

Effectively Evaluating Programs (evaluating effective programs?)

Presented at:
National Low Income Energy Conference
Nashville, TN June 5, 2007

Presented by:
Michael Blasnik
M. Blasnik & Associates · Boston, MA
michael.blasnik@verizon.net

Some semi-random thoughts...

In theory, there's no difference between theory and practice. In practice, there is.

All models are wrong, but some are useful
(G. P. Box, Albert Einstein?)

It's amazing what you can see just by looking
(Yogi Berra?)

There was no problem until you started looking
(as heard by Jack Laverty)

Why Evaluate?

- Assess impacts of program
 - compare to expectations, prior years, other programs
- Assess performance of specific treatments
 - Find out what works (or when it works) and what doesn't
- Identify opportunities for improvement
 - based on energy savings analysis
 - based on technical review & field inspections
 - based on interviews and surveys of stakeholders
- Demonstrate value of program
 - to funders, regulators, and others

Types of Evaluation

- Impact Evaluation
 - Assesses impact of program in meeting objectives
 - Energy savings (and electric baseload and water?)
 - explore patterns in savings: by measure, house type, provider, etc
 - Other impacts, non-energy benefits: bill payment, health and safety, environmental, economic/ job creation, etc.
- Process Evaluation
 - Assesses other aspects of program
 - technical review of procedures, field tools, training, QC
 - field visits using in-depth diagnostics
 - administrative/logistical systems of program
 - survey clients to get feedback on program implementation, marketing, education, etc.

Measuring Energy Savings

- Energy usage changes from year to year due to
 - Program Treatments, and...
 - Weather, behavior, and other changes in the home
- Evaluation goal: measure change due to the program
 - Adjust usage for weather variation from average year
 - Degree day methods (e.g. PRISM) work well for heating, but big changes still cause problems. Works worse for cooling loads.
 - Use statistics to adjust for other non-program factors
 - Large groups average out the random non-program changes
 - Comparison group should reflect trends that don't average out
 - Net Savings = Participant savings minus Comparison savings
 - Must be similar to participants, often use later participants
 - Comparison group "savings" typically <3%, unless price spike

Impact Evaluation Bias

- Experimental Design (rare for Wx evaluation)
 - random assignment of treatments to units
 - Untreated units = true control group
 - Clear basis for cause and effect: treatment -> impact
 - Statistical tests are primarily designed to assess whether "random chance" is a plausible explanation
- Observational Study (typical)
 - Assess on-going program - treatments not randomly assigned
 - No control group, instead create a "comparison" group
 - Bias should be assumed and explored
 - Statistical tests generally don't address this problem
 - Reliability depends on representative groups
 - Attrition bias in treated group, comparison group not comparable

Impact Evaluation Bias (2)

- Most program evaluations are observational studies, so representative groups are key
 - Participants analyzed may not be typical of program
 - less likely to include renters, occupant turnover, shut off or intermittent utility service, supplemental heat
 - more likely to include homeowners and seniors
 - Comparison group may not be comparable
 - HEAP population differs from WAP (esp. owner/renter)
 - Even the use of future participants is no guarantee
 - Example: many studies find HEAP declines after WAP. But WAP and HEAP are often applied for together, so HEAP is very high in WAP application year but reverts back.

Energy Savings Evaluation Methods: Billing Data Analysis

- Typically full year pre/post with a comparison group
 - evaluate each home, use break-outs to explore patterns
 - Large samples allow detailed assessments
- Need cooperation and data from utility companies
- Costs vary widely: \$20k -> \$150k+
 - Depends on data collection and depth of research questions (e.g., non-energy benefits)
 - Can save \$ by using staff for data collection
- Fancy pooled econometric models are used by some
 - “Black box” models and jargon mystify, but may be useful cross-check

Energy Savings Evaluation Methods: Run-Time Metering (ASAP)

- Meter daily usage for a few winter weeks pre & post
- Provides quick feedback, avoids problems with utility data, changes in household unlikely in short period
- Drawbacks
 - weather/timing problems common, esp. in mild climates
 - labor intensive - small samples, no comparison group
- Excellent way to compare two treatment approaches
- Poor way to evaluate overall program
 - metered jobs never seem to represent average work on average houses – everyone knows who is being evaluated

Energy Savings Evaluation Methods: Projected Savings

- “Calculate” savings from energy audit software or engineering algorithms
 - Not reliable: Measured savings are typically just 50%-70% of projected savings
 - Flaws in assumptions, inputs, and the engineering models themselves all tend to over-predict savings for virtually every measure
 - If you want to learn about actual savings you need actual usage data

Learning from Energy Savings Results

- Average savings don't tell you much
 - Patterns of savings often matter more in trying to improve program and identify what works best
 - Group savings break-outs
 - average savings by measure (or combo of measures), house type, provider are interesting but simple group comparisons don't show cause and effect because other factors may also differ between groups
 - Statistical modeling (regression)
 - tries to account for multiple factors at once to provide better estimate of what drives savings
 - be cautious of "black box" fancy statistics

More Evaluation Problems

■ Interpretation Issues

- Results for any one house are unreliable - stuff happens
- Savings vary between groups of homes for many reasons
 - Opportunities (housing or appliances), work selected, and work quality all play a role
 - Only field observations by experts can assess quality

■ Evaluator Issues

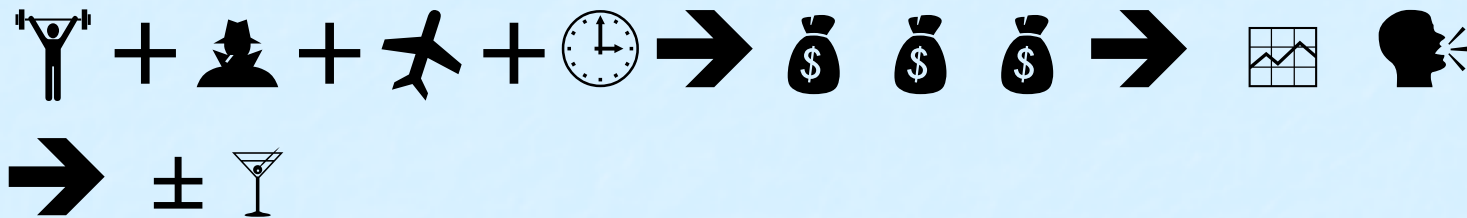
- Good evaluations recognize and try to deal with potential bias, don't over-state certainty
- Many evaluators have little building science knowledge, so be wary of strange results that could be precisely wrong
- Technical mistakes go unnoticed or unchallenged, hidden behind complex jargon and equations

Fancy Evaluations: beware

- 4 out of 5 econometrician and statistician PhDs recommend...that you hire them!
 - fancy statistical models are often touted by people who make a living doing fancy statistics
 - pooled time series cross-sectional regressions, a.k.a. ANACOVA, SAE, CDA, fixed effects models
 - Most try to estimate “the answer” as a single coefficient in a complex model that is often inappropriate
 - Assumptions not tested or met
 - “Black Box” + jargon = mystified audience, big \$ contract , limited critique or discussion
 - Lack of knowledge about building science can lead to strange and even absurd theories and conclusions

Your Savings = $\Omega/\partial^2 \Delta^* \Pi + \Sigma \int x \approx 25\% (\pm \infty)$

umm, what kind of number did you have in mind?



Good Evaluation as Storytelling

with numbers...

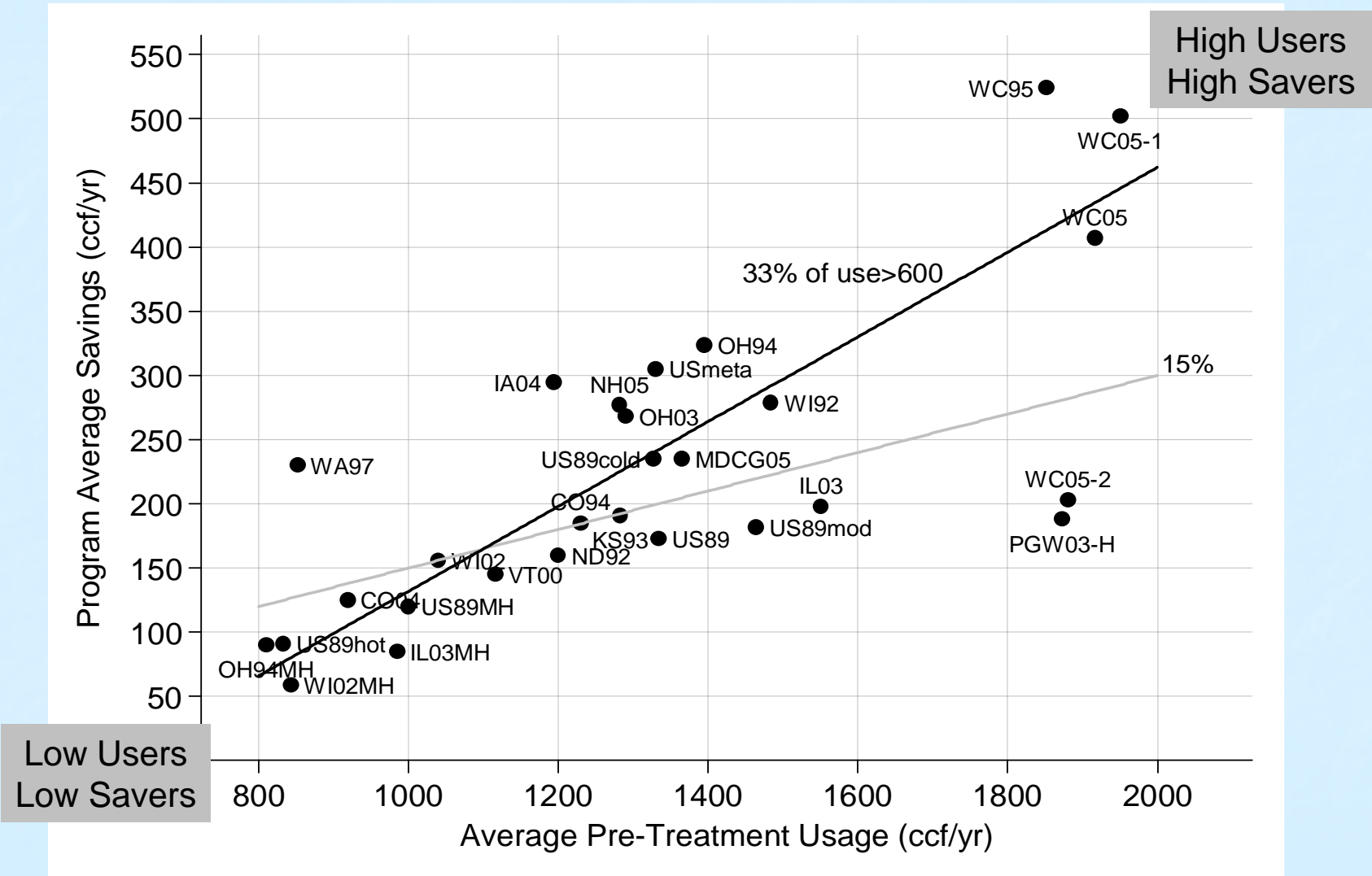
- Must be able to identify and assess potential sources of bias and factors that affect outcome
- Subject matter knowledge is more important than statistical knowledge
- Must develop a story that ties the data and analysis results to the real world
 - If a result seems unbelievable, don't just blindly believe it
- Use simple statistics to summarize and test, use more complicated statistics as needed
 - But be careful to understand the assumptions and tie it all into a narrative that makes sense

Wx Gas Savings Results

average therms/year/home

Study	Year	# Units	Pre_use	Savings	Save%	Cost \$/job
National: All	1989	3,873	1,334	173	13%	\$1,550
National: Cold	1989	1,040	1,327	235	18%	
National: Hot	1989	590	833	91	11%	
Nat'l Meta-eval	'93-'03	n/a	1,330	305	23%	\$2,913
Ohio	1994	2,209	1,395	324	23%	\$2,547
Iowa	2004	633	1,194	295	25%	\$5,682
New Hampshire	2005	27	1,282	277	22%	\$4,253
Ohio	2003	1,625	1,290	268	21%	\$4,800
Washington	1997	71	852	230	27%	\$3,132
Illinois	2003	2,056	1,551	198	13%	\$4,337
Kansas	1993	165	1,283	191	15%	
Colorado	1994	3,431	1,230	185	15%	
Wisconsin	'01-'03	8,252	1,040	156	15%	\$6,800
Vermont	'98-'00	25	1,116	145	13%	\$3,227
Colorado	2004	1,557	919	125	14%	\$2,850
Col Gas PA: Cont 1	2005	108	1951	502	26%	\$5,626
Phila Gas: Cont 2	2003	671	1873	188	10%	\$816

Wx Program Gas Savings Results

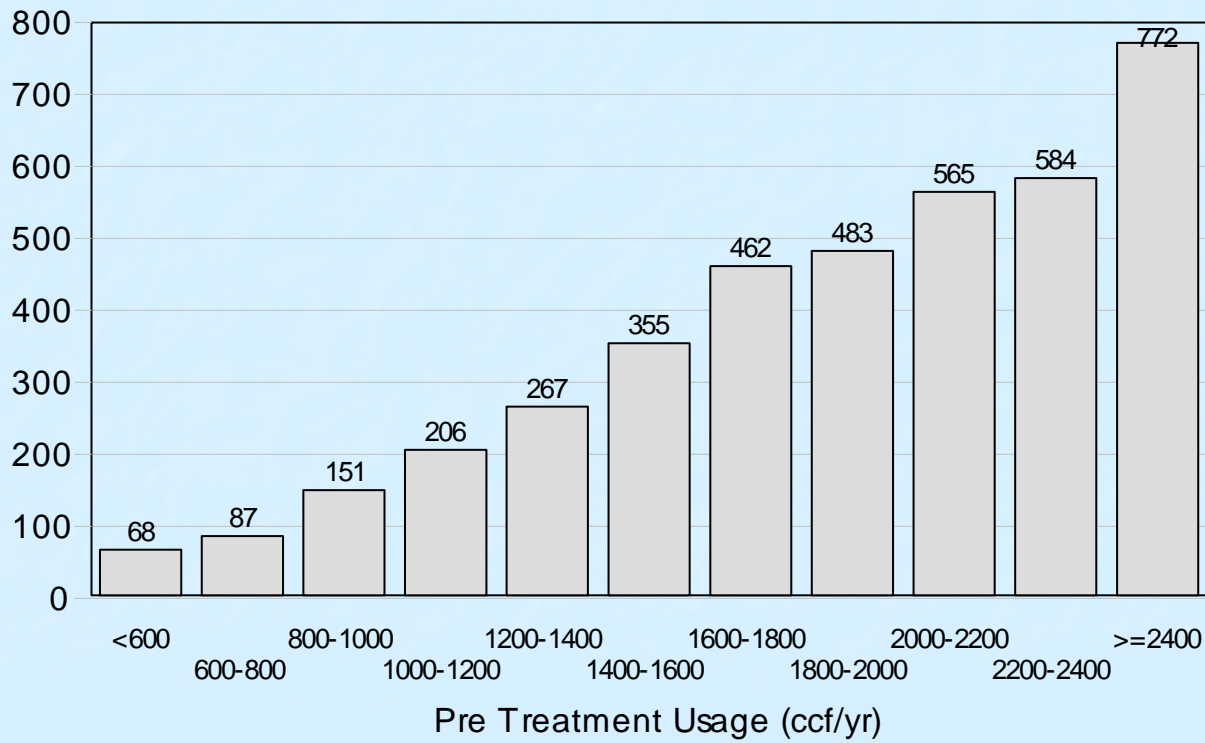


Gas Savings and Pre-Treatment Usage

You can't save what you don't use

(graph source: Ohio HWAP '94)

Average Savings (ccf/yr)



Wx Gas Savings: Mobile Homes

usage and savings in therms/year/home

Study	Year	# Units	Pre_use	Savings	Save%
National	1989	346	1000	120	12%
Colorado	2004	282	828	129	16%
Ohio	1994	176	810	90	11%
Illinois	2003	170	985	85	9%
Wisconsin	'01-'03	1,039	843	59	7%

- Mobile home gas savings much lower than site-built homes: avg. ~100 therms
 - Smaller, newer, better insulated than site-built leads to lower pre-treatment usage
 - Treatments more difficult, health & safety needs substantial
 - Untapped potential for ducts? Electric baseload?

Wx Electric Heating Savings

usage and savings in kWh/year/home

Study	Year	# Units	Pre-Use	Savings	Save%
National WAP	1989	426	14,972	1,830	12%
National Meta-eval	2000	n/a	19,919	2,153	11%
Washington WAP	1997	114	22,261	2,991	13%
Ohio TEE	2004	233	25,083	2,913	12%
Ohio WAP	1994	150	21,542	2,002	9%
Ohio WAP	2003	213	22,282	1,879	8%
New Hampshire	2005	26	21,556	2,182	10%

- Savings 8%-13%, 2,000-3,000 kWh
- Generally newer homes with fewer insulation and air sealing opportunities than gas-heated homes. More duct sealing opportunities?

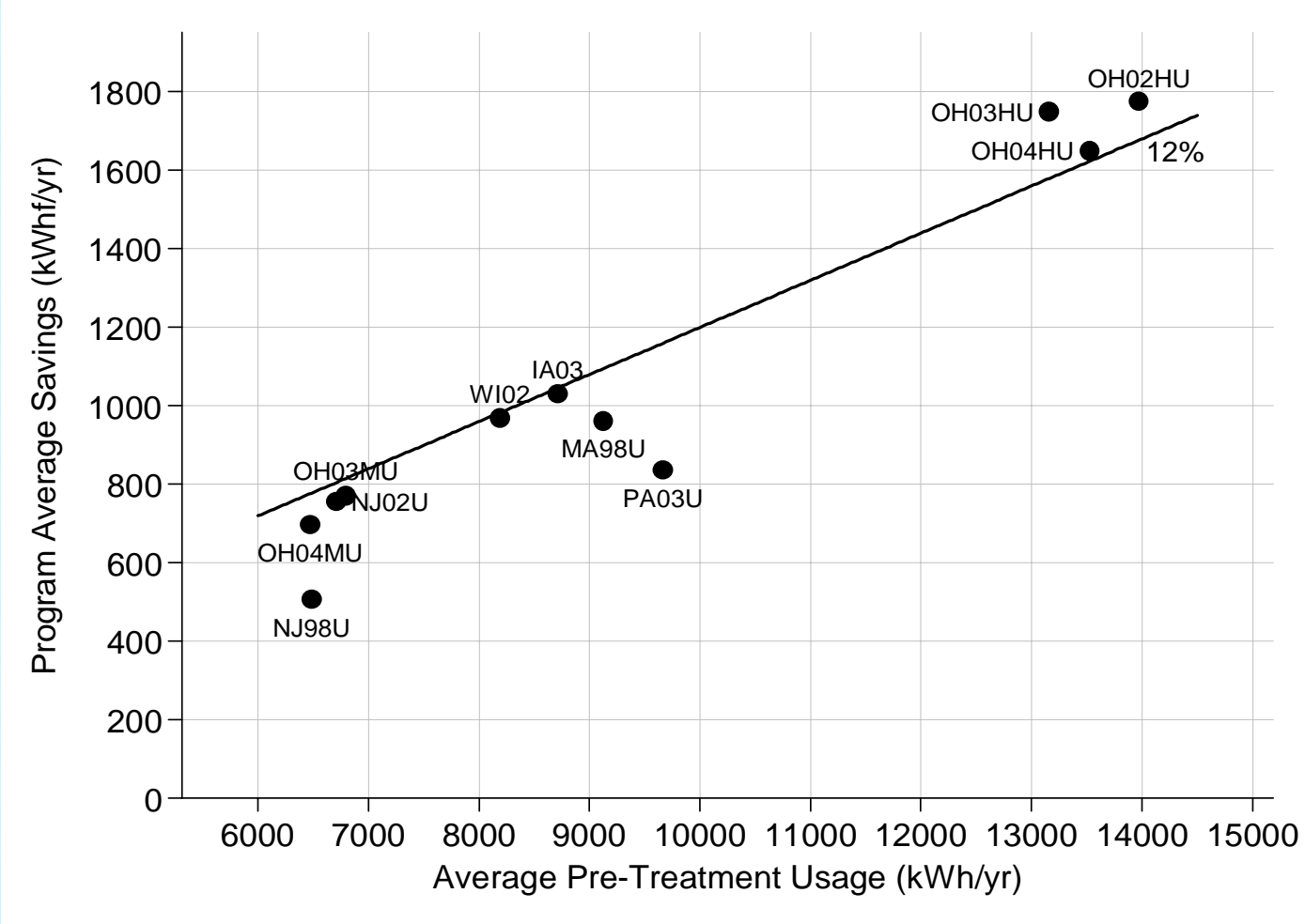
Baseload Electric Savings

usage and savings in kWh/year/home

Study	Year	# Units	Pre-Use	Savings	Save%
Ohio HiUse	2004	4,789	13,525	1,650	12%
Ohio HiUse	2003	4,525	13,159	1,750	13%
Ohio HiUse	2002	2,194	13,975	1,775	13%
Ohio ModUse	2004	1,355	6,468	697	11%
Ohio ModUse	2003	2,585	6,791	772	11%
Iowa	2003	294	8,711	1,031	12%
Wisconsin	'01-'03	8,675	8,191	970	12%
MA utility	1998	~2,000	9,125	962	11%
NJ Utility	2002	756	6,705	756	11%
PA Utility	2003	659	9,661	836	9%

- Savings ~ 12%, but varies 772-1750 kWh depending on usage
- Payback generally quicker than most gas heating measures
- Refrigerators, lighting produce most of savings, except some hot water (and fuel switching in WI)

Electric Baseload Program Evaluation Results



Lessons Learned: What Works (1)

- Wall Insulation
 - big savings, ~ 190 th/yr (0.2 th/yr/ft²), usually most cost-effective major measure
- Heating System Replacement
 - 100-180 th/yr, 12%-20% of heating, higher savings for high-efficiency units (and costs have dropped for them)
 - now cost-effective in high use homes due to gas price increases, otherwise a safety measure
- Attic Insulation
 - Save 65-150 th/yr, vary widely with conditions and quality
- Refrigerator replacement (and freezer replacement, and removals)
 - 600-900 kWh/yr avg. savings, very cost-effective in most places
- Duct Sealing in attics, garages & crawlspaces
 - 7%-13% savings for reasonable cost, quite cost-effective

Lessons Learned: What Works (2)

- Blower-door guided air sealing
 - Savings average 50-100 th/yr or ~ 7 th/yr/100CFM50, 70%-100% of projected
- Hot water measures (showerheads, aerators, wraps, etc)
 - 45 th/yr for combined measures (Iowa) =60% projected. Many devices are already low flow and people overstate showering time.
- Lighting
 - 25-50 kWh/yr/bulb, worthwhile but only $\sim 50\%$ of projected savings due to overstated hours, premature failure, removal
- Set-back thermostats?
 - utility program evaluations have found 4%-7% savings (70-100 th/yr). Excellent payback but do savings last?

Lessons Learned: What Doesn't Work?

- Window and Door replacement
 - Savings very small, negative correlation with savings in 1 study
- Duct Sealing in Basements
 - Multiple studies show 0%-3% savings. Still fix big holes & IAQ.
- Floor Insulation
 - Save <20 th/yr in OH, IA, 39 th in CO, but warmer floors, ducts?
- Heating System Tune-ups
 - OK for safety reasons, no savings?
- Mobile Home Weatherization
 - Cost-effective savings? Ducts? Or focus on electric baseload?
- Energy Education? – jury is still out
 - Works well in pilots, hard to replicate large scale.
 - At least as hard to do right as the technical stuff
 - Low cost makes it most likely worthwhile in any program.

Lessons Learned: Big Picture

- Measured savings typically 50%-70% of projections
 - Don't blame the occupants: behavior is not the culprit
 - Poor assumptions create shortfall for most measures including insulation, air sealing, and even lighting and hot water
- Higher savings are associated with
 - Targeting high use homes
 - Employing diagnostic approach (blower doors, etc)
 - Insulating lots of walls
 - Replacing Heating systems, extra utility/gov't funds
- Lower savings associated with
 - Wx as a housing program -- windows & doors
 - Low usage / Mild climate
 - Newer housing and/or many mobile homes