



complex world | CLEAR SOLUTIONS™

# Massachusetts Energy Efficiency Programs

## C&I Free-Ridership and Spillover Methodology Study

Prepared for: The Regional EM&V Forum Seminar, July 28, 2011

Prepared by:





# Agenda

1. Introductions
2. Presentation objectives
3. Background and study objectives
4. Methodology study activities
5. Participating end-user self-report design and testing
6. 2011 participant self-report study
7. Questions



complex world | **CLEAR SOLUTIONS™**

# Presentation Objectives





## Presentation objectives

- Summarize study objectives
- Summarize C&I free-ridership and spillover methodology development activities
- Summarize participating end-user self-report survey development
  - Integration with 2011 C&I non-energy impact (NEI) activities



complex world | **CLEAR SOLUTIONS™**

# Background and Methodology

## Study Objectives



# Background

- **A variety of methods have been used to estimate what would have happened in the absence of energy efficiency programs**
  - **Net effect is the observed effect, less the estimate of what would have happened in the absence of the program(s)**
  - **Program-attributable savings = net energy effects attributable to program efforts**
  - **Program attribution or Net-to-Gross (NTG) ratio = Program attributable savings as a fraction of total savings incented by the program**

# Background (cont'd)

## ➤ NTG Terminology

- **100% or Pure free-riders**
  - Would have adopted exactly the same quantity of the energy efficient end-use equipment at that time, absent the program.
- **Partial free-riders**
  - Would have adopted some of same end-use equipment on their own, but of a lesser efficiency or lesser quantity, or at a later time.
- **Participant spillover**
  - Energy efficient end-use equipment adopted by a participant due to program influences, but without any financial or technical assistance from the program.
- **Nonparticipant spillover (free-drivers)**
  - Energy efficient end-use equipment adopted by program non-participants due to the program's influence.
- **Net (Attributable) Savings = Gross Savings – Free Rider Savings + Participant Spillover + Nonparticipant Spillover**



## Background (cont'd)

- Prior to 2003, MA PAs independently quantified
- In 2003, PAs funded a study to develop standardized techniques for estimating free-ridership and spillover using participating customer and vendor self-report methodology
  - Sampling
  - Data collection approaches
  - Survey questions
  - Analysis methodology

## Background (cont'd)

- In 2010, MA PAs funded a study to review methodologies used across country and recommend approaches
  - High profile, given ramp-up in savings goals in 2010
  - Advances in NTG measurement nationally
  - Differences between the MA approach and other leading states
  - Benefits of common methodologies
    - Economies of scale
    - Consistent comparisons across territories, programs, and time

# Debate surrounding NTG

- Extensive debate across the country regarding need for and measurement of what would have happened in the absence of energy efficiency programs
  - Arguments why not needed
    - Free-ridership and spillover cancel each other out, **although there is little quantitative evidence that this is true**
    - Too difficult to reliably estimate, **although sound data collection techniques can mitigate many of the potential biases**
    - Funds are better spent on program implementation, **although NTG measurement are needed to improve program design**
    - Forward capacity market only looks at gross savings, **although this may not be true in the future**

## Debate surrounding NTG (cont'd)

- **Extensive debate across country regarding need for and measurement of what would have happened, absent energy efficiency programs**
  - **Arguments why needed**
    - Understand program and portfolio cost-effectiveness
    - Improve portfolio design and resource allocation
    - Refine program design
    - Understand market transformation
    - Align PAs' financial interests with societal interests
    - Understand how programs affect baseline load forecasts and short-term power procurement decisions.

## Methodology study objectives

- Offer guidance on when different methods should be used, including what types of surveys with what market actors will be used to get at what
- Standardize questions and scoring for participating C&I customer and vendor self-reports
  - Achieve consistency over time in regulatory reporting
- Achieve cooperation among PAs
  - Avoid re-creating the wheel, cost efficiencies

## Methodology study objectives (cont'd)

- **Devise improved methods for measuring spillover**
  - Measuring unlike spillover
  - Measuring spillover at later date to allow time for it to happen
- **Recommend situations where participating end-user self-reports are sub-optimal or can't provide accurate NTG by themselves**



complex world | **CLEAR SOLUTIONS™**

# 2010 Methodology Activities





## 2010 free-ridership and spillover methodology development activities

- **Develop taxonomy of PA programs**
- **Best practice review**
- **Develop decision framework for selecting appropriate methodology**
- **Re-examine self-report methodology**



## Develop taxonomy of PA C&I programs

- **Develop a complete understanding of the C&I programs**
  - Review 3-year plans
  - Review previously collected information from the PAs as part of on-going 2010 evaluations
  - Interview PAs as needed to fill in gaps
- **Results used to inform decision framework for selecting appropriate methodology**

# Best practice review

- **Identify and classify variations of NTG methodologies used across country**
  - **Review and synthesize recent attribution research and compare with existing methodology**
    - CPUC (using CALMAC)
    - Wisconsin Focus on Energy
    - IEPEC, ACEEE, and AESP conference proceedings
    - Publicly available evaluation reports
  - **Identify best practice survey design, data collection and analytic elements**
- **Results used to develop decision framework**

# 2010 free-ridership and spillover methodology development activities

- Develop taxonomy of PA programs
- Best practice review
- **Develop decision framework for selecting appropriate methodology**
- Re-examine self-report methodology

# Variety of methods available for estimating net savings:

- Customer/vendor self-reports (most common)
- Market sales data analysis
- Retailer/contractor surveys
- Econometric modelling
  - Pricing and elasticity analysis
  - Stated preference (choices respondents would make)
  - Revealed preference (choices respondents did make)
- Structured expert judging (e.g., Delphi panel)
- Historical tracking
- Billing analysis
- Program delivery staff interviews
- On-site data collection M&V contractors



## Method selection factors

### ➤ Driven by structure of program and type of information that can be obtained

- Availability of market sales data and meaningful comparison group
- Homogeneity of the measure and the customers
- Likelihood of substantial upstream effects unknown to end-use participants
- Cost/value tradeoffs



## Method characterization

- **Describe each survey based method in terms of:**
  - Who are the respondents or data providers?
  - What types of information are they providing?
  - How are the data analyzed
- **For each method, describe key factors affecting each method's validity and key data issues**

## Not all self-reports are created equal

- **Most Maligned:**
  - Asking participants what they would have done hypothetically absent the program
  - Even this question can be framed to get useful answers
- **Much of the information we base market studies on is self-reported**
  - Vendor sales data
  - Participant and nonparticipant preferences (stated and revealed)
  - Participant and nonparticipant actual purchases
- **Some is more easily corroborated by records and direct observation than others**

## Survey-based method characterization

Surveyed Group	Types of data collected	Data Description	Analysis	Includes Like Spillover	Includes Unlike Spillover	Includes Nonparticipa nt Like Spillover
Participating End-users	Post Hoc counterfactual:	Self-reported likelihood of buying absent program assistance	Scoring and averaging	Optionally	Optionally	No
Participating and nonparticipating end-users	Revealed preferences	Actual purchases, prices, and customer characteristics	Discrete choice analysis or simple average adoption rates	Optionally	No	Optionally
Nonparticipating end-users	Stated preferences	conjoint or double bounded exercises	Conjoint or double bounded specialized analysis	Yes	No	Yes
Nonparticipating end-users	Stated preferences	likelihood of purchase at varying conditions	Scoring and averaging	No	No	No
Participating customers and comparison group customers	Billing data	Consumption data for roughly one year pre- and post-participation	Weather normalization and change analysis	Yes	Yes	Counted negatively

## Survey-based method characterization (cont'd)

Surveyed Group	Types of data collected	Data Description	Analysis	Includes Like Spillover	Includes Unlike Spillover	Includes Nonparticipant Like Spillover
Retail store managers and contractors	Sales data	Sales of efficient and standard equipment when program is and isn't present	Weighting and/or averaging	Yes	No	Yes
Retail store managers and contractors	Sales and counterfactual	Promotional activity and sales with and without program	Weighting and/or averaging	Yes	No	Yes
Retail store managers and contractors	Customer-specific influence	Customer-specific influence of program and supplier	Weighting and/or averaging	Typically not	No	No
Retail store	Shelf and stocking observations	Observed shelf volumes and prices	Modeling or averaging	Depends on accompanying elasticity analysis	No	Depends on accompanying elasticity analysis
Manufacturers & Regional buyers and distributors	Market sales data	Sales of efficient and standard equipment in program and non-program areas over time		Yes	No	Yes

## Survey-based method characterization (cont'd)

Surveyed Group	Types of data collected	Data Description	Analysis	Includes Like Spillover	Includes Unlike Spillover	Includes Nonparticipatory Like
Experts		NTG estimates from multiple methods	Delphi process	Depends on input methods	Depends on input methods	Depends on input methods

## Design factors and data collection issues

Surveyed Group	Types of data collected	Validity: Depends on—	Data collection issues
Participating end-users	Post Hoc counterfactual:	Scoring; accuracy of self-reported hypothetical action absent the program	Requires well designed surveys to mitigate response bias
Participating and nonparticipating end-users	Revealed preferences	Comparable conditions between purchases in presence of program and those in absence of program; accuracy of purchase and pricing data obtained from customers	Typically requires in-store intercept or onsite observation
Nonparticipating end-users	Stated preferences	Match of stated preferences for hypothetical purchases to actual experience; customers in exercise see and consider information more systematically than in actual purchases	Requires recruitment of customers willing to engage in extended exercise. Requires companion study to determine change in prices and stocking absent the program
Nonparticipating end-users	Stated preferences	Scoring; accuracy of self-reported hypothetical actions if purchasing with and without program	Requires companion study to determine change in prices and stocking absent the program
Participating and comparison group customers	Billing data	Valid comparison group with minimal self-selection effects	

## Design factors and data collection issues (cont'd)

Surveyed Group	Types of data collected	Validity: Depends on—	Data collection issues
Retail store managers and contractors	Sales data	Market coverage and weighting; accuracy of supplier recall if not based on sales records	Often some key suppliers don't cooperate, "holes" need to be plugged
Retail store managers and contractors	Sales and counterfactual	Market coverage and weighting; accuracy of supplier-reported actual and hypothetical activities	
Retail store managers and contractors	Customer-specific influence	Accuracy of supplier's report on influencing factors for individual customers	
Retail store	Shelf and stocking observations	Stores that represent a meaningful comparison set to those with active programs	Need access to stores; some companies deny access
Manufacturers & Regional buyers and distributors	Market sales data	Ability to construct non-program baseline from pre-program and/or comparison area market coverage	Often some key suppliers don't cooperate, "holes" need to be plugged



## Design factors and data collection issues (cont'd)

Surveyed Group	Types of data collected	Validity: Depends on—	Data collection issues
Experts		Well documented methods Effective iteration process Experts' expertise	

# Summary of methods applicable to different conditions

Surveyed Group	Types of data collected	Applicability				Typical Cost or Complexity
		Custom Measures	Measures with few, diverse participants	Prescriptive Measures with large numbers of similar participants	Measures with substantial upstream influence invisible to customers	
Participating End-users	Post Hoc counterfactual:	Good	Good	Good	Poor unless combined with retailer or contractor surveys	Medium
Participating and nonparticipating end-users	Revealed preferences	Poor	Poor	Good	Good if combined with pricing study	High
Nonparticipating end-users	Stated preferences—experiment	Poor	Poor	Good	Good if combined with pricing study	High
Nonparticipating end-users	Stated preferences—likelihood	Poor	Poor	Good	Good if combined with pricing study	High
Participating and comparison group customers	Billing data	Poor	Poor	Good if have valid comparison group	Good if have valid comparison group	Low

## Summary of methods applicable to different conditions (cont'd)

Surveyed Group	Types of data collected	Applicability				Typical Cost or Complexity
		Custom Measures	Measures with few, diverse participants	Prescriptive Measures with large numbers of similar participants	Measures with substantial upstream influence invisible to customers	
Retail store managers and contractors	Sales data	Poor	Poor	Good	Good	Medium
Retail store managers and contractors	Sales and counterfactual	Poor	Poor	Good	Good	Low
Retail store managers and contractors	Customer-specific influence	Good	Good	Medium	Good	Medium
Retail store	Shelf and stocking observations	Poor	Poor	Good if combined with stated or revealed preference study	Good if combined with stated or revealed preference study	High
Manufacturers & Regional buyers and distributors	Market sales data	Poor	Poor	Good	Good	Low

## Summary of methods applicable to different conditions (cont'd)

Surveyed Group	Types of data collected	Applicability				Typical Cost or Complexity
		Custom Measures	Measures with few, diverse participants	Prescriptive Measures with large numbers of similar participants	Measures with substantial upstream influence invisible to customers	
Experts		Depends on quality of input methods				Low
None	Comprehensive market sales data for program area and comparison area	Poor	Poor	Good	Good	Low if data are available, High or not possible if data need to be developed

# Recommended approaches for MA C&I programs

Program	Importance to Portfolio	Measure Homogeneity	Customer Homogeneity	Likelihood of important influences unknown to customer	Likelihood of large NP effects	Market data availability	comparison area/Market baseline	Recommended Methods				
Large C&I retrofit	H	Custom and/or few	L	M	L	L	N/A	Participating End-users and Design Team Members	Post Hoc counterfactual			
Large C&I Retrofit	H	Prescriptive, large numbers	L	M	M	L	Possible	Participating End-users	Post Hoc counterfactual	+	Retail store managers, contractors	Sales and counterfactual

# Recommended approaches for MA C&I programs (cont'd)

Program	Importance to Portfolio	Measure Homogeneity	Customer Homogeneity	Likelihood of important influences unknown to customer	Likelihood of large NP effects	Market data availability	Comparison area/Market baseline	Recommended Methods				
Small C&I retrofit	M	Prescriptive, but few	M	M	S	None	None	Participating End-users	Post Hoc counterfactual:	+	Retail store managers and contractors	Customer-specific influence
Small C&I retrofit	M	Prescriptive, large numbers	M	M	M	Possibly for smaller gas measures	Possible	Participating End-users	Post Hoc counterfactual:	+	Retail store managers and contractors	Sales data

# Recommended approaches for MA C&I programs (cont'd)

Program	Importance to Portfolio	Measure Homogeneity	Customer Homogeneity	Likelihood of important influences unknown to customer	Likelihood of large NP effects	Market data availability	comparison area/Market baseline	Recommended Methods					
C&I New Construction/ Major Renovation	M	Custom and/or few	L	L	L	L	N/A	Participating End-users	Post Hoc counterfactual:				
C&I New Construction/ Major Renovation	M	Prescriptive, large numbers	M	L	M	L	N/A	Participating End-users	Post Hoc counterfactual:	+	Retail store managers and contractors	Sales data	



complex world | **CLEAR SOLUTIONS™**

# Participating end-user self-report design and testing



# Recommendations for improvements to 2003 standardized methodology

- 2003 self-report methodology contained many best practice survey design, data collection, and analytic elements
- For current programs, recommendation to rely mostly on participating end-user and trade ally interviews (self-report approach)
- Suggestions to further improve reliability and validity of estimates *(A “reliable estimate” means that if measured again, the result would be the same as the first result. A “valid estimate” means that the measurement accurately corresponds to what we’re conceptually trying to measure.)*
- Incorporate acceleration into the NTG calculation
  - Annual net savings
  - Life-cycle savings

## Recommendations for improvements to 2003 standardized methodology (cont'd)

- Where feasible, more timely measurement of free-ridership after participation (e.g., 3 months prescriptive, 6 months custom)
- Include more probing on alternative hypotheses
- Test adjustments to the NTG score using consistency questions
- Test different methods to deal with missing data (DK)
- Employ a preponderance of evidence approach for large, complex cases
- Measure spillover at a later date
- Estimate both like and unlike spillover

# Free-ridership calculation

## ➤ Scoring of survey responses

- Calculate efficiency and quantity scores to create initial free-ridership score
- Adjust initial score by timing factor, influence of past participation, and influence of any technical assistance to create adjust free-ridership
- If vendor was rated influential, substitute vendor responses

## ➤ Consistency checks

- Review programmed consistency check questions and adjust responses as needed
- Review open-ended “in your own words” responses for 100% and 0% free-riders
- Perform “gut check” with outside analyst

# Spillover calculation

- **Scoring of survey responses**
  - Calculate amount of spillover equipment
  - Verify efficiency level of spillover equipment
  - Calculate influence of program on spillover equipment decision
- **Consistency checks**
  - Review 0-10 scale consistency check question and adjust responses as needed
  - Review reasons given for not going through the program to ensure that equipment qualifies as spillover
- **Nonparticipant spillover is calculated in a similar method but with responses from vendors**

# Test recommended self-report approach

- Pre-test the survey instrument with sample of 2010 program participants
- Explore different ways of handling missing data
  - Removal of records with missing from analysis
  - Median vs. mean substitution for quantity and efficiency questions
- Apply 2003 algorithm to 2010 data

# Testing of recommended approach (cont'd)

- **Test for reliability and validity**
  - **Test the internal reliability of the free-ridership score and two independent consistency checks**
  - **Compare prior year's rates with current rates**
  - **Compare NTGR to warm-up questions  
(corporate policy, motivations for purchase)**

# Testing of recommended approach (cont'd)

For example:

## Free-ridership by Corporate Policy Regarding New Equipment

R3 - Company has energy efficiency policy when purchasing new equipment	Free-ridership Rate	Number of Participants
Yes	32%	19
No	10%	60
Don't know	0	2
<b>Total</b>	<b>16%</b>	<b>81</b>

# Testing of recommended approach (cont'd)

## ➤ Simulated dataset analysis

- Create simulated set of purchase scenarios with motivations and expected NTGR
- Create simulated dataset based on scenarios
- Apply algorithm and compare results (computed NTGR) to expected NTGR

# Recommend final changes to self-report methodology

- Clarify some of question wording
- Include receipt of interest-free financing in description of program services
- Add questions to assess program's influence on unlike spillover
- Add consistency check scale items to like and unlike spillover batteries
- Add questions regarding expected non-energy impacts
- Remove a couple process related questions
- Add pre- and post-participation sales questions to the vendor survey
- Incorporate expected non-energy impact questions in free-ridership survey
- Develop in-depth spillover questions to be administered later in 2011 as part of the non-energy impact study
  - To be conducted by an expert interviewer
  - Conducted with a subset of 2010 free-ridership survey respondents



complex world | **CLEAR SOLUTIONS™**

# 2011 participant self-report study



# 2011 free-ridership and spillover activities

- **Collect electric and gas 2010 participant sample data for programs included in study**
- **Field free-ridership data collection**
  - **Electric – April-May**
  - **Gas – June-July**
- **Free-ridership analysis and reporting**
  - **Electric - June**
  - **Gas – August**
- **Integration with 2011 C&I non-energy impact (NEI) activities**



complex world | **CLEAR SOLUTIONS™**

# Questions?

