

M E M O

DATE: DECEMBER 11, 2012
TO: CUSTOM GAS EVALUATION TEAM
FROM: HASELHORST
RE: COMPARISON OF 2011 DESK REVIEWS WITH DECISION CRITERIA

This memo compares the results of the desk reviews of PY2011 participants with the Decision Criteria developed from benchmark sites (evaluated sites from PY2009 and PY2010 impact evaluations). Please refer to the previous memo for the development of the Decision Criteria.

Table 1 summarizes the statewide results. The Benchmark Value, No Action Range, and Weighting Factor are the previously agreed upon Decision Criteria values. Note that the savings fraction (previously 8.8%) and evidence of bills (previously 24%) criteria changed from the December 7th memo, due to a spreadsheet error. The PY2011 Desk Review Value is the result of the desk review of a sample of PY2011 project files. This column has been color coded to indicate where the value is within the ‘No Action Range’ (green) and where it falls outside of that range (rose).

Table 1. Statewide Decision Criteria Comparison

Benchmark	Benchmark Value	PY2011 Desk Review Value	No Action Range	Weighting Factor
Baseline is appropriate	75% of the time	79%	60% - 89%	40%
Savings method was appropriate	47% of the time	61%	38% - 57%	10%
Savings fraction	8.2%	6.8%	6.6% - 9.8%	10%
Document inventory	44% of documents found	42%	35% - 53%	10%
Evidence of bills in the file	35% in agreement	45%	28% - 42%	10%
Savings was reproducible	54% of the time	39%	43% - 65%	10%
Quality of the estimate	67% reasonable quality	71%	54%-81%	10%

These findings indicate that a significant change in practice is not indicated broadly enough to warrant another statewide impact evaluation. As noted in Table 1, four of the seven criteria fall within the ‘No Action Range’ including the most heavily weighted criterion characterizing the baseline.

The results were also examined by PA. Table 2 uses the same highlighting scheme to indicate whether a PA's results were within or outside of the No Action range. NSTAR is the only PA where the sum of the weighted criteria indicates further study is warranted.

Table 2. Comparison of Benchmarks by PA

Benchmark	Columbia			NGRID			NSTAR		
	2011	2010	2009	2011	2010	2009	2011	2010	2009
Quality of the estimate	80%	91%	53%	65%	72%	64%	78%	42%	18%
Savings method was appropriate	85%	54%	63%	47%	56%	52%	72%	25%	6%
Savings was reproducible	27%	66%	35%	23%	60%	42%	78%	51%	36%
Savings fraction	6.8%	10.4%	7.1%	6.7%	5.9%	7.9%	7.6%	11.6%	20.4%
Bills factored in estimate	71%	39%	95%	38%	19%	37%	42%	9%	85%
Baseline is appropriate	74%	76%	75%	78%	79%	75%	87%	69%	74%
Document inventory	47%	41%	46%	43%	51%	58%	48%	50%	63%

Because NSTAR is the only PA clearing falling outside of the 'No Action' range and the NSTAR indicators are fairly strong for the NSTAR PY2011 data set, no further statistical testing was conducted.

These conclusions are aligned with the PA reports of process changes. NSTAR reports a significant and definitive change occurred in the late 2010 timeframe. Prior to the change, the gas program manager conducted the savings estimate review; after that date, staff engineers were assigned responsibility to review custom estimates of savings. The other PAs have not identified any such sharp change in practice.

PY2011 DESK REVIEW RESULTS AND DECISION CRITERIA

This section presents a breakdown of the PY2011 desk review results. If the desk review results are sufficiently close to the benchmark result, they are within the "No Action Range" (color coded green), indicating that the savings methods have not changed. However, if the desk review results fall outside of the No Action Range (color coded rose), it implies significant changes to savings estimation methods.

The Decision Criteria are presented by each of the three largest PAs and statewide, consistent with realization rate reporting protocols.

Quality of the Savings Estimate

We proposed a Decision Criterion which compares the PY2011 desk review sites to the benchmark sites in the quality of the savings estimates. We have characterized that quality of the savings in three assessments:

- Quality of the overall estimate
- Appropriateness of the algorithm employed
- Reproducibility of the savings

Table 3 tabulates the reviewer's judgment of the quality of the engineering estimate. The engineer assigns the site estimate one of five grades, as indicated in Table 3. The values in the table are in therms and represent the portion of savings at the quality level indicated.

For criterion of this type, we have assigned responses as falling in the Green Zone – the quality is acceptable, or in the Red Zone - where it is not.

Table 3. Benchmark Results: Overall Quality of the Estimate

Quality of Estimate	Columbia	NGRID	NSTAR	State-wide
Native files, reasonable, some field measurements, clear documentation	344,962	383,996	1,070,660	1,882,564
Evidence of good estimation, but no native files to verify	514,348	1,848,103	56,695	2,431,111
Algorithm with some site based information, but poor assumptions	116,863	451,800	221,670	790,332
Use a fixed savings fraction with no site based data	69,545	86,473	230,766	386,784
No calculations apparent	178,988	1,335,474	148,011	1,677,470
Savings in the Green Zone	80%	65%	78%	71%
Desk review result is GREATER than	> 66%	> 56%	> 32%	> 54%
Desk review result is LESS than	> 99%	> 83%	> 48%	> 81%

The next table, Table 4, summarizes methods used to estimate savings and whether the method employed was appropriate for that measure.

Table 4. Benchmark Results: Applicability of the Estimation Method

Method Used is Most Applicable	Columbia	NGRID	NSTAR	State-wide
Building simulation	109,312	696,707	453,279	1,259,298
Proprietary method	250,000	305,050	0	555,050
8760 or bin spreadsheet	390,502	522,563	570,465	1,483,530
Factor driven, one-line calcs	291,285	391,127	225,003	1,086,192
No calculations in the file	178,988	1,335,474	148,011	1,677,470
Total savings	1,224,705	4,105,845	1,727,801	7,168,261
Sites in the Green Zone	85%	47%	72%	61%
Desk review result is GREATER than	> 45%	> 43%	> 18%	> 38%
Desk review result is LESS than	> 67%	> 65%	> 27%	> 57%

Table 5 summarizes how often the engineer assessed that the savings estimates could be reproduced.

Table 5. Reproducibility of the Applicant Savings

Savings is Reproducible	Columbia	NGRID	NSTAR	State-wide
Yes	332,318	897,238	1,317,262	2,614,714
Partial	0	55,766	36,461	200,717
No	892,388	3,139,117	374,078	4,434,936
Sites in the Green Zone	27%	23%	78%	39%

Desk review result is GREATER than	> 47%	> 43%	> 37%	> 43%
Desk review result is LESS than	> 71%	> 65%	> 55%	> 65%

Savings Fraction of the Billed Usage

The savings fraction is the ratio of the savings to the pre-installed weather normalized bills and provides another method for comparing tracked savings.

Table 6. Benchmark Results: Savings Fraction by PA

Savings Fraction	COLUMBIA	NGRID	NSTAR	Statewide
Number of measures	31	39	25	102
Tracking Savings	1,224,705	4,105,845	1,727,801	7,264,092
Pre-installed Billing	18,052,862	61,186,209	22,690,880	106,297,072
Tracking Savings Fraction	6.8%	6.7%	7.6%	6.8%
Savings Fraction NO ACTION				
Desk review savings fraction GREATER than	> 7.5%	> 5.2%	> 9.8%	> 6.6%
Desk review savings fraction LESS than	> 11.3%	> 7.7%	> 14.6%	> 9.8%

Billed Usage Factored in Savings Estimates

It is relatively easy to account for the few end-uses on the gas bill and to weather normalize weather dependent bills. For these reasons, most estimates should include an examination of the gas bills to at least sanity check the estimates.

This benchmark identifies how often gas billing appears to be factored into the applicant savings estimates, as shown in Table 7.

Table 7. Benchmark Result: Evidence of Bills

Project files with billing data	COLUMBIA	NGRID	NSTAR	State-wide
Referenced billing usage is reasonable	864,258	1,315,409	725,483	2,905,149
Applicant did not account for other end uses	-	-	-	-
Appear to be missing accounts or account mismatch	-	256,473	-	321,602
No reference billed use	97,404	2,489,214	937,190	3,633,717
Savings with billing usage	71%	38%	42%	45%
Desk review savings fraction GREATER than	> 42%	> 20%	> 14%	> 28%
Desk review savings fraction LESS than	< 63%	< 30%	< 21%	< 42%

Baseline is Appropriate

Changes to baselines by the evaluator accounted for about 5% of the 30% discrepancy in realization rate observed in the previous evaluations. A frequent source of the baseline change occurs when the applicant installs a large capital piece of equipment, such as a boiler, where code is the likely baseline.

Table 8 compares the applicant and evaluator identification of the baseline.

Table 8. Benchmark Result: Baseline Agreement

Baseline is appropriate	Columbia	National Grid	NSTAR	TOTAL
Agreement between applicant and evaluator	549,201	1,378,604	449,226	2,444,927
All Savings	737,215	1,772,144	515,744	3,107,997
Realization rate, 2009-2010	74%	78%	87%	79%
Desk Review No Action Range				
GREATER than	> 61%	> 62%	> 55%	> 60%
LESS than	< 91%	< 93%	< 83%	< 89%

Document Inventory

The applicant file contains copies of administrative documents, such as the application and offer letters and documents that support the savings, such as technical assistance studies. Table 9 tabulates the frequency with which typical documents appear in the project file delivered to the evaluation team.

Table 9. Benchmark Result: Document Inventory

Document	Columbia	NGRID	NSTAR	Statewide
Application	8	33	23	64
Technical assistance study	24	15	15	55
Customer offer letter	29	23	16	68
Cut sheets	6	5	6	17
Invoice	9	14	12	35
Post inspection	11	11	0	22
TOTAL	87	101	72	261
Savings Fraction NO ACTION				
Benchmark fraction	47%	43%	48%	42%
Document fraction GREATER than	> 35%	> 43%	> 43%	> 35%
Document fraction LESS than	> 52%	> 64%	> 64%	> 53%