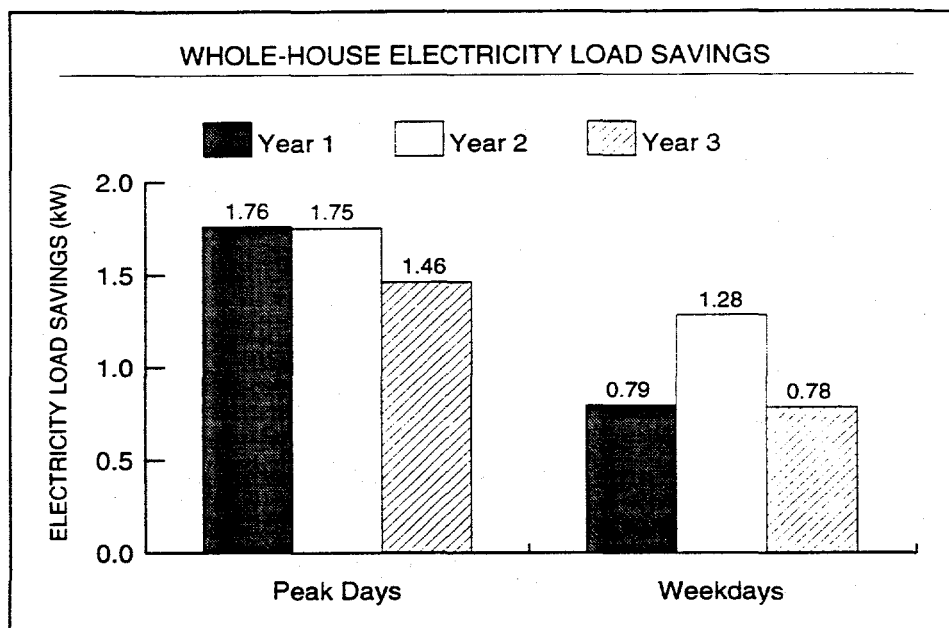
Narum, Pigg, and Schlegel (1992) performed a variety of analyses of data collected from two utilities. For Wisconsin Gas Company, they found that, although post-weatherization gas consumption decreased immediately, gas consumption generally increased over time. Nevertheless, net savings persisted (and actually increased) throughout the eight-year study period. Analyses of subgroups found the estimated magnitude of those net savings to be highly uncertain. The results of data analyses for Madison Gas & Electric presented net persistence of savings trends similar to those for Wisconsin Gas Company (see below), but the magnitude of savings differs. The authors did not explain the magnitude differences between the utilities, but the differences may have been the result of such factors as different housing stocks, different proportions of shell and heating system measures, and different occupant characteristics.



Example case: Hood River Conservation Project (HRCP)
(White, Stovall, and Tonn 1992)

Selected Findings

- load savings were persistent across the sample, one to three years after weatherization
- the analysis of peak days indicated that whole-house and space-heating electricity load each were reduced by 0.8 kW/house
- the analysis of peak days also indicated that water heating electricity load increased by 10% and baseload electricity load did not change



The electricity load of interest was the peak hourly load. Data generally "were averaged across households and then other arithmetic operations were performed to obtain sample values in loads, interior temperatures, and other group values" (p.7).

Example case: Bonneville's Residential Weatherization Program (Brown and White 1992)

Selected Findings

- net energy savings (first year) generally declined for successive cohorts of participants
- net energy savings (from the first to the third post-retrofit year) generally declined over time
- patterns of net energy savings were variable among different cohorts of participants

	Net Energy Savings (kWh/Year)		
	Year 1	Year 2*	Year 3*
Pilot (1980-82)	3,840	3,790 (1.3%)	3,410 (11.2%)
Interim (1982-83)	4,200	3,600 (14.3%)	2,500 (40.5%)
Long-Term RWP (1985)	2,610	2,565 (1.7%)	2,600 (0.4%)
Long-Term RWP (1986)	3,060	2,112 (31.0%)	2,140 (30.1%)
Long-Term RWP (1988)	2,180	2,000 (8.3%)	
Long-Term RWP (1989)	1,330		

* Percentages of Year 1 net savings are presented in parentheses.

6. DISCUSSION AND CONCLUSIONS

The ultimate purpose of this literature review is to serve as the foundation for a handbook on the measurement of persistence. Therefore, it has focused much more on the methods used for collecting and analyzing data than on estimates of persistence. The review discussed the underpinnings of persistence studies; namely the definitions of persistence and the purposes of persistence studies. Then, it described issues relevant to both the collection and analysis of data for persistence studies. Findings from persistence studies also were summarized. Throughout the review, four studies were used repeatedly as illustrations of different methodological and analytical approaches to persistence so that readers can track the data collection, data analysis, and findings elements of a set of comprehensive studies that represent alternative approaches.

In part because of the inconclusiveness of findings from persistence studies, one recommendation is to use methods that will provide a basis for statistically strong results. Along these lines, it is important to report levels of precision and variability along with the point estimates of savings over time. At the same time, costs must be considered. Increasingly, utilities are looking for relatively inexpensive and straightforward methods for evaluating their DSM programs. No single method best meets the variety of users' information needs while remaining within their budgetary constraints.

Beyond these methodological considerations, the future may present a different type of challenge to the conduct of persistence studies. We do not know how utilities will address persistence as they adapt to a competitive world. On the one hand, accurate information about persistence or how to improve persistence may enhance a utility's competitive edge. On the other hand, expensive persistence studies may be an unaffordable luxury. Competition also may result in multiple utilities cooperating to conduct persistence studies.

7. GLOSSARY

carryforward: an adjustment to cumulative savings that consists of estimated annual energy savings after year one of a measure's operational life (Parker 1993)

free-drivership: the adoption of program measures by non-participants as a consequence of the program

free-ridership: the adoption of program measures by participants who would have adopted those measures in the absence of the program

gross energy savings: total energy savings attributable to a measure or program, without regard to the actions people might have taken in the absence of the program

market transformation: a program's ability to accelerate market trends toward increasing energy efficiency, above and beyond the investments that are caused by the program's incentive payments

measure efficiency or performance: "the actual energy efficiency performance of the measure" (Jeppesen and Rudman 1993, p. 521)

measure life: "the time during which the measure is installed and is accruing energy efficiency benefits" (Jeppesen and Rudman 1993, p. 521)

metering: the collection of whole-building or end-use energy consumption data via meters (Pacific Gas & Electric *et al.* 1993)

monitoring: the use of devices to collect data related to energy consumption, such as weather conditions, duty cycles, and hours of equipment operation (Pacific Gas & Electric *et al.* 1993)

net energy savings: the savings attributable to a measure or a program, taking into account what participants might have done in the absence of the program or measure

persistence: the long-term temporal pattern of energy savings and load reductions from DSM investments

stimulated savings: savings due to installations of additional energy conservation measures by participants or non-participants as a result of a program, but without program incentives

surge effects: "the tendency for some consumers to react to the savings realized by the initial conservation measure installation by expanding their adoption to other conservation measures" (Jeppesen and King 1993, p. II-6); these behaviors are adopted by individual program participants

survival energy savings: the gross energy savings from measures installed by a program

8. REFERENCES

Baxter, Les, 1995 (January 19). Personal communication.

Braithwait, Steven D., Arthur Maniaci, J. David Glycer, and David Attenberger, 1994 (June). *Persistence and Snapback Measurement Methodologies: An ESEERCO Scoping Study*. Draft. ESEERCO Project EP 93-04. Prepared by Laurits R. Christensen Associates, Inc., Madison, Wisconsin for Empire State Electric Energy Research Corporation. New York, New York.

Brown, Marilyn A., Linda G. Berry, Richard A. Balzer, and Ellen Faby, 1993 (May). *National Impacts of the Weatherization Assistance Program in Single-Family and Small Multifamily Dwellings*. ORNL/CON-326. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

Brown, Marilyn A. and Dennis L. White, 1992 (December). *Evaluation of Bonneville's 1988 and 1989 Residential Weatherization Program: A Northwest Study of Program Dynamics*. ORNL/CON-323. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

Coates, Brian, 1992. "Persistence of Energy Savings in a Commercial Conservation Program." *Proceedings: ACEEE 1992 Summer Study on Energy Efficiency in Buildings, Panel 7, Program Evaluation*. Washington, D.C.: American Council for an Energy-Efficient Economy, pp. 7.37-7.44.

Delaware Office of the Public Advocate and Delmarva Power & Light Company, 1993 (March 17). *Settlement Agreement Before the Public Service Commission of the State of Delaware In the Matter of the Application of Delmarva Power & Light Company for Approval of the New Conservation Programs and Revisions to the Commercial and Industrial Indoor Lighting Program and the Commercial Industrial Air Conditioning Program*. DPSC Docket No. 92-73. Submitted to The Honorable G. Arthur Padmore, Hearing Examiner, by the Office of the Public Advocate, State of Delaware and the Delmarva Power & Light Company.

Eto, Joseph, Edward Vine, Leslie Shown, Richard Sonnenblick, and Christopher Payne, 1994. *The Cost and Performance of Utility Commercial Lighting Programs*. LBL-34967. [A Report from the Database on Energy Efficiency Programs (DEEP) Project.] Berkeley, California: Lawrence Berkeley Laboratory.

Fleming Group, 1994 (June). *Persistence of Commercial/Industrial Non-lighting Measures*. Appendix IV-8. Prepared for New England Power Service Company by The Fleming Group, HEC Inc., Albany, New York.

Goeltz, Richard, Eric Hirst, and David Trumble, 1986. *Electricity Savings One to Three Years After Participation in the BPA Residential Weatherization Program*. ORNL/CON-194. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

Guyant, George, Mark Hopkins, and Michael W. Reid, 1989 (July). *Longevity of Energy Savings: An Inspection and Analysis of Oilheat Systems Five Years After Retrofit*. The Alliance to Save Energy, Washington, D.C.

HEC Inc., 1993 (May). *Executive Summary for New England Power Service Company, Persistence of Savings from Lighting Measures Installed in the 1990 and 1991 Energy Initiative and Small Commercial and Industrial Programs*. Appendix M, Commercial/Industrial Persistence Studies. HEC Inc., Natick, Massachusetts.

HEC Inc., 1994 (May). *Final Report for New England Power Service Company, Persistence of Savings from Lighting Measures Installed in the 1990 and 1991 Energy Initiative and Small Commercial and Industrial Programs*. Appendix IV-7. HEC INC., Natick, Massachusetts.

Hickman, Curtis R. and Pamela Brandis, 1992. "Building Site Visits, A Look at Measure Retention as an Element of Program Evaluation." *Proceedings: ACEEE 1992 Summer Study on Energy*

Efficiency in Buildings, Panel 7, Program Evaluation. Washington, D.C.: American Council for an Energy-Efficient Economy, pp. 7.101-7.108.

Hirst, Eric and Richard Goeltz, 1985 (January). *Energy Savings One and Two Years After Participation in Minnesota Home Energy Audit and Retrofit Loan Programs.* ORNL/CON-168. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

Hirst, Eric and Carol Sabo, 1991 (October). *Electric-Utility DSM Programs: Terminology and Reporting Formats.* ORNL/CON-337. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

Hirst, Eric, Dennis White, and Richard Goeltz, 1984 (January). *Energy Savings Due to the BPA Residential Weatherization Pilot Program Two Years After Participation.* ORNL/CON-146. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

Hirst, Eric, Dennis White, and Richard Goeltz, 1985 (January). *Three Years After Participation: Electricity Savings Due to the BPA Residential Weatherization Pilot Program.* ORNL/CON-166. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

Jacobson, David, Meredith Miller, and Scott Englander, Bob Reals, and Tony Colonnese, 1993. "Use of On-Site Surveys to Assess Commercial and Industrial Measure Persistence," *Energy Program Evaluation: Uses, Methods, and Results. 1993 International Energy Program Evaluation Conference, August 25-27, 1993.* Proceedings, CONF-930842, pp. 614-620.

Jeppesen, John C. and Michael J. King, 1993 (January 15). *Effective Measure Life and Other Persistence Issues in DSM Programs, Final Report.* SRC Report No. 7729-R4. Prepared for California Conservation Inventory Group by Synergic Resources Corporation, Oakland, California.

Jeppesen, John C. and Monica Rudman, 1993. "Measure Life Research: Options and Lessons." *Energy Program Evaluation: Uses, Methods, and Results. Proceedings, 1993 International Energy Program Evaluation Conference, Chicago, Illinois, August 25-27, 1993,* pp. 520-527.

Keating, Kenneth M., 1991 (December). "Persistence of Energy Savings," In Eric Hirst and John Reed, eds., *Handbook of Evaluation of Utility DSM Programs.* ORNL/CON-336. Oak Ridge National Laboratory, Oak Ridge, TN, pp. 89-99.

Massachusetts Department of Public Utilities 1993. Letter (dated March 16) regarding expected Conservation and Load Management Annual Report with accompanying enclosure (revised May 27, 1993), "Definitions for C&LM Annual Report Tables."

Miller, Meredith B., William R. Blake, Sarah G. Dagher, Jeremy M. Schutte, and Roger L. Wright, 1992. "Development of Measured Savings for a Small Commercial/Industrial Lighting Program." *Proceedings: ACEEE 1992 Summer Study on Energy Efficiency in Buildings, Panel 7, Program Evaluation.* Washington, D.C.: American Council for an Energy-Efficient Economy, pp. 7.143-7.154.

Minnesota DSM Evaluation Consortium Working Document, 1993. Draft.

Misuriello, Harry and Mary Ellen F. Hopkins, 1992 (March). *Research Opportunities to Improve DSM Impact Estimates.* Prepared for the California Institute for Energy Efficiency by The Fleming Group, Washington, D.C.

Narum, David, Scott Pigg, and Jeff Schlegel, 1992 (September). *Looking Past the First Year: Do the Savings Last? A Study of the Persistence of Energy Savings in Low-Income Wisconsin Residences.* Wisconsin Energy Conservation Corporation, Madison, Wisconsin.

New Jersey Board of Regulatory Commissioners, 1993 (April 28). *Measurement Protocol for Commercial, Industrial and Residential Facilities.* Prepared in the context of the implementation of

New Jersey's Demand Side Management Rules. Newark, New Jersey: Board of Regulatory Commissioners.

Niagara Mohawk Power Corp, 1994 (July). *Annual Evaluation Report: 1993 Demand-Side Management Program*. Volume 1. Syracuse, New York: Niagara Mohawk Power Corporation.

Oswald, Kim Johnson, 1995. Personal communication.

Pacific Gas & Electric, San Diego Gas & Electric, Southern California Edison, Southern California Gas, California Energy Commission, Division of Ratepayer Advocates (CPUC), and Natural Resources Defense Council, 1993. *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs*. Adopted by the California Public Utilities Commission Decision 93-05-063, Revised November 30, 1993.

Parker, Lee, 1994 (May 16). Personal communication.

Parker, Lee, 1993 (July 2). *Final Report. Persistence Study of Energy Conservation Measures Implemented in LILCO's Commercial Audit and Dollars & Sense Programs*. Prepared by Applied Energy Group, Hauppauge, New York for Long Island Lighting Company, Economic & Management Planning Department, Hicksville, NY.

Pennsylvania Public Utility Commission, 1992 (December). *An Examination of Key Demand-side Management Evaluation Issues with a Suggested Evaluation Report Format*. Pennsylvania Public Utility Commission, Bureau of Conservation, Economics, and Energy Planning.

Public Service Commission of Wisconsin (PSCW), ca. 1992. *AP6 [Advance Plan 6] Requirement for Annual Plans to Document and Attain Persistence and Reliability: PSCW Staff Issue Paper*. Unpublished white paper.

Robinson, Donna M., 1992. "Comprehensive Residential Lighting Retrofits: A Case Study." *Proceedings: ACEEE 1992 Summer Study on Energy Efficiency in Buildings, Panel 7, Program Evaluation*. Washington, D.C.: American Council for an Energy-Efficient Economy, pp. 7.193-7.202.

Saxonis, William, 1991. "Free Riders and Other Factors that Affect Net Program Impacts," in Eric Hirst and John Reed, eds., *Handbook of Evaluation of Utility DSM Programs*. ORNL/CON-336. Oak Ridge National Laboratory, Oak Ridge, Tennessee, pp. 119-134.

Schweitzer, M., M. A. Brown, and D. L. White, 1989 (September). *Electricity Savings One and Two Years After Weatherization: A Study of 1986 Participants in Bonneville's Residential Weatherization Program*. ORNL/CON-289. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

Skumatz, Lisa A., Karen M. Lorberau, Ronald J. Moe, Robert D. Bordner, and Robert D. Chandler, 1991 (December). *Bonneville Measure Life Study: Effect of Commercial Building Changes on Energy Using Equipment. Final Report*. SRC Report No. 7619-R2. Prepared for Bonneville Power Administration, Portland, Oregon, by Synergic Resources Corporation.

Synergic Resources Corporation (SRC), 1992 (February 13). *Effective Measure Life and Other Persistence Issues in DSM Programs. Literature Review, Interim Report #1.2*. Prepared for the California Conservation Inventory Group by Synergic Resources Corporation, Oakland, California.

Train, Kenneth E., 1994. "Estimation of Net Savings from Energy-Conservation Programs." *Energy* 19 (4): 423-446.

Trumble, David and Michael MacDonald, 1994 (August). *Extention to Phase 2 Energy Savings Evaluation of the Commercial Direct Investment Program*. ORNL/CON-360. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

U.S. Environmental Protection Agency, 1993a (March). *Conservation Verification Protocols: A Guidance Document for Electric Utilities Affected by the Acid Rain Program of the Clean Air Amendments [sic] of 1990*. EPA 430/8/B-92-002. Acid Rain Division.

U.S. Environmental Protection Agency, 1993b (April). *The User's Guide to the Conservation Verification Protocols*. EPA 430-B-93-002. Acid Rain Program.

Velcenbach, Paul and Lee Parker, 1993. "Can DSM Survive Persistence?" *Energy Program Evaluation: Uses, Methods, and Results. Proceedings, 1993 International Energy Program Evaluation Conference*, Chicago, Illinois, August 25-27, 1993, pp. 506-517.

Vine, Edward L., 1992. "Persistence of Energy Savings: What Do We Know and How Can It Be Ensured?" *Energy* 17(11): 1073-1084.

Vine, Edward L., 1995 (January 17). Personal communication.

Violette, M. Ozog, M. Keneipp, F. Stern, and P. Hanser, 1991 (February). *Impact Evaluation of Demand-Side Management Programs. Volume 1: A Guide to Current Practice*. EPRI CU-7179. Prepared by RCG/Hagler, Bailly, Inc., Boulder, Colorado for Electric Power Research Institute, Palo Alto, California.

White, Dennis L. and Marilyn A. Brown, 1990. *Electricity Savings Among Participants Three Years After Weatherization in Bonneville's 1986 Residential Weatherization Program*. ORNL/CON-305. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

White, Dennis L. and Marilyn A. Brown, 1992. "The Reliability of Residential Energy Conservation Resources." *Proceedings: ACEEE 1992 Summer Study on Energy Efficiency in Buildings, Panel 7, Program Evaluation*. Washington, D.C.: American Council for an Energy-Efficient Economy, pp. 7.267-7.275.

White, Dennis L., Therese K. Stovall, and Bruce E. Tonn, 1992 (February). *Persistence of the Impact of the Hood River Conservation Project on Typical and Peak Loads Three Years After Weatherization*. ORNL/CON-321. Oak Ridge National Laboratory. Oak Ridge, Tennessee.