# Counterfactuals, Damned Lies, and Statistics: Impact Evaluation in the Era of Big Data

*Emily L. Cross, P.Eng, CEM, Navigant Consulting, Inc., Burlington, Vermont, USA*

## Abstract

This paper proposes a constructive path forward toward a future energy efficiency (EE) impact evaluation paradigm in the era of "Big Data".[1] Firms that offer advanced real-time data analytics for EE initiatives envision a new paradigm in which current incentive and evaluation structures are either no longer necessary, or will be radically transformed.

This paper argues, with supporting references, that in the electric utility of the future, accurate measurement of energy and demand savings using a variety of methods will remain necessary to provide accountability for energy efficiency investments and provide a firm basis for the potential commoditization of EE.

While current industry literature agrees on the potential of high-frequency smart meter data to transform impact evaluation, and agrees on many of the barriers to realization of this potential, the near-term path forward for smart meter data as a high-accuracy impact evaluation tool remains unclear.

The paper explains (1) the requirements for leveraging high-frequency data for an automated analysis of EE savings on entire populations of EE initiatives, and (2) how the IPMVP could be expanded to explicitly include high-frequency data in Option C. The author also supports (3) measurement of first year impact savings focusing only on whole building savings for projects above a signal-to-noise ratio in the range of five to ten percent of annual pre-installation consumption.

The paper concludes that accurate, direct, census measurement and evaluation of entire populations of energy and demand impacts is consistent with the broader industry goals relative to climate change, emerging 'white certificate'[2] policies, and the sustainable utility of the future, and can be achieved at lower cost than current impact evaluation methods by using advanced accelerated analytics, with relatively modest changes to data tracked by programmes.

## Introduction

Beginning with the end in mind, what would it look like if we wanted to leverage accelerated data analytics in support of sustainability, in the paradigm of the transitioned utility of the future? To maintain accountability for investments in Energy Efficiency (EE) and the lay the groundwork for white certificate policies in support of commoditization of EE initiatives, this paper answers the following questions about energy savings: 1) What should be measured? 2) How should it be measured? 3) Who should measure it?

The purpose of this paper is three-fold:

1) **First**, identify essential energy efficiency programme tracking data that will pave the way for accurate automated impact evaluation using high-volume, high-frequency utility smart meter data.

---

1 Big Data refers to the increasingly rapid availability of high-frequency electric utility data such as from smart meters, and the use of such data by advanced data analytics firms to support the advancement of energy efficient infrastructure.

2 "A white certificate, also referred to as an Energy Savings Certificate (ESC), Energy Efficiency Credit (EEC), or white tag, is an instrument issued by an authorized body guaranteeing that a specified amount of energy savings has been achieved. Each certificate is a unique and traceable commodity carrying a property right over a certain amount of additional energy savings and guaranteeing that the benefit of these savings has not been accounted for elsewhere." https://en.wikipedia.org/wiki/White_certificates

2) **Second**, propose an example update to the International Performance Measurement and Verification Protocol (IPMVP) which supports automated evaluation and shorter reporting periods by addressing cutting edge data gathering methods and analytics.
3) **Third**, estimate the cost savings associated with automated evaluation, based on the proportion of evaluations that could be addressed using accelerated, automated analysis methods.[3]

## The Utility of the Future

Why would the utility of the future (the "transitioned utility") benefit from the more widespread use of high-frequency energy data, leading to energy savings, and lower energy sales? The answer: electric power systems in many countries are transitioning from one-way, highly-regulated power systems to an emerging 'energy cloud' characterized by widespread adoption of Distributed Energy Resources (DER), two-way energy flows, digitalization of a flexible, dynamic, resilient grid, and complex market transactions supported by advanced business models based on new product and service revenues (Vrins, 2015) which could replace revenue lost due to energy savings.

The future utility business model is shifting from energy sales to energy services, with the utility focusing on the capacity of their transmission and distribution infrastructure, while allowing customers to more directly impact their energy flows and cost of energy.[4] Utility data infrastructure and the information communications backbone will necessarily be integrated with the day-to-day operations of the transformed utility. Revenue lost in annual energy sales will be replaced by energy services and real-time measurement capability, resulting in lowest environmental and economic costs of energy use. Wholesale power markets and an increasingly customer-centric paradigm, in terms of program design, implementation, and evaluation is described in (Rogers et al, 2015). Customers may "purchase energy efficiency as a commodity that can then be traded in regional capacity markets. More sophisticated customers may bypass programs and monetize that value themselves." (Rogers at al., 2015).

## Accountability for Energy Conservation Investments

Accountability for investments in energy efficiency and demand-side management (DSM) will remain an area of interest for regulators, utilities, and customers in the transitioned utility. As the smart grid begins to accommodate two-way energy flows, measurement of energy, and measurement of changes in energy and demand, will remain essential for the smooth functioning of the grid itself. Therefore, utility-owned smart meters, and the associated communications backbone will naturally be developed and leveraged by the utility and its subcontractors, even in an entirely unregulated market. Cloud-enabled computing capacity will ensure transmission and distribution capacity requirements which can be easily quantified via accelerated analysis of entire populations of buildings without resorting to sampling. This will provide a rigorous basis for tasks ranging from load forecasting and accountability for energy conservation investments, to policies needed for commoditization of EE.

In his 2011 presentation "Assessment and Experience of White Certificate Schemes in the European Union", Bertoldi emphasizes the "crucial importance of measurement and verification, and strong focus on standardised saving values" supporting white certificates. White certificates embody an obligation to achieve a certain target of energy savings via deemed, "subsidy" measures and standardized savings factors. The availability of high-frequency (quarter-hourly or hourly) electric smart meter data (in contrast to the monthly billing data more currently commonly available), and accelerated data analysis using high-computing capacities offered by cloud computing, raises the possibility of producing high-accuracy validation at lower cost per EE transaction.

3 Using recent savings data for current U.S. EE programmes.
4 Remarks by Alice Jackson, Xcel Energy. Navigant Forum, Denver Colorado, March 4, 2016.

Per Slote et al., 2014 "In the European Union, energy efficiency EM&V is largely driven by the necessity to estimate the energy savings achieved by those energy suppliers subject to energy efficiency obligations (EEOs) in some Member States. The need for robust and stringent EM&V will increase as all Member States respond to the requirement in the 2012 Energy Efficiency Directive that they implement EEO schemes or alternative energy efficiency policies and programs that deliver 1.5-percent energy savings each year."

"One goal [of EM&V] applicable in all regions [China, the European Union, India, and the United States], is to maintain or instill confidence that the energy and demand savings claimed by various types of energy efficiency activities are valid and reliable. When this is accomplished, energy efficiency can be more confidently incorporated into resource and reliability planning, as a strategy in pollution emissions reduction, and as a measure to achieve an increasingly efficient and competitive economy. Achieving that goal requires developing and/or maintaining sophisticated, robust, and stringent EM&V protocols, methodologies, and practices."

## The Path Forward

In the U.S., "*Meters with the ability to provide interval data have been around for some time but were previously restricted to research projects and to larger customers that had special time-of-use rates that justified their installation.*" (Rogers et al, 2015). As utilities and evaluation contractors have been leveraging utility-owned data, similar to smart meter data, for many years, leveraging higher volumes of such data represents a nominal change to the existing paradigm, and is a natural progression building on prior utility meter technologies.

In the U.S. many 'software as a service' (SaaS) and 'disaggregation as a service' (DaaS) firms[5] have sprung up (see also DNV GL, 2015) capitalizing on this increase in high-frequency data, and while they have proven their software is successful at identifying potential energy efficiency opportunities (DNV GL, 2015) (Rogers et al, 2015), the software do not appear to be as well suited for retrospective impact evaluation with the level of accuracy normally expected by regulators with respect to regulatory accountability for meeting energy and demand savings targets (DNV GL, 2015).[6]

Lawrence Berkeley National Labs (LBNL) and others in the U.S. are in the process of researching and quantifying the potential of such Remote Building Analysis (RBA) and Non-Intrusive Load Monitoring (NILM), machine-learning software tools to meet accuracy standards typical for Evaluation, Measurement and Verification (EM&V) savings validation.[7] Such tracking of cost effective, persistent investment in energy conservation will ensure we continue to do the right thing well.

Sections below discuss the specific case of what should happen for impact evaluators to capitalize on the increasing availability of smart meter energy and demand data, outlining specific programme data to be tracked. A review of recent industry references suggests an approach such as the following could be successful in reducing impact evaluation costs:

1) Impact evaluation will primarily calculate annualized energy savings and coincident peak demand savings at the facility level for the first year after measure installation.

---

5 U.S. firms such as EnergySavvy, First Fuel, Retroficiency, EEme.

6 "Current research has not yet addressed the ability of these predictive models to measure energy savings *after measure installation*, but future research planned by LBNL will address this issue" (DNV GL, 2015).

7 LBNL definitions of 'big data' broadly include studies that include at least one of the following (though ideally two or more): very large data sets, e.g. use of hourly or higher resolution usage data; accelerated data acquisition; data from equipment-embedded sensors or equipment/building control systems; some form of automated and/or pre-specified analytics and reporting; some form of modern data analysis or visualization; and very very large samples or even population-scale inputs for the analytics (ref. personal communications between Navigant and LBNL).

2) High-quality utility grade electric meters, such as smart meters, should be the sole basis for energy flow measurements and savings calculations, for the purpose of supporting white paper market mechanisms, and rigorous validation of realized energy savings.[8]
3) Impact evaluators in the future could create the conditions for their own success by focusing on measuring energy savings above the signal-to-noise ratio at the facility level (bare minimum electric savings in the range of five to ten percent[9] of the baseline electric energy usage).
4) Energy efficiency programmes could deliberately incentivize projects, or groups of projects, in such a way that the savings are easily measurable at the facility level in the first year after installation.
5) Energy efficiency programmes could accurately track the following essential information for each EE initiative:
    - project initiation and completion dates
    - measure installation initiation and completion dates
    - baseline type
    - baseline annual electrical usage per site
6) Accelerated computing using high-frequency smart meter data and algorithms for routine adjustments, such as weather, can eliminate human-in-the-loop analysis for some types of energy efficiency programme evaluation, using largely automated accelerated cloud-enabled computing.[10,11]
7) Industry standard resources, such as Option C in the International Performance Measurement and Verification Protocol (IPMVP), can be modified to explicitly address the availability of high-frequency smart meter data, providing a policy basis for essential energy and demand savings measurements in the transitioned utility.

## Impact Evaluation in the Era of Big Data

Pursuit of high-impact energy savings makes sense for grid capacity and emissions reasons, economies of scale, evaluation cost savings, and cost effectiveness for programme administration and evaluation. As an industry, we create the conditions of our own success by investing in projects with an impact that can easily be seen in the data with the naked eye.

The right problem to solve when faced with inefficiencies and high costs of implementing and evaluating per-site energy savings too small to measure is—how to increase the per-site savings so that they are measurable. This thinking is in keeping with the urgency of climate change issues, particularly to the extent that electric generation remains carbon intensive.

The results below suggest an estimated 35% of U.S. EE programme savings could be evaluated via accelerated analysis methods using smart meter data. In a preliminary pilot study Navigant found (Navigant, 2015) when essential data such as the actual initiation and completion dates of measure

---

8 In the U.S., many third-party devices have been developed, raising the issue of interoperability of technologies and barriers to data access due to proprietary technologies (DNV GL, 2015). For example, DNV GL also states "Information provided by [Home Energy Monitoring Systems] HEMS and other smart devices can enhance EM&V or reduce additional data collection requirements, but these tools do not provide enough information on their own to sufficiently estimate a program's impact."

9 "For individual customers, annual prediction accuracy based on 12-monthly observations is on the order of +3 to 7% at 90% confidence. This means a pre-post change on the order of 5 to 10% would be considered statistically significant for an individual building." (DNV GL, 2015).

10 See the results section for the project information that should be tracked by the programme for such an approach to be successful.

11 That said, periodic boots-on-the-ground audits will remain an essential component of energy savings validation, due diligence, and persistence of conservation and efficiency efforts over time.

installation are accurately tracked by the EE programme, routine adjustments such as weather normalization could be automated for a range of different project types, resulting in successful, automated validation of energy savings.[12] A statistically rigorous estimate of the uncertainty in savings can be generated by the analyst's computer code[13] which can later be reviewed for reasonableness, and questionable results can be reviewed on a case by case basis prior to final certification of savings.[14]

This section provides an overview of current thinking on selected topics: the IPMVP, measurement uncertainty targets, "EM&V2.0" advanced analytics, and baselines. Referenced source documents contain detailed studies including remote building analysis (RBA), impact evaluation methodologies, and baseline characterization. Except where noted, no distinction is made with respect to residential, commercial, and industrial projects. It is assumed each facility on the smart grid of the future will have at least one smart meter, regardless of the type of facility.

## The IPMVP

The International Performance Measurement and Verification Protocol (IPMVP) was first developed in 1997 by EVO. EVO (then IPMVP®), was "a committee of volunteers who came together under a U.S. Department of Energy initiative to develop an international monitoring and verification protocol that would help determine energy savings from energy efficiency projects in a consistent and reliable manner."[15]

"IPMVP Concepts and Options for Determining Energy and Water Savings Volume 1" 2012 is a seminal industry reference which is the industry-standard technical basis and justification for the cost-effective determination of energy impacts for energy efficiency projects. It is referenced in countless impact evaluation guidance documents, including the Uniform Methods Project (UMP) (Kurnik and The Cadmus Group, 2013), and many others too numerous to mention. The "IPMVP Core Concepts" 2014 document is the latest version of "the IPMVP", and provides a framework of options for measuring energy savings ranging in complexity from sub-metering of key variables (Option A) through calibrated whole building energy modeling (Option D). Figure 1 presents an overview of the IPMVP Options A through D, and Figure 2 presents the current language for Option C (IPMVP, 2014), which deals with impact evaluation using whole-building utility data.

---

12 Navigant found that in some cases, the savings obtained using several months of pre and post utility data were arguably more accurate than the traditional calculation methods which relied on assumptions about the baseline equipment which in some cases did not appear to be consistent with the overall usage at the facility. Therefore, estimated savings uncertainty for each method is thought to be a better way to validate the EM&V 2.0 approach, rather than relying on a comparison of the EM&V 2.0 method to the standard evaluation method as a measure of success. That said, in cases where the on-site ex ante assumptions were well documented, the energy realization rates (ex post / ex ante) for the standard and the EM&V 2.0 methods were reasonably close. Four projects were studied, including lighting, industrial variable frequency drive, and data center energy savings.

13 For example, "R" is a free software environment for statistical computing, offered by The R Foundation, that can be used to program population-level econometric energy savings and uncertainty analyses. https://www.r-project.org/

14 Per (Granderson et al., 2015): "Collectively, our results suggest that modern tools, with their automated baseline models and savings calculations can at a minimum, provide significant value in streamlining the M&V process, providing results that could be quickly reviewed by an engineer to determine if adjustments and further tailoring are necessary. They also suggest, that savings can be reliably quantified at the whole-building level, using the interval data-based models that are available today. Depending on the level of confidence required, and the precise depth of savings expected, these savings might be quantified in a fully automated manner, or with some engineering intervention."

15 http://evo-world.org/en/about-en/history-mainmenu-en

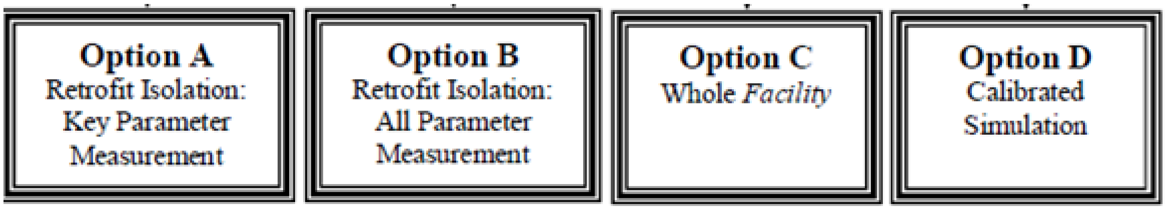


**Figure 1.** IPMVP Options A through D

| IPMVP Option | How Savings Are Calculated | Typical Applications |
|---|---|---|
| **C. Whole Facility**<br><br>*Savings* are determined by measuring energy use at the whole *facility* or sub-*facility* level.<br><br>Continuous measurements of the entire *facility's energy* use are taken throughout the *reporting period*. | Analysis of whole *facility baseline* and *reporting period* (utility) meter data.<br><br>*Routine adjustments* as required, using techniques such as simple comparison or regression analysis.<br><br>*Non-routine adjustments* as required. | Multifaceted energy management program affecting many systems in a *facility*. Measure energy use with the gas and electric utility meters for a twelve month *baseline period* and throughout the *reporting period*. |

**Figure 2.** Current IPMVP Option C language

## Measurement Uncertainty

The author proposes that the target uncertainty for evaluated savings on a per-site basis should be ±5% of the evaluated savings for project level evaluated impact savings, as a guideline for the design of automated algorithms. While perhaps only truly achievable for a warehouse dominated by 8760 hour lighting loads, the proposed target is based on the policy for forward capacity market requirements for post-installation power use at the project level of 2.5%[16] for short duration onsite power metering equipment, and the assumption the utility of the future will be more focused on demand savings and capacity transactions than it is today.

Utility-grade meters are already capable of providing an acceptable instrumentation accuracy.[17] While the actual calculated power in each of the baseline and efficient cases, and therefore the savings, may be expected to be outside this target after site routine adjustments, such as for weather, the 2015 study, by Granderson et al, shows promising results, particularly when the results for populations of similar buildings are analyzed together after site-by-site analysis (Granderson et al, 2015).[18]

This approach to measure uncertainty is therefore consistent with our desire as an industry to

16 Based on the ISO New England (ISO-NE) forward capacity market (FCM) resource validation manual ISO-NE M-MVDR.

17 Capability of 0.3% for existing utility grade meters (Rogers et al, December 2015).

18 "Using actual field data sourced from hundreds of interval meters, the research team found that for a quarter of the population of buildings in the data set, the energy savings resulting from program activities could be determined within a 6.5% margin of error, and that was without close inspection of the facilities or adjustments for nonroutine variations in energy use (Granderson et al. 2015)." (Rogers et al, 2015).

capitalize on the best of our prior achievements in terms of high-accuracy instrumentation, and should realize cost reductions for implementation, evaluation, and enforcement of policies. In contrast, methodologies relying on a largely unregulated, new, untested mélange of proxy sensors and third-party submetering devices (Rogers et al, 2015) are bound to be less accurate, and raise barriers to cost effectiveness, reliability, and defensibility of energy savings impacts due to the lack of interoperability of metering technologies, and issues around proprietary access to customer data by third-party firms.

Programme incentives could be designed to encourage projects which can be measured on the facility-level energy bill. This may mean bundling projects, or handling projects below the facility measurement limit as prescriptive (see Table 1 for more details on how certain types of projects and programmes could be handled).

In summary, impact evaluators could measure the **measureable** first year savings with the most accurate instrumentation available, and then submit the rest for longer-term study using a variety of methods (such as Options A and B of the IPMVP).

## EM&V2.0: Advanced Analytics Using Smart Meter Data

We are entering ‘‘The Analytics Age of Efficiency’’ (Grueneich and Jacot, 2014) where networked devices are clients on a ‘cloud’ of information (Rogers et al., 2015), giving us the ability to continuously monitor savings, persistence of energy conservation, thereby transforming utility EE-DSM programs, with the prediction that “regulators will come to rely on analytics to ensure the proper level of savings are delivered and counted”, and where “advancing baseline” (Grueneich and Jacot, 2014) models could be automatically normalized for weather and occupancy. However, “primary uses currently being promoted for automated M&V are services provided to the program administrator, and are not used to support claimed savings reported to commissions” (DNV GL, 2015).

Discussion in the literature of advanced analytic methods using large quantities of data generally falls into two categories: 1) characteristic profile development and disaggregation, versus 2) accelerated population-scale analysis. Each method uses the data in different ways. Data can be pooled to inform a representative model then disaggregated based on trained computer algorithms, or, it can be used to analyze data in bulk, facility by facility. This paper focuses on the second method, where accelerated analysis of interval data of the type researched by Granderson et al, 2015, is performed on entire populations of energy efficiency projects supported by cloud computing.

## Baselines and Counterfactuals

For logistical reasons, an impact evaluator will group projects within a programme based on evaluation type, such as desk review, onsite verification, onsite verification with metering, or billing analysis. It is up to the evaluator to determine, with the help of the IPMVP, which method of evaluation is appropriate depending on the amount of savings and uncertainty in the variables used to determine ex-ante savings. Likewise, the evaluator must confirm the appropriate baseline for a project. Baseline determination can add additional cost and complexity to the evaluation, and is one of the largest sources of discrepancy between *ex-ante* and *ex-post* savings estimates for specialized ‘custom’ projects where the energy and demand savings are calculated on a site specific basis, rather than using deemed or prescribed savings estimates (Maxwell et al, 2011).

The author proposes deliberately grouping programme information, or programmes themselves, by baseline type, in order to align programme data with impact analysis methods to more easily leverage smart meter data. This section provides an overview of baseline types used to develop Table 1 below which is organized by baseline type.

**Definitions.** A seminal paper on baseline concepts (Maxwell et al, 2011), states “The baseline is the least efficient, non-regressive, code or regulations-compliant option specific to a particular facility and application that the customer technically, functionally, and economically could have alternatively considered to deliver the post-retrofit level of production or service”.

In this paper, the term ‘counterfactual’ refers to any baseline that cannot be measured as pre-existing, because it is ‘made up’. The baseline energy use and equipment to be compared with the efficient case never existed. With a counterfactual baseline, pre-installation hourly energy data may be available, however it does not provide an appropriate baseline comparison. An example of this situation is when a code baseline is appropriate due to a major renovation of an existing facility, subject to energy code-minimum requirements.

The author notes that usage data for a pre-minus-post custom analysis are usually adjusted or constructed in some way, such as for weather or occupancy,[19] however such data are not considered counterfactual in terms of the baseline equipment installed. Many of these routine adjustments could be included in automated econometrics-based computer code. Such automated adjustments, once validated according to future versions of industry standard protocols such as the IPMVP (EVO, 2014) and ASHRAE Guideline 14 (ASHRAE, 2002), would provide the basis for automated gross impact evaluation.

This paper is concerned with technical, or gross savings baselines, and not net-to-gross (NTG) baselines. As summarized in (DNV GL, 2015): “The baseline represents what would have occurred without the installation of the program measure (gross savings) or without the influence of the program (net savings)”.

**Baseline Types.** The fundamental baseline types considered in this paper are as follows:

- **Pre-existing equipment** baselines
  Example project types:
  - Early replacement of working equipment not otherwise being replaced
  - Add-on controls
- **Counterfactual** baselines (applicable codes or standard practice)
  Example project types:
  - New construction
  - Deep energy retrofit[20]
  - Equipment failure[21]

# Results

## Annual Cost Savings for Evaluation

By designing energy efficiency programmes in alignment baseline types I through IV shown

19 Per Slote et al., 2014 “There is no direct way of measuring energy or demand savings, because (1) it is not possible to measure a participant’s energy use, at the same time, with and without the program; and (2) one cannot measure the absence of energy use. Consequently, the energy and demand savings values (and any associated non-energy benefits) for an energy efficiency program that are produced by EM&V are always going to be estimates… The use of these estimates as a basis for decision-making can be called into question if their sources and level of accuracy are not analyzed and described.”

20 While deep energy retrofit of an existing building may be a combination of measure types, each with a different baseline, for the purpose of developing Table 1 below, this category assumes that the majority of savings in a deep energy retrofit are generated by measures where the prior systems are completely removed, and a code or standard practice baseline is appropriate.

21 In some cases, failed equipment may be replaced by identical equipment to the pre-existing equipment. In this case the pre-EE measure utility data could be considered an appropriate baseline.

in Table 1 below, deliberately incenting projects with high energy savings relative to the baseline facility energy use, and leveraging high volumes of high accuracy smart meter data, the author sees a potential for high cost savings for the type of annual impact evaluation typically required for regulatory compliance, via automation of impact savings validation. An estimate of this potential cost savings, based on the current paradigm of U.S. EE programme expenditures, is as follows:

| **Cost savings** = Cost.*Evaluation Type I and II* x Cost Reduction Fraction = \$42M x 0.5 = \$21M |
|---|

where the fractional energy project savings in the final column of Table 1 below is used as a proxy for fraction of expenditures, such that

**Cost.*Evaluation Type I and II*** = (0.30 + 0.05) x Total Impact Expenditure
= 0.35 x Total Impact Expenditure = \$42M

where the total annual impact evaluation expenditure is \$120M is derived as follows

**\$120M** = \$200M x 0.6, where
**\$200M** ≡ the total estimated U.S. annual evaluation expenditure (estimated as 2% of the total U.S. EE programme expenditure), and
**0.6** is the fraction of the total estimated evaluation expenditure for impact evaluation (with the remaining 0.4 attributable to process evaluation).[22]

and where

**Cost Reduction Fraction** is the estimated cost reduction factor for accelerated evaluation of Evaluation Type I and II projects, CR = 0.5.[23]

Thus, although currently the industry generally agrees "The cost-effectiveness of these methods as part of overall program delivery or evaluation is not yet established" and "Because of the current challenges of data processing, data management, data access, and data cleaning, use of [Advanced Metering Infrastructure (AMI)] AMI data is not necessarily a lower cost evaluation option compared to traditional evaluation approaches, even where AMI is already in place" (DNV GL, 2015), the above results suggest promising potential cost savings relative to the current U.S. approach.

**Methodology.** In developing Table 1 as the basis for the above cost savings calculation, the author considered the following baseline types:[24]

**Type I Existing Equipment.** Examples of projects with this type of baseline include early replacement for energy efficiency, addition of commercial and industrial controls.

**Type II Counterfactual.** Examples of projects with this type of baseline include new construction, deep energy retrofit, failed equipment replacement.[25]

**Type III Empirically-Derived Baseline.** Examples of projects with this type of baseline include deemed or prescriptive projects for which the savings do not need to be measured on an annual basis.

22 Personal correspondence from Stu Slote, Navigant Consulting, Inc.
23 Personal correspondence from Nishant Mehta, Navigant Consulting, Inc.. This estimate is conservative, and could be lower once algorithms and protocols are established by the evaluator.
24 Energy savings data for various programme types utilized in this study was publicly available data for selected U.S. utilities as listed in the references.
25 Also called 'replace on burnout' or 'market opportunity'.

**Types IV and V Special Consideration.** Very large capital commercial and industrial, emerging technologies, data centers.

**Table 1.** Matrix of Evaluation Methods

| **Baseline Label** | **Baseline Type / Program Type** | **Site Level Measurement Method for Project Savings[26] < 5% of Whole Facility Annual Usage** | **Site Level Measurement Method for Project Savings >5% of Whole Facility Annual Usage** | **Estimated Percentage of Expenditures[27, 28, 29] (based on U.S. programmes)** |
|---|---|---|---|---|
| **I** | Existing Equipment | Will not be measured.[30,31] | Evaluation Type I | 30% |
| **II** | Counterfactual (Code or Standard Practice) | Not applicable.[32] | Evaluation Type II | 5% |
| **III** | Empirically Derived Blended Baseline[33] | Not applicable. | Evaluation Type III | 62% |
| **IV** | Special Consideration—Custom | Evaluation Type IV | | 1% |
| **V** | Special Consideration—Emerging | Evaluation Type V | | 2% |

Table 2 provides additional information for the proposed evaluation types recommended in Table 1.

**Table 2.** Evaluation Type Legend

26 Ideally the ex-ante and ex-post project savings are the same, therefore no distinction is made here.
27 Where available, reported *ex-post* evaluated savings were utilized. Where not available, *ex-ante* savings were utilized.
28 Where it was not clear from the description of the programme which type of baseline was appropriate, the author assumed an even split between the types, generally allocating the savings by thirds to the Existing Equipment, Counterfactual, and Empirically Derived Baseline categories.
29 This column is based on the estimated fraction of energy savings for selected U.S. energy efficiency programmes (see References section), and assumes that, on average, expenditures track savings.
30 This encourages bundling of measures, and incentives being applied strategically. While the absolute savings of projects such as these could be large, it is assumed they would not have sufficient societal benefits to incentivize, relative to the potential savings at the facility. Additionally, if the absolute savings were large, there could be an economic incentive for the customer to do this project without the assistance of a programme, anyway.
31 California legislation "Programs are to support improvements in existing buildings "taking into consideration the overall reduction in normalized metered energy consumption as a measure of energy savings." Recently enacted legislation in California, AB802, per (DNV GL, 2015).
32 This situation is not expected to occur due to the fact that new construction, major renovations, and large capital retrofits with incentivized energy efficiency measures are expected to exceed code baseline or standard practice by significantly more than 5%.
33 This baseline type is typically utilized for 'prescriptive' or 'deemed' measures that do not need to be measured based on first-year annual savings.

| Evaluation Type (Program Type) Key | Evaluation Description |
|---|---|
| **Type I – Accelerated Analysis** | Accelerated automated first year impact evaluation using utility smart meter interval data.<br><br>Use IPMVP Option C. |
| **Type II – Accelerated Analysis with Deemed Baseline** | Accelerated automated first year impact evaluation using smart meter interval data with a deemed baseline.<br><br>The nature of the deemed baseline would need to be determined, and could be based on outcome-based or asset-based codes and standards by building type, for New Construction projects.[34]<br><br>Use IPMVP Option C. |
| **Type III – Periodic Study** | Study performed every three to five years.[35]<br><br>Based on a combination of Smart Meter, Information Communications Technologies (ICT), and Internet of Things (IoT data) proxy variables, Remote Building Analytics (RBA) analysis, pooled analysis methods, Randomized Control Testing (RCT), Non-Intrusive Load Monitoring (NILM) and other types of comparison testing to assess market effects, including free ridership.<br><br>Most likely to use IPMVP Options A or B. |
| **Type IV – Stratified Random Sample (SRS) of project population** | Sample of project population (90% confidence, 10% relative precision) with a combination of onsite data gathering, customer interview, and 15 minute utility interval data if available.<br><br>Most likely to use IPMVP Option B. |
| **Type V – Census sample** | Census sample with a combination of onsite data gathering, customer interview, and 15 minute utility interval data if available.<br><br>Most likely to use IPMVP Options B, C or D, and ASHRAE Guideline 14-2002. |

## Proposed IPMVP Option C Revision

Finally, in support of realizing the above potential impact evaluation cost reduction, the author proposes that industry standard foundational guidelines, such as the IPMVP, ASHRAE Guideline 14, could be revised to explicitly include high volume, high accuracy, high-frequency (hourly, half-hourly, or quarter-hourly) smart meter data. For example, IPMVP Option C could be revised, as follows or

34 PNNL Commercial Building Energy Asset Score (https://buildingenergyscore.energy.gov/), ASHRAE Building Energy Quotient (bEQ) http://buildingenergyquotient.org/

35 For example, the ISO New England M-MVDR reference manual, used for rigorous capacity market validation, states "All reports, studies, specifications and other documents referenced in the Project Sponsor's Measurement and Verification Plan shall have been prepared and published within five years of the Measurement and Verification Plan's submission date to the ISO."

using similar language, where example revisions are in **<u>bold underline type</u>:**

> "*Savings* are determined by measuring energy use at the whole *facility* or sub-*facility* level. Continuous measurements of the entire *facility's energy* use are taken throughout the *reporting period*.
>
> Analysis of whole *facility baseline* and *reporting period* (utility) meter data. *Routine adjustments* as required, using techniques such as simple comparison or regression analysis. *Non-routine adjustments* as required.
>
> **<u>This method is recommended wherever high-frequency (hourly) smart meter data is available for both the pre and post reporting period.</u>** Multifaceted energy management program affecting many systems in a *facility*. Measure energy use with the gas and electric utility meters for a twelve month *baseline period* and throughout the *reporting period*. **<u>Where the measure in question generates savings in one season only, or operates on the same schedule throughout the year (such as facilities with continuous lighting schedules), and high-frequency smart meter data is available, a smaller reporting period may be utilized to shorten reporting timelines</u>**."

## Conclusions

This paper concludes the following:

1. In the utility of the future, the U.S. impact evaluation industry expenditure savings is estimated to be in the range of $21M USD annually, an estimated 17.5% of estimated annual U.S. impact evaluation costs, with a target evaluated impact savings uncertainty of ±5% of evaluated savings at the smart meter.

2. The goal of supporting sustainable, persistent electrical infrastructure throughout the world could be realized by making the following relatively modest EM&V policy updates:
   a. IPMVP Option C, and other foundational industry evaluation guidelines such as ASHRAE Guideline 14, should be updated to explicitly include language addressing high-volume, high-accuracy, high-frequency smart meter data.
   b. Programmes should accurately track the following information in their databases for each EE initiative:
      i. project initiation and completion dates
      ii. measure installation initiation and completion dates
      iii. baseline type
      iv. baseline annual electrical usage per site
   c. Utilities and evaluators should only measure first year impact for projects with high enough savings to be seen in the (weather adjusted) facility baseline consumption data (savings are at least 5% to 10% of the baseline facility consumption data).
   d. Programmes could deliberately incent projects, or bundles of projects, having savings at least 5% to 10% of the facility baseline. This is consistent with the goal of reducing emissions associated with electric generation.
   e. Evaluators should use cloud computing to perform accelerated analysis of utility-grade smart meter power data for entire populations of energy conservation projects.
   f. Evaluators or third-party SaaS firms should leverage methods based on emerging information technologies (ICT, IoT) to standardize and evaluate prescriptive deemed savings on a three- to five-year timeframe for projects below the signal-to-noise level relative to their facilities' baseline usage data.

3. Further:

a. Rapid advances in policy, emerging information technology, and advanced analytics, should be achievable in the near term by capitalizing on existing high-quality smart meter and intelligent communications infrastructure.
b. The methods are equally applicable in any country with a smart grid infrastructure.
c. Accelerated site-by-site data methods represent a low-risk, low cost, rigorous system of accountability to support innovative policies, such as white certificates, paving the way for commoditization of EE initiatives in the transitioned future utility
d. Entire populations of EE savings initiatives could be validated, rather than relying on sampling approaches.

## References


AEP Ohio, Navigant 2015. "2014 Portfolio Status Report of the Energy Efficiency and Peak Demand Response Programs, Volume I 2014 Portfolio Status Report Affidavit of Jon Williams Appendices A – E", May, 2015.

Ameren, Opinion Dynamics, Cadmus, Navigant, Michaels Energy 2016. "Ameren Illinois Company Electric & Gas Residential, and Commercial and Industrial Energy Efficiency Portfolios - PY6 Integrated Evaluation Report", February 16, 2016.

ASHRAE 2002. "Measurement of Energy and Demand Savings", ASHRAE Guideline 14-2002, ISSN 1049-894X, June 27, 2002.

Bertoldi 2011. "Assessment and Experience of White Certificate Schemes in the European Union", presentation for the Joint Research Center (JRC) of the European Commission, December 2011.

Commonwealth Edison Company, Navigant 2014. "Review of EPY5 Total Resource Cost Test Assumptions, Energy Efficiency / Demand Response Plan: Plan Year 5", November 8, 2014.

DNV GL 2015. "The Changing EM&V Paradigm, A Review of Key Trends and New Industry Developments, and Their Implications on Current and Future EM&V Practices", Northeast Energy Efficiency Partnership (*NEEP)*, December 2015.

DTE Energy, Navigant 2015. "DTE Energy's Energy Optimization Portfolio, PY2014 Evaluation Report: Executive Summary", June 30, 2015.

Duke Energy Progress, Navigant 2015. "EM&V Report for the Residential New Construction Program Program Years 2013 and 2014", October 2, 2015.

Duke Energy Progress, Navigant 2015. "2014 EM&V Report for the Energy Efficiency for Business Program", March 1, 2015.

Efficiency Valuation Organization (EVO) 2014. "International Performance Measurement and Verification Protocol Core Concepts", *EVO,* June 2014. EVO 10000 – 1:2014

GDS Associates, Inc., Statewide Evaluation Team, Research Into Action, Apex Analytics, PECO, Navigant 2016. "Act 129 Statewide Evaluator Annual Report, Program Year 6: June 1, 2014 – May 31, 2015", January 15, 2016.

Granderson, J., P., Price, D., Jump, N., Addy, and M. Sohn. 2015. “Automated Measurement and Verification: Performance of Public Domain Whole-Building Electric Baseline Models.” *Applied Energy* 144: 106-113.

Grueneich, Dian and Jacot, David 2014. “Scale, Speed, and Persistence in an Analytics Age of Efficiency: How Deep Data Meets Big Savings to Deliver Comprehensive Efficiency”, *The Electricity Journal,* April 2014: Vol. 27, Issue 3.

The Indiana Statewide Core Program Evaluation Team, TecMarket Works et al., 2012. “2012 Energizing Indiana Programs, EM&V Report”, Prepared for The Indiana Demand Side Management Coordination Committee, June 20, 2013.

ISO New England, Inc. 2012. “ISO New England Manual for Measurement and Verification of Demand Reduction Value from Demand Resources, Manual M-MVDR”, *ISO-NE*, Revision 4, Effective Date: June 1, 2012.

Kurnik and The Cadmus Group 2013. “The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures”, NREL/SR-7A30-53827, April 2013.

Maxwell, John, Gandhi, Nikhil, and Coito, Fred 2011. “Establishing Baselines for Industrial Energy Efficiency Projects”, ACEEE Summer Study on Energy Efficiency in Industry, July 2011.

Navigant Consulting, Inc., Cross, Emily and Mehta, Nishant 2015. “DTE Energy Commercial and Industrial Non-Prescriptive Programs PY2014 Evaluation Report, Section 6 C&I EM&V 2.0 Pilot Results”, June 1, 2015.

Rogers, Ethan A., Carley, Edward, Deo, Sagar, and Grossberg, Frederick 2015. “How Information and Communications Technologies Will Change the Evaluation, Measurement, and Verification of Energy Efficiency Programs”, *American Council for an Energy-Efficient Economy (ACEEE),* December 2015.

Slote, S., Sherman, M., and Crossley, D. 2014. “Energy Efficiency Evaluation, Measurement, and Verification: A Regional Review of Practices in China, the European Union, India, and the United States”, RAP Global Power Best Practice Series, March 2014.

Vrins, Jan. 2015. “Distributed Energy Resources: Lead or Follow.” *Aspen Institute Energy Policy Forum,* July 18, 2015.