Massachusetts Commercial/Industrial Baseline Framework

Massachusetts Program Administrators and Energy Efficiency Advisory Council

Report No.: Final Report
Date: April 26, 2017

Acknowledgments

This baseline framework is the result of a collaborative effort. The DNV team led by ERS wishes to recognize the following active participants in this document's development:

Scott Dimetrosky, Apex Analytics

Mike Sommer, Berkshire Gas

Gail Azulay, Cape Light Compact

Meredith Miller, Cape Light Compact

Margaret Song, Cape Light Compact

Monica Cohen, Columbia Gas

Jennifer Chiodo, Cx Associates

Alec Stevens, DMI

Jessica Baldic, DNV

Dan Barbieri, DNV GL

Chad Telarico, DNV GL

Wendy Todd, DNV GL

Andrew Wood, DNV GL

Sara Akashian, ERS

Kerry Burke, ERS

Manuel Fontan, ERS

Sue Haselhorst, ERS

Jon Maxwell, ERS

Betsy Ricker, ERS

Erik Mellen, Eversource

Tracy Redmond, Eversource

Ryan Willingham, Eversource

Sam Alpert, GDS

Mike Rufo, Itron

Dave Jacobson, Jacobson Energy

Laurie Pereira, Liberty Utilities

Tricia Walker, Liberty Utilities

Greg Brown, Navigant

Marie Abdou, National Grid

Bill Blake, National Grid

Whitney Brougher, National Grid

Tony Larson, National Grid

Ezra McCarthy, National Grid

Mark Sevier, National Grid

Ralph Prahl, Ralph Prahl Associates

Bing Tso, SBW

Mary Downes, Unitil

Table of contents

Tables and Figures

1 PURPOSE

This document articulates a statewide framework for evaluators to consistently characterize the baseline of commercial or industrial (C&I) measure selected for evaluation measurement & verification (EM&V) in an impact evaluation. The guidance applies to both electricity- and natural gas-saving measures, both prescriptive and custom measures, and all Massachusetts (MA) Program Administrator (PA) programs. The focus is on characterization, that is, assessing the appropriate baseline system type as part of an impact evaluation. Judgment regarding hours of use, load factors, and related considerations is not included in this document. The document also does not provide specific technology baselines.

The primary audience for the document is C&I evaluators: Independent evaluation contractors, PA evaluation staff and advisors, and those who oversee evaluation on behalf of the Energy Efficiency Advisory Council (EEAC). PAs' C&I implementation staff and contractors who develop savings estimates using baseline estimates are a secondary but important audience, as their project will be assessed using these guidelines. Residential program evaluators have contributed to its development and it is hoped that they will follow a common protocol, although it is not their obligation to do so.

2 BASELINE DEFINITION

Baseline is the condition that would have existed absent the installed measure. [1] It is characterized differently depending on whether the measure's combination of technology and application is unique or not. Most common measures offered in a prescriptive program and higher-volume commodity measures offered through the custom path are non-unique.

Defining baseline for non-unique measures. MA program electricity savings claims must be consistent with the requirements of ISO New England for the bidding of demand reductions into the Forward Capacity Market. ISO New England requires that baseline for efficiency projects be defined by the applicable efficiency code or standard or "industry standard practice" (ISP) if there is no code or standard or if the ISP is more stringent than the relevant code or standard[2]. The MA C&I evaluators adopt this definition for defining both electric- and gas-saving measure baseline for non-unique measures (defined as those with relevant code, standard, or ISP).

This framework defines ISP as the equipment or practice specific to the application or sector that is commonly installed absent program intervention. This requires theoretical judgment, as market behavior without programs cannot be observed directly. In many cases, the ISP can be characterized as the most commonly installed relatively lower efficiency option without program intervention.

To clarify the definition by exception, ISP is not defined by singular use of any of the following:

- "Mean efficiency installed," which could be a relatively high efficiency option for measure in mid-stage technology adoption, and/or which could reflect program intervention effects

[1] What the customer would have done absent the measure differs from "what that customer would have done absent the program." The latter would draw in free-ridership consideration.

[2] Section 6 of the *ISO New England Manual for Measurement and Verification of Demand Reduction Value from Demand Resources Manual M-MVDR* Available at http://www.iso-ne.com/participate/rules-procedures/manuals. This framework interprets that "applicable" could in some cases include pre-measure existing conditions.

- “Most commonly installed” option for the entire market, for the same reasons
- “Least cost” option, which could be less efficient but rarely if ever used
- “Least efficient” option possible, which could cost more than a more efficient option and therefore not be a plausible baseline or could be so cost-ineffective that the market does not recognize it as a viable option; or
- What a particular customer otherwise would have done without program intervention, for two reasons: (1) Using such would risk incorporating the effects of free ridership into what is intended to be a gross, not net, savings estimate and (2) ISP should represent the market, not a particular customer’s choice.

Some of these characteristics are likely to be true for a particular measure’s baseline, but no one of them always defines it. Project-specific exceptions to the use of industry standard practice are permissible when particular circumstances render standard practice irrelevant and evidence is provided to justify it.

Defining baseline for unique measures. The ISO’s definition does not cover unique technology-application combinations for which there is neither code, standard, nor broader market from which to establish an ISP. For unique applications the baseline must directly reflect what the particular customer otherwise would have done absent the measure.

The designation of a measure as being “unique” has ramifications described later in this framework and bears a burden of demonstration. The guiding principle is that unique measures have a very small or no standard, recognizable market. The evaluator should explain the technical characteristic(s) that make the combination unique and note unsuccessful efforts to identify the market. A measure-application combination that saves energy and is larger, faster, more precise, at higher pressure, cleaner, uniquely controlled, or fits into a smaller space than has been applied elsewhere in the market before is likely to be unique. Examples of typically unique measures include: A comprehensive integrated project, most industrial process measures (excluding balance-of-plant equipment such as compressors and boilers), and retrofit of an HVAC system from air-cooled to water-cooled where the baseline is pre-existing condition. Examples of characteristics that by themselves do not define a measure as unique include: custom-built equipment, unusual interconnections, or uses of an emerging technology. Examples that are not unique include: A custom-sized and configured process boiler economizer, an industrial compressed air system upgrade, and a packaged CHP system.

Defining baseline when it changes over time. Baseline can impact both first-year and lifetime savings and realization rates. In the case of measures that replace working equipment, <u>the evaluator must consider the baseline both at the initial time of replacement system specification and at the time the replaced working equipment would have reached end of life</u>. This dual baseline principle accounts for lifetime savings over the participant population.

3 KEY MASSACHUSETTS DOCUMENTS RELATED TO BASELINE[3]

Massachusetts has not previously developed a consistent statewide baseline policy for evaluation. However, the Energy Efficiency Advisory Council (EEAC) consultants issued a memo to the Council, *Improving Savings Estimates for C&I Custom Projects – Draft*[4], in February 2016 that featured baseline guidance based on evaluation experience in the state, which was followed by the MA Program Administrators and the EEAC joint white paper, *Approach to Baseline Determination* in April 2016.

National Grid maintains a table of ex ante assumptions regarding measure-specific baseline for certain technologies installed through the custom track that is periodically updated. The current version as of the writing of this Framework is *the Baseline Document: Massachusetts & Rhode Island, 2015*.

The Massachusetts Technical Reference Manual (TRM) assumes a first-year baseline for each measure and implicitly assumes a more efficient later-year baseline that is taken into account by discounting the effective measure life for early replacement or retrofit measures, an issue discussed in more detail in Section 4.4.1. The MA TRM does not give specific baseline information for any custom measures.

Chapter 3 of the *Savings & Evaluation Methodology for Codes and Standards Initiative*, which was submitted to MA Department of Energy Resources (DOER) on behalf of MA PAs on October 20, 2015, addresses evaluation of code compliance, stretch codes, and baseline ramifications of code compliance initiatives.

4 BASELINE CATEGORIES AND PROTOCOLS FOR SELECTION

The first step in characterizing baseline is to classify the event type for the measure, as a function of conditions at the time of installation. This framework defines five different baseline categories that are introduced in Table 1.

---

[3] Documents referenced in this section with web availability are: Massachusetts TRM: http://ma-eeac.org/wordpress/wp-content/uploads/2016-2018-Plan-1.pdf and *Improving Savings Estimates for C&I Custom Projects – DRAFT*: http://ma-eeac.org/wordpress/wp-content/uploads/MA-CI-Custom-Methods-memo.2016.03.04.pdf. The 2014 version of National Grid's *Baseline Document: Massachusetts & Rhode Island* is available at https://www.nationalgridus.com/media/trade/2014-baseline-document-for-ma-and-ri.pdf. The memorandum, Approach to Baseline Determination in April 2016, is not available on-line.

[4] http://ma-eeac.org/wordpress/wp-content/uploads/MA-CI-Custom-Methods-memo.2016.03.04.pdf

Table 1: Baseline Event Types

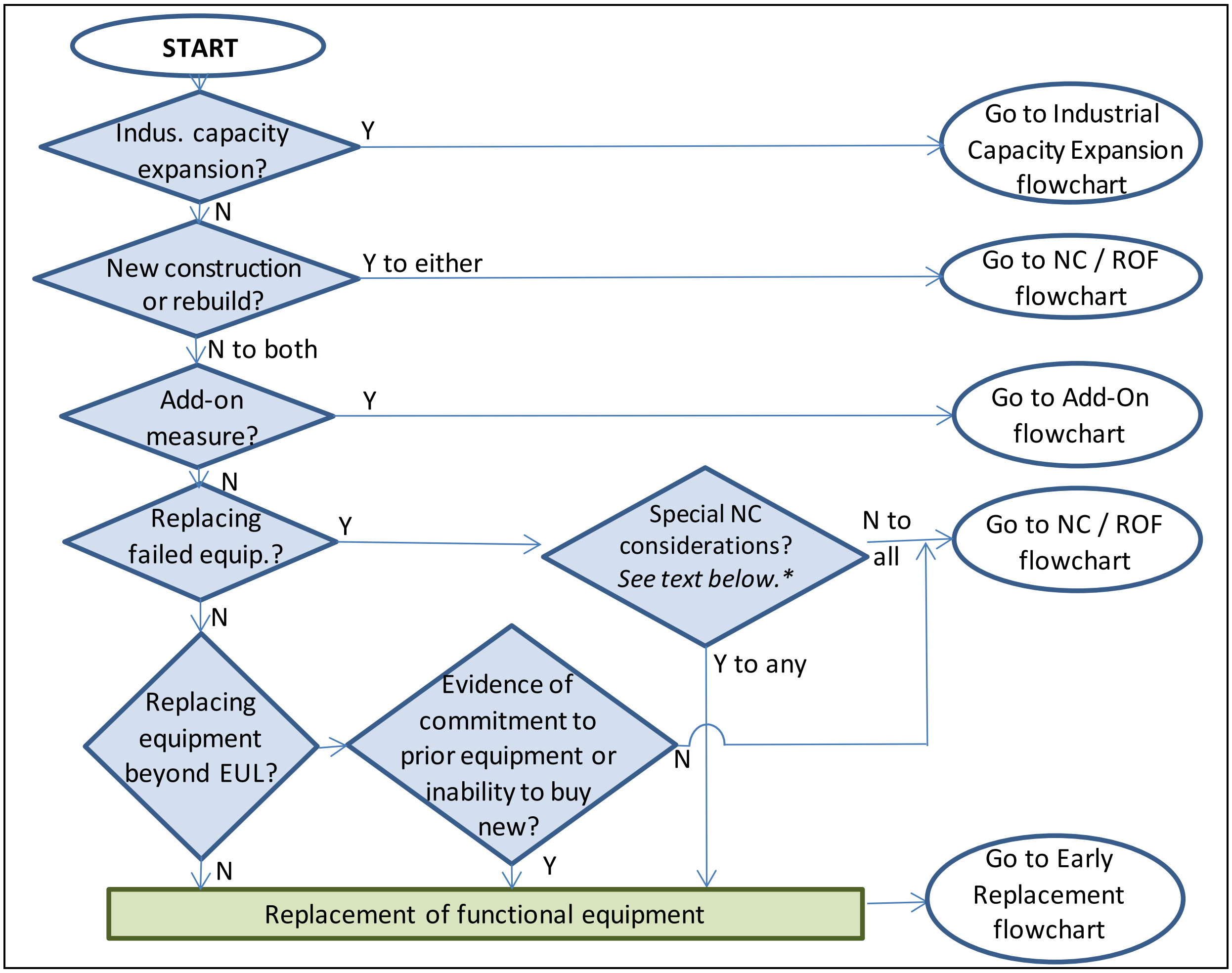
Event Type *Alternate terms used*	**Application**	**Examples**
New construction (NC) **or major renovation** *"lost opportunity" in MA TRM*[5]	For any measure installed at the time of new ground-up facility construction, or as part of a major renovation. Manufacturing capacity expansion is a special subset of this category with unique considerations.	A building design team specifies a high efficiency rooftop unit (RTU) rather than the industry standard practice, which is minimally code-compliant RTU. *or* A leased building receives a gut renovation. Current code and ISP define baseline for systems affected by the renovation, not the conditions prior to gutting.
Replace on failure (ROF) *"end of useful life" and "lost opportunity" in MA TRM. Also "market opportunity" "natural replacement" "natural turnover" "normal replacement" "replace on burnout"*	For any measure that is installed in response to the failure of a previously functioning system. ROF also applies if a customer is compelled to replace an old working system. The ROF baseline applies in this case even if the old system's performance had deteriorated.	A restaurant's oven fails. The owner buys a new efficient convection oven. The baseline is the efficiency of a new less-efficient convection oven, not the failed oven. *or* A working 60-year old boiler requires increasingly expensive yearly maintenance or can no longer meet emissions requirements. The owner replaces it with a new unit. A new boiler is the baseline, not the old boiler.
Add-on	Applies to measures that improve the efficiency of an existing system but do not replace it. In most cases the baseline is the preexisting system without the efficient measure. Some add-on measures require dual baseline consideration.	Pre-rinse spray valves, demand controlled ventilation, energy management systems (EMSs), pipe insulation measures, and variable speed drives installed on previously constant speed systems all are examples of add-on measures.
Early replacement (ER) with remaining useful life (RUL) consideration *"retrofit" in MA TRM*[6] *"early retirement" "accelerated replacement" "advancement"*	ER means the replaced system was fully operational. With RUL consideration means that the replaced system had a definable remaining life and the baseline efficiency for that system is certain to be different (usually higher) at the end of that remaining life than it was when it was replaced.	It is 2018. A customer replaces an operational 10-year old rooftop unit (RTU) with a high efficiency RTU. RTUs have an average effective useful life (EUL) of 15 years. The federal efficiency standard increases in 2023. The baseline is the replaced RTU's efficiency for the remaining 5 years of measure life, and the 2023 code efficiency in the 10 years after that.
Early replacement – without remaining useful life consideration	ER means replaced system was fully operational. Without RUL consideration means that either the replaced system had no definable period for end of life or the baseline efficiency for that system is not expected to be different at the end of its remaining life than it was when it was replaced.	A high efficiency fractional hp motor replaces a working motor installed in 2016. There is no standard that is expected to increase the minimum efficiency of this equipment compared to the pre-existing efficiency, so the baseline efficiency is that of the replaced motor for the entire measure life.

The flowchart in Figure 1 illustrates the basic decision-making regarding baseline event type. More detailed decision-making discussion and flowcharts follow in subsequent sections.

[5] The MA TRM Glossary defines "lost opportunity" as "a measure being installed at the time of planned investment in new equipment or systems. Often this reflects either new construction, renovation, remodelling, planned expansion or replacement, or replacement of failure."

[6] In most cases. The MA TRM Glossary defines "retrofit" as "the replacement of a piece of equipment or device before the end of its useful or planned life for the purpose of achieving energy savings. 'Retrofit' measures are sometimes referred to as 'early retirement' when the removal of the old equipment is aggressively pursued." While the TRM typically associates retrofit with RUL-type measures, there are some measures where the term is used in the context of measures without RUL or discounted EULs.

Figure 1: General Event Type Selection Flowchart



** Special considerations for repair over replace: Repair cost <20% of replacement cost, has run in failed mode >2 yrs., or other preponderance of evidence that repair is a viable option.*

It can be challenging to distinguish between two pairs of the major categories in Figure 1: Renovation versus early replacement, and replace on failure versus early replacement. The next two paragraphs address these decisions.

Renovation versus early replacement. The choice to rebuild a facility or major system typically forfeits the grandfathered status of facility systems relative to current codes and ISP. If a portion of a building is subject to gut rehabilitation, the evaluator should follow the NC/ROF logic flowchart, as the local MA building department reviews the project in the context of current code. The preexisting systems are not the baseline. There is one exception to this general rule: If the sole purpose of the event that triggers the loss of grandfathering is the efficiency measure itself, then the preexisting condition is the baseline and the measure baseline should be determined using the early replacement flowchart. This is because the baseline at the time of the decision-making was in fact the preexisting condition.

Two examples: (1) A customer decides to renovate the majority of an older office building before new tenants move in. The renovation includes refinishing several surfaces and replacing energy-using systems, including lighting and HVAC, at the same time and as part of the same overall project. The customer receives an incentive for high efficiency lighting. The NC baseline applies to the lighting project, not the pre-existing condition. (2) A customer decides to replace all of the lighting fixtures in a 10-year-old building and nothing else, and receives an incentive for the upgrade. For building officials, according to IECC 2015 this triggers application of current code requirements regarding lighting power density (LPD). For energy efficiency program evaluation, however, this is ER not NC/ROF, because the measure itself is what triggered application of the new code.

Replace on failure versus early replacement. When deciding whether a measure is ROF or ER, the evaluator must consider the plausibility of the customer continuing operations in the pre-retrofit state. The framework directs the evaluator to use the preponderance of evidence basis to determine if early retirement is a viable baseline. "Preponderance of evidence" means that "the greater weight of evidence" favors one condition or the other.[7] The quality of evidence is more important than the volume of it. Figure 1 illustrates the two circumstances where the approach is applicable: When assessing whether or not failed equipment was reparable, and when replacing equipment that was still functional but in excess of its EUL. Table 2 provides examples of evidence that support determination of each of the two conditions.

[7] For further discussion on this methodology, see *Early Retirement Using Preponderance of Evidence*, v1.0, July 16, 2014. Prepared for the California Public Utilities Commission. Downloaded 9/1/16 from www.energydataweb.com/cpucFiles/pdaDocs/1035/Project Basis_EUL-RUL _ Evidence v1.docx.

Table 2: Evidence of Replace on Failure and Early Replacement Conditions

Evidence of Early Retirement	Evidence of Replacement on Failure
Evidence that the prior equipment was functional such as: • A video of the equipment in operation before removal • Dated maintenance logs that describe functionality • Customer interview by evaluator during which functional operation is confirmed • Documentation by implementer describing functionality	Evidence that the prior equipment was not functional such as: • Dated maintenance logs that describe failure • Customer interview by evaluator during which lack of functional operation is confirmed • Documentation by implementer describing lack of functionality
Documentation that shows the replaced equipment was less than 2/3 through its standard EUL, such as: • A photographed nameplate with date of manufacture • Equipment logs showing installation date • Invoices showing the purchase date • Catalog data showing the years the model number was offered to the market The facility age, for newer sites	Prior to measure implementation the replaced equipment was facing a repair, and that customer describes the prospective repair cost as being significant (>10%) relative to the replacement cost.
The replaced equipment was beyond 2/3 of its EUL but there is documented evidence of commitment to long-term maintenance to the prior equipment, such as: • Invoices showing marked investment (>10% of replacement cost) on maintenance in the period prior to replacement • Copies of maintenance logs showing major maintenance soon before replacement • Photographs of recently installed major new components prior to system replacement (e.g., re-tubed boiler) • An ISP study includes life research and shows that for this particular customer's market, EULs are markedly longer than the standard EUL Relevant documented formal corporate policy on repair/replacement practices.	Evidence of actual or impending expected catastrophic failure of equipment that is less than 2/3 through its EUL, such as: • Interview description • Maintenance logs showing gradually escalating repair costs • An ISP study includes life research and shows that for this particular customer's market, EULs are markedly less than the life of the replaced equipment Relevant documented formal corporate policy on repair/replacement practices
The replaced equipment was beyond 2/3 of its EUL but there is documented evidence of facility's inability to make the capital commitment necessary to replace it, even if major repairs are needed, such as: • Corporate/institutional information showing history of deferred investment	Documentation showing the replaced equipment was more than 2/3 through its standard EUL and there is no exceptional evidence as described in text to the left.
The facility managers had an inventory of back-up equipment similar in efficiency to the removed equipment that they could have used to replace the old equipment had it failed	The evaluator observes that other similar equipment on-site is conspicuously newer than is typical and reflects a policy of proactive replacement before failure *(valid as corroborating evidence only)*.
	There is compelling evidence that this particular customer's market functions with relevant equipment EULs that markedly differ from the standard.
	Simple payback calculations show that the benefit of replacing the old equipment with new baseline equipment is compelling: The annual savings approaches cost of replacement or incremental cost compared to repair.
	The system has a primary and back-up design (e.g., pumps), and had been operating with one of the two components broken.

Since this is a propagation of error-based framework, the numeric values in the table should be considered indicative, not binding.

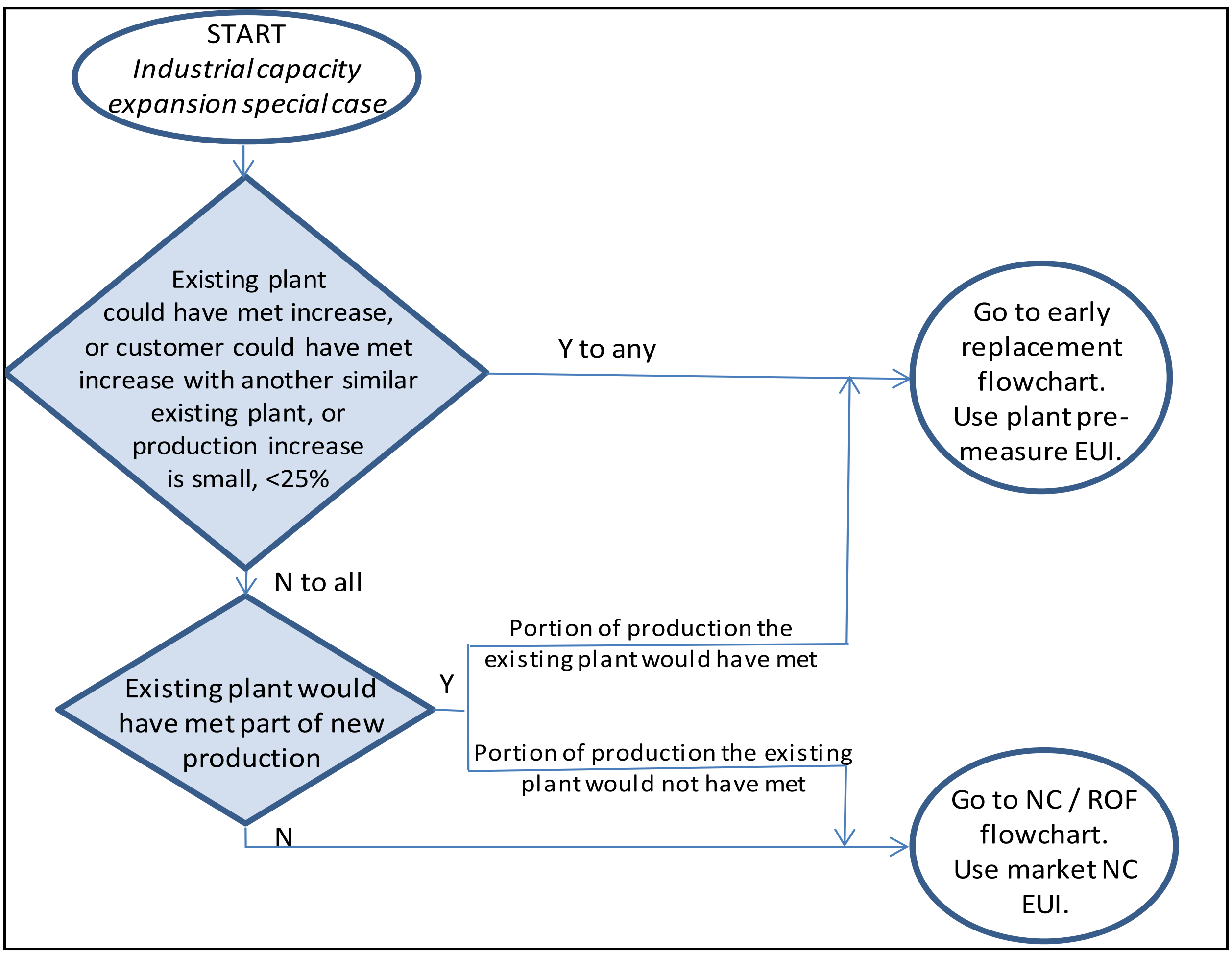
Note that some of these elements can only be documented during the application process, typically by the PA. If there is no evidence for either ROF or ER, in most cases the evaluator should define the baseline using the ROF condition.

Exceptions to the ROF default baseline need not be demonstrated site by site if the program delivery mechanism or technology is inherently associated with an existing condition baseline and this is determined at the program or ISP study level before site-specific study begins. For example, evaluation of a direct install program with predominantly cold-call marketing may not require that the evaluators demonstrate a preponderance of evidence in favor of existing conditions at each evaluated facility.

4.1 Industrial Process Expansion

Industrial process measure baselines can be challenging to classify regarding the new construction/early replacement decision when capacity expansion is involved. If the expansion is small (less than nominally 25% of production capacity), or if the added capacity could have been realistically reached by reconfiguring the plant, adding shifts, or similar, or if the customer could have met the capacity by increasing production at another similar plant, then the individual customer's pre-project energy use per unit production or equivalent is the baseline. Otherwise, the new construction market-wide baseline governs for the portion of capacity the existing facility would not have met. Figure 2 illustrates the logic.

Figure 2: Industrial Process Capacity Expansion Event Type Selection Flowchart



The baseline (and as-built) efficiency for industrial measures related to changes in production almost always should be expressed in normalized terms: Energy use per unit production. Evaluators should use the post-installation production rate[8] to annualize the use.

4.2 New Construction and Replace on Failure

For new construction measures the relevant code[9] or regulated standard defines the baseline unless research finds that a preponderance of evidence exists to the contrary.[10] The requisite evidence may vary but generally should follow the protocols described in Section 6 of this document. When standard practice is less efficient than code[11] and code compliance program efforts exist, installation program evaluators should use code as the baseline, as other code compliance efforts in the state are expected to claim savings associated with raising performance to code levels. When standard practice is worse than code, and no code compliance program efforts exist, installation program evaluators should use standard practice (below code) as the baseline. When standard practice exceeds code, the higher level of efficiency is the baseline. These scenarios most often occur when intermediate efficiency levels are present in the market.[12] If they are found, the evaluator should consider commissioning a separate market effects study to investigate possible spillover. The higher-level baseline typically is a blended market average of either discrete efficiency increments (e.g., market shares of CEE efficiency tiers) or weighted averages of values in a numerical continuum (e.g., LPD). For prescriptive measures and non-unique custom measures, it is preferable that the baseline be determined through a market-level ISP study. When this has not been done in advance and is necessary because prior code research or measure-specific evaluation uncovers a likely market deviation from code, a lower-level rigor study or other exceptional approach may be necessary as part of a sample site evaluation. The options for this are discussed more in Section 6. Figure 3 illustrates the logic.

[8] In most cases the post-installation production rate should be based on "observed" production after any commissioning. In exceptional circumstances, the observed production rate may be modified to reflect projected long-term production rates. This exception must be supported by evidence.

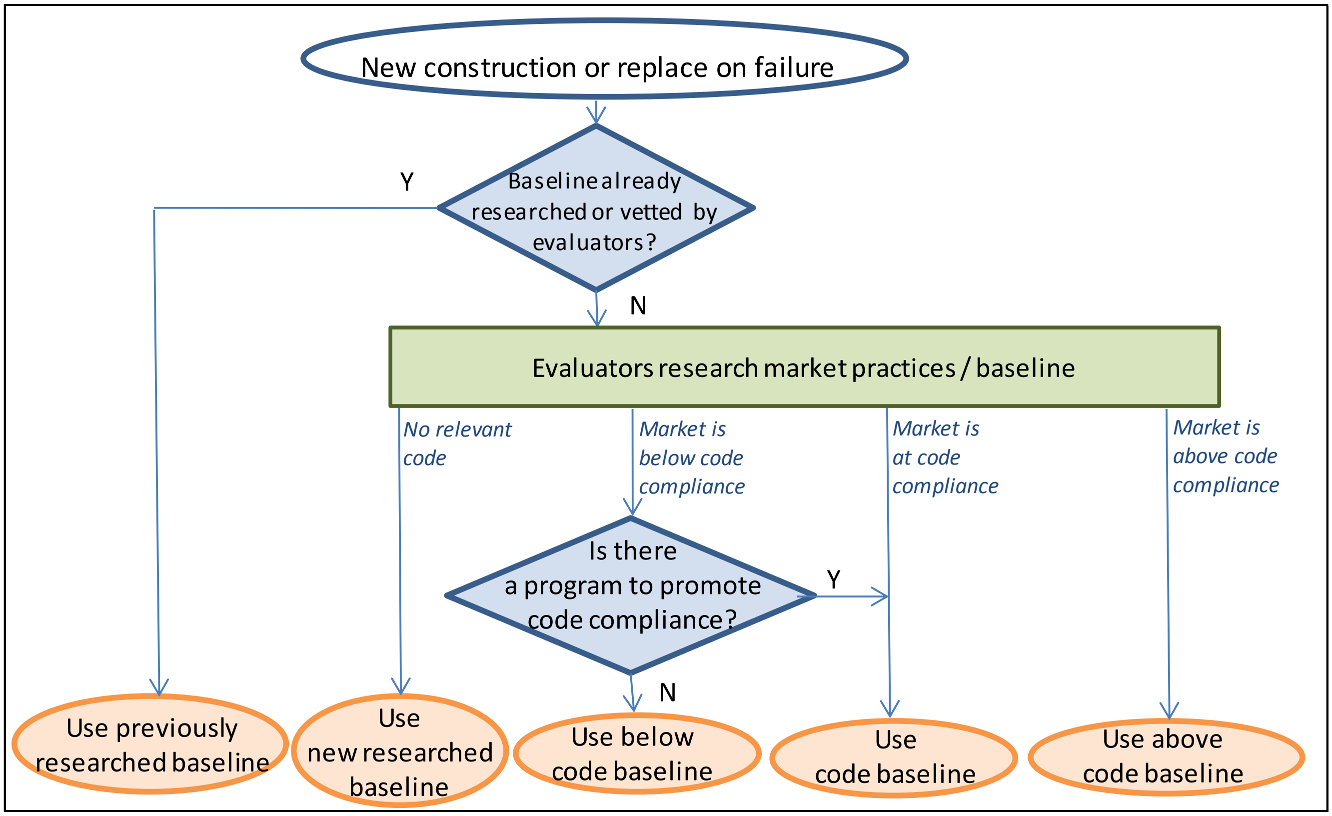
[9] If a Massachusetts city or town has adopted a more efficient "stretch" code, the statewide code and not the stretch code is the baseline, due to prior state policy decision-making.

[10] The evaluator must vet the research if it was not evaluator-led. Massachusetts code compliance research has found such evidence in recent research. For an example of ISP exceeding code, see *Upstream Lighting Program Baseline Adjustments for New Commercial Construction LPD*, memo from Ari Michelson and Sue Haselhorst, ERS, to MA PA Research Team and the EEAC Evaluation, Measurement & Verification (EM&V) Consultants, July 15, 2016. Recent MA research on daylighting controls has found evidence of market-wide noncompliance.

[11] Lack of code compliance typically would lead to efficiency levels that are below code. "Lack of code compliance does not always mean that manufacturers or retailers are breaking the law. For example, some codes and standards apply a phase-in where they allow for a sell-through period of existing stock. In these cases equipment may not meet existing standards but may be legally sold, sometimes for a number of years (e.g., EISA 2007 prohibits the import and manufacturer of certain incandescent lighting products, but not the sale of such products)."

[12] "Intermediate efficiencies refer to efficiency levels that exceed federal or state code, but are below program eligibility. This would typically be applicable for programs that incent only higher tiers of efficient measures (e.g., ENERGY STAR most efficient appliances, furnaces at or above 95% AFUE). In these cases, research would need to be conducted to determine the efficiency levels for units that are sold outside the program, and the baseline should reflect this blended average. Blended baselines due to intermediate efficiency levels are expected to exceed federal or state codes." From *Approach to Baseline Determination White Paper*, prepared by the MA PAs and EEAC, April 15, 2016.

Figure 3: Measure Type Flowchart – New Construction/Replace on Failure[13]



The logic used to assess ROF is the same as the new construction baseline, but savings and baseline cost conclusions can differ if, for example, the project does not trigger application of newer energy code requirements.

Other NC/ROF considerations:

- **System context.** NC projects require consideration of context that is not relevant to other measure types, particularly for HVAC system baseline characterization. When assessing system baseline, the evaluator must consider the building's market and design context and configure the baseline accordingly.

  Example: If an application claims high efficiency due to use of an efficient water-cooled central hotel cooling system and uses a packaged air-cooled system as baseline, the evaluator must assess whether or not the packaged system was a viable option for the building's overall design. It may require judgment regarding the physical viability of ductwork installation, and whether or not this clientele in this type of market would bear the extra compressor noise or the aesthetic of packaged units hanging out of each wall. If not, a less efficient central system may be the baseline instead of an entirely different system type.

- **Code compliance path.** If the host facility uses a comprehensive building design or performance rating basis for code compliance (e.g., ASHRAE 90.1 Appendix G), then the baseline too should be defined in this context.

[13] Expanded from *Approach to Baseline Determination White Paper*.

Example: It is inappropriate to credit savings for low LPD in one area of a building if another area of the building has a high LPD and they average out to code compliance. The baseline would need to be based on the whole-building LPD.

- **Timing**. The evaluator's objective is to characterize the baseline code, standard, or standard practice at the time the decision was being made on the measure-related system's specifications. For new construction and major renovations typically the most appropriate, identifiable, replicable, and defensible point in time to use as the basis is the time the building permit is pulled.[14] If that date cannot be identified or is not relevant,[15] the baseline should consider conditions in effect no later than the published date of the technical assistance study, or if that is not applicable, the initial date of the project application. If the measure efficiency changes through the iterative design process, the appropriate date to use is the one where the efficiency last changed due to non-program factors.
- **Non-regressive.** For site-specific evaluation of ROF measures, a regressive baseline generally is not allowed, that is, the installed measure's baseline should be at least as efficient as the efficiency of the system it replaces, even if ISP indicates a lower baseline. There are three specific allowable exceptions: (1) If restaurant cooking equipment is replaced as part of gut rehabilitation, the pre-existing equipment efficiency need not be considered the minimum standard. (2) If a variable frequency drive (VFD) replaces a failed VFD but the prior drive is documented to have failed more than two years prior to replacement and the system is documented not to have been controlled via speed modulation during this time, then the non-VFD baseline is allowable. (3) The preexisting condition was program-funded. Other exceptions may be allowed, with a high threshold required to demonstrate plausibility.
- **Fuel switching**. If evaluating a ROF project with fuel switching in aggregate the evaluation must account for the net impact of the measure on all fuels. [16,17] If both electric and gas PAs claimed savings for separate aspects of the measure, evaluation will require the use of two baselines. The first baseline is the new code, standard, or ISP baseline for the measure with the old fuel, as if there had been no switch to a new fuel. The second baseline (which is also the theoretical as-built basis for the original fuel savings) is the new code, standard, or ISP baseline with the new fuel.

  Example: An electric resistance process heating system fails. Instead of replacing it with the same or with an indirect-fired gas system at 80% efficiency, the customer installs a direct-injection burner at 100% efficiency. The first baseline, for the electric PA application, is a new electric resistance system. The second baseline, for the gas PA, is the 80% efficient gas system. The electric program savings is based on the switch from electricity to an 80% efficient gas system. It also must account for the increase in gas usage or negative gas savings with that switch. The gas program savings is based on the upgrade from an 80% to a 100% efficient gas system. As with any measure there could be auxiliary impacts on other fuels associated with the upgrade from a lower to higher efficiency system – i.e., an air-to-air heat exchanger, which saves gas heating fuel, might have a small electric penalty to run an additional blower.

[14] If there are multiple permits pulled, use the date of the one most relevant to the measure. For example, if the measure relates to the HVAC system, and there are separate demolition, foundation, structural, and balance of building permits, then the last one is the most relevant, even though the applicant may have been considering HVAC options at earlier stages.

[15] Examples could include a measure implemented in stages over a long period, or a targeted renovation that is too small to trigger a requirement for new code compliance.

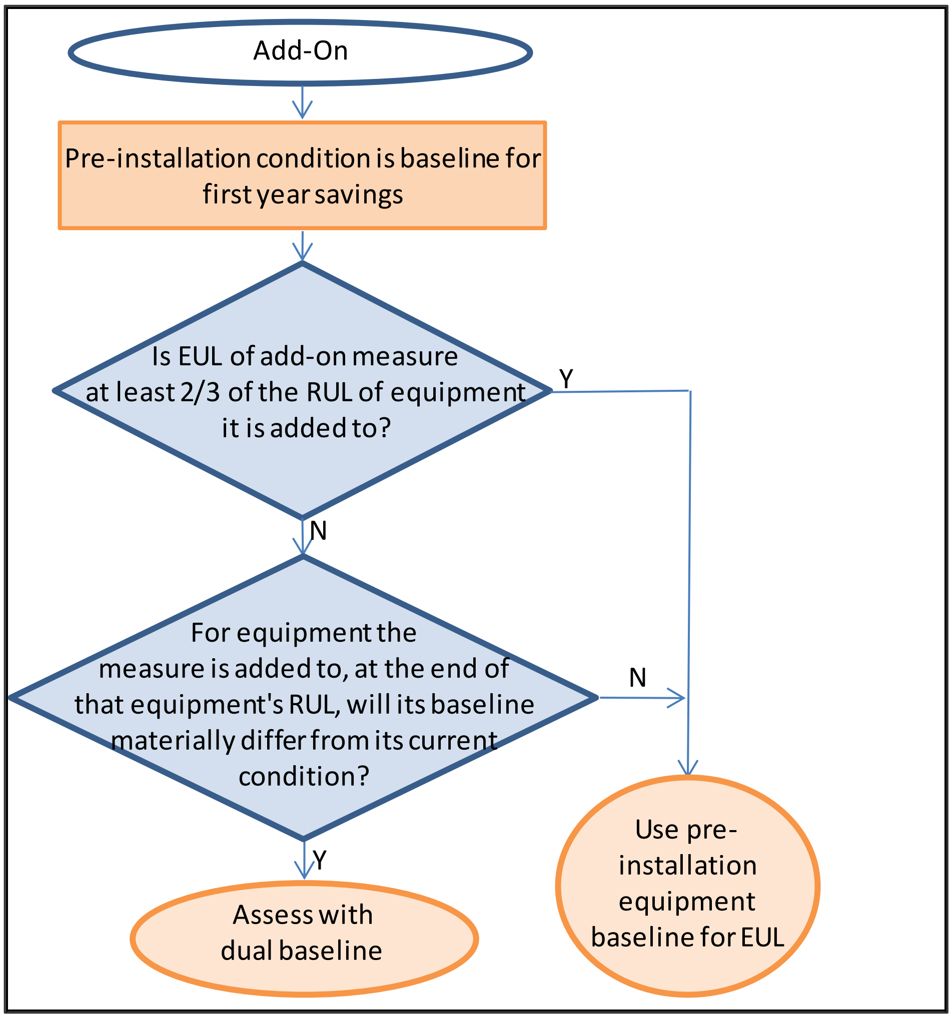
[16] For reference, see the memo *Fuel-switching policies, cost-effectiveness, savings and cost tracking*, from EEAC Consultants to PA CI Governance Team, April 22, 2010.

[17] Exception: If the customer was obliged to make the fuel switch for reasons unrelated to the program, then the baseline is defined as the ISP for the measure with the new fuel.

4.3 Add-On Measure

For measures that are added to existing systems the preexisting system is the baseline for at least the first-year savings, after which a dual baseline (discussed below) may be necessary. If the measure is an operational improvement it also is considered an add-on measure. The evaluator must consider if the measure life of the equipment that the efficiency measure is added to affects the long-term savings, as illustrated in Figure 4.

Figure 4: Measure Type Flowchart – Add-On Measure



As shown in the figure, the evaluator must consider whether the system that receives the add-on measure has less of a remaining life than the add-on measure has full life, and if that will have a material effect on lifetime savings. The issue is most relevant for controls-related add-on measures. If the energy use of the equipment that is being controlled will predictably decline in the future for reasons unrelated to the add-on measure, then the lifetime savings should be adjusted accordingly.

Examples:

(1) A VFD is installed to control an older 5 hp motor with an 89% efficiency. Once the existing motor fails, the new motor efficiency baseline will be 89.5%. This is not materially different and need not be considered.

(2) Lighting controls (currently EUL=9 years for retrofit) are installed to control old T-12 fixtures with a 5-year RUL. In 5 years, new T-8 fixtures will be the baseline, reducing the power by 30%. The evaluated lifetime savings is based on 5 years of full savings and then 4 years at 30% less savings.

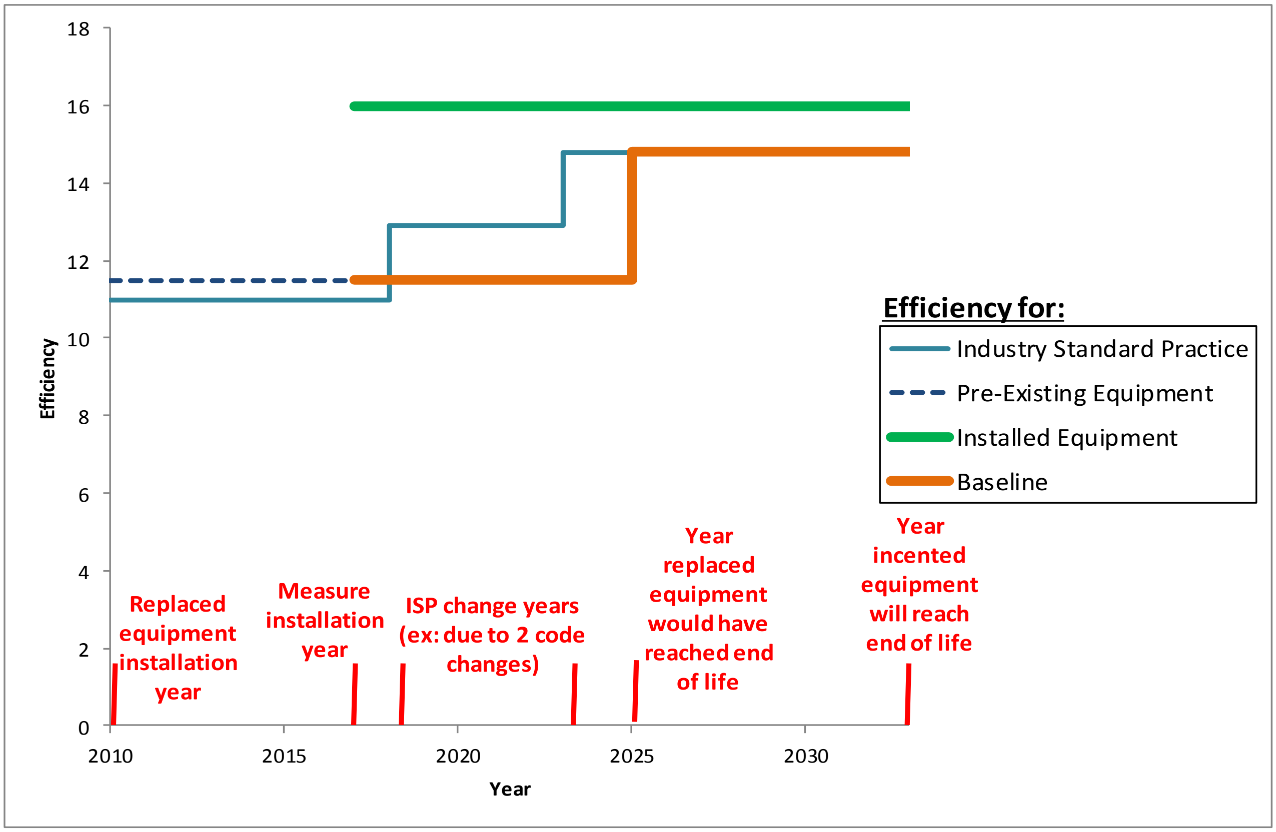
4.4 Early Replacement

For early replacement projects, the pre-installation condition is the baseline for at least the first-year savings. The evaluator must assess whether or not savings will vary as a function of time due to changing codes, standards, and/or ISP after the remaining useful life of the replaced equipment has passed. If so, dual baseline principles apply. This is the principle decision to address within the early replacement category. Dual baseline first is defined, and then the corresponding evaluation flowchart is introduced and reviewed.

4.4.1 Dual Baseline

Evaluators assessing early replacement measures that meet the criteria for dual baseline must consider both the market baseline at the time of measure installation, and the projected baseline at the time the replaced equipment would have naturally failed. Figure 5 illustrates the dual baseline principle for a hypothetical packaged rooftop air conditioning unit. In the example, the original unit with an 11.5 EER was installed in 2010 (indicated by the dark-blue dotted line), and replaced with 16.0 EER equipment in 2017 (the solid green line). For the first several years of measure life, the first period, the baseline is of the efficiency of the replaced equipment, 11.5 EER (shown as the solid gold line). By the time the preexisting equipment would have reached its natural EUL, 2025, the industry standard practice (solid blue line) will have increased to 14.8 EER due to increases in federal efficiency standards. If the program had not existed, therefore, the customer would have been obliged to purchase at least a 14.8 EER unit at that time. This becomes the baseline for later years, the second period, and is why the orange baseline line increases. It remains the baseline for the rest of the measure's life. The two-tiered orange line should be used as the baseline to calculate lifetime savings. The difference between the green and gold lines represents the savings.

Figure 5: Example of Dual Baseline



The 2016–2018 Program Years Plan Version of the Massachusetts Technical Reference Manual (TRM) recognizes the dual baseline phenomenon. For ER with RUL measures[18] ("retrofit" in the TRM), the manual in most cases discounts the measure's effective useful life when compared to a similar measure offered as an ROF/NC measure. The MA Dual Fuel Custom Screening Tool does as well[19]. The TRM recognizes the possibility of either retrofit or lost opportunity scenarios for custom measures, but does not list specific measure lives.

4.4.2 Determining if a Project is Single or Dual Baseline

Dual baseline should be considered for early replacement. Early replacement measures should be evaluated as single baseline measures without dual baselines if the preexisting equipment likely would have been used over the full EUL of the measure had it not been replaced with the incentivized equipment. Specifically, a single baseline should be used for:

[18] Commercial measures with both full EULs for "lost opportunity" and discounted lives for "retrofit" are: lighting systems and controls, enthalpy and building EMS controls, compressed air low pressure drop filters, zero loss condensate drains, motors, low-flow faucets and aerators. TRM also recognizes the possibility of dual baseline for custom measures.

[19] The MA TRM Glossary defines retrofit as "the replacement of a piece of equipment or device before the end of its useful or planned life for the purpose of achieving energy savings. "Retrofit" measures are sometimes referred to as "early retirement" when the removal of the old equipment is aggressively pursued." It defines "lost opportunity" as "a measure being installed at the time of planned investment in new equipment or systems. Often this reflects either new construction, renovation, remodeling, planned expansion or replacement, or replacement of failure."

Measures for which there is compelling evidence of commitment of long-term maintenance to the prior equipment. Evidence could include documentation such as invoices or observation of new parts from a recent overhaul or other substantive investment in the removed equipment (e.g., re-tubing a boiler two years previous) or maintenance logs that illustrate the system was maintained to a "like new" condition, for example. Interviews alone may not be sufficient. The topics and questions used to assess ROF versus ER that are presented in Table 2 are also relevant for assessing single versus dual baseline ER. Illustrative questions for a maintenance superintendent could include, for example:

- About how old was the replaced equipment when it was removed? (Compare to EUL). If equipment is at or much older than its EUL an ROF decision should be considered but considerations below should be taken into account that would warrant an extension of the life beyond the standard EUL
- What were the maintenance requirements like in the three years prior to its replacement? (Normal, flat, escalating, low, high, prohibitive)?
- Had the replaced equipment required a major repair in the three years prior to its replacement that was done? Describe. How much did it cost? (Compare to replacement cost)
- Are there maintenance logs or purchasing records showing maintenance activity or costs?
- How do you keep track of equipment maintenance requirements for planning purposes? Generally, does your business tend to replace older equipment pre-emptively on a schedule, as maintenance costs creep up, or only on catastrophic failure?
- How would you describe the condition of the equipment prior to replacement??

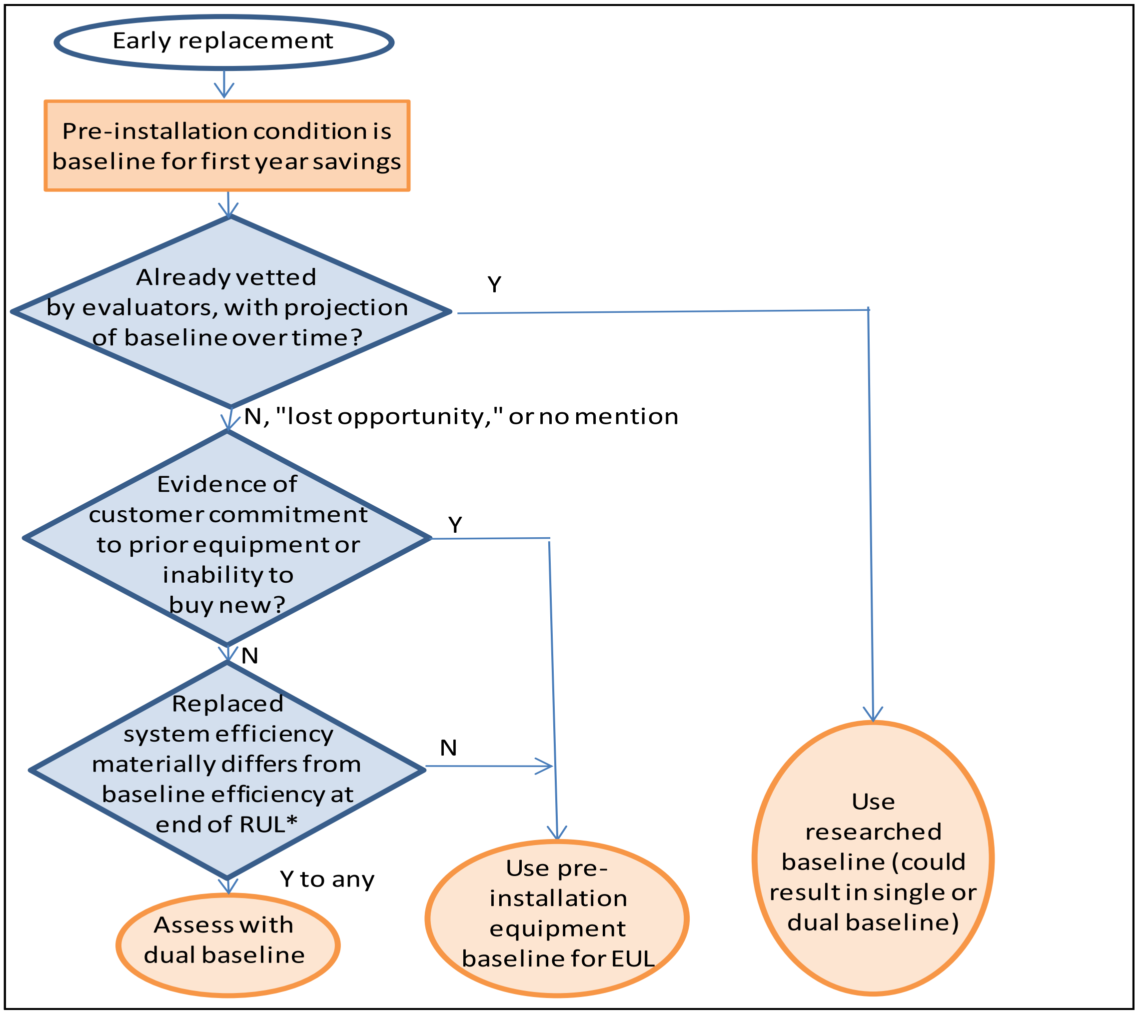
- Individual high capital cost measures for which there is documentation of history and expectation of lack of access to capital in the future that demonstrates an inability to buy new equipment. Interviews alone may not be sufficient.

If the replaced system efficiency is substantially the same as the projected baseline efficiency at the end of the replaced equipment's RUL, the measure is effectively a single baseline measure even if it is dual baseline in principle.

If the future baseline has been researched and projected by evaluators in an ISP study or similar, the research should be used to decide if dual baseline is applicable and, if so, what the out-year baseline level should be.

illustrates the decision-making logic regarding whether or not an evaluator should consider an early replacement measure as having a dual baseline.

Figure 6: Measure Type Flowchart – Early Replacement



**For any reason including:*

- *Known code or standard change will occur before the end of RUL*
- *Trends in standard practice will change the baseline efficiency*
- *The baseline efficiency will not change but already is materially different from pre-installation efficiency*

4.4.3 Characterizing the Second-Period Baseline in Dual Baseline Scenarios

Evaluating the baseline at a future time is difficult and requires special consideration. This framework identifies five distinct bases of estimating baseline that are a function of study type, measure type, and existence of relevant codes and standards. The logic is shown in Figure 7.

If the evaluator is conducting an industry standard practice (ISP) study of a measure, as opposed to performing site-specific evaluation of a single application of the measure, the evaluator should use any relevant codes or standards as the initial basis for the future baseline of the second period. In such an intensive investigation, that baseline may be refined by supplemental market research. If, for example, the MA C&I market has a consistent history of exceeding or falling short of code and the code is known to be changing in the future, the ISP may project the second-period baseline to follow a similar pattern. In the event that the evaluator concludes that the future baseline will deviate from both the current baseline and the future code or standard, the ISP study should document this distinctly. Importantly, the projected deviant ISP should be applied for measures starting at the beginning of the three-year cycle that follows ISP study completion. Until then, PAs and the evaluator should use the current ISP or future code efficiency as the future baseline.[20]

If the measure has no relevant code or standard, the second-period baseline should be developed in the same fashion as the baseline at time of research.

If the combination of measure and application is unique, use the first-year baseline as the second-period baseline.

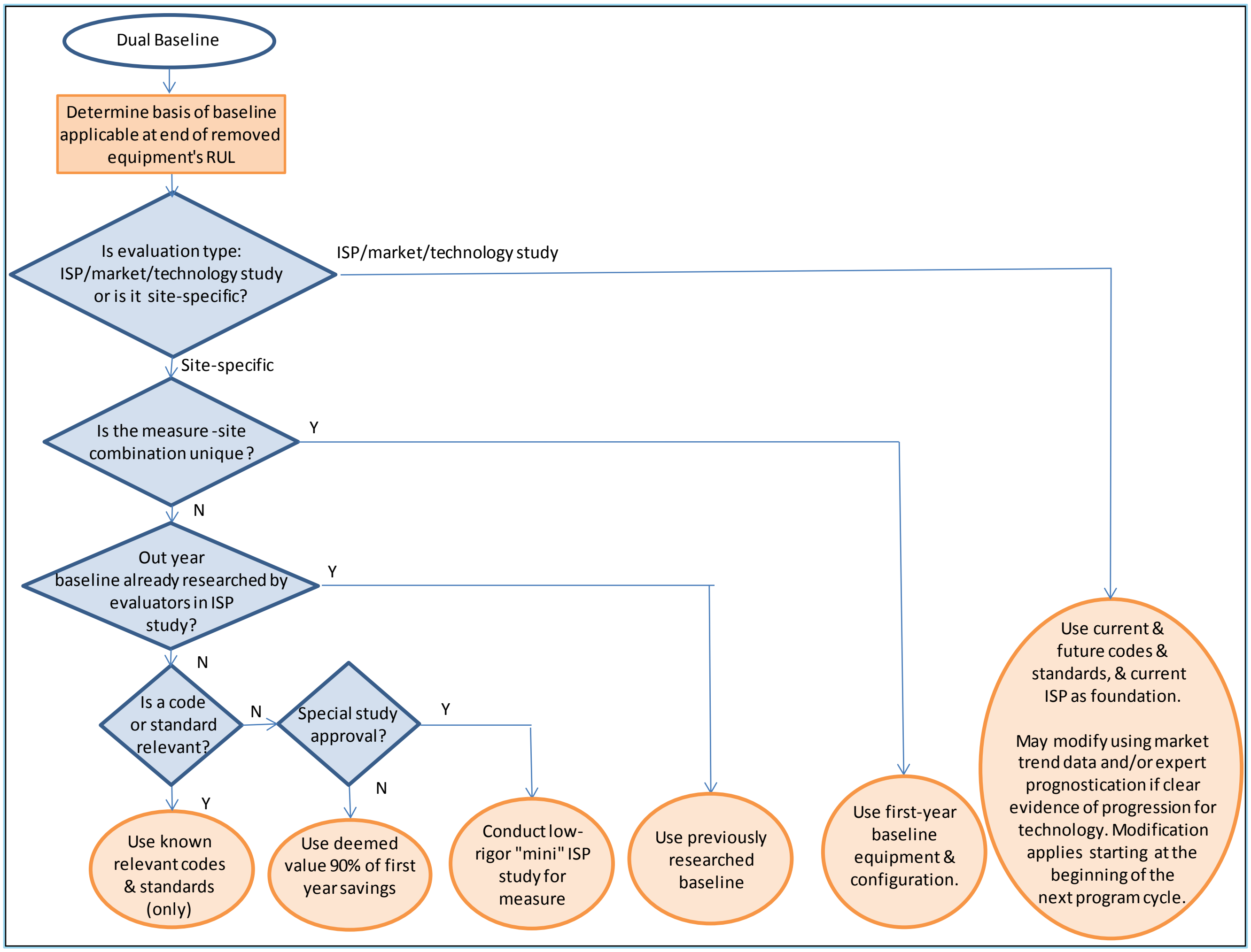
Where no future standards have been published, the evaluator can apply current standard practice as baseline at the time the equipment would have reached end of life.

If the evaluator faces a non-unique dual baseline measure for which there is neither relevant code or standard, an already-researched basis for projecting the second-period baseline nor a special study approved for the typically extra effort required to do so for the measure, the default is that the second-period savings is 90% of the first-year energy savings and if appropriate, demand.[21]

[20] The delayed application is made for practical considerations, to allow programs to plan for the change, and in recognition of the relatively high uncertainty associated with projecting baselines that are not defined by codes, standards, or current ISP. To illustrate the delayed application by example: A 2017 study predicts ISP baseline efficiency will increase in the future, even without known future code changes. Between the time of completion of the study and the end of the active three-year period, December 31, 2018, the second-period (future) baseline should not exceed the greater of the future code efficiency and current ISP. Evaluators should only apply the increased baseline efficiency to measures completed after January 1, 2019.

[21] Consider the early replacement of an ice rink chiller as an example. Code does not apply because a rink’s operating temperature requirements are outside the range of those for which chiller equipment efficiency standards apply. Rink chillers are equipment for which a general market exists. ISP could theoretically be established. Massachusetts has not (as of January 2017) funded an ISP study on them. The savings at stake does not warrant commissioning an ISP study. The evaluation engineer develops a first-year baseline based on preexisting equipment, as with any early replacement evaluation. It is a dual baseline scenario. Without research we know that ice rink chillers sold now generally are more efficient than those sold 25 years ago. Rather than speculating on the continuation of this trend regarding ISP efficiency 5 to 10 years in the future, this guidance deems the second-period annual savings to be 90% of the first-period savings.

Figure 7: Dual Baseline Basis



5 CALCULATING LIFETIME SAVINGS REALIZATION RATES

Dual baseline directly affects measure lifetime savings and in turn measure cost-effectiveness.[22] Evaluators are not responsible for cost-effectiveness analysis but are responsible for lifetime savings evaluation.

For all measures except early replacement with dual baseline, the first-year and lifetime RR are the same, absent a correction to the EUL, and separate lifetime RR calculation is not necessary. For dual baseline measures, however, the lifetime savings realization rate will be affected by the two baselines and the number of years applied to both.

[22] PAs will be responsible for adjusting the incremental cost to account (and be consistent with) the dual baseline savings estimates.

Early replacement measures for which either the tracking or evaluated annual energy savings estimate uses dual baseline principles or for which the tracking savings uses a discounted “retrofit” EUL to account for dual baseline require that the evaluator calculate and report a lifetime savings realization rate.

Site-specific EUL is impossible to observe for most early replacement projects, as it would depend on projecting future failure of working equipment. The evaluator should use the technology standard EUL in the above calculations.

Given the deemed EUL, estimating site-specific RUL depends on knowing equipment age at replacement (*RUL = EUL – replaced equipment age*). The estimated age at replacement is vulnerable to substantial measurement error, as it typically depends on recollections of installations five to twenty years prior, absent the production year being stamped on the nameplate, and the implementer having saved documentation of it such as by photographing the nameplate prior to removal or copying the original filed sales invoice. The evaluator should attempt to collect site-specific age at replacement to inform future research on measure EULs and RULs, but should only provide the estimate if it is definitive and documented. It should not use it for project retrospective gross savings evaluation even if provided. For retrospective use in impact evaluation the evaluator should use the RUL value of one-third of the EUL unless evaluators previously have developed a program- or measure-specific RUL or the evaluation is of a unique measure that has exceptional available RUL data.[23]

As with first-year savings RRs, the numerator (**green**) is the evaluation calculated lifetime savings and the denominator (**blue**) is the tracking lifetime savings.

If the program tracking savings uses dual baseline values:

$$RR_{lifetime} = \frac{\left[(EU_{baseline} - EU_{ee})_{RUL} \times RUL + (EU_{baseline} - EU_{ee})_{postRUL} \times \ (EUL - RUL)\right]_{evaluated}}{\left[ES_{RUL} \times RUL + ES_{postRUL} \times \ (EUL - RUL)\right]_{tracking}}$$

where,

$RR_{lifetime}$	=	Lifetime savings realization rate
$EU_{baseline}$	=	Evaluated annual energy use of the baseline equipment (kWh or MMBtu per year)
EU_{ee}	=	Evaluated annual energy use of the installed energy efficient configuration (kWh or MMBtu per year)
RUL	=	Remaining useful life of the replaced system (years)
EUL	=	The NC/ROF-based effective useful life of the system (years).
ES_{RUL}	=	Tracking savings during the period during the remaining useful life of the removed equipment (first period of dual baseline life)
$ES_{postRUL}$	=	Tracking savings during the period between the end of the RUL and the end of the EUL (second period of dual baseline life)

[23] As cited in the TRM, the Massachusetts Common Assumptions default remaining useful life (RUL) is one-third of the effective useful life (EUL). This is a reasonable compromise to balancing research cost and improving lifetime savings accuracy. This basis also has been used in California. See *Summary of EUL-RUL Analysis for the April 2008 Update to DEER*, KEMA Inc., and more recently, SCE/CPUC’s *Early Retirement Using Preponderance of Evidence*, v1.0, July 14, 2014. The MA TRM uses the default for most retrofit measures. Selected measures use other adjustments based on technology-specific research.

If the program tracking savings is a single annual energy savings value with a single, possibly blended retrofit, EUL, then the evaluation-based numerator of the formula stays the same and the tracking-based denominator simplifies to $ES_{tracking} \times EUL_{retrofit\ tracking}$.

6 RESEARCHING AND DOCUMENTING THE BASELINE

To characterize baseline of a measure that is neither the preexisting condition nor the relevant code/standard, evaluators should follow one of two distinct research paths:

(1) If the combination of measure and application is unique, site-specific data must be the basis of the baseline, with assessment regarding the options available to the particular applicant. The gross baseline is the less efficient alternative option the participant otherwise most likely would have done absent the measure.

(2) If there is a recognizable market for the measure, it is not unique, and there is no relevant code or standard, the evaluation should rely on a population-based ISP study to define baseline. The only use of site-specific data in baseline characterization should be to affirm that the application is indeed common to the broader market and not a unique measure-application combination. Thus, if the implementer correctly used an evaluator-vetted baseline for a measure and it is in fact an ISP-appropriate application, the baseline should not change as a result of the ex-post impact evaluation.

    If no prior ISP study exists that evaluators have completed or endorsed and is not obsolete, evaluators should conduct one if at all possible. The techniques to use in conducting an ISP study are addressed in the next section.

Custom measures often but not always are unique. As an example of an exception to this, a custom program might fund measures such as injection molding machines or electronically commutated motors that are not common enough to have prescriptive incentives but that are common enough in MA for there to be a recognizable market for them. They should be evaluated as non-unique measures.

Conversely, prescriptive measures typically but not always are non-unique. As an example of an exception, an apparently common application of an air compressor control measure in manufacturing might earn a prescriptive incentive, but if the equipment using the air has special requirements regarding pressure or air quality, the air compressor ISP may not be relevant, requiring site-specific baseline assessment.

Evaluators must cite the basis for characterizing each measure baseline when the evaluation baseline deviates from the tracking baseline.

The balance of this section details ISP research method options and how to determine the appropriate level of rigor for such research.

6.1 Industry Standard Practice Research Methods

For defining ISP, if the technology has been researched in a recent MA evaluation through a focused research activity or otherwise documented and reviewed by evaluators[24], the prior work should be used and referenced. If not, the evaluator likely will need to conduct original ISP research on the technology baseline. The need may be identified during either portfolio- or program-level evaluation planning or during site-specific review.

Adjustments other than those based on direct codes and standards interpretations should be based on empirical research of market practices. Baselines should represent population characteristics absent the program, and not the particular decision-making observed at the participant's site. Examples of methods that could be used to determine the market baseline include:

- A survey of equipment suppliers and/or system designers on standard practice, relative cost, and applicability
- Interviews with multiple national technology experts on the same
- Analysis of manufacturer or distributor shipment volumes by efficiency tier
- Analysis of a sample of recently filed new construction drawings
- Survey of customers, likely program nonparticipants, that have taken similar actions near the time of application
- Citation of recent relevant secondary research

6.2 Level of Rigor

The rigor of new research should be dictated by the amount of savings at stake and the range in possible baseline uncertainty. This framework describes two tiers of rigor.

A high rigor protocol requires at least two of the first five research methods discussed above, including one survey-based approach, plus secondary research. A high rigor method is a major undertaking and will typically be commissioned as a standalone study, outside of a typical measure or program-specific impact evaluation. A study of this scale should be considered for a technology with a significant contribution to MA portfolio savings (nominally 3% of program-wide savings or $5 million or more in statewide incentives over three years) and moderate to high uncertainty in baseline characterization. Ideally, a study of this rigor will result in a baseline adoption curve which identifies how the baseline will evolve from now until market transformation or will identify a timeframe for a study update. This enhanced research will both help the study remain relevant longer and aid in dual baseline evaluation. As described in Section 4.4.3, the study authors may conclude that a future ISP is likely to both change from current practice and differ from future known codes or standards. If so, the study should emphasize that evaluators are to use the current practice or future code baseline and not the researched projected baseline until they are analyzing measures completed in the next three-year program cycle.

Low rigor research is appropriate when potential savings variance is more moderate but still material. It could apply either to a standalone ISP measure study or to site-specific evaluation conducted as part of a

[24] MA evaluators previously have researched baseline for thermal oxidizers and currently are doing so for injection molding machines, for example. *Baseline Document: National Grid Massachusetts & Rhode Island © 2015* lists one PA's judgment of baselines for about 100 measures, but it has not been vetted by evaluators.

typical impact evaluation. A low rigor protocol includes at least one of the first four research methods and secondary research. It should be sufficient for estimating the current market baseline but may not be as precise as the high rigor result and is unlikely to result in projections in baseline changes over time. Anecdotal evidence such as a single interview or the evaluator's expert judgment alone generally is not sufficient for low rigor. While site information alone typically is insufficient for the reasons described in the definitions section,[25] it should be included for low rigor site-specific studies. Typical site-specific information would include design documents or price quotes regarding the alternative baseline equipment, to demonstrate it really was a technically viable and materially less expensive option. Low rigor research could serve as a precursor to high rigor research and may but will not typically assess future baseline.[26]

The evaluation engineer should describe the basis of the baseline in the research report. See Section 7.2 for details or reporting expectations.

An implementer that follows the above evaluation research protocols and provides this information to evaluators will substantially reduce the risk of a low realization rate due to baseline re-characterization.

7 COORDINATION WITH IMPLEMENTATION

Evaluator changes to baseline definitions are not, historically, the most common reason for C&I savings deviation from tracking estimates but they can have the most significant effect on overall program realization rates[27] because when they do occur such changes tend to result in more dramatic adjustment than measured hours-of-use differences, for example.

7.1 Concurrent Review

Because of the large stakes involved with baseline assumptions, PA implementers may desire to engage evaluators in pre-installation assessment of baselines for specific prospective projects. This concurrent review process is allowable and encouraged for large unique projects with uncertain baselines. In such cases the evaluators will work with the implementer and independently characterize the measure baseline. Evaluation assessment may include any of the steps described in the prior section, and/or site visits and measure-specific interviews with the prospective applicant, their designer, or the technical assistance provider. Typically, such investigation focuses on baseline characterization rather than hours of use, load factors, proposed equipment configuration, and other variables that may change after installation but is not limited to that topic.[28] At the end of the concurrent assessment, the evaluator should write a short memo summarizing the conclusions and recommendations. The PA is not bound to accept them but has gained clear advance indication of the baseline characterization[29] if the measure later is selected for evaluation.

---

[25] The exception to this case is in that of a truly unique application, typical industrial process, where there is no market-wide alternative to a customer-specific baseline. When this situation arises, the gross savings evaluator must communicate with the net savings evaluator to ensure that free ridership interview questions and subsequent calculations complement and do not overlap with each other.

[26] The recurring need for site-specific assessment of thermal oxidizer baselines in custom project evaluations led to a low rigor (less than $5k) ISP study, for example. Similarly, the repeated need to evaluate baseline for injection molding machines led to a high rigor ISP for that technology

[27] For example, see *2013 Custom Impact Evaluation Industrial, Agricultural, and Large Commercial Final Report*, submitted to California Public Utilities Commission, Figure 1-3, by Itron with assistance from ERS, Energy Metrics, Michaels Engineering, and Katin Engineering Consulting, www.calmac.org/publications/IALC_2013_Report_Final_071715.pdf.

[28] Non-baseline considerations could be reviewed for reasonableness in the concurrent review scope.

[29] Absent the occurrence of extraordinary new information between the concurrent and ex post evaluation reviews.

Concurrent review adds cost to application review for the PAs and the evaluators and may require a distinct evaluation sample design. If the review materially delays PA response to potential applicants there could be a lost opportunity cost as well. The evaluator cost is partially offset directly by reduced evaluation review time during ex post evaluation. Program evaluations may ultimately cost less with this approach due to reduced realization rate variability and thus lower sample sizes going forward. The non-cost benefits include reduced uncertainty of evaluation outcomes for PAs, education, and (importantly) faster feedback from evaluators to implementers.

7.2 Communication of Results

The primary means of communicating site-specific results of evaluator research regarding baseline characterization is the evaluation summary and in particular the one-page summary to each report. During research the findings may be shared directly with the PAs and their TAs during review and before site reports are finalized.

For ISP studies of prescriptive or non-unique custom measures the research report should be concise. The content will vary as a function of level of rigor and typically should include:

1. Background on why the measure was selected
2. Methodology
    a. Research methods used, level of rigor, sample sizes, and reason selected
    b. Configuration alternatives if they are discrete, or the range of performance levels including any known third-party classification systems (e.g., CEE tiers) if they are continuous
    c. How program participation effects were considered
    d. The basis of estimating commonly installed baseline is the mean (blended), median, mode, or other measurement of ISP
    e. Citation of relevant sources
    f. Economic analysis, if applicable
3. Baseline conclusions, as a function of application, if necessary, and exceptions
4. Historic and projected baseline trends, if possible
5. The results' sunset date, the time after which the findings are likely to be obsolete and should be updated

If new baseline trends emerge from evaluation activity or a need for a high rigor technology baseline is identified, the evaluation team can bring the topic to the MA C&I Management Committee to gather input on their perspective for consideration.

8 APPLYING EVALUATION BASELINE RESULTS

Applying the baseline principles described in this document will affect processes in place prior to its publication and require the introduction of new mechanisms to PA and evaluator reporting. This section addresses these issues.

8.1 Application of Baseline Findings to Evaluated Gross Savings

Evaluator results associated with first-year baseline research affect first-year savings of all measures. The resulting realization rate should be applied retrospectively when PAs compute evaluated gross first-year savings. This is continuation of past practice.

Computing lifetime savings and corresponding realization rates as a function of dual baseline principles or of evaluation of tracking lifetimes is not part of past MA CI evaluation scope and reporting protocol. Evaluators should start computing such results and reporting them in site-specific and program-level evaluation reports in 2017. Similarly, ISP studies initiated after the finalization of this framework should include lifetime consideration in the scope.

The evaluators' lifetime-oriented findings are to be used by PAs for informational purposes through the end of the 2018 program year. Lifetime savings associated with all measures completed after January 1, 2019 are subject to retrospective adjustment to account for the evaluated second-period baseline efficiency, evaluated lifetimes, and the resulting evaluated lifetime savings and realization rate results described in Section 5.

ISP baseline study results should not prompt restatement of savings for which evaluation has been completed and applied.

8.2 Application of Baseline Findings to Avoid Interaction with Net-to-Gross Savings

For non-unique measures, all baselines used for evaluated gross savings must be based on market-level research and not depend on a single participant's perspective. Free ridership questions should be based on the premise of this same market-defined baseline. As such there should be inherent separation of baseline and free ridership effects that avoids overlap or double-counting of any free ridership effect. Past studies have not used such measure-specific baselines in their wording. Going forward, additional care must be taken in the phrasing of free ridership questions.

In the case of unique measures by definition, no market baseline exists. As stated in Section 6, the gross baseline is the less efficient alternative option the participant otherwise most likely would have done absent the measure. Free ridership questions for this unique project must clearly use the gross baseline in the proper context.

For example, a vendor offered a participant three distinct configurations for their new process: 100,000, 90,000, and 80,000 kWh/yr alternatives. All were plausible. The program funded the 80,000 kWh/yr option. The customer tells the evaluation engineer that the company would have chosen the 90,000 kWh/yr option without the program. This is used as the baseline for evaluated gross savings. The free ridership battery must clearly use the 90,000 kWh/yr configuration as the starting point when asking about program influence relative to the move to the 80,000 option. If the free ridership questions posed allow the customer to consider the 100,000 kWh/yr configuration as an option at all, it will double-penalize the program, by up to 50%.

Massachusetts may embark on a process to integrate gross and net impact research for custom and possibly prescriptive measures. There are challenges associated with such an approach, including those associated with data collection timing and sampling, but it would have the significant benefit of making the example scenario easier to manage. Further, if “straight to net” research occurs the importance of avoiding overlap becomes less significant. Setting the precisely correct gross baseline will guide the allocation of savings but will not define it. Research would emphasize customer-specific baselines and program effects.