Prepared for

Southern California Edison

and

California DSM Measurement
Advisory Committee (CADMAC)

June 1994

Final Report

DSM Program Spillover Effects

Review of Empirical Studies and Recommendations for Measurement Methods

Prepared by

Cambridge Systematics, Inc.
222 Third Street
Cambridge, Massachusetts 02142

Principal Investigators

Glen Weisbrod
Kenneth Train
Andrew Hub
Peter Benenson

Table of Contents

List of Tables

List of Figures

1.0 Introduction

1.0 Introduction

■ 1.1 Background

The evaluation of DSM program benefits can have consequences for program design, cost recovery, rate setting and public policy. The methods that have been used for measuring net benefits have been, and are, continuing to evolve over time. The measurement of spillover impacts is one aspect of that continuing evaluation in measuring net program benefits.

The term "spillover" refers to additional program benefits that are beyond those normally counted as the gross program benefits. Normally, in DSM program evaluations, the gross program benefits are measured as the energy savings (i.e., reduction in energy consumption) or peak demand savings associated with the introduction of program measures by program participants.

Net program benefits are benefits that can be attributed to the program. There are two types of adjustments possible in going from gross benefits to net benefits: subtraction and addition. There can be subtractions from gross benefits to the extent that some of the observed energy savings would naturally have occurred even without the program. This is sometimes referred to as "free ridership." There can also be additions to gross benefits to the extent that there are energy savings beyond the program measures installed by participants. This is sometimes referred to as "spillover." These differences between net program savings and gross program savings are illustrated in Figure 1.1.

There is, of course, also the possibility of additional program benefits other than energy savings. Examples of these "societal benefits" are listed in Table 1.1. Many of these societal benefits are typically difficult to measure in monetary terms, and some correction for double-counting must be made if any societal benefits are to be added to energy savings benefits in any measure of overall net program benefit.

While the impacts of a DSM program can "spill over" to any and all of the classes of benefits listed in Table 1.1, this study focuses specifically on spillover benefits that are energy savings impacts.

Figure 1.1 Relationship of Net and Gross Savings

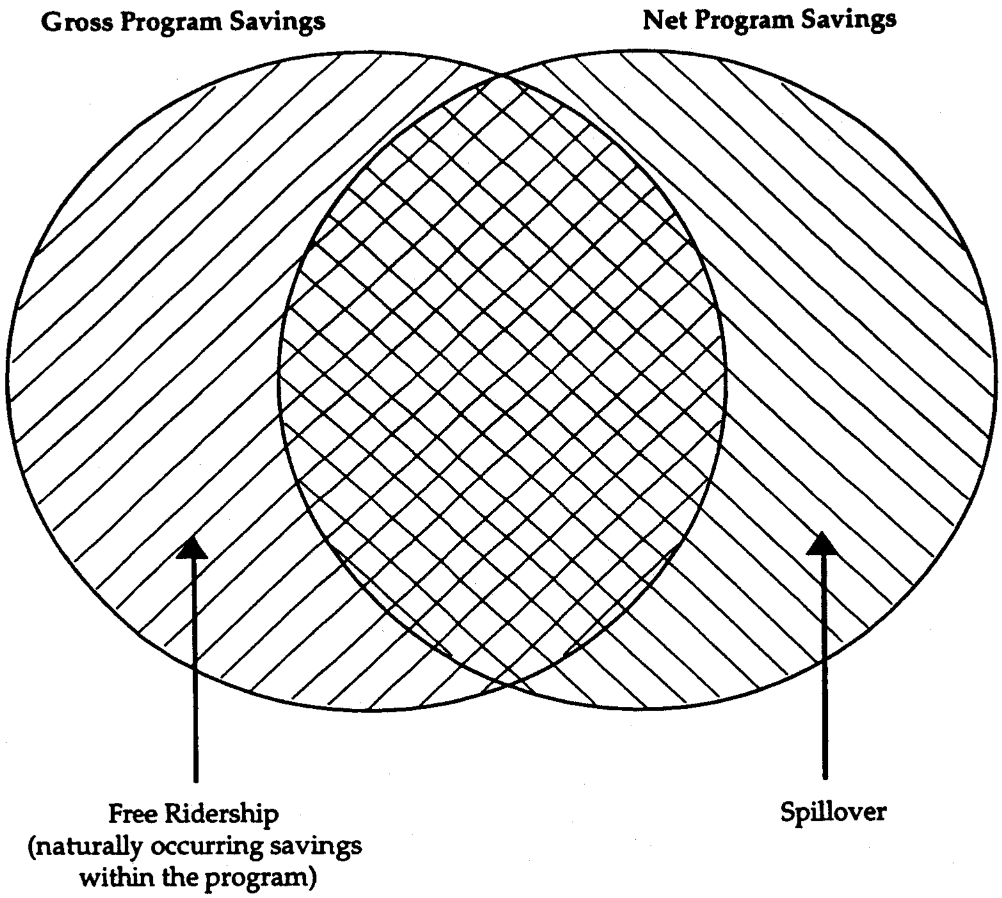


Table 1.1 Types of DSM Program Benefits Beyond Gross Energy Savings Impacts

Additional Energy-Saving Benefits

- Participants energy saving outside the program
- Non-participants energy saving outside the program

Non-Energy-Saving Benefits for Participants and the Rest of Society

- Environmental benefits (air and water quality)
- Safety benefits (from improved customer equipment and/or reduced power plant needs)
- Economic security benefits (e.g., reduced dependence on oil imports)
- Equipment and capital operating cost savings benefits (beyond savings in cost of energy)
- Employment and labor income benefits (from jobs created)
- Social equity benefits (cost saving and quality of life improvements for low-income households)

Thus, for purposes of this study, we narrowly define spillover as follows:

> Spillover is the reduction in energy consumption and/or demand in a utility's service area caused by the presence of the DSM program, beyond program-related gross savings of participants.

This is a conceptual definition of energy-saving spillover impacts; a more useful operational definition for purposes of actual measurement is provided in Section 2.0.

■ 1.2 Objective

The California Demand-Side Management Measurement Advisory Committee wants to improve the understanding and measurement of spillover effects, as part of an effort to improve impact measurement for DSM programs. The objective of the present study is to provide a critical review of past empirical studies that included spillover effects, and to recommend approaches for measurement of spillover in future DSM program evaluations.

To meet this objective, the following questions are addressed:

1. What definitions of spillover have been used in other studies, and what are their strengths and shortcomings?
2. What definition of spillover will yield the most accurate estimate of the overall net effects from a DSM program?
3. How should spillover be measured in terms of data sources and analysis methods?
4. What are the tradeoffs in terms of cost and usefulness of measuring spillover?

■ 1.3 Issues

In reviewing the literature and recommending approaches for measuring spillover, this study addresses a set of nine key issues. These issues are summarized as follows:

Relation of Spillover to Net Impact

The fundamental interest in DSM program evaluation is in accurately estimating the overall net impacts of DSM programs. Spillover is merely a component of overall net impacts. This point is important, for as discussed in Section 2.0, the methods for assessing spillover and overall net impacts are related. Accordingly, this study lays out a unified set of operational definitions and methodologies for evaluating both spillover by itself and overall net impacts.

Breadth of Spillover Definitions

For any given DSM program, users of energy can be divided into two groups: program participants and non-participants. Non-participants can even include customers of other nearby utilities. Likewise, for any given DSM program, energy savings actions can be divided into two groups: program-eligible measures and non-eligible measures. The possible combinations of user categories and measure categories is shown in Figure 1.2. Essentially, program savings is measured as the savings associated with implementation of rebated program measures by program participants. Any other category of energy savings shown in Figure 1.2 can, in theory, be classified as a program spillover benefit, as long as it can be shown that implementation of those measures is attributable to the DSM program. There is, however, a distinction between theory and practice. In practice, broader claims of spillover benefits can be harder to support and are more likely to raise issues of double-counting or over-counting. Accordingly, this study notes the differences between broader definitions of spillover in theory and what has tended to be narrower definitions of spillover in actual empirical analysis.

Evolution of Spillover and Free Driver Concepts

While the concept of spillover is generally understood among researchers, most of the relevant methodology and practice relating to spillover has actually taken place through the concept of "free drivers" as a counterpart to "free riders." (These terms are defined and discussed in Section 2.0.) In fact, originally free driver impacts of DSM programs were seen as merely a source of bias in non-participant control groups for program impact studies. Only more recently has spillover been recognized as a potentially legitimate element of additional program benefit. Thus, what appears at first to be wide differences in the literature on how to define free driver and spillover impacts can actually be more accurately portrayed as an evolution of theory and practice. Accordingly, this study addresses these measurement issues in the literature review (Section 4.0), and also attempts

Figure 1.2 Types of Spillover Benefits

Types of Measures	Program Participants	Non-Participants
Program Eligible	Program Savings	B
Non-Program Eligible	A	C

Potential Spillover Savings

to reduce confusion by first presenting a general perspective for defining the terms (in sections preceding the literature review).

Behavioral Factors Affecting Spillover

DSM programs can lead to behavioral changes affecting spillover benefits in many ways. As discussed in the conceptual framework and literature review of this report, we see that spillover impacts may result from actions taken by manufacturers, distributors, dealers, contractors, energy auditors, participating customers, non-participating customers, colleagues and friends, and/or customers of other utilities. These differences lead to a confusing (and overlapping) array of different aspects of spillover effects covered in different studies. Accordingly, this study addresses these behavioral complexities by presenting a general framework intended to cover all types of behaviors, and then noting how they map to specific types of programs and situations in Section 5.0.

Relationship of Spillover to Free Ridership

While this study focuses specifically on the concept of spillover, it is shown in this study that the measurement of spillover is in fact related to the measurement of free ridership (again, these terms are defined in Section 2.0). First of all, both are factors affecting the measurement of overall net program impacts. Second, both can only be measured through some technique to assess what would have happened if the DSM program had not existed. Third, the extent of spillover may affect the usefulness of some techniques for measuring free ridership. Accordingly, this study addresses this interrelationship in the context of definitions (Section 2.0), methods (Section 3.0), and recommendations (Section 5.0).

Confusion in Terminology

The field of DSM program impact evaluation contains a number of metaphorical jargon terms which can be misleading in the context of attempts to clarify and define spillover impacts. A prime example of this is the concept of "free driver," a label originally intended to be juxtaposed with that of a "free rider," and which would seem to be applicable to one who freely drives a series of energy-saving behaviors. However, as previously noted, spillover impacts can be "driven" by (i.e., lead by the actions of) many different actors. In actual practice, the term is used differently in various studies. Accordingly, this report recognizes and notes differences in definitions of terms (in Sections 2.0 and 4.0), and yet tries to adopt a consistent set of definitional concepts.

Oversimplification of Concepts

The measurement of net program impacts, including spillover impacts, necessarily involves some explicit or implicit estimation of what would have happened if the DSM program had not existed. Terms such as "free rider" and "free driver" are applied to various sets of customers on the basis of whether they would or would not have implemented certain energy-saving measures in the absence of the program. Yet in reality we know that this is an oversimplification, as some customers would have implemented alternative actions with some, but not as much, energy saving as occurred with the program in place. The twin concepts of partial free riders and partial free drivers can substantially complicate the definition and measurement of net impacts, and are seldom both used in empirical studies. However, this report does recognize and address these complexities, and provides an overall framework that includes them.

Measurement Cost and Accuracy Tradeoffs

A range of analysis methodologies, ranging from simple survey questions to complex econometric models, may be employed to measure spillover and overall net impacts of DSM programs. There are also significant differences in cost and accuracy associated with the these methodologies. This study reviews the results of limited studies to date which have compared multiple methods, and recommends general guidelines for applying alternative methods (in Section 5.0).

Tailoring Approaches to Different Types of Programs

There are several different types of DSM programs. One difference is their applications focus, e.g., new construction equipment applications, retrofit equipment applications, or behavior modification applications. Another difference is sectoral focus, e.g., industrial, commercial, agricultural or residential. Yet another important difference is program technique used: e.g., customer information, on-site audits or financial incentives (rebates). Measurement definitions and analysis methods must be tailored to address specific types of programs. While it is premature for this study – based on an analysis of the literature – to provide a definitive set of rules for all such situations, some general guidelines are provided in the recommendations (Section 5.0).

1.4 Approach

This study consists of two steps: (1) a review of the literature and (2) the development of general recommendations concerning definitions and measurement of spillover impacts. The literature review clearly shows a diverse and potentially confusing array of alternative definitions and methodologies which have been employed in various prior studies.

Instead of walking the reader through all of the differences taken in different studies, this report adopts a more streamlined form of presentation, starting with the study conclusions on appropriate definitions of spillover and net impact concepts. These definitions, contained in Section 2.0, are intended to provide an internally consistent framework for defining spillover impacts, and for understanding how spillover relates to other aspects of overall net program impact. The relationship of this definitional framework to other terms and definitions found in the literature review is also noted there.

This is followed by a general overview and classification of alternative methods used for estimating spillover and overall net impacts of DSM programs. This review of methods, contained in Section 3.0, is intended to provide a comprehensive overview of alternative methodologies, their data requirements, strengths and weaknesses. Particular empirical studies utilizing each of the various methods are also noted there.

The literature review itself is contained in Section 4.0. It is divided into three parts. The first part is a discussion and presentation of alternative definitions utilized in various studies. The second part is a discussion and summary of empirical studies and articles relating to spillover effects. The third part is a list of documents reviewed.

Finally, Section 5.0 presents the methodological recommendations. This includes recommendations regarding appropriate definitions, analysis methods and issues remaining for further study.

2.0 General Definitions

2.0 General Definitions

The literature on net impact measurement in general, and that on spillover measurement in particular, presents a widely variant array of alternative approaches and methods. Before launching into a review of them, it is useful to adopt a general set of concept definitions and a general classification of methodologies. This is intended to provide a universal framework from which all prior studies and current methods can be reviewed. Accordingly, this section lays out that framework in two parts:

1. Definitions of terms relevant to spillover; and

2. Identification of the components of spillover.

A classification of methods for estimating spillover is later provided in Section 3.0.

■ 2.1 Components of Net Impact

Overview

Spillover is most readily, and usefully, defined in the context of appropriate and consistent definitions for other terms that are used in program evaluation.

The measurement of DSM program impacts is ultimately provided in terms of energy savings and demand reduction. For purposes of understanding spillover and net program impacts, however, it is first useful to consider the basic unit of analysis as an energy-saving measure taken by an individual customer. A measure is defined broadly as any energy-saving action including retrofits, behavioral changes (such as turning off lights or adjusting the thermostat), and installation of efficient equipment. Major categories of energy-saving actions are summarized in Table 2.1. There is a reduction in energy consumption that results from the customer implementing the measure (relative to the consumption that would have occurred if the customer had not taken the measure). This reduction is called the "gross savings of the measure." The gross savings of the measure is specific to the particular measure (in that different measures have different savings) and to the customer (in that different customers taking the same

Table 2.1 Types of Energy-Saving Measures

A. Install High Efficiency Equipment in lieu of otherwise-planned new equipment

B. Install High Efficiency Equipment as a retrofit to replace otherwise-functioning equipment

C. Install Energy Conservation Materials (e.g., insulation)

D. Change Patterns of Use (e.g., length, frequency or thermostat settings) measure may obtain different savings from the measure.) Thus, gross savings from the measure is defined as follows:

> The "**gross savings from the measure**" is the difference between the consumption that the customer would have without the measure and the consumption that the customer has with the measure.

This definition of gross savings incorporates any snapback effect (i.e., changes in equipment usage patterns). That is, if the measure induces the customer to change its usage patterns, then the change in consumption that results from this change in usage patterns is incorporated into the gross savings from that measure. Such effects are also reflected in pre/post billing analysis comparisons of gross savings.

The various terms that are used in program evaluation – namely, gross program savings, net savings, naturally occurring savings, and spillover effects – are similar in that each is based on a sum of gross savings of some energy-saving measures. They differ in which measures are included in the sum. This fact, when recognized, greatly facilitates understanding of each term and the relations among terms.

Based on a review of the literature and current state-of-the-art, this section provides a set of standard definitions which together comprise an internally consistent approach for understanding and measuring spillover effects. It is important to acknowledge, however, that not all other studies have defined or measured spillover effects or net impacts in ways that fully match the definitions provided here. Those variations from these general definitions and measurement approaches are addressed in the literature review (Section 3.0).

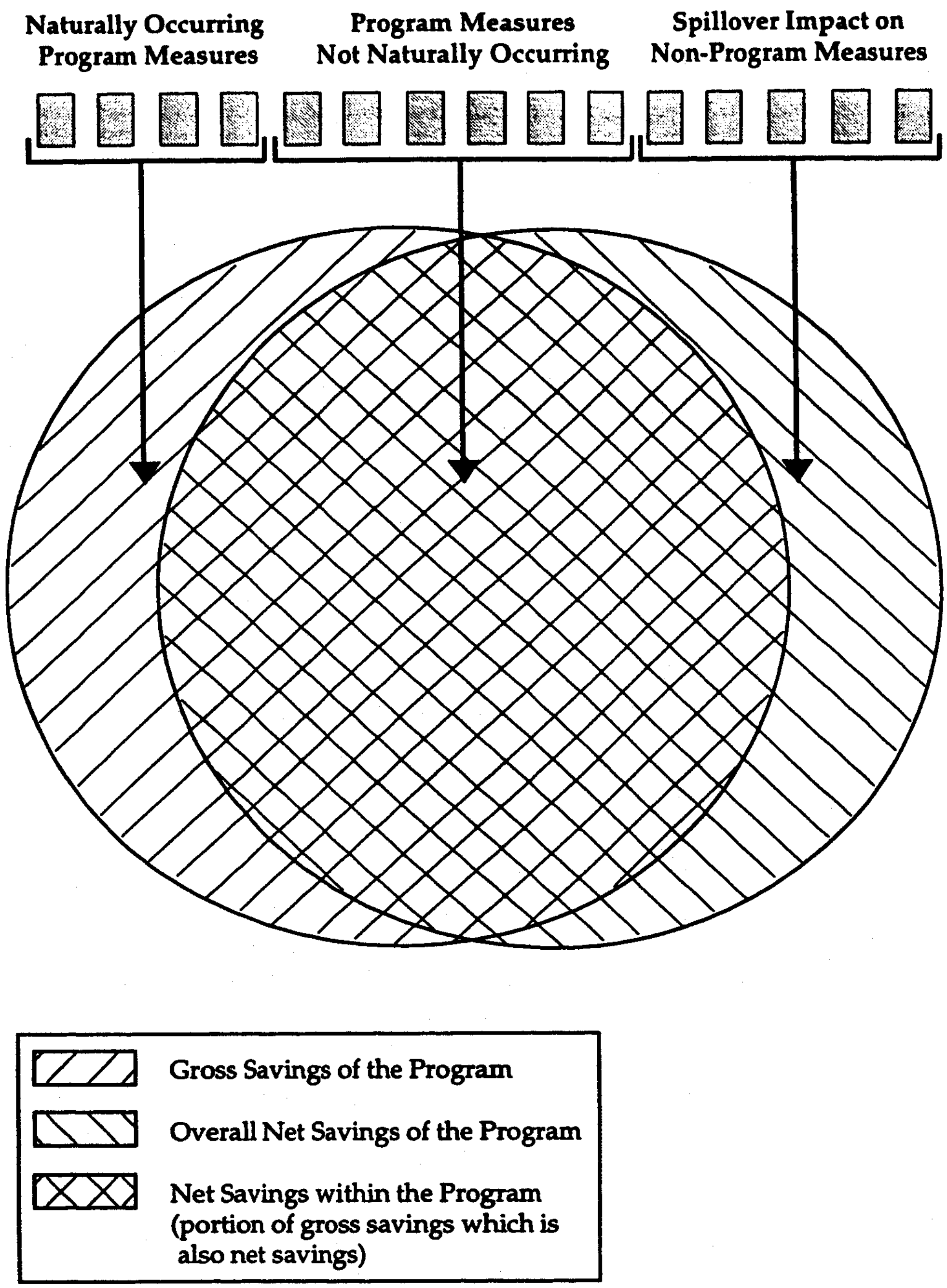
Gross and Net Program Savings

The relationship between gross and net savings is illustrated in Figure 2.1 below. Gross Program Savings is defined as follows:

> The "**gross savings of the program**" are the gross savings of all the measures that are counted as being taken under the program.

For example, for a rebate program, the gross savings of the program are the gross savings of all the measures on which rebates were paid. For an audit program, the gross savings are usually the gross savings for all measures that were recommended to a customer during an audit and were subsequently taken by the customer. Note that the term "gross savings" has two distinct uses: the gross savings of a measure and the gross savings

Figure 2.1 Gross Savings of Energy-Saving Measures



of the program. This dual use of one term is unfortunate, but will be maintained in our discussion since both uses are well established in the field.

Net Program Savings is defined as follows:

> The "**overall net savings of the program**" are the gross savings of all measures that were taken and would not have been taken if the program had not existed, minus the gross savings of any other measures that were **not** taken and would have been taken if the program had not existed.

This definition of net savings adjusts for the fact that some measures would not have been taken at all without the program, while in other cases actions would have occurred without the program, but they would have saved less energy (for example, a less efficient model of equipment would have been purchased or the equipment purchase would have occurred later on). Another important aspect of this definition is that it does not restrict net savings to include only measures that were counted in the program; measures that are not counted as part of the gross savings of the program can be included in net savings as long as the measures would not have been taken if the program had not been offered.

Relationship of Spillover to Net Savings

Two other terms allow us to relate the net savings of the program to the gross savings of the program: "naturally occurring savings" and "spillover." They are defined as follows:

> The "**naturally occurring savings within the program**" are the gross savings of any measures taken under the program, or less efficient substitute measures, that would have been taken if the program had not existed. (This value is always a portion of the gross program savings.)
>
> "**Spillover**" is the gross savings of measures that are **not** counted as part of the gross program savings and yet would not have been taken if the program had not existed.

With these definitions, the net savings of the program are equal to the gross savings of the program, minus the naturally occurring savings within the program, plus the spillover. In other words:

> **Overall net savings of program** = Gross savings of program - Naturally occurring savings within program + Spillover.

This relation can be expressed in an even more concise fashion that relates directly to the methods that are used to estimate spillover. Define "the net savings within the program" to be the gross savings of the program minus the naturally occurring savings within the program. This "net savings within the program" is what most evaluations estimate; it does not incorporate spillover. With this term, the above equation becomes:

Overall net savings = Net savings within the program
+ Spillover.

The term "overall net savings" is used rather than simply "net savings" to emphasize that it includes spillover and to distinguish it from the "net savings within the program, " which does not include spillover. The net savings within the program accounts for free ridership but not spillover, while the overall net savings accounts for both free ridership and spillover.

This relation is fundamental to accurately measuring spillover and to incorporating spillover into calculations of net savings. In discussions of spillover, several methods for estimating spillover have been proposed. Some of these methods estimate spillover while others actually estimate overall net savings, which incorporates spillover but also incorporates the net savings within the program. The distinction has not always been clear, and the advantages and limitations of different methods have been compared without recognizing that the methods are estimating different things. These issues are discussed in Section 3.0, along with other reasons for keeping this basic relation in mind when examining spillover.

Classification of Customer Behavior

It is useful to review some other terms that are common in program evaluation and relate them to the definitions provided above for gross, net, naturally occurring savings and spillover.

Participants

A **"participant"** is a customer who participated in the program, although it must be noted that different programs define participation differently. An audit program, for example, would usually consider a customer to be a participant if the customer was audited, whether or not the customer installed any recommended measures. In a rebate program, the customer is generally considered a participant only if the customer receives a rebate for purchasing or installing an eligible measure. In such a rebate program, each participant takes at least one measure that is included in the gross savings of the program, whereas this is not necessarily true for an audit program.

Free Riders

The term "**free rider**" is often used to refer to participating customers who receive rebates even though they would have implemented the same measures without the rebate; hence they are getting a "free ride" on the incentive program. In the context of the terminology used here for net saving calculations, a participant may be called a "free rider" if that participant implements a measure that is included in the gross savings of the program, and would have implemented the same measure even if the program had not existed.

We can also define a "**partial free rider**" as a participant who would not have implemented the exact same program measure, but would have implemented a lesser-savings substitute action, if the program had not existed. (This is sometimes also broken down into "incremental" and "deferred" free ridership.)

Free Drivers

The term "**free driver**" has been used to refer to the opposite of free ridership – i.e., non-participating customers who are affected by the program's existence, but implement program-eligible measures without officially joining and receiving the rebate. The term comes from the idea that these customers, rather than going for a free ride with the rebate, freely drive themselves to take actions without the rebate. The term has been undergoing significant evolution over time, as it has been expanded by some writers to also include customers implementing other types of energy-saving measures beyond those covered by the rebate. Use of this term is discussed further in Section 3.2.

For purposes of theory, it is useful to adopt an expansive rather than a restrictive definition of free drivers. In the context of the terminology used here for net savings calculations, customers thus may be called "free drivers" if they take any measure that is included in spillover. That is, a "free driver" is a customer who takes a measure that is not counted in the gross savings of the program but would not have taken the measure if the program had not existed. In this context, such a customer can be either a participant or a non-participant.

Customers who take multiple measures can fall into different categories for each measure. For example, customers can be free riders of some measures (e.g., if they received a rebate for measures that they would have taken without the program) and not on others (e.g., if they took other measures outside the program, or took other rebated measures that they would not have taken without the program). Similarly, a customer can be a free driver on some measures and not on others. In fact, it is possible for a customer to be both a free rider and a free driver. For example, a customer might receive a rebate for a measure that it would have implemented even without the rebate, and yet that same customer might also implement another, non-rebated measure because of the general awareness of energy efficiency which the rebate program engendered.

For definitional clarity, we find it more useful to consider the basic unit of analysis to be a measure installed by a particular customer, rather than the customer itself (who might have taken several measures). This is the reason for defining gross, net, naturally occurring savings, and spillover in terms of measures rather than customers.

In this regard, we can refer to the naturally occurring savings within the program as the savings from free riders, and the spillover as the savings from free drivers. This usage is intuitively meaningful, but can be misleading since it does not explicitly recognize that a customer can be a free rider for some measures, a free driver for others, and neither a free rider or free driver for other measures. Again, we find it more useful to think in terms of categorizing measures rather than customers.

■ 2.2 Components of Spillover

The definition of "spillover" given above is broad: it is defined as "the gross savings of measures that are not counted as part of the gross savings of the program, but would not have been taken if the program had not existed." Many types of spillover are incorporated in this definition. These types are usually distinguished on the basis of how the spillover effect is generated. In the following paragraphs, we discuss general ways that spillover can occur (see Table 2.2).

A customer who participates in a program might install measures outside of the program because of the interest or understanding of conservation that the program induced. For example, a customer might install a measure under a program and learn, by observing the savings from this measure, that energy efficient equipment can indeed be cost effective. This information might then induce the customer to take other measures outside the program. The savings from these non-program measures are part of the spillover of the program, since they would not have been taken if the program had not existed. This type of spillover has been called the "surge effect," since it evidences itself as a surge in savings in a period after the program measure was installed. This corresponds to item "A" in Table 2.2.

Table 2.2 Types of Spillover

A. Participants Adopt Non-Program Measures

- Program participants may gain energy awareness and knowledge of further energy-saving opportunities from the program, leading them to adopt additional energy-saving measures outside of the program.

B. Non-Participant Awareness of Program Measures

- Non-participants may gain knowledge of energy-saving opportunities from the program, leading them to also choose to adopt program-eligible types of energy-saving measures that they had not previously been planned, although they did so without becoming a program participant.

C. Non-Participant Awareness of Non-Program Measures

- Non-participants may gain energy awareness and knowledge of energy-saving opportunities from the program, leading them to also choose to purchase, install or adopt other types of energy-saving measures that they had not previously been planning to do.

D. Moving the Market

- Manufacturers, dealers and contractors may change the array of energy-using equipment that they offer to all customers, leading to increases in purchases of energy-saving equipment by non-participants.

A non-participant might take a program-eligible measure because of interest or awareness induced by the program. For example, a customer might install a rebate-eligible measure because of learning about the measure from the program's advertising campaign, but then not apply for the rebate. There could be any number of reasons for the customer not applying for the rebate, e.g., the customer could have forgotten to apply, or might have considered the hassle or time required to apply not to be worth the value of the rebate. In any case, the savings from the measure are not counted under the gross savings of the program (since the customer did not apply for a rebate) and yet the savings are attributable to the program. The term "free driver" has been used by some writers to refer to these customers, namely, non-participants who install program-eligible measures because of the program. (This corresponds to item "B" in Table 2.2.)

However, it is useful here to define "free driver" more generally as a customer who takes a measure outside the program that it would not have taken if the program had not existed. The more general definition includes both participants and non-participants and both program-eligible and non-program-eligible measures, while the more restrictive definition includes only non-participants and program-eligible measures. (This includes item "C" in Table 2.2.) Of course, either definition of "free driver" can be acceptable in a particular situation, as long as the definition is clear. Regardless of whether or not it is referred to by the term "free driver," the phenomenon described in the more general definition can exist.

A program might induce dealers to stock energy efficient equipment that they would not otherwise stock. Customers might purchase this equipment without participating in the program, and yet would not have had the option to purchase it if the program had not existed. Similarly, the program might serve to educate contractors about energy efficient equipment, make them more adept at installing this equipment, and more willing to recommend the equipment to their clients when bidding on work. This shift in contractors' practices can result in more widespread installation of the equipment, even among customers who do not participate in the program. These program-induced changes are called "moving the market," and this could also induce customers to become "free drivers." (This corresponds to item "D" in Table 2.2.)

Program-Specific Differences

The general forms of spillover described above may or may not apply for any given DSM program, depending on the type of program. For example, impacts associated with increased awareness of energy efficiency could be classified as spillover impacts for an equipment rebate program, but classified as direct program impacts for an information awareness program.

Similarly, "moving the market" may be classified as a spillover impact of a rebate program, but classified as a program impact for a market intervention program. Table 2.3 lists typical ways in which a DSM program may cause customers to adopt energy-saving measures. Whether or not a customer action is classified as a program impact or as a spillover impact depends on the type of DSM program and how it defines participation, which then also determines how its gross impacts are measured.

Spillover impacts such as energy cost awareness can also be a double-edged sword. For instance, if an equipment retrofit program has the spillover impact of leading non-participants to reduce their use of existing heating or hot water equipment, then the energy savings in their existing equipment is an additional (spillover) benefit for the program. On the other hand, if the same program also leads participants to reduce their use of newly-installed retrofit equipment, then the net effect is a reduction in the measurement of net savings associated with that equipment installation.

Definitional Complications

It has also been suggested by some writers that a customer can appear to be a "free rider" and a "free driver" even for the same measure. The scenario goes as follows: A customer installs a rebate-eligible measure and receives a rebate for doing so, such that the savings from the measure are counted under the program. When asked, the customer says that it would have installed the measure even if the rebate had not been offered; the customer is therefore considered a free rider. But without the program, the measure would not have been stocked by dealers or offered by contractors. Therefore the savings from the measure are indeed attributable to the program; and since the installation is due to program-induced movement in the market, the customer who took the measure is considered a free driver.

This is an interesting and quite plausible scenario. However, using the definitions that we developed above, the customer is not a free rider or a free driver. The savings from the measure are counted under the program (so the customer is not a free driver) and would not actually have occurred if the program had not existed (so the customer is not a free rider). The reason the customer appears to be a free rider is that the customer said that it would have taken the measure without the program; however, in fact, the customer would NOT have taken the measure if the program had not existed. The problem is in measurement: the self-report of participants is not, in this case, sufficient to distinguish which measures would have been taken without the program and which ones would not have been. The customer identified itself as a free rider when indeed it was not. Of course, if the customer is labeled a free rider, and the savings from the measure

Table 2.3 Reasons for Adopting Energy-Saving Measures

A. Financial incentives offered by DSM program

B. Awareness of energy efficiency and potential cost savings (provided by DSM Program)

C. Exposures to information about DSM program

D. Awareness of the energy-saving behavior of others (neighbors, colleagues and competitors)

E. Changes in the array of equipment stocked by distributors and retailers

F. Changes in the types of equipment recommended by dealers or contractors are subtracted from the gross savings of the program, then it is necessary to also consider the customer a free driver, adding the savings from the measure back in. Therefore, if the customer is seen as a free rider, it must also be seen as a free driver to obtain the correct estimate of net savings. However, if the customer is not seen as a free rider in the first place, then the customer need not be seen as a free driver either; its savings will be included in net savings as a participant whose action was induced by the program.

The issue of how to define terminology in this situation is not so critical, as long as the definitions are clear and.consistent. We feel that it is more useful to define terms independent of methods. Then each potential method for identifying a term can be assessed for how closely the method correctly estimates or measures the concept indicated by the term. In the situation described above, some writers have defined a free rider as a customer who says that it would have taken the measure without the program, whether or not this is true. We find it more useful to call the customer "an apparent, but not real free rider", learning from this situation that there is a real limitation of self-report data. However, both sets of definitions will result in the same estimate of net savings. Either way, the savings from the measure will be included, despite what the customer said in the self-report.

3.0 Estimating Methodological Approaches for Spillover

3.0 Estimating Methodological Approaches for Spillover

■ 3.1 Alternative Approaches

The definitions and decomposition of terms given in Section 2.1 and 2.2 facilitate analysis of methods for estimation of spillover effects and the spillover contribution to the net-to-gross ratio. As stated in Section 2.1, the basic relation is:

Overall Net Savings = Net Savings Within the Program + Spillover

Different methods are capable of estimating each element of this equation, as described below. By recognizing which methods provide estimates of which elements, appropriate choices can be made by the analyst, given the goals of the evaluation. For example, suppose the goal is to estimate overall net savings. The most prominent way to estimate the overall net savings is to estimate net savings within the program, estimate spillover separately, and then add the two. However, it is also possible to utilize a method that estimates overall net savings directly; that is, a method can be used that estimates overall net savings with the combined effects of free ridership and spillover both incorporated automatically. This latter procedure will not allow the analyst to know the size of spillover effects per se, since the effects of spillover and free ridership are both incorporated into the one estimate. However, if the goal is simply to estimate overall net savings, then this limitation might not be important.

Sometimes, the analyst will need to obtain an estimate of spillover by itself. In this case, spillover can be estimated directly with a method that estimates only spillover. Or, using the above relation, spillover can be estimated through subtraction by (i) estimating overall net savings with a method that automatically incorporates the effects of free ridership and spillover, (ii) estimating net savings within the program using a method that estimates net savings without spillover, and then subtracting the two to obtain an estimate of spillover. Given the difficulty (described below) of estimating spillover directly, this procedure might be preferable in some settings.

The above relation, and recognition of which analytic methods provide estimates which elements of this relation, are also useful in avoiding potential mistakes in the calculation of overall net savings. For example, it is incorrect to estimate overall net savings with a method that incorporates spillover and then add a separate estimate for spillover. In that case, spillover would be counted twice. It is also incorrect to use a method for estimating net savings that incorporates spillover and add it to an estimate of net savings within the program (as if the former estimate were an estimate of spillover only). This procedure counts the net savings within the program twice. Though these errors seem obvious, both have been made in previous evaluations. The problem seems to arise because the analyst is not clear about which methods are capable of estimating which elements of this relation. For example, an analyst might believe that his/her method provides an estimate of spillover when in fact it provides an estimate of overall net savings, including spillover and the net savings within the program. Or, the analyst might mistakenly think that his/her method estimates net savings within the program (i.e., ignoring spillover) when in fact the method estimates overall net savings (including spillover), making it redundant to add spillover effects separately.

In the paragraphs below, we describe the methods that can be used to estimate (a) overall net savings, (b) net savings within the program, and (c) spillover. As stated above, the analyst can estimate any of these terms directly or indirectly. Overall net savings can be estimated by using any of the methods in (a). Or, overall net savings can be estimated by pairing any method from (b) with any method from (c) and adding the two results. Similarly, spillover can be estimated directly with a method from (c); or it can be estimated by pairing any method from (a) with any method from (b) and subtracting to obtain the difference between the two. In determining which methods to apply in any given evaluation, the analyst must examine the data requirements, advantages and limitations of each method, as well as the budget, time frame, and goals of the evaluation (see Figures 3.1 and 3.2).

3.2 Estimation of Overall Net Savings (Incorporating Spillover)

Overall net savings is the savings due to the program. By definition, it is the difference between the consumption that occurred with the program and the consumption that would have occurred if the program had not existed. Stated equivalently, it is the savings from all measures that would not have been taken if the program had not existed. Estimating overall net savings therefore requires a determination of what measures would have been taken without the program or what consumption would have been without the program. This estimate necessarily requires some form of extrapolation. Two generic procedures have been suggested, one based on

Figure 3.1 Summary: Methods for Estimating Spillover and Net Savings

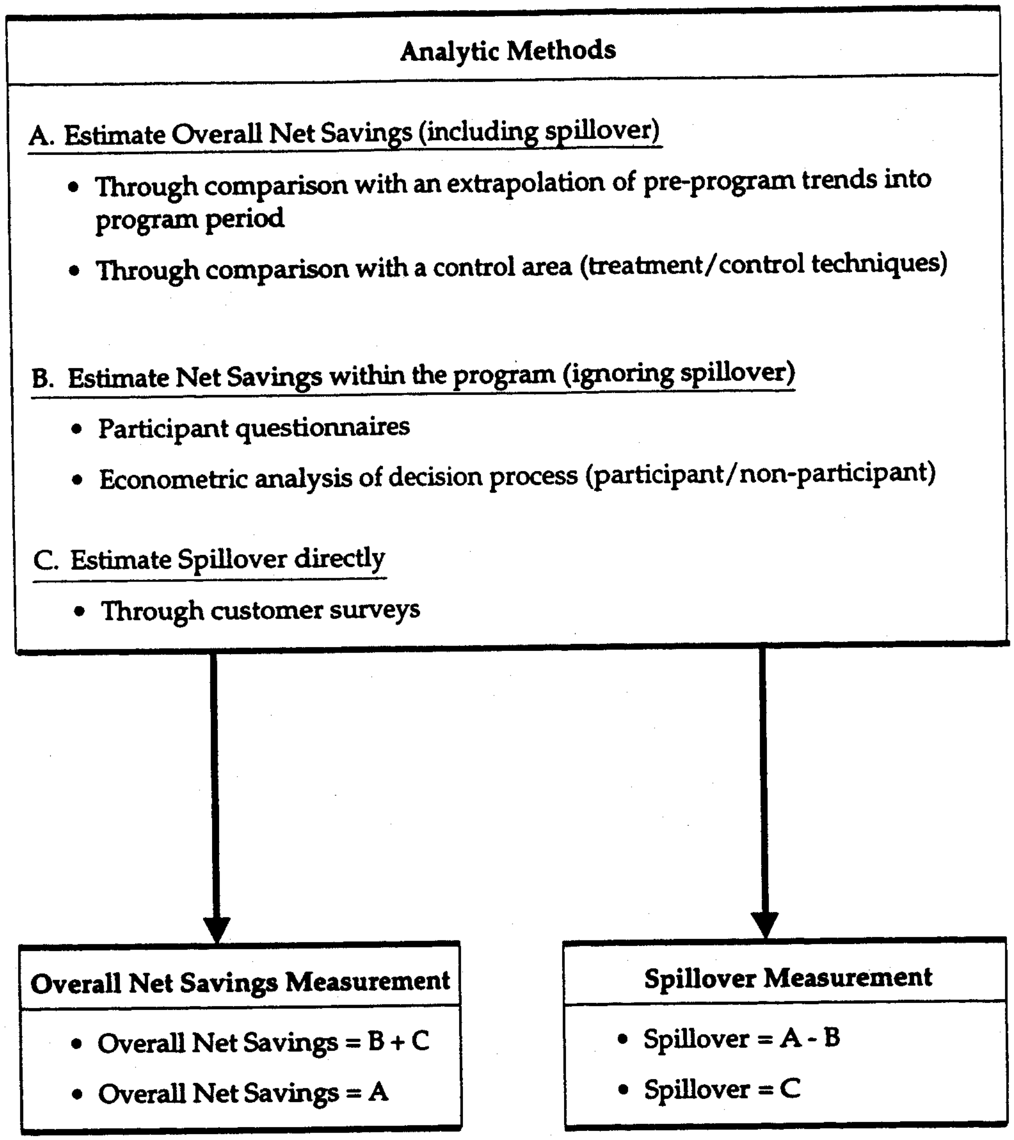


Figure 3.2 Sources of Data and Techniques for Estimating Spillover and Net Savings

Source of Data

1. Implementation Rates for Adoption of Energy-Saving Measures (Customer Surveys)
2. Energy Consumption Data (Billing/Meter Data)
3. Equipment Sales Data (Distributor/Dealer Data and Surveys)

Techniques for Overall Net and Spillover Savings

1. Regressions to Extrapolate Pre-program Trends
2. Comparison of Rates (Treatment/Control)
3. Discrete-Continuous Models (Treatment/Control) (Regressions with Discrete Choice Adjustment for Self-Selection Bias)
4. Discrete Choice Decision Models (Participation/Adoption of Measures)
5. Direct Survey Response Indices

extrapolation over time and another based on extrapolation over geographic areas. Each is considered below.

Extrapolation of Pre-Program Trends into Program Period

One procedure is to observe pre-program trends in penetration rates of energy efficient equipment, or trends in consumption levels, and to extrapolate these trends into the period after the program began. Under this approach, these extrapolations are considered to represent the penetration rates or consumption that would have occurred without the program. Comparison with the actual penetration rates or consumption that occurred with the program then provides an estimate of the impact of the program.

The easiest version of this approach is to observe the penetration of the relevant measures before the program and to assume that any change in the penetration after the program started is attributable to the program. That is, the penetration before the program began is assumed to be the naturally occurring penetration during the program period. For example, survey data can indicate that 10 percent of customers had energy efficient lighting before the program began. If the penetration rises to 12 percent during the program period, then the program is considered to have raised the penetration rate by two percent. Alternatively, customers who took the measure during the program period but had not previously taken the measure are assumed to have been induced to take the measure by the program, even if the measure was taken outside of the program. This procedure is similar to that employed by Horowitz and Spada (1992).

The difficulty with this approach, of course, is that it assumes that, without the program, customers would behave the same as before the program started. In many situations, however, customers might have taken the measure more readily during the program period than before even without the program. That is, the penetration might have risen even without the program due to changes in energy costs, equipment pricing and technology availability that have occurred in the meantime.

If it is expected that penetration would have risen anyway, without the program, then a method for predicting this rise is needed. There are several possible approaches. First, in addition to looking at the penetration rate before the program started, the analyst can also look at the annual implementation rate before the program started. That is, in addition to determining that 10 percent of customers had efficient lighting before the program began, the analyst can determine the proportion of customers who convert annually to efficient lighting each year before the program started. The analyst might find, for example, that one percent of customers converted annually to efficient lighting. This can be considered the naturally occurring annual implementation rate. Extrapolating this into the program period means that, without the program, penetration would be 11 percent in the first year of the program. If the actual penetration by the end of this year was 12 percent, the program is considered to have induced the extra one percent rise. Stated alternatively: if one percent of customers converted to efficient lighting in the year before the program began, then it is assumed that one percent would have converted in the first year after the program began, even if the program had not existed. If in fact two percent of customers converted in the first year after the program, then the extra one percent is attributed to the program.

More sophisticated extrapolation methods can be employed if data are obtained on the penetration or implementation rates for a number of years before the program began. The time trend in penetration or implementation rates can then be continued into the program period. Econometric methods can be used to account for changes in factors, such as prices, that might affect the pre-program rates. And variables can be included in the analysis to reflect awareness of the measure, or how long the measure has been available. In general, the pre-program penetration or implementation rate is modeled with time-series methods, such as:

$$R(t) = bx(t) + R(t-1) + e(t)$$

where $R(t)$ is the implementation rate in year t, $R(t-1)$ is the rate in the previous year, $x(t)$ is a set of variables that relate to the rate in year t, b is a set of parameters to be estimated, and $e(t)$ is an error that captures the effect of unincluded variables. This model is estimated on pre-program rates and then used to predict what the rate would be during the program period if the program had not existed. Essentially, the predicted rate in the program period is an extension of pre-program trends, adjusted for whatever explanatory variables are included in the model.

This approach – modeling pre-program trends with time-series methods and extending the trend into the program period – has not been applied in the literature reviewed here and seems worthy of consideration. The only evaluations that used pre-program data to predict what would have happened without the program have used only the time period immediately before the program began. By using data from several years before the program and examining the trend in these data, a better estimate can be obtained of what would have occurred in the program period if the program had not existed.

The limitations of this approach have been pointed out by several writers. First, the approach works better for technologies that have been available for a while, such that the pre-program data provide some meaningful information about customer adoption of the measure without the program. For new technologies, little, if any, pre-program data are available. Second, the method becomes less and less meaningful, the longer a program (or set of programs) has been in place. For example, in examining the effects of a 1992 program, data from 1991, 1990, etc., can be used. However, to examine the effects of a program that started on a limited basis in, say, 1989, and expanded over the period 1990 to 1992, data are required for 1988, 1987, etc. Data of such age might not be available; and, if it is, the difficulty of adjusting for the various non-program factors that have changed since 1988 becomes greater than when examining more recent time periods.

Comparison with a Control Area

In the classical statistical concept of "control/treatment comparisons," the treatment for overall net savings is the utility's implementation of the program. The relevant comparison is therefore between the area in which the program was implemented and an otherwise comparable area in which the program was not implemented. After accounting for the effects of other differences between the control and treatment areas, the remaining difference is attributable to the program. It reflects the complete effect of the program, and as such incorporates implicitly (or automatically) the effects of free ridership and spillover.

Several types of control/treatment comparison are possible, depending on which variable is being compared:

Comparison of Implementation Rates for Relevant Measures in Each Area

The easiest procedure is to observe the penetration or implementation rate for relevant measures in the control area and the treatment area. If there are no other differences between the two areas other than the program, then the difference in penetration or implementation rates are an estimate of the impact of the program. Generally, however, there will be other factors that differ across the two areas, and a method is needed that accounts for these non-program differences. A viable approach in this situation is to obtain data on a sample of customers in each of the two areas. (In the treatment area, it is important that that sample be of both participants and non-participants, since the goal is to estimate overall net savings which includes spillover.) For each sampled customer, information is collected on which measures the customer could have taken during the program period, the costs and savings associated with these measures, and which of the measures that customer actually took. Using these data, binary choice models (such as logit or probit) are estimated in which the dependent variable is whether the customer took the measure. Explanatory variable include the characteristics of the customer, the cost and savings of the measure, and a variable (perhaps a dummy variable) that identifies customers in the treatment area. The coefficient of this treatment variable captures the effect of the program on customers' decision to take the measure, after accounting for the effects of other variables in the model. Insofar as the other variables capture the differences between the control and treatment areas that are not due to the program, then this method provides a meaningful estimate of the impact of the program.

This procedure has been applied for PG&E's incentives programs in the commercial and industrial sectors (Cambridge Systematics, 1993). Results were reasonable, and conformed closely to results obtained with other methods. There are, however, important limitations of this method that will make it inappropriate in some settings. First, the data requirements are severe. Data need to be obtained on the measures that could have been taken by a sample of customers in each area, along with the costs and savings of these measures. Second, it is difficult in the modeling of customers' implementation choice to distinguish differences that are attributable to the program and those that are not. For example, if the average cost of installing a particular measure is different in the control area than the treatment area, is this difference because the program caused the costs to change in the treatment area, or is it because the customers in both areas have different characteristics, such as building configurations, that make the costs different without the program? The difficulty of making these distinctions arose in the PG&E evaluation and can be expected to arise in most control/treatment comparisons. Third, as DSM programs become more widespread, it becomes more difficult to identify a meaningful control area that has not had similar programs. This limitation is already fairly constraining; areas that have not had programs are usually so different in other ways also that meaningful comparison, controlling for these other differences, is difficult. Finally, the control/treatment comparison incorporates the impact of all the programs in the treatment area, and therefore cannot be used to estimate the individual effect of each program within a package of programs. For example, if a utility offers an audit program and a rebate program, the control/treatment comparison can potentially reveal the combined impact of the two programs, but not the separate effect of each.

Comparison of Change in Energy Consumption in Each Area

Rather than examining the implementation rates of customers, the analyst can compare the energy consumption of customers in each area, or the changes in consumption over time. After adjusting for non-program differences between the two areas, the remaining difference in consumption, or the difference in changes in consumption over time, can be attributed to the program. The key here, as in the analysis of implementation rates, is the adjustment for non-program factors. This adjustment can potentially be accomplished with regression analysis on the billing data for a sample of customers in each area. (Again, it is important to compare a sample of customers from the treatment area – including both participants and non-participants – with the sample from the control area. Simply comparing participants with customers from the control area is inappropriate and can be expected to result in large biases. (See Train, 1993.) The regression model takes as its dependent variable the customers' energy consumption, or change in consumption. Explanatory variables include customer characteristics and other factors, such as weather, that affect the customer's consumption, as well as a variable (perhaps a dummy) that identifies customers who are in the treatment area. The coefficient of this treatment variable captures the impact of the program.

The same limitations that apply to the control/treatment comparison of implementation rates also apply to the comparison of consumption, with one exception: unlike implementation data, billing data for a sample of customers is usually comparatively straightforward to obtain. The major potential difficulty is gaining the cooperation of the utility in the control area, so as to obtain the billing records for a sample of this utility's customers.

Comparison of Sales, or Installations, of Energy Efficient Equipment in Each Area

An approach that has been applied in several evaluations is to survey dealers and contractors in the control and treatment areas. In the survey, questions are asked that are designed to determine the growth in sales or installations of energy efficient equipment in each area. The difference in the growth in sales or installations between the two areas is taken as the impact of the program.

It is important to note that dealer/contractor surveys provide information on the overall impact of the program – including spillover <u>and</u> the net savings within the program. Some of the increased sales attributable to the program are due to participants who would not have taken the measure without the program. To obtain spillover per se, the sales to these participants must be subtracted from the total increase in sales attributable to the program. (That is, net savings within the program must be subtracted from the estimate of overall net savings to obtain an estimate of spillover.)

This approach has been applied to several programs, including Wisconsin Electric's dealer-incentive program (Brooker and Fitcher, 1991), BC Hydro's motor program (Nelson and Ternes, 1993), PG&E's refrigerator program (Van Liere et al., 1993), residential appliances (Van Liere et al., 1992), PG&E's commercial and industrial rebate program (Cambridge Systematics, 1993), and the commercial lighting programs of NYSEG (Freeman and Vinhage, 1993) and NY State ESEERCO utilities (Cambridge Systematics, 1994).

The limitations that apply to the other control/treatment comparisons also apply to the dealer/distributor analysis. In particular, it is difficult to find a meaningful control area, and the comparison provides an estimate of the combined impact of all the programs that the utility offers rather than each program separately.

There is an additional limitation that does not appear (or at least not as severely) in the comparison of customer implementation rates or consumption. Non-program factors that differ across the two areas might also affect the growth rates in sales and installation. In theory these should be incorporated into the analysis. In practice this has not, and in fact cannot, be done. The surveys essentially provide two data points: growth in sales in the study area and growth in sales in the treatment area. With only two observations, statistical analysis cannot be performed to determine and adjust for the effects of other factors.

■ 3.3 Estimation of Net Savings Within the Program (Ignoring Spillover)

The net savings within the program are the gross savings of the program minus the naturally occurring savings within the program, where the naturally occurring savings are the savings from measures that were counted in the gross savings of the program but would have been taken anyway even if the program had not existed. Essentially, estimation of net savings within the program consists of: (i) estimating the gross savings of measures counted under the program and (ii) estimating which of these measures were induced by the program and which would have been taken anyway.

The gross savings of measures taken in the program are estimated with billing data, with metered data, through engineering methods, or through some combination of these. We will not discuss these methods in any detail here, since they are well known and documented elsewhere. Essentially, with billing or metered data, gross savings of a measure are estimated by determining the decrease in consumption or load from before the measure was installed to after, controlling for other factors (such as changes in weather). Engineering methods derive, from engineering principles and data on operating hours and other factors, the decrease in consumption or load that is expected from the measure.

The greatest difficulty in estimating the net savings within a program, and the part of the analysis least understood, is how to determine which measures within the program would have been taken without the program – that is, estimating the naturally occurring savings within the program. In several evaluations, analysts have mistakenly used methods that they believed estimated the net savings within a program, but in fact are not actually designed for that purpose. Train (1993) describes these inappropriate procedures and why they should be avoided. So as not to dwell on the negative, we describe below only appropriate methods. Two such methods are available:

Questionnaires Applied to Participants

Perhaps the most common method for determining the net savings within the program is to survey a sample of participants and ask them whether they would have taken the measure without the program. This procedure is called "self-report." The savings from measures that customers say that would not have taken if the program had not existed are included in the net savings within the program, and the savings from other measures are not. Examples of self-report participant surveys include evaluations of Central Hudson commercial and industrial program (Applied Management Sciences, 1990), and the NY State appliance program (Saxonis, 1991).

Actually, questionnaires that are designed to determine what participants would have done without the program are usually considerably more complex than simply asking the customer directly. The impact of the program is determined in various ways, such as, through asking questions regarding how important was the rebate in the decision to install the measure, and when the customer heard about the program relative to the time the customer decided to install the measure. Consistency checks can also be included with this method. The analyst combines all the responses of a given participant and assigns the customer a score that represents the likelihood that the customer took the measure because of the program. The design of these types of surveys, and the methods for combining responses, has been studied extensively in the analysis of PG&E's incentives programs (Cambridge Systematics, 1993) and New York utilities lighting programs (Cambridge Systematics, 1994).

Self-report data have important, and well-known, limitations. Essentially, customers might not be able to accurately report what they would have done if the program had not existed, or they might not be willing to state what they would have done. This standard problem is greater if the program has moved the market, making measures available through contractors and dealers that would not have otherwise been available. In this case, the customer might think that it would have taken the measure without the program, not realizing that the measure would not be carried by dealers or handled by contractors if the program had not existed.

Econometric Analysis of Decision Process of Customers

Customers who are offered a program have a choice of whether to participate in the program and whether to install measures. Sometimes participation is only possible if a program measure is taken; for example, for most rebate programs, a participant necessarily installed at least one rebate-eligible measure. For other programs, like audit programs, participation is possible without installing a measure. Regardless of the type of program, the two choices can be interrelated. A customer might install a measure because of the rebate, or it might apply for the rebate since it is installing the measure (and would have installed the measure even if the rebate had not been available).

Models of the choice process of customers can be estimated on a sample of customers who were offered the program. For each measure and each customer, data are obtained on whether the customer could have taken the measure, the cost and savings of the measure, whether the customer took the measure (either within the program or outside of it), and whether the customer participated in the program. Discrete choice modeling methods are used to describe the interrelated choice of whether a customer implements the measure and whether the customer participates in the program. Once the model is estimated, it is used to forecast what the customer would have done if the program had not been available. If a customer who took a measure under the program is forecast not to take the measure without the program, then the savings of the measure is counted in the net savings of the program; otherwise the savings is counted in the naturally occurring savings.

This method is similar to the modeling analysis described in Section 3.2, in that both types of analysis utilize discrete choice models to estimate whether the customer took particular measures, with the cost and savings of the measures entering as explanatory variables. The difference, however, is critical. The analysis in Section 3.2 is a "control/treatment comparison," in which the observations consist of customers who were offered the program and a sample of customers from a control area. The critical explanatory variable is a dummy that identifies the customers in the area with the program. The estimated coefficient of this variable captures the overall impact of the program, including spillover, since the comparison is between customers in the area with the program and customers in an area without the program. By contrast, the analysis described in the current section, is a "participant/non-participant comparison," in which the sample consists only of customers in the area offered the program and not customers from a control area. The analysis essentially compares participants with non-participants, accounting for the fact that participation is endogenous – that is, be caused by the decision to take a measure. Since the comparison is between participants and non-participants, the analysis necessarily excludes spillover, since spillover affects both participants and non-participants.

This method has been applied for Niagara Mohawk's residential program (Regional Economic Research, 1991), SCE's Commercial Rebate Program (Pacific Consulting Services, 1993), PG&E incentives programs, and ESEERCO's lighting programs (Cambridge Systematics, 1994). It has been found to provide reasonable estimates that conform closely to estimates for the same programs found with other methods.

There are several potential difficulties with the method. First, the data requirements are somewhat stringent. In particular, for a sample of participants and non-participants, data are needed on which measures the customer could have taken during the program period, the cost and savings of these measures, whether the customer took each measure, and whether the customer participated in the program. These data can be obtained through on-site visits by qualified energy engineers (as was done for the PG&E analysis) or through telephone surveys with building owners or managers (as done for the SCE and ESEERCO analyses). In either case, data collection is expensive and time consuming. Second, it can be difficult for the analyst to estimate a model that disentangles the directions of causation, correctly identifying whether participation in the program caused the customer to take the measure or whether the customer's decision to take the measure induced the customer to participate in the program. However, it is necessary to disentangle these two directions of causation order to estimate the net savings within the program.

■ 3.4 Direct Estimation of Spillover Effects

As discussed above, spillover can be estimated by using a method from Section 3.2 to estimate overall net savings, using a method from Section 3.3 to estimate net savings within the program, and subtracting to obtain the difference between the two. This is the procedure that was used in the evaluation of PG&E's incentives programs for the commercial, industrial, and agricultural sectors. In particular, the net-to-gross ratio for the program was found to be around 0.75 with methods that accounted for spillover and to be 0.65 for methods that do not include spillover (that is, that look at the net savings within the program only), meaning that spillover effects constitute a 0.10 contribution to the net-to-gross ratio.

A method exists for estimating spillover directly. This method is useful if the analyst has a estimate of net savings within the program to which he/she wants to add spillover, to obtain a direct estimate of overall net savings. However, as discussed below, the method is difficult and subject to potentially large errors. The analyst needs to decide, if the goal is to estimate overall net savings, whether it is preferable to estimate overall net savings directly with a method from section (a), rather than estimating spillover directly and adding it to an estimate of net savings within the program to obtain overall net savings.

Questionnaires Applied to Participants and Non-Participants

In standard self-report studies (as described in Section 3.3 above), participants are asked whether they would have taken the program measure even if the program had not existed. This procedure can be expanded to estimate spillover. In particular, participants can be asked about measures that they took outside the program, and non-participants can be asked about the measures that they took. In each case, the questions can be designed to determine whether the customer would have taken the measure if the program had not existed. For example, participants might be asked whether they took measures outside of the program because of the success that they experienced with the measure that they took in the program. Non-participants might be asked whether they heard about the measures though program advertising.

Self-report regarding spillover is even more problematic than self-report for net savings within the program. Customers are being asked to explain the effect of the program on their decision to install measures outside the program – which, in itself, is a fairly subtle concept. Furthermore, a large component of spillover might consist of the program moving the market. Customers generally would not know the extent to which the options available to them are attributable to the program; they might therefore state that measures would have been taken without the program when in fact they would not have been. Their responses would thus tend to underestimate spillover. The responses could, however, provide a lower bound to the size of spillover effects.

There are studies that have asked participants and non-participants to report non-program measures that they have taken on their own. An example in the evaluation of Puget Power's commercial rebate program (Cambridge Systematics, 1993). To the best of our knowledge, though, no one has estimated spillover through self-reporting of whether those non-program measures would have been taken if the program had not existed. However, several writers have mentioned it as a possibility. Trying it in a particular setting would seem worthwhile.

Comparison of Implementation Rates of Non-Program Measures With a Control Area

It is possible to survey participants and non-participants in the utility service area about implementation of non-program measures (as above) and then compare the findings with those of a parallel survey conducted in an otherwise comparable area where the program was not implemented. This essentially follows the same logic and approach as presented for analysis of overall net savings in Section 3.2, except that the analysis is restricted to just non-program measures. This approach has three problems. First, it is difficult to distinguish whether differences in rates of implementation between the two areas are due to DSM program spillover effects or due to other differences between the two areas. Generally such comparison requires a complex decision choice model, incorporating differences in energy and equipment costs, in order to control for any such differences. Second, it is becoming increasingly difficult to find appropriate control areas where no DSM programs or energy awareness campaigns have ever been implemented. Finally, such comparison omits changes in implementation of program measures among non-participants, and hence only addresses a portion of potential spillover effects.

Questionnaires Applied to Involved Parties

As noted previously, customers are not always aware that they are implementing energy-saving measures that they otherwise would not have done. This is most obviously the case when there is a non-participant who is unaware of the DSM program and is not aware that retailers' offerings or contractors' suggestions have been affected by a DSM program. In that situation, the customer is not able to reliably self-report incidence of spillover. To address this issue, DSM evaluations can include surveys or interviews of manufacturers, distributors, dealers, contractors and other trade allies. These surveys can ask those other parties who are directly affected by the DSM program to report on how they have changed their product lines, stocking, sales and/or installation patterns in response to the DSM program. These results can then be used to infer the indirect effect on implementation of energy-saving measures by customers, who themselves may not even be aware of that effect. Examples include Saxonis (1991). This approach can be a useful adjunct to customer surveys, and a way to confirm the existence of this type of spillover effect when it has been so indicated by other methods.

There are two potential problems with this approach. The first one is the danger of double-counting. Estimates of spillover impacts obtained from trade ally surveys and estimates of spillover obtained from customer data may be compared but not added, insofar as that would be double-counting of impacts. The second problem is the difficulty of sorting out what portion of the changes in dealer or contractor sales (for program-eligible measures) are attributable to the actions of program participants and what portion are attributable to the actions of non-participants. Unless the survey of trade allies is comprehensive and sufficiently detailed to extrapolate up to area-wide patterns, it may be difficult to accurately separate program impacts from non-program impacts on sales patterns. In addition, dealers and contractors located in many utility service areas also sell to customers of adjacent utility areas. In such cases, the determination of what portion of their sales changes are due to a particular DSM program becomes even more indeterminate.

Summary

In the above discussion of methods, citations are given for studies that applied each method. As a summary, we tabulate in Table 3.1 the number of studies, of those reviewed for this report, that utilized each method. This categorization is, by necessity, somewhat arbitrary, since the approaches used by some studies contained elements of two or more methods. However, the table nevertheless provides some useful guidance.

Two implications are clear. First, manufacturer/dealer surveys have been the most popular method by far. Second, very few studies have actually applied any method to capture spillover. Of the 38 studies cited in the literature review of Section 4.0, only 11 actually estimated spillover or overall net savings with spillover. The remainder discussed issues and described methods without applying any methods for empirical testing. Since the proof of any method is in its application, the need for more applications is evident.

Table 3.1 Number of Cited Studies Using Each Method

Method	Number of Studies
Estimation of Overall Net Savings Including Spillover	
Extrapolation of pre-program behavior of customers into program period	1
Comparison with control area of customer's implementation rates	1
Comparison with a control area of aggregate sales data from manufacturers, distributors, and/or dealers	6
Estimation of Spillover Itself	
By estimating overall net savings (including spillover) and subtracting out an estimate of net savings within the program (which excludes spillover)	2
Survey of non-participants	0
Comparison with a control area of customer's implementation of non-program measures	0
Survey of dealers and manufacturer to assess impact of program on sales outside of program	1

Source: Literature Review (Section 4.0).

4.0 Literature Review

4.0 Literature Review

■ 4.1 Definitions of Spillover Effects Used in Other Energy Program Studies

Free Driver Concept

Most of the recent literature regarding spillover has focused on "free driver" behavior. This generally refers to program impacts on energy saving by non-participants, in a manner consistent with the concept defined in Section 2.1. Beyond that simple concept, though, various analysts have each suggested different sets of additional restrictions on who should be accepted as a free driver. These various restrictions include the following:

- Customer must have knowledge of the existence of the program;
- Measure must meet program requirements for a rebate;
- Measure must contribute to the goals of the program; and/or
- Customer must be a non-participant (in the control group).

For each of these restrictive additions to the definition of a free driver, there are some analysts who would include it in a definition of "free driver" and there are some who would not. Some other analysts would include none of these restrictions. In general, the purpose of such restrictions is not necessarily to deny the possibility that DSM programs can have broader spillover effects, but rather to focus on narrower study issues.

There has also been a clear evolution in the concept of a free driver. Until recently, free drivers were seen primarily as a source of bias in control groups for estimating free ridership. Only recently has the concept arisen that free drivers may be considered as an added benefit.

A selection of examples of free driver definitions offered in print during the 1991-1993 period is provided here. (These definitions are organized alphabetically by author. Further document citations follow in Section 4.3.)

1. Horowitz, Spada in "Energy Savings for Residential Lighting Programs: There's More to It Than Just Counting Lamps," ACEEE Conference Proceedings, 1992.

   Free drivers are those households that knew of the existence of the program but purchased qualifying light lamps on their own, without benefit of rebate.

2. Kushler, Keating, Schlegal, Vine in "The Purpose, Practice, and Profession of DSM Evaluation: Current Trends, Future Challenges," 1992 ACEEE Conference Proceedings.

   A free driver contributes to the goals of the program (e.g., reduced energy consumption) but is not formally a program participant (Saxonis 1991). A free driver is affected by the program either through a conscious awareness of the program or because of program-induced changes in the marketplace (e.g., a customer who purchases a product that qualifies for a rebate but does not claim the rebate, or a builder who constructs a home to program standards but does not choose to participate in the program).

3. Buller, Miller in "How Should We Treat Factors Contributing to Uncertainty in Measurement and Evaluation of DSM?," 1992 ACEEE Conference Proceedings.

   Free drivers are people who installed energy efficiency measures without participating in the program, but who were influenced by the program's existence.

4. "ADSMP Evaluation Workshop: Experimental Designs and Techniques" workshop guide presented at the International Energy Program Evaluation Conference, 1991.

   Free drivership is a type of bias within the control group. If there are free drivers, the control group is 'contaminated,' i.e., it is not a true control group.

5. RCG/Hagler Bailly, "Impact Evaluation of Demand-Side Management Programs," Volume 2: Case Studies and Applications, EPRI, 1991.

   Free driver savings is the decrease in energy consumption from customers who do not participate in the program, but where the utility's DSM program can be viewed as causing the savings.

   Moving the market is when the behavior of an entire customer segment, both program participants and non-participants, is influenced by a utility's DSM program, resulting in improved efficiency for the entire sector.

6. RCG/Hagler Bailly, "Impact Evaluation of Demand-Side Management Programs," Volume 2: Case Studies and Applications, EPRI, 1992.

   Free drivers, or moving the market, relates to the DSM program indirectly causing conservation in the control group due to the market responding to the program by shifting towards energy efficient technology. As with free riders, free drivers also involve a potential bias within the control group. In this case, the issue involves the potential for having a 'contaminated' control group.

7. RCG/Hagler, Bailly, "DSM Process Evaluation: A Guidebook to Current Practice," EPRI, 1992.

   Free drivers are DSM actions undertaken by non-participants due to the existence of the program.

8. Hirst, Sabo in ORNL's "Electric-Utility DSM Programs: Terminology and Reporting Formats," 1991.

   Free drivers are customers who adopt measures as a result of the program's influence, but who do not directly participate (e.g., they do not apply for rebates even though the appliance they purchased qualifies for the utility rebate). They are not normally identified at the time of purchase and are not treated as participants. ...The savings or load reductions these customers experience should be credited to the program and included in net savings. DSM programs that aim to transform the market implicitly seek to make everyone a free driver.

9. Keating, "Persistence of Energy Savings," in ORNL's Handbook of Evaluation of Utility DSM Programs, 1991.

   Free driver effect is a spillover effect of the program when the comparison group takes action because of the program. Free drivers are customers who adopt program-recommended actions without participating officially in the program. They may take action because:

   - The program changed the stock of equipment available;
   - The program changed perceptions among customers about the usefulness of conservation;
   - They know about the program, but do not want the hassles of formal participation;
   - The service industry they deal with is now interested in efficiency; or
   - Competitive pressures from participants have influenced them.

Free drivers are certain to accompany programs that aim to change construction practices or the basic marketplace. The dilemma is that free drivers are interpreted as free riders when billing data are used to estimate net program savings.

10. Saxonis, in chapter "Free Riders and Other Factors that Affect Net Program Impacts" in ORNL's Evaluation Handbook, 1991.

    A free driver contributes to the goals of the program (e.g., reduce energy consumption) but is not formally a program participant. A free driver is affected by the program either through a conscious awareness of the program or because of program-induced changes in the marketplace. An example of a free driver is a customer who purchases a product that qualifies for a rebate but does not claim the rebate.

11. Saxonis, in "Measuring Free Riders: Does the Economic Climate Make a Difference?," 1991 Energy Program Evaluation Conference Proceedings.

    A free driver is a person who was influenced to take an action by a program, but not identified as a program participant.

12. Violette, Ozog, Wear in "Measuring Free Riders: Do Some Experimental Designs Control Twice for Free Ridership?," 1991 Energy Program Evaluation Conference Proceedings.

    The free driver effect is when the program has "moved the market." ...The program is inducing some "non-participants" into taking conservation actions.

13. "Free Riders, Free Drivers – Who Needs Them?," Evaluation Exchange, Vol. 1, No. 1, July/August 1991, interview with Cummings, Quigley, and Goett.

    Quigley – Free driver is a customer who takes an action as a result of a utility program, but without participating in the utility's program.

    Goett... recommends excluding from free drivers the effects of the adoption of ineligible measures by program participants. Programs often have an effect of inducing customers to buy additional measures outside the program.

    Cummings – The definition starts getting fuzzy around the edges if you start pursuing it. Free driver customers are official non-participants, they have in effect been influenced in some way by the utility program. It's important to distinguish between the free drivers and market-induced conservation.

14. Vine, in "Social Dimensions of Program Evaluation," 1993 Energy Program Evaluation Conference Proceedings.

    A free driver contributes to the goals of the program but is not formally a program participant. A free driver is affected by the program either through conscious awareness of the program or because of program-induced changes in the marketplace.

15. Nelson, in "Phantoms of the Program: In Search of Measurable Free Riders and Free Drivers," 1993 Energy Program Evaluation Conference Proceedings.

    A free driver is a non-participant who acquired the technology or energy conservation measure and did not receive an incentive, but was influenced by the program.

    Free drivers include legislated savings where the DSM program has directly influenced the timing of energy efficiency legislation.

Net Impact and Spillover Concepts

In contrast to the growing literature on "free drivers," there has been very little mention of the term "spillover" in published articles and conference proceedings. When the term "spillover" does appear, it is sometimes (but not always) equated with "free drivers." There has only recently been published acknowledgement that participants too may engage in additional energy savings that is outside of the DSM program and yet is initiated as a result of the program. Most DSM impact evaluations continue to limit analysis of net impacts to "within program" effects. Recent examples of definitions of these concepts are provided below:

A. Kitchin, in "The Impact of Market Transformation on DSM Evaluation Techniques," 1993 Energy Program Evaluation Conference Proceedings.

   "Free driver" or "spillover" effect is when customers who are non-participants are affected by the program and implement some efficiency measures because of the impact of the program on the market.

   Three sources of the difference between gross and net savings include: ...'free rider' effect... 'free driver' or 'spillover' effect... and 'take-back.' Unfortunately, free drivers or spillover effects have often been ignored because of difficulty in measuring their impacts.

B. Megdal, in "Estimating Take-back For a Low-Income Program, a Loan Program, and a Single Family Rebate Program," 1993 Energy Program Evaluation Conference Proceedings.

Spillover is additional energy savings from participants and non-participants outside of the program but due to the program. This includes additional actions taken by participants, non-participant actions taken due to program marketing impacts on awareness of energy efficiency, and from non-participants who purchase greater efficiency than they otherwise would have due to differences in dealer and contractor actions. The latter is sometimes referred to as savings induced by "free drivers," or market transformation.

C. RCG/Hagler, Bailly, "Impact Evaluation of Demand-Side Management Programs," Vol. 1: A Guide to Current Practice, (Electric Power Research Institute, 1991).

Net impact is "the change in participants' electricity consumption which is directly attributable to the program."

D. Vine, in "Social Dimensions of Program Evaluation," 1993 Energy Program Evaluation Conference Proceedings.

Net savings are the difference between gross savings and the change in consumption and demand that participants would have achieved had the program not existed.

E. Eckman, Benner and Gordon, "It's 2002: Do You Know Where Your Demand-Side Management Policies and Programs Are?," Proceedings of ACEEE 1992 Summer Study on Energy Efficiency in Buildings, Asilomar, CA.

"In some cases... conservation actions taken by non-participants (free drivers) should count as additional program benefits, not as a reduction in program impacts."

F. Kushler, Keating, Schlegal, Vine in "The Purpose, Practice, and Profession of DSM Evaluation: Current Trends, Future Challenges," 1992 ACEEE Conference Proceedings.

..."Clear definitions of what is increased in 'net savings' or 'net benefits' will be needed. In particular, the question of how (or whether) to account for the benefits due to market transformations resulting from utility DSM programs needs to be addressed."

G. Vine in "Persistence of Energy Savings: What Do We Know and How Can It Be Ensured?," 1992 ACEEE Conference Proceedings.

"Surge effect" is when program participants add additional energy efficiency measures after initial participation in the program.

H. Prahl, Schlegel in "Evaluating Market Transformation," 1993 Energy Program Evaluation Conference Proceedings.

Market transformation occurs when DSM programs induce a lasting, beneficial change in the behavior of some group of actors within a market system.

I. Train, "Estimation of Net Savings from Energy Conservation Programs," University of California, Berkeley, 1993; forthcoming in Energy 1994.

Net savings are the savings from measures that would not have been installed if the program had not been offered. Net savings differs from gross savings in two ways. First, measures that were implemented under the program might have been implemented even if the program had not been offered... The savings from these measures is called 'naturally occurring savings within the program'... Second, measures that were implemented outside the program might have been due to the program in the sense that they would not have been implemented without the program... The savings from measures such as these are called 'spillover effects.'

J. Train, "Estimation of Net Savings from Energy Conservation Programs," University of California Berkeley, 1993; forthcoming in Energy 1994.

Net savings are the savings from measures that would not have been installed if the program had not been offered. Net savings differs from gross savings in two ways. First, measures that were implemented under the program might have been implemented even if the program had not been offered... The savings from these measures is called the 'naturally occurring savings within the program'... Second, measures that were implemented outside the program might have been due to the program in the sense that they would not have been implemented without the program... The savings from measures such as these are called 'spillover effects.'

K. Xenergy, Inc., Evaluation of the CIA Retrofit Rebate Program, Final Report, Pacific Gas & Electric Co., San Francisco, 1993.

The net-to-gross ratio is the ratio of net program benefit to gross program benefit. Essentially, the "net savings" attributable to the program is calculated as the "gross savings" attributable to the program measures, minus "free ridership" (i.e., naturally occurring conservation) plus "spillover benefits" (effects of induced actions taken outside of the formal program).

■ 4.2 Summaries of Selected Energy Program Studies Relating to Net Impacts and Spillover

Applied Management Sciences, "Central Hudson Gas & Electric's Commercial and Industrial Incentive Program Final Evaluation," 1990.

This program evaluation included a survey of participating customers, which asked whether they had intended to take the program-eligible conservation actions before or after they had joined the program. The study results indicated that, because of the various influences of the program on trade allies, customers were not always able to reliably know whether or not the program affected what they would have done. The report notes that a contractor or dealer who is aware of the DSM incentives may persuade a customer to invest in energy efficient equipment. Yet that customer may think that he had intended to invest in efficient equipment from the beginning because that was the only type of equipment the dealer tried to sell him. In such cases, a self-reported free rider may not really be a free rider.

Brooker and Fichter, "Measuring Program Impact in a Commercial and Industrial Lighting Program: The Case of Wisconsin Electric's Smart Money Program," Proceedings of the 1991 International Energy Evaluation Conference, 1991.

In order to evaluate net impacts of Wisconsin Electric's commercial and industrial lighting program, information was collected on sales of energy efficient and standard efficiency fluorescent lamps and ballasts over the 1986 to 1990 period. Data was collected from dealers in Milwaukee and a selected comparison area with similar population and electric rates which had no incentive programs – Cincinnati. The study found that sales of efficient lighting remained roughly constant in Cincinnati, while an increase occurred in Milwaukee during the program period. This difference was attributed to the program.

Battles, S. and Thompson, W., "Are Participants of Demand-Side Management Programs Different?" Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, (September 1992), Asilomar, CA.

The authors outlined some systematic differences between participants and non-participants of utility sponsored DSM programs. Results were from a sample of 5,095 households under the 1990 Residential Energy Conservation Survey (RECS). The survey covered housing unit and household characteristics, as well as billing data. Energy suppliers were also surveyed. The article reported that, by 1990, only five percent of the U.S. households had reported participating in a DSM program. It was reported that DSM participants tend to be owners rather than renters, more affluent, better educated, and older. Participants are also reported to be more likely to undertake conservation actions on their own. The implication of this analysis is that non-participants are not a good control group for determining how participants behave in the absence of the program.

Buller, S. and Miller, W., "How Should We Treat Factors Contributing to Uncertainty in Measurement and Evaluation of DSM?" Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, (September 1992), Asilomar, CA.

The authors explain that in California a "net-to-gross" ratio is a way to account for 'what would have happened without the program.' Currently, this analysis has focused on free riders, but in the future free drivers, measure persistence, and changes in patterns of consumption must also be included. Without a "time machine," other methods must be developed to measure these other factors affecting the net-to-gross ratio.

Free drivers are defined as people who install energy efficient measures without participating in a program but were influenced by the program's existence. For example, in PG&E's service territory, nearly all floor models of refrigerators carried by retailers qualify for a PG&E rebate. Therefore, nearly all customers will buy an energy efficient model, regardless of whether they receive a rebate.

The authors then briefly discuss the advantages and disadvantage of metering and billing analysis, engineering estimates, and customer surveys in measuring factors affecting the "net-to-gross" ratio. Customer surveys can be used to measure free drivers, although not without some difficulty. Customers do not know how a program affected stocking patterns, and energy efficient equipment might not have been available without the program. Customers may not hear about the program and be influenced by it. Another important point made is that some participants identified as free riders might not have been able to purchase the equipment without the program moving the market. These participants would therefore not be free riders. Identifying these customers as free riders, as well as not properly identifying free drivers, underestimate net program impacts.

Cambridge Systematics, Net to Gross Ratios for PG&E's CIA Rebate Program: Study A – Participant Survey, Study B – Manufacturer, Vendor and Contractor Survey, Study C – Treatment/Control Comparison, and Study D – Decision Analysis Model, Pacific Gas & Electric Co., San Francisco, 1993.

This series of four net-to-gross studies explored four approaches for estimating net impacts of PG&E's Commercial, Industrial and Agricultural Rebate program. Studies A and D focused on measuring free ridership. The participant survey (Study A) measured free ridership by comparing a variety of different types of survey questions and methods for combining survey information. The decisions analysis model (Study D) utilized a nested logit model framework to estimate the relationship between measure adoption and program participation, in a dataset comparing participants and non-participants.

Studies B and C both utilized treatment/control approaches to estimate overall net impacts, including spillover impacts. Spillover is defined in Study B as "the additional energy-savings behavior induced in participants and/or non-participants as a direct or indirect result of the program." The trade ally survey (Study B) examined trends over time in sales of energy efficient lighting, motors, refrigeration and HVAC, comparing the PG&E territory with a non-DSM control area (which was Birmingham, AL for lighting measures and Reno, NV for other measures). The study concluded that this approach could be used to indicate net impacts, but that additional assumptions were necessary to develop a separate estimate of spillover.

Study C utilized a discrete choice model methodology, employing nested logit, to compare the interrelationships of program offer, program participation and measure adoption. It utilized on-site survey data on PG&E program participants, non-participating PG&E customers and a control group of non-PG&E electric customers. This information was used together with engineering models to develop estimates of the potential costs, potential rebates and potential energy-saving benefits of installed and non-installed measures. The discrete choice model was then developed to estimate overall net-to-gross impact ratios for different types of measures. The study concluded that there was indeed evidence of spillover benefits.

Cambridge Systematics, CEMS Program Evaluation: Task 2: Billing Analysis and Task 8: Telephone Survey Findings, Puget Sound Power & Light, Bellingham, WA, 1993.

These reports examined spillover effects and net impacts of Puget Power's "Commercial Energy Management Services" program, which is a customized rebate program. The Task 8 report used a telephone survey to examine differences in rates of adoption of program and non-program measures among: (1) participants in Puget Power's commercial retrofit rebate program, (2) non-participants receiving a Puget Power audit, and (3) non-participants who received neither an audit nor any rebate from Puget Power. Respondents were asked about rates of adoption of all types of energy-saving measures, distinguishing those that were covered by the rebate program, those that were recommended by auditors although not covered by rebates, and other non-program measures. It was concluded that energy-saving measures that were implemented but not rebated could potentially be a "spillover benefit," regardless of whether or not the applicant subsequently joined the program to receive rebates on other types of measures.

The Task 2 report compared results of econometric analyses of net and gross energy-savings impact. A discrete-continuous model was developed to estimate net program impacts, by combining a binary logit participation model with a participant/non-participant regression model of energy savings. A regression model of energy consumption was developed to estimate gross program impacts. The results of the two models were found to yield very similar findings of energy-savings impacts, implying that the free ridership found in customer surveys was being countered by additional spillover benefits.

Cambridge Systematics, ESEERCO Free Ridership Study: Final Report, Empire State Electric Energy Research Corporation, New York, 1994.

For a consortium of eight electric utilities, this study examined three alternative approaches for estimating net-to-gross ratios for DSM programs. The three approaches were: (1) direct participant surveys, (2) manufacturer and distributor surveys and (3) discrete choice models.

The manufacturer and distributor surveys examined trends in sales of energy efficient lighting, motors and HVAC, comparing New York State trends with national trends. This could be done on a statewide basis because all of the major electric utilities of NY State had initiated equipment rebate programs at roughly the same time. The surveys also asked questions about the extent to which utility DSM programs were affecting manufacturer and distributor behavior. The survey showed some evidence that sales of energy efficient equipment was growing faster in NY State than nationally, and faster than could be explained by program participants alone. Additional assumptions were necessary, however, to isolate free ridership and spillover effects.

The other two approaches were focused exclusively on fluorescent lighting. The participant surveys asked for self reporting of free ridership, with additional consistency checks. The discrete choice model estimated differences in technology adoption between participants and non-participants, yielding estimates of free ridership. Overall, it was concluded that all three methods can be useful, but that trade ally surveys alone were insufficient. The appropriate selection of survey or modeling approaches depends on the size, breadth and type of DSM program.

Eckman, T., Benner, N., and Gordon, F., "It's 2002: Do You Know Where Your Demand-Side Management Policies and Programs Are?" Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, (September 1992), Asilomar, CA.

The authors provided an overview of the dynamics influencing the nation's use of energy. For two decades, states, local governments, utilities, the private sector, and the federal government have all taken action to affect energy consumption. Government actions have included weatherization of homes of low-income citizens, tax credits for conservation, and revised building codes and appliance efficiency standards. Utilities have conducted consumer education, energy audits, and rebates for purchasing energy efficient equipment. Private sector actions included development of national energy efficiency building standards, the expansion of the energy service industry, and the development of more efficient products. Understanding these actions and dynamics is necessary to comprehend how the market has moved over time, both from utility and non-utility initiatives. Understanding these factors helps determine program and non-program effects on individual behavior and presence of free drivers.

Guidelines were presented for the evaluation of the next generation of DSM programs. It was noted that indices for monitoring conservation program progress should be consistent with those used to establish conservation targets, i.e., if the target is stated in terms of the gross penetration achieved by a specific technology, then progress should be based on the total penetration achieved by both program participants and non-participants. In some cases, free riders may be more appropriately viewed as "early adopters," and conservation actions taken by non-participants (free drivers) should count as additional program benefits, not as a reduction in program impacts.

"Free Riders, Free Drivers – Who Needs Them?" A discussion with Martin Cummings, Dan Quigley, and Andy Goett, Evaluation Exchange, Vol. 1, No. 1, (July/August 1991).

In this discussion the interviewees discussed a number of issues regarding the definition and measurement of free riders and free drivers. For free drivers, some indicated that they would prefer to use a term other than free drivers, such as one involving "net-to-gross" used in the California collaborative. Also, Goett would not include the adoption of ineligible measures by participants as free drivers effects. Cummings noted that it is important to separate free driver effects from market-induced conservation. As programs shift the market, doing this is quite difficult. He wrote the "classic" example of free driver effects was a commercial/industrial customer that participates in an audit program, installs some quick payback, low rebate measures, but does not claim a rebate.

Measurement of these effects is quite difficult, and some solutions may need to be developed at the program design stage, according to Goett. Before implementing a program a utility could measure pre-program penetration rates for technologies to use as a basis of comparison. Goett also wrote that properly designed survey questions can limit the problems with self-reported free ridership rates. Screening questions help limit the biases from these questions.

Quigley wrote that he had been trying to use sales data to estimate free ridership. The problems he has encountered has been finding a control area and collecting all of the data. He has attempted to work with national trade associations that have some of this data, but indicated that he had not had much success with these efforts.

Quigley also discussed PG&E's refrigerator rebate program that has been running the past 10 years through trade allies. In order to measure free ridership, they have to not only measure the customer's decision-making process but also the entire distribution chain. Perceptions of these broad, market effects have gone into the negotiation of free ridership figures.

When asked about the certain market sectors being more susceptible to free drivers than others, Quigley responded that he did not think this was true. Instead, he saw it as a function of technology. Free drivership is probably higher with newer technologies, in his view.

Freeman, L. and Vinhage, W., "Investing in DSM Lighting Programs: Are Dealers' Stocks Up?" Proceedings of the 6th National Demand-Side Management Conference, (March 1993), Miami Beach, FL.

The authors examined the extent to which commercial equipment dealers have been affected by utility DSM rebate programs. The program evaluated was NYSEG's MaxiMiser program for commercial and industrial customers. Dealers in NYSEG's service territory and those in Pennsylvania, where no utility rebate programs were offered, were compared. The findings from a telephone survey indicated that dealers in Pennsylvania were as knowledgeable, or more so, about rebated equipment, as were NYSEG dealers. Also, the Pennsylvania dealers were more likely to have qualifying lighting measures in stock than were the NYSEG dealers. Two explanations were that the utility programs had no effect on the market (thus reducing any free driver effects) or that the programs have strong spillover effects between states within the same regional market.

Fryer, L. and Stone, C., "Establishing Baseline Practices in the Industrial and Commercial Motor Market: Findings from the New England Motor Baseline Study," Proceedings for the 6th National Demand-Side Management Conference, (March 1993), Miami Beach, FL.

The authors examine how existing market behavior in New England can be used to promote energy efficiency in the market for electric motors. Most motor sales are for failed equipment, so distributors must stock efficient motors in order to substantially increase sales of efficient motors. The study was conducted by New England's three largest utilities to establish a baseline for motor DSM program design and evaluation.

As part of this study the team collected 1991 sales data from manufacturers that represent approximately 90 percent of motors sold in four New England states. They were successful in conducting in-person interviews with 12 of the country's 14 major manufacturers or importers of motors. Interviews were also conducted with 50 motor distributors. These interviews helped establish relationships so that the interviewees would provide sales data.

The distribution structure for industrial motors sold in New England was also depicted. Knowledge of this structure and the baseline developed will help the utilities measure free riders and drivers in future years.

Grott, R. and Hessen, J., "Customized Incentives for Industrial DSM Projects," Proceedings of the 1993 International Energy Program Evaluation Conference, (August 1993), Chicago, IL.

This paper evaluated BC Hydro's Bonus Partners Program for industrial customers in which they implement their own energy-savings ideas. The authors provide a warning for this and other programs regarding free drivers and free riders. They fear that people at the utility may be motivated to give customers a rebate in order to "get them on the books," thus encouraging free riders. If the customers do not receive a rebate and install the measures, they are free drivers (having received an audit and recommendations). However, the program does not get credit for the savings from these customers. The authors suggest that documented free drivers are acknowledged and credited to the program. One interesting aspect of this program is that rebates are negotiated between the utility and the customer. The negotiations and discussions between the utility and customer helps limit free riders, if handled properly.

Hirst, E. and Sabo, C., "Electric-Utility DSM Programs: Terminology and Reporting Formats," Oak Ridge, TN, prepared for Oak Ridge National Laboratory, (1991).

Free drivers were defined as customers that do not directly participate in a program, but are influenced by the program and adopt program recommendations anyway. The authors provided a diagram to illustrate the effects of free riders and drivers on a DSM program. Customers who purchase qualifying measures were grouped into four categories:

- Those who would purchase the measures without the program and did not request a rebate;
- Those who would purchase the measures without the program and did request a rebate;
- Those who would not purchase the measures without the program and did request a rebate; and

- Those who would not purchase the measures without the program and did not request a rebate (free drivers).

Horowitz, M. and Spada, M., "Energy Savings for Residential Lighting Programs: There's More to It Than Just Counting Lamps," Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, (September 1992), Asilomar, CA.

The authors estimated the lifetime savings from Boston Edison's Residential Lighting program using sales data and customer surveys. As part of their savings estimates, they examined the effects of free drivers. They defined free drivers as those households that knew of the existence of the program but purchased qualifying lamps on their own without the rebate.

In the participant surveys, 6.8 percent of the respondents were designated free drivers since they had not purchased qualifying lamps prior to the program and purchased lamps during the program without receiving a rebate. (Some may categorize this as spillover or surge effect, not free drivers.)

The non-participant surveys indicated that 5.7 percent of the respondents were free drivers. These respondents had heard of the program and purchased qualifying lamps without participating and receiving a rebate. In addition, 17.4 percent of the respondents claimed to have never heard of the program but still purchased qualifying lamps. These customers were designated as "market driven." The authors recognized that some of the free drivers may have been motivated by factors other than the program. The 5.7 percent was assumed to be the upper bound of non-participant free drivers, with 1.0 percent used as a more reasonable estimate. The remaining 4.7 percent were designated as market driven, for a total of 22.1 percent of the total residential customer base. (These market driven estimates assume that the efficient bulbs would not have been available and purchased by customers without the program.) The participant and non-participant free drivers were then factored into net energy savings estimates.

Hummel, P. and McMenamin, S., "Residential Technology Scenario Analysis: Defining the Role of Efficiency Standards, DSM, and Market Forces," Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, (September 1992), Asilomar, CA.

The paper discussed the numerous influences on future markets, including technological change, efficiency standards, utility DSM programs, and market forces. Analysis was conducted on three residential appliances – air conditioners, water heaters, and refrigerators – using REEPS. For each appliance, it was reported that there are evolving national efficiency standards, particularly for refrigerators. Each appliance has a range from the least to most efficient appliance, and DSM programs target the upper portion of this range. Refrigerators have national standards which approach the upper range of efficiency. Air conditioners have a greater range of efficiency, and DSM programs can have an impact on moving purchases towards the upper end. For water heaters, 1990 efficiency standards limit the efficiency range. DSM programs are limited to increasing efficiency beyond these standards. The authors concluded that market forces are rapidly changing the stock of efficient appliances. Changes in stocks of efficient equipment result from not only DSM program reactions but also these changing standards. Using an historical (or pre-program) baseline of consumption to measure free riders and drivers may be problematic for those appliances with rapidly evolving efficiency standards.

Keating, K., "Persistence of Energy Savings," in Hirst and Reed, eds, Handbook of Evaluation of Utility DSM Programs, Oak Ridge, TN, prepared for Oak Ridge National Laboratory (1991).

A free driver effect is defined as when the comparison (non-participant) group takes action because of the program. The author claimed free drivers are interpreted as free riders when billing data are used to estimate net program savings. The issue is particularly important for persistence research, because this effect (confusing free riders and drivers) is more likely to occur the longer a program exists. To identify free riders and free drivers, the author suggests using surveys.

Kitchin, D., "The Impact of Market Transformation on DSM Evaluation Techniques," Proceedings of the 1993 International Energy Program Evaluation Conference, (August 1993), Chicago, IL.

The spread of DSM in many areas, as well as the longevity and high penetration rates of some programs, has lead to market transformation in some areas. This transformation is defined as making new products, processes, and practices widely available and usable, whereas before their availability was limited. The transformation makes using non-participants as control groups more problematic. Many factors are involved in transforming the market. Examples given are building codes in the Pacific Northwest, and the industrial motor market in BC Hydro's service territory. There has not been a way to test and control for differences in markets resulting from a program.

Free driver or spillover effects have been ignored because of difficulty in measuring their impacts. If market transformations occur, net savings can not be determined by comparing participants to non-participants. Several approaches for analyzing market transformation and savings estimates are discussed. Pre- and post-program market surveys in which the baseline for comparison is developed from pre-program conditions is sufficient only when few market changes have occurred. A multivariate regression-based approach is using a control group from outside the program area, and explanatory variables are included to explain differences in service territories. This approach is difficult because many areas have had DSM programs implemented. Combination approaches include simulation models which predict energy use in the absence of the DSM program, supplemented with market surveys. These models require extensive data and may involve value judgements in specifying the model. However, they may offer the best platform for evaluation changes in markets.

Kushler, M., Keating, K., Schlegel, J. and Vine, E., "The Purpose, Practice, and Profession of DSM Evaluation: Current Trends, Future Challenges," Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, (September 1992), Asilomar, CA.

In reviewing some of the key issues and challenges facing DSM evaluators, the authors briefly touch upon free drivers. They report that free drivers are not formal participants, but contribute to the goals of the program. They are affected by the program through conscious awareness of the program or through program-induced changes in the marketplace. The authors cite the three methods discussed by Saxonis for addressing this issue. The first is using a historical baseline of consumption from early years of the program. However, they report the problem is likely more significant for programs that have been in place for several years and have high participation levels. During the life of the program, efficiency standards and the market may change, which would bias estimates developed from a historical baseline. The second methods is to use surveys of non-participants or trade allies to determine impacts on non-participants and the market. The third approach is to compare the distribution of efficient equipment between the program area and a comparable area in which no DSM program is offered.

Nelson, D., "Phantoms of the Program: In Search of Measurable Free Riders and Free Drivers," Proceedings of the 1993 International Energy Program Evaluation Conference, (August 1993), Chicago, IL.

This paper sought to develop a market-based approach for measuring free riders and free drivers for net savings analysis. A free driver was defined as a non-participant who acquired the technology or energy conservation measure (ECM) and did not receive an incentive, but was influenced by the program. Identifying these customers and measuring the impact of the program on them was acknowledged to be difficult. In contrast to free drivers, "ongoing adopters" were defined as non-participants who acquire the technology and did not receive an incentive, but would have adopted the technology in the absence of the program. It was concluded that if free drivers are not understood and identified, using control groups that include free drivers to estimate net impacts will understate those impacts.

Staff at BC Hydro also developed a "technology tree" to understand the paths of different customers in the DSM programs. Customers were classified as free riders, program participants, ongoing adopters, free drivers, or program non-participants. Using this structure, the utility attempted to forecast free riders and drivers for its programs. To use this tree, a baseline of the market without the program must be estimated. Using this information and penetration rates of technology enabled the estimation of free riders and drivers. In order to establish the baseline, pre-program measurements were taken of the acceptance of technologies (termed the technology adoption curve, which was composed of potential free riders and ongoing adopters). BC Hydro did this by surveying distributors and using national sales data to determine sales levels before launching the program.

Once establishing the estimate of the status quo, total market sales after the program was underway had to be determined, again using surveys and national sales data. The difference between the total sales and the status quo was interpreted to be program impacts on participants and free drivers. Program records identified the impact of participants, so the impact on free drivers could thus be isolated.

Nelson, D. and Ternes, M., "Flipping the Industrial Market: The Move to High Efficiency Motors," Proceedings of the 6th National Demand-Side Management Conference, (March 1993), Miami Beach, FL.

The authors reviewed BC Hydro's DSM efforts in industrial motors from 1977 to 1991. In 1988, after eight years of education and promotion, the market share of high efficiency motors (HEM) was 3.5 percent. Most vendors did not carry HEM in stock. After three years of incentives, the 1991 HEM market share was over 60 percent. Participants received rebates of $400 per kilowatt saved. Vendors were also offered incentives, and a buy-back program was also conducted.

A baseline of the HEM market without the program was developed using surveys of BC motor vendors and other utilities in Canada. This baseline was termed the "status quo," which is made up of potential free riders and ongoing adopters, who would purchase the technology on their own and not claim a rebate. Using manufacturer surveys on sales and program records for the number of rebates paid, they determined that 29 percent of purchasers of HEM did not apply for rebates. Free rider and ongoing adopter estimates were then developed. Free drivers were calculated as the total HEM sales minus the program activity, free riders, and ongoing adopters. The free driver percentage grew through 1988 and 1989 but shrunk during 1990. At the beginning of 1990 free drivers were over one-fourth of the HEM motor market in BC. The authors did not mention any other factors besides the incentive programs which would have shifted the BC motor market towards HEM.

Pacific Consulting Services, 1990 Southern California Edison Energy Management Services and Hardware Rebate Program Evaluation, Vol. 6 (September 1993).

The authors estimated net savings excluding spillover for SCE's commercial programs. Discrete choice models were developed of customers' licenses to participate in the audit and rebate program, and to implement energy-saving measures. The simultaneous nature of these decisions was explicitly recognized in the estimation procedure. Net savings estimates are obtained for audit and rebate programs combined, as well as for each program separately.

Prahl, R. and Schlegel, J., "Evaluating Market Transformation," Proceedings of the 1993 International Energy Program Evaluation Conference, (August 1993), Chicago, IL.

Noting that market transformation has become an increasingly important goal of DSM programs, the authors sought to provide a framework for understanding and evaluating market transformation. They noted that DSM programs can influence this transformation through customers, retailers, or even manufacturers. They defined market transformation as when DSM programs induce a lasting, beneficial change in the behavior of some group of actors in the market system. (Most other definitions are wider in scope than this one in that the change may be caused by factors other than DSM programs.)

Their conclusions and observations were as follows: Insofar as evaluating market transformations, current methodologies are insufficient. New methods are required to measure changes in attitudes, incentives, and behaviors of market actors. A longer perspective will be necessary to capture effects as changes can occur slowly over an extended period of time. A number of variables may be necessary to capture program savings. Even with new methods, a single answer may not be apparent. Instead, a range of answers may be developed which frame the overall program impacts. A different approach may be necessary for each component of the market system affected by the transformation. To study the transformations, one must measure market baselines, attitudes and values, and sales data. These efforts may require collaboration between organizations and utilities, as the effects of programs may extend beyond that utility's service territory.

Regional Economic Research, Inc. An Evaluation of Niagara Mohawk's Low Cost Measures Program, Niagara Mohawk Power Corp., Syracuse, NY, June 1991.

Niagara Mohawk's Low Cost Measures Program offered self-install measures to residential customers. The evaluation of this program included a survey of participants and non-participants, from which a binary logit model was developed of the likelihood of adopting the various measures, and the impact of program participation on that likelihood. The estimated model was then used to predict the relative change in adoption rates depending on program participation. A Mills ratio was used in an attempt to correct for self-selection bias in the participant group. While the survey had a relatively small sample size, the model results did yield an estimate of free ridership.

Saxonis, W., "Free Riders and Other Factors that Affect Net Program Impacts," in Hirst and Reed, eds, Handbook of Evaluation of Utility DSM Programs, Oak Ridge, TN, prepared for Oak Ridge National Laboratory (1991).

The author stated that free drivers can be affected by a DSM program through either a conscious awareness of the program or through changes in the market created by the program. One example cited was Northern States Power Company's appliance rebate program in which only 40 percent of the customers who purchased qualifying appliances applied for the rebate.

Three methods of measuring free drivers were outlined:

- Surveys of non-participants, to find out if the program influenced their adoption of efficiency measures;
- Comparisons with areas outside the ones in which the program is offered, to develop some overall measure of the effect of the program; and
- Surveys of trade allies, to find out how stocking patterns may have changed.

Saxonis, W., "Measuring Free Riders: Does the Economic Climate Make A Difference?" Proceedings of the 1991 Energy Evaluation Conference, (August 1991), Chicago, IL.

In discussing free rider measurement issues, the author provided some insight into free driver effects and the tendency of programs to "move the market." The paper centered on the New York State Appliance Rebate Program, which was run in two stages, the first from April 1987 – May 1988 and the second from December 1989 – December 1991. Under this program customers received rebates for buying qualifying energy efficient appliances.

A survey of participating dealers revealed that 63 percent indicated they increased the availability of appliances meeting the requirements of the program. Approximately 50 percent of these dealers claimed to increase their inventories of qualifying models by at least 30 percent. According to the author, this occurred during a period when the national shipment of energy efficient refrigerators and room air conditioners remained constant. Also, 50 percent of the dealers who increased their inventories of energy efficient appliances cited the program as the primary reason for doing so. The state office that conducted this research also wanted to compare sales data for participating counties and a similar region. However, collecting sales data, considered proprietary information by many dealers, was unsuccessful. One means of avoiding this problem is for utilities to require dealers to provide sales data in order to participate in the utility-sponsored program.

The author briefly addressed the issue of the free driver, which he defined as a person who was influenced to take action by a program but not identified as a program participant because he does not receive a rebate.

Train, "Estimation of Net Savings from Energy Conservation Programs," University of California-Berkeley, 1993 (forthcoming in Energy, 1994).

This paper compared several popular modeling approaches for estimating overall net savings, which adjust for free ridership. The methods were alternative formulations of billing analysis regressions and discrete choice participation models. These included approaches comparing participants and non-participants with corrections for self-selection bias, as well as control group comparisons. A simulated data set was used to test the accuracy and bias of actual methods.

Van Liere, K., Feldman, S., and Brugger, D., "The Design and Structure of a Statewide Sales Tracking System for Residential Appliances," Proceedings of the 1993 International Energy Program Evaluation Conference, (August 1993), Chicago, IL.

This paper summarized a study which tracked the sales of five residential appliances in the state of Wisconsin. A sales tracking system was designed as part of this effort. During the study the evaluators conducted an inventory of utility appliance sales-tracking efforts in the U.S. Most efforts attempted to collects sales data from retailers within a utility's service territory. Few efforts had been successful with distributors or manufacturers. Success differed by appliance type, with refrigerators, room air conditioners, and water heaters being problematic, in part because of a lack of cooperation of major retailers. The likelihood of success increased if data providers are recruited through a combination of letters, phone calls, and personal visits. Financial incentives had little effect upon willingness to participate, although incentives recognize that retailers must spend time to compile data. Also, some utility representatives were skeptical of the sales data provided, and felt that data from only a portion of retailers may bias overall results. Finally, collecting data on an ongoing basis generally provides better data than trying to collect retrospectively.

Van Liere, K., Vig, K. and Feldman, S., "DSM Programs and the Residential Appliance Distribution Systems in Wisconsin," Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, (September 1992), Asilomar, CA.

The authors conducted a study of the distribution system for five residential appliances in the state of Wisconsin. They noted that DSM programs interact with the distribution in many ways and the interactions vary by appliance. The authors concluded the following: DSM programs do not have much influence on the total number of high-efficiency units produced by manufacturers. Instead, manufacturers shift stock around to meet different demands. Also, if non-efficient equipment is not selling in an area because of utility rebates, manufacturers may offer their own incentives or push the equipment in other areas. Trade allies will also push non-efficient or non-qualifying equipment so that they are not stuck with it.

Van Liere, K., Winch, R., and Quigley, D., "The Impact of PG&E's Refrigerator Incentives on Refrigerator Sales," Proceedings of the 1993 International Energy Program Evaluation Conference, (August 1993), Chicago, IL.

This report summarized the results of a study of PG&E's refrigerator incentive programs for 1991 (the refrigerator programs started in 1982). Data was collected from customers and retailers, including sales data from areas outside of PG&E's service territory. Sales data was collected from a sample 16 (of 32 attempted) of the 284 retailers in PG&E's service territory. Control data was collected from a national association of appliance manufacturers. This data was adjusted for size of refrigerator before comparisons were made. Savings were estimated for every refrigerator sold in PG&E's service territory compared to national averages. These savings captured free driver and free rider effects. Top-mount refrigerators sold in PG&E territory used 46 kWh less per year than those in the rest of the U.S.; side-by-side models used 23 kWh less than comparable national models.

Vine, E., "Persistence of Energy Savings: What Do We Know and How Can It Be Ensured?" Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, (September 1992), Asilomar, CA.

The author termed participants adding additional energy efficiency measures after initially participating in a DSM program as the "surge effect." They defined "replacement effect" as when participants replace efficient equipment with less or more efficient equipment. They noted that some studies indicate that this "replacement effect" should receive more attention. In one study in the Pacific Northwest, half of the participants had undergone renovation or remodeling since participating in a commercial incentive program two years ago. Measures were also removed or deactivated within half of the buildings. Certain building types were also more likely to be remodeled. Conclusions were as follows: measuring "surge" or "spillover" effects requires detailed analysis of what happens and when. The literature does not focus on measuring these effects since, to some extent, they are measured in gross savings analysis and utilities receive credit for the savings.

Vine, E., "Social Dimensions of Program Evaluation," Proceedings of the 1993 International Energy Program Evaluation Conference, (August 1993), Chicago, IL.

The author asserted that it is necessary to integrate behavioral research into program evaluations in order to understand certain evaluation issues such as free riders and drivers. He speculated that there will be more research in this area if utilities receive incentives for the savings achieved by free drivers. One study cited on free drivers was on the Davis Energy Conservation Building Code, one of the oldest residential conservation programs in the U.S. The study revealed a 15 percent reduction in electricity consumption that was attributed to the structural features of the code and to the energy-conserving behavior of residents (free drivers). He summarized three approaches for addressing free riders (historical baseline, nonparticipant and trade ally surveys, and community comparisons).

Also discusses are market transformations. It was noted that these impacts are likely to be larger over time, but more gradual, less documentable, and harder to predict and control (compared to DSM resource acquisition). The author referred to a framework developed by Schlegel (Wisconsin Energy Conservation Corporation) for understanding market transformation. The framework used three perspectives: economic, social-psychological, and sociological.

Violette, D., "Analyzing Data," in Hirst and Reed, eds, Handbook of Evaluation of Utility DSM Programs, Oak Ridge, TN, prepared for Oak Ridge National Laboratory (1991).

Two approaches for measuring free drivership were noted. The first used pre-program data as a baseline for measuring program impacts. The second approach used surveys that determine whether non-participants have changed their behavior as a result of a program. The author believed free drivers are more likely to be a significant problem for programs that have been in existence for several years and have achieved high participation levels.

Violette, D., Ozog, M., Keneipp, M. and Stern, F., Impact Evaluation of Demand-Side Management Programs Volume 1: A Guide to Current Practice, Palo Alto, CA, prepared for the Electric Power Research Institute (1991).

The authors briefly discussed the impact of utility sponsored conservation programs on the behavior of non-participants. They noted that advertising and other information put out by utilities may induce customers who are not program participants to take conservation actions. One example cited was the Good Cents Home program (Wisconsin Public Service Corporation 1989), which was aimed at the residential new construction market. In one urban area over 50 percent of the new homes were certified under this program. Some area builders did not seek certification through the program yet built homes to the program standards in order to remain competitive. Effects such as these, which are difficult to measure, bias savings estimates downward.

A number of approaches were discussed which help explain the effects of programs on the behavior of participants and non-participants. Two of the approaches discussed were the use of customer surveys and comparisons of sales data between areas with and without these programs. While the authors did not directly address the use of these approaches for estimating program effects on non-participants, they did outline the advantages and disadvantages of each for assessing free ridership. Comments on these methods were applicable to using the same methods for measuring non-participant effects.

Violette, D., Stern, F., and Ozog, M., Impact Evaluation of Demand-Side Management Programs Volume 2: Case Studies and Applications, Palo Alto, CA, prepared for the Electric Power Research Institute (1992).

Violette, D., Ozog, M., and Wear, G., "Measuring Free Ridership: Do Some Experimental Designs Control Twice for Free Ridership?" Proceedings of the 1991 Energy Program Evaluation Conference, (August 1991), Chicago, IL.

In these reports the authors provided information on the impacts of free drivers and possible approaches for measuring these impacts, although noting few have been implemented. The primary concern for analyzing these effects had been to prevent having a "contaminated" control group for net impact estimates in which the control group was not isolated from the effects of the program. Two programs were noted as showing evidence of free driver effects. Both were Good Cents Home programs, one by Wisconsin Public Service Corporation and the other by Central Maine Power Company. While offering anecdotal evidence of free driver effects, the authors did not indicate the magnitude or scope of these effects.

The first approach for measuring these impacts was to use as an estimate of baseline energy consumption an historical baseline from before the inception of the program. This approach would likely produce biases results because of the overall trend towards energy efficiency. The net impact estimates would likely be overestimated if such an approach were used. However, this bias may be small over a short time period (less than three years), unless significant changes occurred within a market, such as new efficiency standards or upgraded building codes. Since the direction of the bias is known and the effects small, this approach could provide an upper bound of savings estimates.

The second approach was to use surveys to determine whether non-participants have changed their electrical usage or installed efficiency measures as a result of the program. With this approach, non-participants would be asked what effect the program had on their behavior. One problem with this method is that non-participants may not be aware of how the program has influenced their behavior. According to the authors, free driver effects can occur without non-participants knowing about a program. One example would be changes in stocking patterns by retailers as a result of a program. Customers may purchase and use these products without knowing about the program. To analyze this effect, trade allies could be surveyed, similar to surveys for free ridership. While not free of biases, these survey approaches provide additional insight regarding free driver effects. It was concluded that quantifying these effects is still problematic.

"Watch Out, Free Riders! Here Come Free Drivers!" Strategies, Volume 2, No. 1, (Winter 1991).

Manitoba Hydro conducted an evaluation of its rebate program for outdoor timers for block heaters. (Block heaters are used to warm car engines before starting the engines on cold winter mornings.) The timers promoted under the program set the heaters to come on a few hours before starting the engine instead of running all night. A survey of manufacturers and retailers indicated 17,000 of these timers were sent to Manitoba during the 1989-1990 program year. The utility issued rebates for only 6,000 timers. The remaining timers were either purchased without a rebate or were still in inventory. Manitoba Hydro referred to products purchased without the benefit of a rebate as "tag-ons," although it was noted that other utility evaluators may refer to this as the 'spillover effect' or as 'free drivers.' After subtracting out current inventory, Manitoba Hydro estimated that two timers were sold for every rebate redeemed.

■ 4.3 Sources Cited for Energy Program Studies

The following is a list of sources used. Specific citations for articles contained in these sources were provided in Section 4.2.

Databases

- EPRINET Computer Search
- Association of DSM Professionals Database
- U.C. Berkeley Computerized Card Catalogue

Energy/DSM Publications

- Evaluation Exchange
- Strategies (newsletter of the Association of DSM Professionals)
- Demand-Side Monthly
- Demand-Side Quarterly
- Energy, An International Journal

Proceedings

- Proceedings from ACEEE 1992 Summer Study on Energy Efficiency in Buildings, at Asilomar, CA, American Council for an Energy Efficient Economy, Washington, D.C.
- Proceedings from ACEEE 1990 Summer Study on Energy Efficiency in Buildings at Asilomar, CA, American Council for an Energy Efficient Economy, Washington, D.C.
- Proceedings from 1993 International Energy Program Evaluation Conference, held in Chicago, IL, August 25-27, 1993.
- Proceedings from 1991 International Energy Program Evaluation Conference, Chicago, IL.
- Proceedings from 6th National Demand-Side Management Conference, held in Miami Beach, FL, March 24-26, 1993.
- Proceedings from 5th National Demand-Side Management Conference, held in Boston, MA, July 30 – August 1, 1991.

Books and Reports

- Impact Evaluation of Demand-Side Management Programs, Volume 1: A Guide to Current Practice. Prepared by RCG, Hagler, Bailly, Inc. for Electric Power Research Institute, February 1991.
- Impact Evaluation of Demand-Side Management Programs, Volume 2: Case Studies and Applications (Revision 1). Prepared by RCG, Hagler, Bailly, Inc. for Electric Power Research Institute, November 1992.
- DSM Process Evaluation: A Guidebook to Current Practice. Prepared by Charles River Associates and Barakat & Chamberlin, Inc. for Electric Power Research Institute, May 1992.
- Economic Analysis of DSM Programs. Prepared for California Public Utilities Commission and California Energy Commission, December 1987.
- Kempton, Willett, and Neiman, Max (editors). Energy Efficiency: Perspectives on Individual Behavior. Prepared for ACEEE.
- Hirst, Eric and Sabo, Carol. Electric-Utility DSM Programs: Terminology and Reporting Formats. Oak Ridge National Laboratory, October 1991.
- Hirst, Eric and Reed, John (editors). Handbook of Evaluation of Utility DSM Programs. Oak Ridge National Laboratory, December 1991.

■ 4.4 Concepts of Spillover Measurement in Other Fields

The concept of spillover is, of course, not unique to the field of energy impact analysis. In fact, it can be found in the general literature on impact analysis associated with a wide range of public policies and programs. In common among these studies is the concept that a policy or program may have broader impacts than just the originally intended impacts on participants, and these other impacts may "spill over" onto parties other than the intended direct participants (or recipients). The analysis of program impact spillover is particularly strong in four specific subjects: education, health care, public facilities and technology investment. The issues and methods used for analysis of spillover in these fields are summarized below. Their parallels with the analysis of spillover associated with energy efficiency and conservation programs are then noted.

Education – Broader Societal Spillover Impacts

In the field of education, there has been a line of studies identifying the additional spillover benefits of public spending on high school and college programs, in terms of the societal benefits over and above the intended direct benefits (which are generally defined as increasing test scores or earning power). The concept behind these studies has been that the public educational spending leads to additional societal benefits in terms of socializing people, increasing voter participation and increasing overall "well-being," as well as ultimately expanding occupational opportunities and incomes, all of which are in addition to the direct effects of learning. The most well known example of this research was a large scale statewide study of college education spending in California (Hansen and Weisbrod, 1969). This study focused on identifying indices which measured the broad value of educational spending as investment in "human capital." Broader indicators of economic well-being resulting from educational spending have since been devised in subsequent studies (see Haveman and Weisbrod, 1975; Haveman and Wolfe, 1984; Hanushek and Taylor, 1990). Other related work has focused on defining broader impact measures of human development benefits (Kelly, 1991). In common among these studies is a focus on identifying classes of impacts on program participants which are beyond the narrow impact measures initially adopted to gauge program performance. Once that has been done, it is shown that new indices can be defined to capture the broader program spillover benefits for persons other than participants.

Health Care – Distributional Spillover Impacts on Non-Participants

In the field of health care, much of the literature on impact evaluation of medical care spending has focused on the spillover impacts of alternative health care pricing schemes, which appear initially to save money but actually have the spillover impact of merely shifting cost burdens to others. This has happened as the public Medicare system adopted a policy of flat fees to hospitals on the basis of the initial admission diagnosis. While the intended impact was to provide incentives for hospitals to economize on costs, what actually happened was that hospitals discharged patients sooner and sicker, so the burden of caring for people recovering from illness and operations was shifted to others as additional in-home care or nursing home care. This can also be referred to as a "contamination effect." These types of issues are discussed further in Altman, 1990. Subsequent work on indices of patient satisfaction and well being have also shown how measures designed to gauge productivity and impact of programs in the medical care sector have had unintended repercussions on incentives to allocate resources (Weisbrod, 1992). In common among these studies is a focus on identifying impacts on non-participants, who may pay higher costs, while the direct participants great a price break.

In the area of mental health policy, there is another line of research which has used "control group/test group" comparison techniques to judge the impacts and cost effectiveness of "test group" policies. These include controlled experiments, which compared jurisdictions with and without transfer of patients from mental institutions to community-based treatment programs (see Weisbrod, 1981; McGuire and Scheffler, 1987). The justification for the comparison/test group comparison was that this allowed for evaluation of overall net costs and benefits associated with a policy tradeoff affecting both participants and non-participants.

Public Facilities, Crime Control and Transportation – Geographic Spillover Impacts

Other research on impacts of public spending have examined geographic spillover issues, in which benefits or impacts spill over from the intended jurisdiction to affect other adjacent jurisdictions. One study in Japan compared the location of public facilities in two neighboring local government jurisdictions, and showed how each jurisdiction's investment in a new public facility could provide land market price impacts on the other jurisdiction, which would "enjoy the spillover effect as a free rider." (Kuroda, 1989).

The geographic spillover concept is a key aspect of a currently on-going Northwestern University evaluation of Chicago's community policing program (by Skogen), which is examining how community policing in some neighborhoods has had broader benefits for consumer satisfaction but also caused unintended spillover shifts in the geographic pattern of crime (as criminals move some of their activities to other districts). Here, time series measurements of crime rates are being compared between treatment and control areas in order to estimate geographic spillover impacts.

Finally, the area of transportation planning has long involved spillover concepts in project evaluation. Inherent in basic transportation planning is an understanding that improvements to a single highway corridor will not only improve traffic flow for vehicles on that stretch of highway (who are the intended recipients), but will also improve traffic flow on other roads feeding into that highway and may also bring improvements in travel conditions for other travelers using alternative modes of transport or alternative routes of travel. This broad set of benefits may, in addition, lead to other impacts on land values and business income as additional economic activity is induced or attracted to the area. One example of this progression of additional benefits was found in a Wisconsin study (Weisbrod and Beckwith, 1992) that utilized traffic simulation and economic simulation models.

Technology Investment – Inter-Sectoral Spillover Impacts

Yet another aspect of spillover is that in which spending by one industry causes additional benefits to other industries. One example is a Japanese study which examined the spillover effect of research and development spending in the Japanese electronics industry on the productivity, growth and quality of other industries. This form of spillover is also referred to as a diffusion effect (see Goto, 1989). Another example is the study of spillover benefits of agricultural research in the U.S. (Evenson, 1989). The geographic spillover and trans-industry sector spillover are combined in a study of rice research in the Philippines (Flores-Moya, 1978). These types of studies generally employ elements of cross-sectional and time series data comparisons.

Relevance for the Evaluation of Energy Programs

Overall, then, we see that spillover is viewed in the other fields within economic literature as: (1) broader benefits for participants, (2) benefits occurring for non-participants, (3) distributional and contamination effects on non-participants, (4) geographic impact effects, and/or (5) trans-industry sector effects. The techniques used for analysis span time series and cross-sectional approaches.

The literature on evaluation of energy conservation programs, on the other hand, has focused almost exclusively on the first two categories: (1) broader energy-saving benefits for participants, and (2) non-participant energy-saving benefits.[1] In particular, there has been little or no attempt to measure cross-jurisdictional geographic spillover (e.g., the benefits of one utility's programs on customers of an adjacent utility) or trans-industry sector effects (e.g., the benefits of an industrial program on commercial and residential sector customers).

1/ The exception to this is the examination of non-energy benefit and cost spillover impacts of low-income energy efficiency programs. The most extensive review of this area to date can be found in the forthcoming 1994 ACEEE Summer Study paper by Megdal and Piper based upon a recently completed study by Cambridge Systematics for the nine electric and gas utilities in New York.

difference in motivation and approach. Most of the energy program evaluation literature has been motivated in large part by the requirements of state regulators for rigorous and supportable impact analysis to justify cost recovery for utility programs. There are no such regulatory requirements underlying public policy impact studies in these other fields, although legislative allocations of budgets for many of the public programs can potentially be affected by these studies. Thus, many of the non-energy studies have examined spillover as part of an effort to broaden understanding of the full impacts of public programs, in order to illuminate policy debate. This is in contrast to the energy program literature, which has focused more attention on dissection and separation of the behavioral effects of free riders, free drivers, technology constraints and distribution market effects – i.e., the separate components of overall impacts, rather than on the impacts as a whole.

The additional distinctions made in the non-energy program evaluation literature may also be relevant for energy evaluation. One is the distinction between "spillover benefits," which are typically defined as affecting persons other than the intended beneficiaries, and "unmeasured benefits," which are broader benefits for program participants. The other distinction is between "real" spillover effects and "pecuniary" spillover. The latter are redistributional, resulting from changes in prices, which always benefit some people (those receiving the higher price) while making others (those who pay it) worse off. The real external benefits, by contrast, truly increase or decrease the aggregate well-being of others.

Bibliography on Spillover Measurement in Other Fields

Altman, Stuart H. "Measuring Real Case Mix Change under Medicare's Prospective Payment System," Journal of Health Economics; 9(4), 1990, pages 499-504.

Evenson, Robert E. "Spillover Benefits of Agricultural Research: Evidence from U.S. Experience," American Journal of Agricultural Economics, 71(2), May 1989, pages 447-452.

Flores-Moya, Piedad, Robert E. Evanson, and Yujiro Hayami. "Social Returns to Rice Research in the Philippines: Domestic Benefits and Foreign Spillover," Economic Development and Cultural Change, 26(3), April 1978, pages 692-607.

Goto, Akira and Suzuki Kazuyuki. "R&D Capital, Rate of Return on R&D Investment and Spillover of R&D in Japanese Manufacturing Industries," Review of Economics and Statistics, 71(4), November 1989, pages 555-564.

Hansen, W. Lee and Burton A. Weisbrod. "Distributional Effects of Public Expenditure Programs," Public-Finance; 27(4), 1972, pages 414-420.

Hansen, W. Lee and Burton A. Weisbrod. "Distributional Effects of Public Expenditure Programs," Public-Finance; 27(4), 1972, pages 414-420.

Hanushek, Eric A. and Lori L. Taylor. "Alternative Assessments of the Performance of Schools: Measurement of State Variations in Achievement," Journal of Human Resources, 25(2), Spring 1990, pages 179-201.

Haveman, Robert H. and Burton A. Weisbrod. "Defining Benefits of Public Programs: Some Guidance for Policy Analysts," Policy Analysis; 1(1), Winter 1975, pages 169-196.

Haveman, Robert H. and Barbara L. Wolfe. "Education, Productivity and Well-Being: The Role of Nonmarket Effects," Journal of Human Resources, 19(3), Summer 1984, pages 377-407.

Kelley, Allen C. "The Human Development Index: Handle with Care," Population and Development Review, 17(2), June 1991, pages 315-324.

Kuroda, Tatsuaki. "Location of Public Facilities with Spillover Effects: Variable Location and Parametric Scale," Journal of Regional Science, 29(4), November 1989, pages 575-594.

McGuire, Thomas G. and Richard M. Scheffler, eds. The Economics of Mental Health Services, JAI Press, Advances in Health Economics and Health Services Research series, Vol. 8, Greenwich, Conn. and London, 1987.

Weisbrod, Burton A. "Benefit-Cost Analysis of a Controlled Experiment: Treating the Mentally Ill," Journal of Human Resources; 16(4), Fall 1981, pages 523-548.

Weisbrod, Burton A. "Productivity and Incentives in the Medical Care Sector," Scandinavian Journal of Economics, 94(s), 1992, pages S-131-145.

Weisbrod, Glen E. and James Beckwith. "Measuring Economic Development Benefits for Highway Decision Making: The Wisconsin Case," Transportation Quarterly, 46(1), January 1992, pages 57-79.

5.0 Conclusions and Recommendations

5.0 Conclusions and Recommendations

5.1 Summary Assessment of Past Studies

Concept of Spillover

There are three ways in which a DSM program can lead to energy savings which are beyond the measurement of gross program savings. These are:

1. The dissemination of program information may affect customers' knowledge, attitudes and values about energy conservation, which may lead participants or non-participants to take actions outside of the program.

2. Program incentives may "move the market," changing the types of equipment offered by manufacturers, distributors and dealers, or the types of equipment recommended by contractors.

3. The process of program participation may encourage participants to take additional actions beyond the program measures.

Spillover impacts may include all three of the above categories of non-program impacts.

Concept of Free Drivers

While spillover savings are seldom added as an additional element of net program savings in actual program evaluation studies, the related concept of "free drivers" has received significantly more attention in research studies. There is no agreement on the exact definition of a free driver, although there is general agreement that a free driver implements additional energy savings outside of the program which would not have occurred if the DSM program had not existed. The disagreement concerns whether or not "free drivers" include some or all of the elements of spillover savings listed above.

Attention to free drivers initially emerged as it was recognized that estimates of program savings which compared participants with non-participants to correct for free ridership would themselves be biased, as some of the non-participants were also saving energy as a result of the DSM program. More recently, it has been recognized in the literature that free drivers may be viewed not merely a source of bias in free ridership estimates, but also as a source of additional net savings.

Separating "Overall" Savings from "Within Program" Savings

There is a crucial distinction between "overall net program savings" (which adjusts for both free ridership and spillover) and "net savings within the program" (which accounts for free ridership but ignores additional spillover savings). While spillover is starting to be recognized in the literature as a potential component of "overall net program savings," DSM program evaluations are in practice nearly always limited to analysis of "net savings within the program." The reason is primarily a lack of access to appropriate data and methodologies to measure spillover or overall net program savings.

5.2 Inclusion of Spillover in the Calculation of Overall Net Savings

Overall Methods

Methodologies that attempt to estimate "overall net program savings" (including spillover) are generally based on some form of treatment/control comparison, in which the treatment group is the set of all utility customers (participants and non-participants) who are offered the program, while the control group is customers of some outside area where the program is not offered. This is in contrast to the estimation of "net savings within the program," which generally utilizes a participant/non-participant comparison (with appropriate adjustment for self-selection bias), or else utilize direct surveys to correct for free ridership.

Insofar as spillover is fully counted in the estimation of overall net program savings, it may not be necessary to separately estimate the magnitude of spillover. It is possible, however, to separately estimate spillover savings by one of two methods. The first way is to calculate the difference between the estimate of "overall net program savings" (which includes spillover) and "net savings within the program" (which ignores spillover).

The second way is to directly estimate spillover through direct surveying of customers and/or trade allies.

Analytic Procedures

There are also several types of analytic procedures. One approach is to rely on **reported perceptions**. For instance, a participant is asked to report any additional equipment purchases or changes in use which are outside of the DSM program but which they feel they would not have done had the DSM program not existed. Similarly, a non-participant is asked if they are aware of the utility's media marketing efforts, and whether they feel that it has affected their behavior. An alternative approach is to rely on **comparisons**. For instance, comparisons are made of the rates of implementation of energy-saving measures which are not eligible for the program, among participants, non-participants and an outside control group. A third approach is **statistical modeling**. For instance, econometric techniques are used to model the effect of program offer on the decision to take other (non-program) energy efficiency actions.

■ 5.3 Key Measurement Issues

The biggest danger with measurement of spillover savings is over or under estimation through inappropriate counting of impacts. Key issues relating to spillover measurement are summarized below:

Need to Estimate Spillover

The fundamental interest is in accurately estimating the overall net savings of DSM programs. In that respect, spillover should be seen merely a component of overall net impacts. Sometimes, it is not necessary to separately estimate spillover savings. Interest in isolating the spillover component appears to comes largely from a desire to test the reasonableness of the net impact claims of different utilities by comparison of relative magnitudes of external (spillover) and internal (program) savings. In such cases, it can be useful to isolate the spillover component.

Over-Measuring Spillover Effects

There are many ways that DSM programs can lead to behavioral changes affecting spillover benefits. As discussed in the conceptual framework and literature review of this report, we see that spillover impacts may result from actions taken by trade allies, participating customers, non-participating customers, and/or customers of other utilities. Spillover may occur and may be defined on the basis of many factors including: prior plans for purchase and installation of energy-saving measures, awareness of the DSM program, patterns of equipment purchases, patterns of equipment use, DSM marketing, DSM audit recommendations, and types of measures covered by DSM incentives. Attempts to address these various facets of spillover, such as by conducting surveys of both customers and trade allies, can yield interesting information but can also inadvertently lead to "double-counting" of spillover savings. Care must be taken to avoid such double-counting, which leads to overestimation of spillover savings.

Importance of Recognizing Partial Spillover Savings

The measurement of net program impacts, including spillover impacts, necessarily involves some explicit or implicit estimation of what would have happened if the DSM program had not existed. Terms such as "free rider" and "free driver" are applied to various sets of customers on the basis of whether they would or would not have implemented certain energy-saving measures in the absence of the program. Yet in reality we know that this is an oversimplification, as some customers would have implemented alternative actions with some energy savings, but not as much energy saving, as occurred with the program in place. The twin concepts of partial free riders and partial free drivers can substantially complicate the definition and measurement of overall net impacts, and are seldom both used in empirical studies. However, failure to recognize this concept can also lead to overestimation of spillover savings. Failure to account for snapback effects (changes in patterns of use) accompanying implementation of spillover measures can also lead to overestimation of spillover savings.

Need to Account for Multiple Program Effects

Multiple types of DSM programs offered by the same utility (e.g., information, audit and rebate programs) may affect an individual customer. Similarly, the publicity and information aspects of one utility's DSM programs may spill over to affect customers of other nearby utilities. All of these situations can present a problem for the measurement of spillover effects, as it is possible to double-count spillover impact among multiple programs. Care must be taken to avoid this problem, which also leads to overestimation of spillover savings.

■ 5.4 Guidelines for Applying Alternative Methodologies

A range of analysis methodologies may be employed to measure spillover and overall net impacts of DSM programs. These are:

1. Direct estimation from customer surveys;

2. Calculations from trade ally data

3. Regression analysis of billing data; and

4. Implementation decision models using discrete choice analysis.

There are also significant differences in cost and accuracy associated with these methodologies. In general, it is important to tailor the methods used to match the specific size and type of DSM program being studied.

In general, we can define several different types of DSM programs, based on various dimensions of difference. One difference is their **applications focus**: e.g., new construction equipment applications, retrofit equipment applications, or behavior modification applications. Another difference is **sectoral focus**: e.g., industrial, commercial, agricultural or residential. Yet another important difference is **program technique** used: e.g., customer information, on-site audits or financial incentives (rebates). Finally, we can distinguish DSM programs in terms of **age and size**. While it is premature for this study – based on a review of the literature – to provide a definitive set of rules for all situations, some general guidelines on data collection and analysis methodologies can be identified.

In comparing alternative methodologies, four specific dimensions must be considered. They are:

- Data requirements and expense;

- Bias and uncertainty;

- Statistical precision; and
- Applicability.

Each of these dimensions is summarized here.

Data Requirements and Expense

Customer surveys can directly ask about implementation of non-program measures, as well about expected intentions of what would have happened if the DSM program had not existed. They can be a simple and straightforward process. Trade ally surveys, in which stocking or sales data is reired, involve significantly greater effort and perseverance. Billing data comparisons with outside comparison areas (i.e., customers of another utility) can be done but data acquisition can be difficult to obtain. Discrete choice models of decisions to implement energy-saving measures require the most data. They require information on the nature of existing equipment, the nature of equipment recently installed (both program and non-program measures), the eligibility for installation of measures that were not installed, plus data to estimate the potential energy savings associated with both installed measures and non-installed (but potentially feasible) measures.

Bias and Uncertainty

Customer surveys concerning what would have occurred without the DSM program are subject to offsetting biases. Stated intentions may under- or overestimate whether non-program measures would have occurred without the DSM program. Trade ally sales data can be used to measure changes in sales patterns, but not all of the observed changes may be due to the DSM program. Other factors occurring over time, such as prices, technologies, regulations and impacts of other DSM programs can also affect sales patterns. Thus, this method can over- or underestimate spillover effects.

Billing data has the advantage of potentially large sample sizes available, but requires correction for factors such as self-selection bias to be used. The econometric approach of decision choice modeling explicitly accounts for the effect of DSM program offer and program participation on adoption of non-program energy-saving measures. On the other hand, new uncertainty is introduced in the construction of technology alternatives available and their relative costs and energy savings.

Statistical Precision

For all methods, surveys or econometric models, statistical precision depends on sample sizes. With a large enough sample size, direct customer surveys can yield narrow statistical confidence intervals for spillover estimates, but those estimates can be subject to bias, as noted previously. The precision of spillover effects estimates from regression models of energy consumption and discrete choice models of measure adoption choices depend on the standard errors of estimated coefficients. These often appear to yield less precise estimates of program impacts, than survey-based methods which update engineering-based calculations. That is a misleading comparison, however, for the comprehensiveness of survey-based methods and the reliability of associated engineering calculations may be unknown. The regression and discrete choice methods can also be subject to less bias.

Applicability

There are distinct limitations on the use of customer surveys. In general, participants can only be asked about installation of non-program measures if the program focused on specific types of equipment. Participants can be asked how the program affected decisions to install non-program measures only to the extent that the program had a clear and limited set of types of measures that qualified under it. It is also difficult to ask non-participants how they were affected by a DSM program, especially if they were not even aware that the program had moved the market.

Trade ally sales data can be collected and applied with reasonable accuracy only for limited types of equipment, and they must be types for which it is easy to distinguish between sales of high efficiency products and the typical standard efficiency product. This can work well for highly standardized products such as fluorescent lamps, but is more difficult for less standardized products such as commercial refrigeration units.

Billing analysis can be appropriate only for types of programs for which the program measures and spillover measures are believed to cause enough change in energy consumption to be accurately reflected in customer bills.

Technology adoption decision models, utilizing discrete choice techniques, also depend on the ability to distinguish between high efficiency and standard efficiency installations. In addition, they require situations in which there are some observations of non-participants who installed program-eligible measures. This may not be easy to find for some types of measures which are seldom taken, or for programs which have achieved high market saturation.

In general, then, we conclude that the selection of appropriate measures for estimating overall net program savings should be made on the basis of the size of the program, the type of energy-saving measures which are covered by the program, the magnitude of their associated energy savings, the length of time the program has been in existence and the availability of data for an outside comparison area (i.e., jurisdiction of another utility). These factors, rather than a classification of types of DSM programs, should be the primary determinant of the methods used to estimate overall net savings and the spillover component of it.

Choice of Method

Figure 5.1 provides a checklist of the available methods and the types of programs for which they can be applied. In this figure, a blackened circle means that the method can be usefully applied, and a clear circle means that the method cannot as meaningfully be applied. For a method to be appropriate for a particular program, the method must receive a blackened circle for each attribute of the program. Conversely, if a method receives a clear circle based on any attribute of the program, then the method cannot meaningfully be applied to that program. The information in this figure is discussed in the following paragraphs.

Customer surveys provide an indication of spillover by asking participants and non-participants how and whether the program affected their decisions to install energy efficiency measures outside of the program. This method works relatively well when high efficiency measures are readily distinguishable from low efficiency measures; if the two types of measures cannot be easily distinguished, then customers will not be able to say whether they installed high efficiency measures outside of the program. The method does not reflect market transformation (that is, changes in stocking patterns and contractor practices) since customers would not know that measures that they took would not have been available to them without the program. Therefore, customer surveys are less appropriate for large programs, or programs with long duration, for which market transformation might constitute as substantial share of the spillover. To be succinct: customer surveys are most appropriate for smaller programs with limited duration which promote measures that are easily distinguishable from lower efficiency measures.

Trade ally data provide information on the aggregate sales of high efficiency measures. For this method to be useful, the program must be large enough to produce a noticeable change in aggregate sales. The collection of trade ally data is therefore not advisable for programs that are so sufficiently small that their effects would not be identifiable at the aggregate level. The collection of useful trade ally data requires that the

Figure 5.1 Checklist of Appropriate Methods for Different Types of Programs

Method	Program Attributes							
	Program Size		Can High-Efficiency Measures be Easily Distinguished?		Are Average Customer Savings Large Enough to Show Up in Bills?		Duration of Program	
	Small	Large	Yes	No	Yes	No	Short	Long
Customer Surveys	●	○	●	○	●	●	●	○
Trade Ally Data								
Before-After Comparison	○	●	●	○	●	●	●	○
Control Comparison	○	●	●	○	●	●	○	●
Billing Analysis								
Before-After Comparison	○	●	●	●	●	○	●	○
Control Comparison	○	●	●	●	●	○	○	●
Implementation Decisions								
Before-After Comparison	○	●	●	○	●	●	●	○
Control Comparison	○	●	●	○	●	●	○	●

● Method can be usefully applied.
○ Method is less meaningful if applied.

trade ally be able to distinguish high efficiency measures from standard efficiency measures. Programs that promote measures that are less easily distinguishable are less likely to benefit from trade ally data.

If trade ally data are going to be collected, there are two ways that the effect of the program can be identified: either by comparing the sales of high efficiency measures before the program to that after; or comparing the sales of high efficiency measures in the program area with that in a control area. The before-after comparison is meaningful only for programs of relatively short duration. With long-term programs, the before-program period is sufficiently distant in the past that trade allies will not be able to provide reliable data. Also, other factors will have changed sufficiently to make the before-after comparison difficult. For long-term programs, comparison of aggregate sales between the program area and a control area is needed.

Billing analysis in the context of net savings estimation attempts to identify the average change in consumption for all customers (participants and non-participants combined) that is attributable to the program. The method can be expected to provide useful information only if the average savings from the program (that is, the total net savings divided by the number of customers in the population, including participants and non-participants) is sufficiently large to be identified in customers' bills. For example, if a program reduces consumption among the eligible population by only one to two percent or less, it is doubtful that billing analysis will be able to identify this effect. (Note that the relevant figure is the change in the average customer's bill, not the change in participants' bills. For example, a program might save participants 20 percent of their consumption and due to spillover reduce non-participants' bills by an average of one percent. However, if only five percent of the population participates, the program reduces the average bill by only 1.95 percent – 20 percent of five percent and one percent of 95 percent – which is probably insufficient to be identified in billing analysis.) Billing analysis of net savings is appropriate for large programs that induce large savings per eligible customer, and less so for programs that are small and induce a fairly small reduction in the average customers' bill.

As with trade ally data, billing analysis can proceed with either of two types of comparisons: comparison of bills before the program to that after, or comparison of bills from customers in the program area with those in a control area. The before-after comparison is useful only when the program is of relatively short duration. With longer programs, a sufficient amount of change in other factors will have occurred to mask the effect of the program. In these cases, comparison with customers from a control area can be useful, provided that a meaningful control area can be identified.

Analysis of implementing decisions is expensive and should not be attempted for small programs, since the cost usually cannot be justified.

The method requires that high efficiency measures be distinguished from standard efficiency ones. It is more applicable, therefore, the more such a distinction can reasonably be made. For large programs where high efficiency measures can be identified, the analysis of implementation rates can be a useful tool.

As with billing analysis and trade ally data, the comparison can be over time (before program to after program) or across areas (program area compared to control area). For short-term programs, the comparison over time is recommended. For long-term programs, a meaningful comparison over time is not possible, and comparison with a control area is needed.

The previous discussion can be summarized as follows:

For small, short-term programs where high efficiency measures can be distinguished from standard efficiency measures, customer surveys are appropriate. For these programs (small programs of short duration), net savings can be expected to be sufficiently small such that trade ally data, billing analysis, and implementation decision analysis will not be able to identify it. Also, the cost of these methods is probably not warranted given the small scope of the program.

For large programs, customer surveys may be useful. Trade ally data, billing analysis, and implementation decision analysis can be appropriate depending on various factors. If high efficiency measures can be distinguished from standard efficiency measures, then collection of trade ally data and analysis of customers' implementation decisions are appropriate. If net savings are large enough on a per eligible customer basis to be identified in average bills (at least three percent savings per eligible customer), then billing analysis is appropriate. For each of these three methods, the type of comparison that is appropriate depends on the length of the program. For large programs of short duration (i.e., programs that lasted only a short period but affected many customers), a before-after comparison is appropriate. For large programs of long duration, comparison with a control area is needed.

■ 5.5 Directions for Future Research

It is clear that there has been very limited research into methods for estimating spillover savings, and additional research is necessary. Three types of studies are worthy subjects for further research. They are:

Direct Survey Estimation of Spillover

This involves asking participants and non-participants to itemize the energy-saving measures which they have initiated in a one or two year period, and then asking them to explain the reasons why they took those actions, with specific attention to whether the DSM program played a major or minor role in those decisions. This would be accompanied by a companion survey of contractors and dealers which probed the extent to which they changed their recommendations, prices and stocks in response to the DSM program. While some studies have addressed elements of the above, there is a critical need here to formally test a methodology which explicitly links the two, so that we can determine the extent to which there were customers who were affected by the behavior of trade allies but were unaware of that fact. Attention must also be given to avoid double-counting of spillover effects for cases where both the customer and trade ally were aware of program impacts on their behavior. Finally, attention must be given to estimate the extent of "partial spillover benefits," in which customers would have taken some actions without the program, but those actions would have saved less energy than the spillover actions caused by the program.

Cross Utility Comparisons

To estimate overall net program impacts, discrete choice models of measure adoption decisions require data on the eligibility, costs and payback associated with alternative program and non-program measures. There must be sufficient variation in observations of measure costs and payback among utility customers and a comparison area. Such data is often difficult to collect. Data collection for a comparison area is particularly problematic. One way around this problem is to assemble data for a set of utilities, which themselves provide variation in energy costs, rebate levels and other program rules. The various utilities are, in effect, comparison groups for each other. A cross utility study involving the major California utilities will thus provide basis for estimating more robust models, which can be used to assess effects of incentive design changes on implementation of program and non-program measures. This approach should start with one type of program and then be expanded later to a broader set of equipment technologies and program types.

Comparison of Methodologies for a Single Program

Several different approaches for measuring spillover savings (and/or overall net program savings) were described in this report. It can be instructive to select one straightforward DSM program for one utility and then explicitly test the measurement of spillover savings obtained from two or more alternative methods. These should include: 1) direct estimation of spillover, 2) indirect estimation through differences in "overall net" and "within program" billing analysis measures of savings, 3) technology adoption decision models, and 4) trade ally sales data. (The latter three both require treatment/control comparison.) Differences in estimates of spillover savings should then be investigated and explained in terms of data accuracy, analytic bias and methodological limitations. If this method is also applied for a similar program at an adjacent utility district, then it will also be possible to examine issues of cross-jurisdictional spillover impacts and how they are reflected in these alternative methods.