AN EVALUATION OF THE PERFORMANCE OF THE NEAT AUDIT FOR THE IOWA LOW-INCOME WEATHERIZATION PROGRAM

Gregory K. Dalhoff, Dalhoff and Associates, Verona, Wisconsin

Introduction

The National Energy AudiT (NEAT) was developed by Oak Ridge National Laboratory (ORNL) under contract with the Department of Energy. The audit is designed to identify cost-effective measures to be installed through the Weatherization Assistance Program (WAP). The audit tool provides a list of recommended measures along with estimates of annual heating and cooling energy savings, bill savings, and installed costs. NEAT's engineering algorithms are based on those originally developed for Lawrence Berkeley Laboratory's Computerized Instrumented Residential Audit (CIRA).

Currently, over half of the nation's WAP programs use NEAT. In the 1994 Iowa WAP, NEAT Version 4.3[a] was used by eighteen weatherization agencies and one local government to identify approximately $10,000,000 in improvements for about 4,000 houses. This study assesses four aspects of NEAT's performance in the 1994 Iowa WAP, including: 1) a comparison of NEAT's predicted gas heating savings with observed savings; 2) whether differences in predicted and observed savings can be traced to procedural errors in implementing NEAT or to limitations in NEAT's algorithms and assumptions; 3) whether these differences are large enough to affect which measures are recommended; and 4) how well NEAT's predicted savings compare to estimates derived using algorithms developed in the Iowa Statewide Low-Income Collaborative Evaluation (SLICE.)

Rationale for the Study

Other studies have found large discrepancies between engineering estimates of energy savings and observed savings, with engineering estimates typically overstating savings, sometimes by 50% or more (see, for example Nadel and Keating[1.].) The audit software that is used for WAP programs are not immune to this problem. For example, a recent evaluation of the Wisconsin Audit, which is used in that state's WAP, found that observed savings were 62% of predicted[2.].

An evaluation of an early version of the NEAT audit by ORNL found results that were in a similar range. That evaluation assessed NEAT's field performance for eighteen houses completed through the North Carolina WAP in 1990[3.]. After correcting for overstated values for the heated square footage in some houses, the study found mean predicted savings of 24.4 MBtu, compared to measured savings of 13.9, yielding a realization rate of 57%.

[a] Version 4.3 is similar to the current version in its thermodynamic calculations for measures installed by the WAP in 1994. Personal communication with NEAT's developer, Mike Gettings of ORNL.

More recent or comprehensive evaluations of NEAT were not available at the outset of this study. This evaluation was designed to assess the performance of a recent version of NEAT for a large enough sample to reflect program-level performance.

Approach

Our basic approach compares NEAT predicted savings with comparison-adjusted savings determined using PRISM. We did this for all treated houses as well as for three subsets of the data. After screening houses with poor PRISM diagnostics, we had an overall study sample of 408 gas-heated houses from the 1994 WAP, and a comparison group composed of 549 participants from the 1995 WAP participants[b].

NEAT provides predicted savings for heating measures. The WAP installs measures for other end-uses, so we had to adjust the NEAT's predicted savings to account for changes in gas consumption from water heating measures (faucet aerators, low-flow showerheads, pipe insulation, water heater blankets, and temperature turndowns) and from waterbed mattress pads (by reducing waterbed heater consumption, heating fuel consumption increases slightly). For the water heater measures, we used estimates from billing analysis adjusted engineering algorithms, developed as part of the Iowa SLICE activities[4.,5. & 6.]. Our adjustments for the indirect impacts from waterbed mattress pads were from engineering estimates. While these corrections add a source of error to our engineering estimates of predicted savings, it is not significant. The average water heater measure and waterbed mattress pad savings are a small percentage of the average NEAT predicted savings in our study sample, about 4% and –0.2%, respectively.

We developed three datasets from our overall study sample of 408 households. These datasets were designed to help separate engineering algorithm error from implementation-based error. For the first, we identified treated houses which had minimal implementation error and for which the NEAT input data reflected actual building conditions. We looked at several factors during this assessment, including: 1) whether installed measures closely matched those recommended by NEAT; 2) whether default values for furnace efficiency were entered into NEAT[c]; and

[b] Our screening process removed houses with greater than 10% uncertainty in NAC at 90% confidence, a coefficient of variation of NAC greater than 10%, fewer than eight observations in the pre-retrofit or post-retrofit periods, and less than approximately 2,000 houses degree days in either the pre-retrofit or post-retrofit periods.

[c] Auditors entered a furnace efficiency rating of 56% in houses where health and safety criteria called for a replacement.

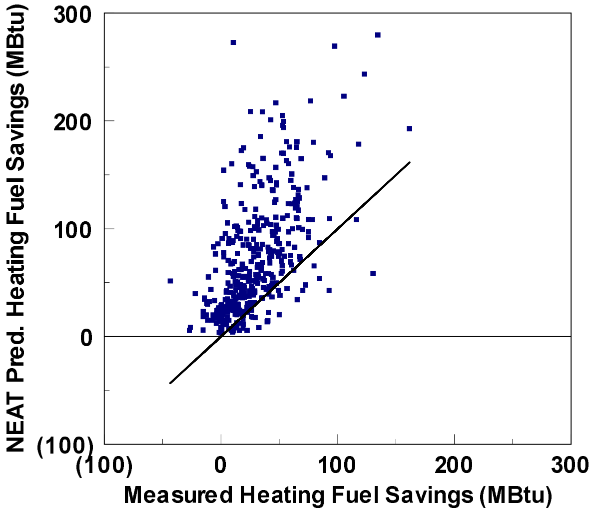

Figure 1. NEAT predicted savings vs. observed heating measure savings for all houses

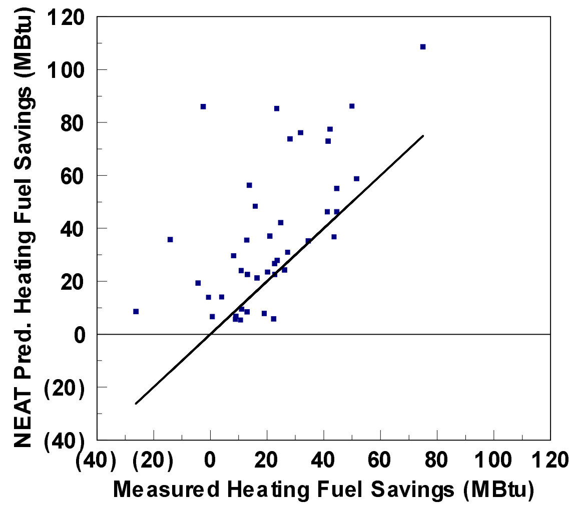

Figure 2. NEAT predicted savings vs. measured heating fuel savings for the refined dataset.

3) whether the air leakage rates entered into NEAT varied significantly what was actually measured in the field [d].

We identified houses with good matches between recommended and installed wall and ceiling insulation by comparing NEAT's total recommended wall and attic insulated area with the tracking system totals (we could not compare whether specific recommended insulation measures were installed, e.g., R-38 in one attic space and R-30 in another, since the tracking system did not collect data in this detail.) We selected houses where the tracking system and NEAT values for attic and wall insulated area varied by less than 5%, and where the data sets were consistent regarding furnace replacement and tune-ups. We screened out houses for which a default value for the furnace efficiency rating was used. Finally, we screened out houses where the NEAT values and the tracking system values of air leakage rates varied by more than 300 CFM_{50}. The forty-two houses in this dataset formed a basis of assessing the engineering algorithm error for this study. We refer to this dataset as the refined dataset throughout this paper.

We developed another dataset to assess the incremental error from using a default values of heating system efficiency. We screened the aggregate program dataset using the same criteria as before, except this time we included households that used the default value of heating system efficiency. This group consisted of 62 houses, including the 42 houses in our refined dataset.

We developed the third dataset to assess the incremental error introduced by poor estimates of the air leakage rates. In this case, we relaxed our initial screening criteria to include houses with large differences between the NEAT input data and the tracking system for air leakage rates. This study sample consisted of 93 treated houses, including the 42 houses in the refined dataset.

[d] NEAT allows entry of the pre-retrofit and post-retrofit air leakage rates. In many instances, this value was estimated in the Iowa WAP's implementation of the NEAT audit.

Results

Figures 1 and 2 show predicted savings plotted against observed savings for the aggregate program and refined datasets, respectively. The line shows a 1 to 1 relationship between predicted and observed savings in these charts. Both groups show that predicted savings tend to exceed observed savings.

Table 1 shows the results of our comparison between NEAT's predicted savings and observed. Measured savings were a fraction of NEAT's predicted savings for all samples, but was highest for the refined dataset: 54% (±11% at 90% confidence[e]). We found a modest drop in the realization rate, from 54% (±11%) to 50% (±7), after we increased our refined dataset by twenty houses that used a default value of steady state furnace efficiency, but were otherwise consistent with the criteria used to develop the refined data.

We found a substantial drop in the realization of predicted savings, from 54% (±11%) down to 44% (±8%), after we added 51 houses with air leakage rates that differed from tracking system values by more than 300 CFM_{50}, but were otherwise consistent with the criteria used to develop the refined data. However, we observed that the reduction in realization rate appeared to be in driven by the inclusion of houses that had very high over-predicted savings and only a moderate difference in air leakage rates relative to the other houses. This suggests that the air leakage rate was not the only factor reducing the realization rate for this group of houses, but it was unclear what other factors were causing the high predicted savings that led to the low realization rate.

We found that the overall realization rate for all houses was lowest: the observed savings were 39% (±3%) of NEAT's predicted heating fuel savings.

[e] We show the 90% confidence interval in parentheses throughout this paper.

Table 1. Results of the PRISM Analysis

Study Sample	N	Mean NEAT Predicted Savings (therms)	Mean Control-Adjusted Savings (therms)	Realization Rate
Refined dataset	42	373 (69)	203 (50)	54% (11%)
Refined dataset & houses which used a default furnace efficiency	62	570 (95)	286 (52)	50% (7%)
Refined dataset & houses with poorly-estimated air leakage rates	93	496 (81)	217 (50)	44% (8%)
All houses	408	725 (43)	286 (24)	39% (3%)

* 90% confidence interval is shown in parentheses

The realization rates for the refined dataset and the set that included default efficiencies are statistically different from the realization rate for the overall program. There is a lot of overlap in the confidence intervals for the other study samples, so that we cannot see statistically significant differences between these groups.

Cost-Effectiveness of Installed Measures

NEAT calculates the ratio of life cycle fuel cost savings to measure costs and reports the results as a Savings-to-Investment Ratio (SIR). NEAT recommends all measures with SIRs above a threshold level--the Iowa weatherization program used a threshold of 1.0. If NEAT's heating fuel savings are overstated, then the cost-effectiveness is likewise overstated and measures which are not cost-effective may be recommended.

Our analysis did not assess measure-level realization rates, so we could not determine what fraction of the total expenditures were spent on measures that were clearly cost-effective. Instead, we tried to get some idea of cost-effectiveness by examining the fraction of expenditures which would be cost-effective across a range of realization rates[f]. Figure 3 shows a plot of those results for wall and attic insulation and infiltration reduction measures. We found that as long as the wall insulation realization rate exceeds 30%, nearly all the expenditures on wall insulation are cost effective. For attic insulation, the fraction of total expenditures that are cost-effective decreases gradually as the realization rates decline; it is approximately 60% at the overall program and 80% for the refined dataset.

We can attribute the clear cutoff in the cost-effectiveness of wall insulation to the range of measures from which NEAT can select. Installations of wall insulation tend to be a single distinct measure (blown cellulose in 3.5-inch cavities) and have approximately the same SIR across all installations. This is in contrast to attic insulation for which NEAT can choose among four levels (R-11, R-19, R-30, and R-38) allowing it to more closely approximate an SIR of 1.0. For the instances where NEAT's overpredictions of attic insulation savings result in non-cost-effective installations, we can assume that a lesser level of insulation would be cost-effective. For these, only the incremental cost from the lower level of insulation to the level actually recommended by NEAT would be a non-cost-effective expenditure.

The cost-effective fraction of total expenditures on infiltration reduction is approximately 75% for the overall program, and 83% for the refined dataset. However, unlike insulation measures, NEAT does not recommend specific infiltration reduction measures. Instead, NEAT calculates the SIR of all installed infiltration measures after blower door directed air sealing has been completed. The SIR is based upon predicted savings (calculated from the values entered for the pre-retrofit and post-retrofit blower door readings) and the actual costs of the infiltration measures which were installed. Consequently, expenditures on infiltration reduction which are not cost-effective are attributable to overly aggressive installation on the agency's part, and not to NEAT's recommendations.

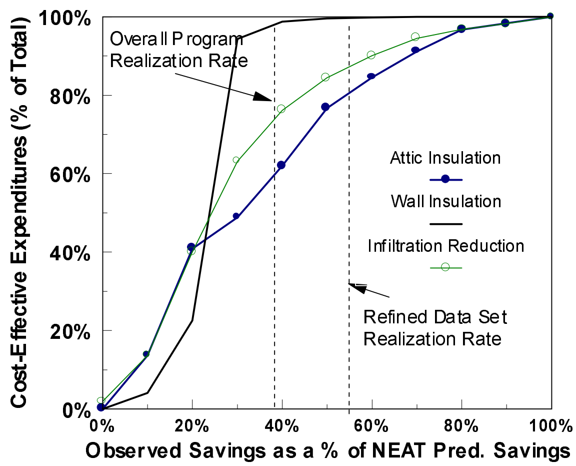

Figure 3. Percentage of expenditures that would be cost-effective at varying realization rates.

[f] The fuel cost savings consider all fuels affected by the measure, thus insulation measures may also have cooling (electricity) savings. Since we did not allocate costs for heating and cooling portions of savings, our discussion implies that the realization rates for heating and cooling are similar for a given measure.

Table 2. Comparison of NEAT Predicted Measure Savings with SLICE Estimated Savings

Measure	Number of installations	NEAT Predicted Savings (therms) Mean	NEAT Predicted Savings (therms) Median	SLICE Estimated Savings (therms) Mean	SLICE Estimated Savings (therms) Median	Ratio of NEAT Predicted Savings to SLICE Estimated Savings Mean	Ratio of NEAT Predicted Savings to SLICE Estimated Savings Median
Attic insulation	30	134	95	61	63	2.2	1.5
Wall insulation	20	191	160	122	109	1.6	1.5
Infiltration reduction	41	99	75	41	24	2.4	3.1
Furnace replacement	9	141	157	120	100	1.2	1.6

Comparison with SLICE Estimates

The Iowa SLICE evaluation has established measure-specific algorithms for estimating the energy and demand savings for measures installed through the WAP[g]. We compaerd the NEAT-predicted savings with the SLICE estimates in order to assess whether NEAT's estimates form a better basis for estimating program impacts.

We found that the SLICE algorithms provided a better estimate of program impacts. For the 408 houses in our study sample, we found that measured savings were 113% of the SLICE estimates, compared with a value of 39% for NEAT's predicted savings.

In order to assess whether the difference between NEAT's predicted savings and the SLICE estimates were due to results for specific measures, we explored differences between NEAT and SLICE savings for the major measures installed in houses of our refined dataset[g] (Table 2).

We found that each of the major measures contributed to the difference between NEAT predicted and SLICE estimated savings. We found the greatest agreement between estimates for furnace replacement savings, and the greatest differences in the infiltration reduction savings. A partial explanation for the greater difference in infiltration reduction may be attributed to the SLICE algorithms, which allocate a portion of infiltration measure savings to wall insulation. This tends to amplify differences between infiltration reduction savings and attenuate differences between wall insulation savings.

Analysis of Factors Contributing to Low Realization Rates

We did not design this study to provide an in-depth analysis of error in the engineering algorithms. Nevertheless, we did examine the data for general indications of the sources of the error. Our comparison of NEAT and SLICE estimates were inconclusive, so we turned our attention toward comparing NEAT's calculations of annual heating consumption[h] with PRISM's estimate of the normalized annual heating consumption[i] (NAHC). NEAT calculates savings as the change between the pre-retrofit and post-retrofit annual heating consumption: if these consumption estimates are inaccurate, then the predicted savings will likewise be in error.

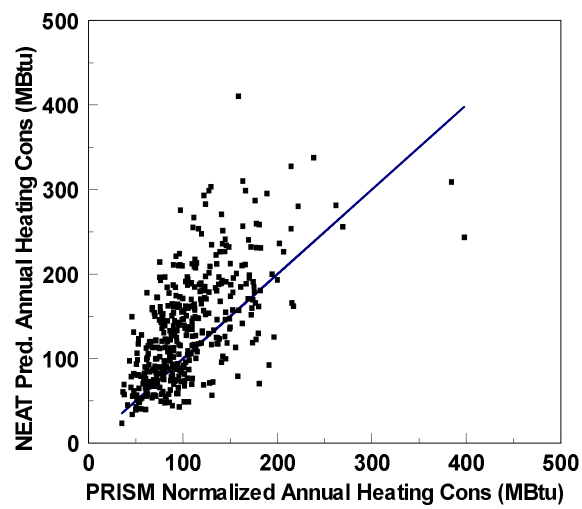

Figure 4. Pre-retrofit NEAT-predicted annual fuel consumption vs. PRISM NAHC, all houses.

Figures 4 and 5 show plots of NEAT's predicted *pre-retrofit* annual consumption compared to the NAHC

---

[g] We could not compare savings for furnace tune-ups--the SLICE algorithms do not predict savings for this measure.

[h] NEAT provides estimates of pre-retrofit and post-retrofit annual heating consumption in its reports. However, the reported pre-retrofit heating consumption is the predicted consumption after infiltration reduction measures have been installed, but prior to the installation of insulation and furnace measures. We calculated the total pre-retrofit heating consumption by adding the total of NEAT's predicted infiltration reduction savings to the predicted post-retrofit annual heating consumption.

[i] The PRISM normalized annual heating consumption is not as robust of an estimator as the normalized annual consumption (NAC) is, however it provides a measure of heating fuel consumption with which to compare to the NEAT-predicted values of annual heating consumption. The NAHC may introduce a small bias due to PRISM's tendency to account for seasonal fluctuations in natural gas water heater fuel consumption as heating fuel consumption, but this should not be a significant factor in this analysis.

for the program and refined datasets, respectively. The line indicates a 1 to 1 relationship between predicted and measured annual heating consumption. We found that NEAT over-predicted the pre-retrofit consumption by about 34% for the full dataset, and by about 8% for the refined dataset.

Figure 6 shows a similar plot for NEAT's predicted *post-retrofit* annual heating consumption from the refined dataset (we did not examine the post-retrofit estimates for the full program group because NEAT's estimates of usage are based upon assumptions of measures that may not have been installed). We found that NEAT tended to under-predict annual heating consumption in the post-retrofit period by about 11%.

If the error in both pre-retrofit and post-retrofit consumption were consistent (i.e. both were over-predicted by 8%), then the predicted savings would be in error by a similar percentage. The relatively higher error that we found was due to the combination of high pre-retrofit and low post-retrofit consumption results.

Given that the only real difference between the pre-retrofit and post-retrofit condition of these houses was the installation of the weatherization measures, we suspected that NEAT tends to over-predict annual consumption in houses with lower overall thermodynamic efficiency, and under-predict annual consumption in houses that have better thermal performance[j].

We examined whether NEAT's assumptions of the initial R-value could explain the over-predicted consumption in the pre-retrofit period, and under-predictions in the post, since these estimates are extremely sensitive to assumptions of low initial R-values. We explored this using regression analysis, by examining models that specified the ratio of NEAT's predicted annual heating consumption as the dependent variable, and parameters relating the thermal performance as independent variables. The independent variables included the thermal conductance of wall and attic spaces (the inverse of the wall or attic R value), and the difference between the value of steady state efficiency entered in NEAT and the value calculated from the output capacity and input energy. This analysis was restricted to the refined dataset. The results were largely inconclusive, although they did suggest that houses with lower initial wall R- values tended to have higher over-predicted annual heating consumption in the pre-retrofit period.

We identified two additional operational aspects of the program that would affect our findings of realization rates. These included higher levels of insulation than NEAT modeled in attic spaces and the installation of high density wall insulation (for which NEAT does not allocate infiltration reduction savings). Given the scope of this project, we could not re-run NEAT to account for the higher levels of installed insulation. If NEAT had been operated to account for these differences, its predicted annual heating consumption in the post-retrofit period would have been even lower, resulting in greater over-predictions of savings than we observed.

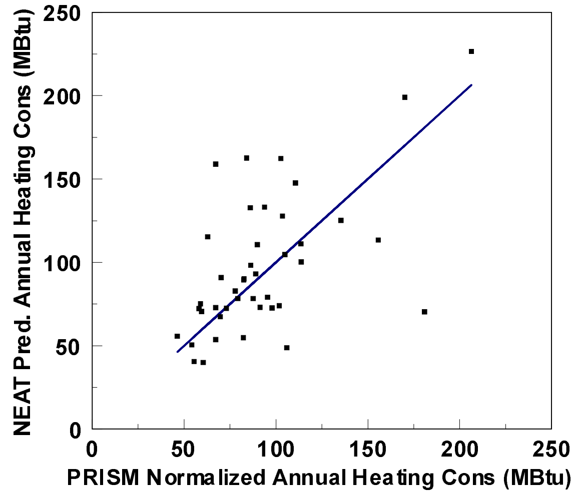

Figure 5. Pre-retrofit NEAT-predicted annual fuel consumption vs. PRISM NAHC, refined dataset.

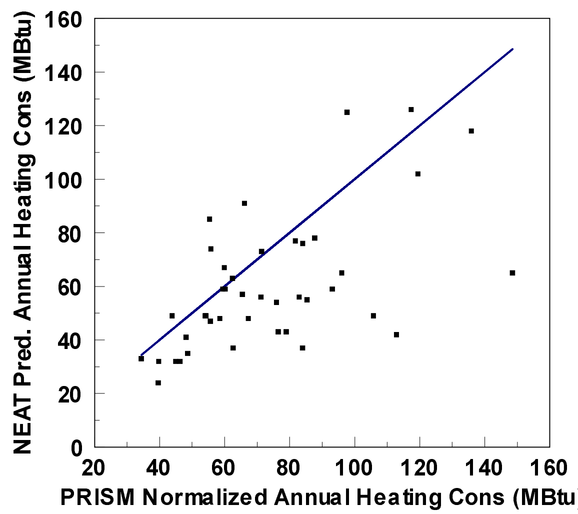

Figure 6. Post-retrofit NEAT-predicted annual fuel consumption vs. PRISM NAHC, refined dataset.

We were concerned that houses that relied heavily on secondary heating fuels would have lower realization rates than those that did not, but we found that the impact of secondary fuel usage in this study was insignificant. Houses with large amounts of secondary fuel usage tended to be removed during our PRISM screens (the exception would be houses where the occupants used secondary fuels in proportion to natural gas consistently). In fact, our study sample showed only eight houses which used secondary fuels in our large study sample. None of these appeared in either our refined dataset or our dataset which used default furnace efficiencies, and only three were in our 93 house group which included those with poorly estimated air leakage rates.

[j] Sharp's study of the North Carolina WAP tends to confirm the difference in error between pre-retrofit and post-retrofit consumption. That study found "predicted consumption rates below 80 MBtu and savings below 40 MBtu are very close to measured values in most cases; when above these limits, predicted consumption and savings often are considerably higher than measured values" (p.103).

Policy Implications

The results of this study (as well as other validation studies of engineering tools) remind us that there are risks associated with the use of sophisticated building simulation models. First and foremost, users should be aware of the tendency for these tools to over-predict savings, and should use conservative assumptions when accurate information is unavailable.

It is also important for users to be aware of the relative great influence that some input parameters and default values have on results. The assumptions of the initial R-values of pre-existing insulation may have a large impact on the prediction of annual energy usage (and predicted savings). Measured air leakage rates should be used whenever possible, and conservative values should be used if these are unavailable. Although this study did not show statistically significant differences between houses which used default furnace efficiencies, users should be aware that savings from all recommended measures are exaggerated when low ratings are entered.

For the Iowa WAP, this analysis suggests that the agencies may be spending more on infiltration reduction measures than is cost-effective, or alternately, may be miscategorizing some measures as infiltration reduction when they may, in fact, be installed for other purposes (i.e., building preservation or repairs).

This study also concluded that the State and utilities involved in the collaborative low-income program should continue to use the SLICE algorithms to estimate overall program impacts rather than rely on totals of NEAT's predicted savings

Although NEAT suffers from the same tendency to over-predict savings as many other engineering building simulation models, it does provide useful information for ranking the relative value of measures to be installed. SLICE evaluations have found that the overall savings from measures recommended by NEAT are cost-effective.

To control for instances where marginally non-cost-effective measures are recommended, operators could consider conservative actions, such as increasing the cutoff value of the SIR in the NEAT setup files so that NEAT will recommend only measures which are clearly cost-effective. However, the analyst should keep in mind that applying a single cutoff factor for the SIR across all measures will almost certainly preclude selection of some cost-effective measures while eliminating others that are not cost-effective.

References

1. Nadel, Steven and Kenneth Keating, (1991). "Engineering Estimates vs. Impact Evaluation Results: How Do They Compare and Why?". Energy 1991 International Energy Program Evaluation Conference, Proceedings, CONF-910807, Fifth International Conference in Chicago.
2. Preliminary results from the executive summary of the untitled draft evaluation of the Wisconsin WAP.
3. Sharp, Terry R.(1994) "The North Carolina Field Test: Field Performance of the Preliminary Version of and Advanced Weatherization Audit for the Department of Energy's Weatherization Assistance Program." Oak Ridge National Laboratory.
4. Dalhoff, Gregory K. (1996). "An Evaluation of the 1996 Iowa Low-Income Weatherization Program". Wisconsin Energy Conservation Corporation.
5. Pigg, Scott and Gregory K. Dalhoff. (1994) "An Evaluation of Iowa's Low-Income Weatherization Efforts." Wisconsin Energy Conservation Corporation.
6. Pigg, Scott and Gregory K. Dalhoff. (1993) "Estimated Low-Income Program Impacts in Iowa". Wisconsin Energy Conservation Corporation.