Title
Data Availability in Appliance Standards and Labeling Program Development and Evaluation

Author
Romankiewicz, John

Publication Date
2013-06-29

ERNEST ORLANDO LAWRENCE

BERKELEY NATIONAL LABORATORY

Data Availability in Appliance Standards and Labeling Program Development and Evaluation

John Romankiewicz, Nina Khanna, Edward Vine, and Nan Zhou

China Energy Group

Environmental Energy Technologies Division

Lawrence Berkeley National Laboratory

May 2013

This work was supported by the China Sustainable Energy Program of the Energy Foundation and Collaborative Labeling and Appliance Standards Program through the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Disclaimer

Data Availability in Appliance Standards and Labeling Program Development and Evaluation

John Romankiewicz, Nina Khanna, Edward Vine, and Nan Zhou
China Energy Group
Environmental Energy Technologies Division
Lawrence Berkeley National Laboratory

Executive Summary

In the fall of 2012, Lawrence Berkeley National Laboratory released two studies that reviewed the international frameworks for a) appliance standard setting and labeling development and b) impact evaluation of appliance standards, labeling, and incentive programs. The scope and rigor of techno-economic and impact analyses which support and enhance standard setting processes are linked to the tools and data available to the standard setting body (Zhou, Khanna, Fridley, & Romankiewicz, 2012). As energy efficiency programs proliferate, there is an administrative and business imperative to evaluate the savings and processes of these programs to ensure that program funds spent are indeed leading to a more energy-efficient economy. The choice of evaluation methodologies is linked to the types and quality of data available (Zhou, Romankiewicz, Vine, Khanna, & Fridley, 2012).

Indeed, governments around the world undertaking energy efficiency standards development and program evaluation have done so with similar motivations but varying levels of budget and data collection. While data are the foundation of development and evaluation, primary data collection can be costly, and many organizations rely on secondary sources for the majority of their data needs. As China ramps up its efforts in energy efficiency in the 12th Five Year Plan and explores different frameworks for standard development and program evaluation, data availability is becoming an increasingly relevant question.

In this report, we describe the necessary data inputs for both standards development and program evaluation and perform an initial assessment of the availability and uncertainty of those data inputs in China. For standards development, we find that China and its standards and labeling program administrators currently has access to the basic market and technical data needed for conducting market and technology assessment and technological and economic analyses. Some data, such as shipments data, is readily available from the China Energy Label product registration database while the availability of other data, including average unit energy consumption, prices and design options, needs improvement. Unlike some other countries such as the United States, most of the necessary data for conducting standards development analyses are not publicly available or compiled in a consolidated data source. In addition, improved data on design and efficiency options as well as cost data (e.g., manufacturing costs, mark-ups, production and product use-phase costs) – key inputs to several techno-economic analyses – are particularly in need given China's unconsolidated manufacturing industry. For program evaluation, we find that while China can conduct simple savings evaluations on its incentive programs with the data it currently has available from the Ministry of Finance – the program administrator, the savings estimates produced by such an evaluation will carry high uncertainty. As such, China could benefit from an increase in surveying and metering in the next one to three years to decrease the uncertainty surrounding key data points such as unit energy savings and free ridership.

The organizations carrying out these development and evaluation studies should place an emphasis on measuring these levels of uncertainty and sensitivity in order to best serve the goals of the energy efficiency standards or programs they are implementing. Budget for these studies is often limited in China. It is therefore critical to establish relationships between the data inputs and algorithms used in standards development and program evaluation and their associated uncertainty and sensitivity. Data inputs come from both primary and secondary sources collected from a wide range of literature and surveys of stakeholder groups, including consumers, manufacturers, and retailers. Each data source will carry with it an associated uncertainty, and each data input will have a relative sensitivity on the output of the evaluation or development algorithm. Quantifying uncertainty and sensitivity can help China prioritize data improvement and increased data collection, and it will also help China justify budget expenditure on such endeavors. Given the scope of China's appliance standards, labeling, and incentive programs and the scale of energy consumption growth in this sector, improvements to practices in development and evaluation should continue to be recommended.

For the purposes of this preliminary study on China's data availability, an effort has been made to categorize the data sources that China would need for standards development or program evaluation as 1) existing data or methodologies, 2) literature, 3) datasets, and 4) survey metering. The sensitivity and uncertainty of each data input has been evaluated using the authors' knowledge of evaluation literature and familiarity with Chinese data availability.

For each section on standards development and program evaluation, we have proposed three options based on China's existing and potential capabilities in this space. The first option in each case uses existing data and literature reviews, and for this reason can be conducted immediately. Indeed, this option may mimic some of the studies and processes China is currently implementing in the energy efficiency space. The second option proposes a small budget increase to conduct targeted data collection (survey/metering) that will improve results and reduce uncertainty for the related analysis, with the idea that this option could be implemented on a short-term time scale of one to three years. The third option involves long-term time scale implementation (five to ten years) of increased metering, surveying, and use of future datasets to achieve the lowest uncertainty with an effective use of budget.

In standards development, we find that a more rigorous standards development framework based on market and technology assessment, technological, economic or life-cycle cost and specific impacts analyses should be supported by improved data collection on design and efficiency options and cost and should be prioritized over the medium to long term. Expanding data collection and widening the scope of analyses will help China refine its technical and cost-related analyses, two areas of relatively high uncertainty and sensitivity, in the medium term. Over the longer term, developing extensive sets of technical, economic and environmental data specific to Chinese products will enable China to develop and consistently use new analytical capabilities in a comprehensive framework for standards development.

In program evaluation for standards, labeling, and incentives, we find that data types with high sensitivity and high or medium uncertainty should be prioritized for increased or improved data collection in the medium term (1-3 years). Both annual energy savings per unit and free ridership fit this description, and China can plan increased surveying and metering to reduce the uncertainty of these values. Data collection and evaluation options in the long-term will be expanded based on data types with medium or low sensitivity and medium uncertainty, including product lifetime, baseline cases, usage adjustment factors, compliance rates, participant spillover, and market effects.[1]

Whether or not the Chinese government seeks to move towards these options that involve increased data collection and may require additional budget depends on their goals in the energy efficiency space. Increased involvement of, and data collection from, manufacturers may aid in improving and increasing the transparency of standards development. Additional budget for surveying and metering may yield more certainty in the energy savings estimates from incentive evaluations. A full assessment of China's data availability will certainly aid the Chinese government's efforts in energy efficiency planning as they continue to improve their rapidly expanding standards, labeling, and incentive programs.

[1] Note that free ridership, participant spillover, and market effects are exclusive to incentive evaluations.

Table of Contents

1. Introduction

Lawrence Berkeley National Laboratory (LBNL) recently completed reviews of methodologies used in different countries around the world (U.S., E.U., and Australia) to develop appliance energy efficiency standards (Zhou, Khanna, Fridley, & Romankiewicz, 2012) and evaluate the impact (energy savings) of standards, labeling, and incentives programs (Zhou, Romankiewicz, Vine, Khanna, & Fridley, 2012). At a fundamental level, these reviews are useful to any country developing appliance standards and evaluating energy efficiency programs in that they describe not only the methodologies used in these processes but also the associated motivations and data requirements.

As the Chinese government continues to implement its standards, labeling, and incentive programs in support of energy efficiency goals in the 12th Five Year Plan, it will seek to improve its development and evaluation processes where possible to achieve higher levels of energy savings. The importance of program evaluation cannot be underemphasized. Not only can program evaluation quantify and verify the energy savings and environmental impacts of a specific energy efficiency program, but it can also help policymakers and program managers improve program design and better understand remaining market potential for energy efficiency.

In the second phase of the project, LBNL was commissioned to develop simplified development and evaluation frameworks for use by the China National Institute of Standardization (CNIS) based on the best practice international methodologies reviewed in previous reports with additional consideration of China's own data availability. Data are the foundation for standards development and program evaluation. This report describes the necessary data inputs for both standards development and program evaluation and performs an initial assessment of the sensitivity and uncertainty of those data inputs in China, using the authors' knowledge of development and evaluation literature as well as familiarity with Chinese data availability.

Section 2 describes the relationship of development and evaluation goals, data uncertainty, and budget availability. As CNIS continues to improve its capabilities in standards development and program evaluation, decisions to reduce data uncertainty will be tied to budget availability and overall programmatic goals. Sections 3 and 4 discuss the following topics for both standards development and program evaluation, respectively:

1. An overview of methodologies from the international framework review report
2. Key data requirements for those methodologies
3. Assessment of China's data availability
4. Three methodology options for China to consider

For those three options, the first will be largely based on China's existing capabilities and data availability. The second option proposes a small budget increase to conduct some data collection that will improve results and reduce uncertainty for the related analysis, with the idea that this option could be implemented on a shorter time scale of one to three years. The third option is a long-term implementation (five to ten years) of a number of improved data collection methods to achieve the lowest uncertainty with an effective use of budget.

2. Data and uncertainty issues for standards development and program evaluation

There are strong relationships between the goals, data availability, uncertainty, and budget resources for any standards development or program evaluation process. These relationships (Figure 1) are described in this section.

The first step in standards development or program evaluation involves setting the goals and scope of those processes. For example, in the U.S., the goal of standards development is to save energy and reduce carbon emissions at the national level while resulting in net savings to the consumer and avoiding major economic impacts to manufacturers. In program evaluations in the U.S., utilities are often seeking to determine energy savings, peak demand reductions, or market transformation and in turn calculate the cost effectiveness of dollars spent on those programs. In the U.S., cost effectiveness mandates have driven evaluators to often spend additional money on data collection and evaluation in order to reduce uncertainty and improve estimates of energy savings. In China, program evaluation does not yet have these cost effectiveness mandates, and while mandatory standards are set in the interest of energy efficiency, they do not yet have formalized parameters for measuring and optimizing fiscal savings for the consumer. China's energy intensity and carbon intensity goals are providing additional impetus for implementation and accurate impact evaluation of a growing number of energy efficiency programs.

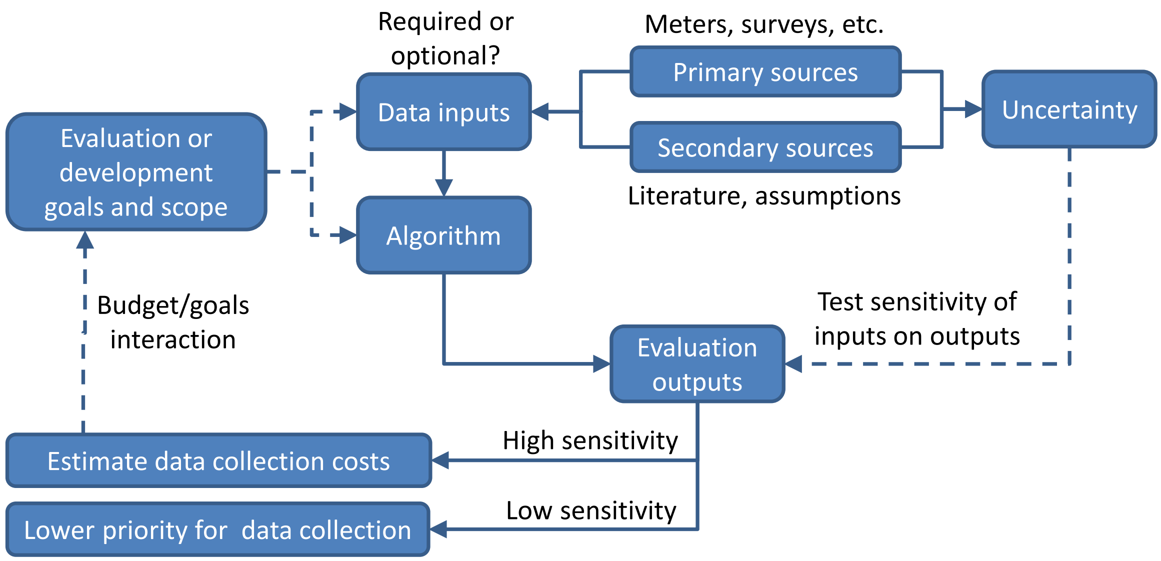

Figure 1. Interaction between goals, data, uncertainty, and budget for standards development and program evaluation

In addition to determining the goals of any program evaluation, the scope of the process should also be outlined in terms of what measures (or products) were covered, what delivery strategy was used, and what impacts were expected in the design of the standard and program. For instance, determining the energy savings from a rebate program for compact fluorescent light bulbs, the market transformation from a voluntary labeling program for refrigerators, or peak demand reduction for a retrofit program for commercial central air conditioning units are all viable scopes for program evaluation. Once the scope has been set, then the required data inputs and appropriate algorithms can be determined as shown in Figure 1. In evaluations, both primary and secondary data sources are used for collecting data: for example, reviewing engineering literature for estimating savings, or conducting surveys to collect building occupancy hours. While secondary sources largely consist of existing literature (including technical resource manuals) and simulations, primary sources involve metering and surveys of manufacturers, retailers, and consumers (participants).

Certain energy efficiency measures often have commonly associated evaluation and data collection methods. Figure 2 shows different types of energy efficiency measures and delivery strategies as well as their associated data sources and collection methods. While some evaluation methods can be entirely prescriptive based on deemed savings estimates, other evaluation methods are custom based on site-specific conditions and employ end-use metering. For example, a CFL buy-down program may simply rely on a prescriptive approach using deemed savings from secondary sources (prescriptive method), while an air-sealing home retrofit program may call for primary data collection via monitoring (custom method). Typically, prescriptive approaches are used for programs where a high number of measures are expected and a significant amount of program experience has already been accumulated, such as CFL and refrigerator rebate programs. Custom approaches are used for measures where installation conditions vary and thus monitoring or modeling is warranted. In between, surveys are often used as a way of reducing uncertainty for deemed savings values or corroborating monitoring results.

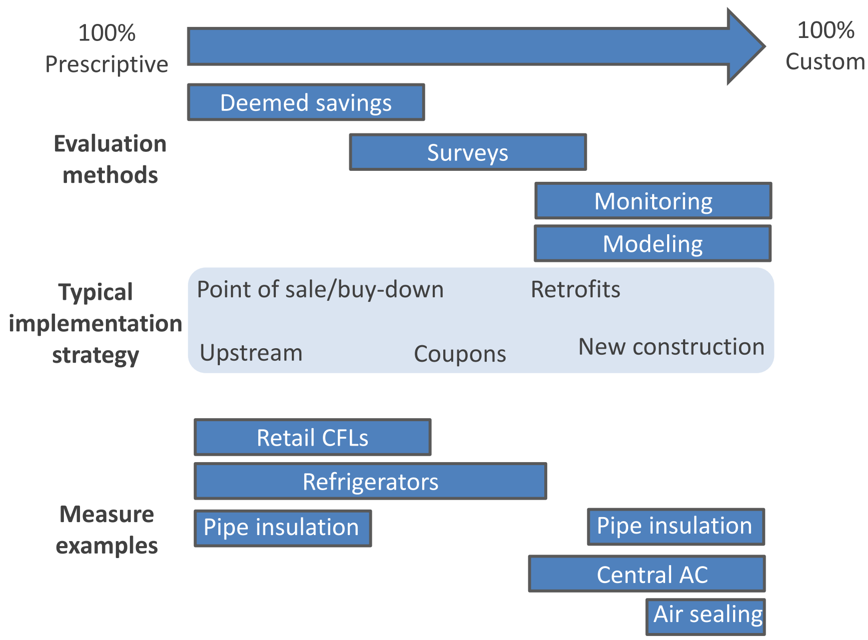

Figure 2. Prescriptive vs. custom evaluation methods and related implementation strategies and measures, Source: Adapted from (Dent & Enterline, 2012)

All data sources, primary or secondary, will carry with them an associated degree of uncertainty. This uncertainty should be estimated. Additionally, the sensitivity of altering each input on the output should be tested using Monte Carlo simulation or some other estimation methodology. Together, the uncertainty and sensitivity of various data inputs can help determine where efforts and budget should be spent to improve data inputs and in turn improve the end accuracy of any evaluation or standards development. If energy efficiency goals are present, then this type of analysis enables the accurate assessment of whether those goals were met or not.

If it is found that a certain data input has high uncertainty and sensitivity on the final output, then funds should be allocated for collecting additional primary data that would reduce that uncertainty. The impetus for that particular budget, however, must originate from the need for accurate and credible evaluations. Likewise for standards development, if there is a goal for highest possible national energy savings or lowest possible life-cycle costs[2] for the consumer, risk analysis can help identify uncertainties for all data inputs in the interest of meeting those goals.

In the U.S., risk analysis has been performed for utility-sponsored energy efficiency programs to identify which programs' energy savings estimates have the highest uncertainty. Commissioned by the California Public Utilities Commission, Jacobs et al. (2006) carried out such a risk analysis for an investor-owned utility in California. This example will help illustrate the interaction of evaluation methodology, data inputs, and uncertainty. Figure 3 shows the central estimate for energy savings of various energy

[2] Life-cycle costs are defined as the sum of upfront purchase costs and operation & maintenance costs throughout the lifetime of a product.

efficiency programs that the utility had implemented, such as residential interior lighting, commercial and industrial processes, and commercial and industrial refrigeration. (Jacobs, Kromer, & Hall, 2006)

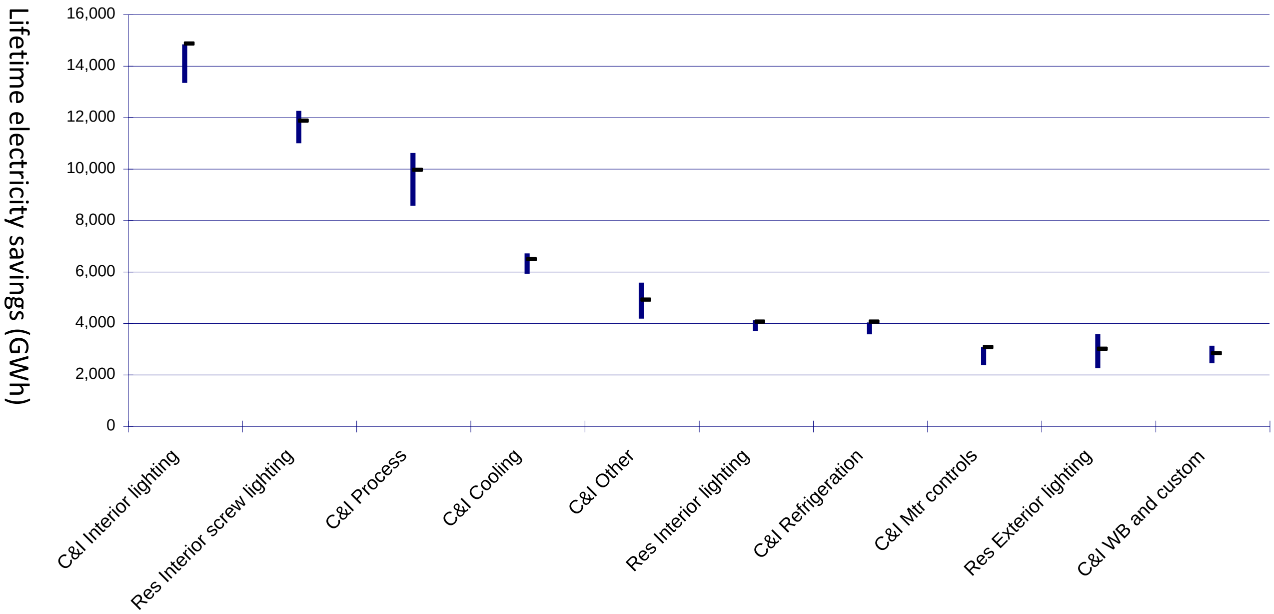

Figure 3. Estimated lifetime electricity savings and uncertainty by measure – California (Jacobs, Kromer, & Hall, 2006), Note: the hash mark represents the estimated lifetime electricity savings for each measure while the bars represent the range of uncertainty at a 90% confidence interval; C&I: commercial and industrial, Res: residential, WB: whole building.

The uncertainty of the savings for each program will depend on the uncertainty for each data input for that program's energy savings algorithm. For example, in this case, the following energy savings evaluation algorithm was used:[3]

$$\Delta kWh_{lifetime,net,measure} = units \times \frac{\Delta kWh_{gross}}{unit} \times NTG \times EUL$$

- $\Delta kWh_{lifetime, net, measure}$= Net energy savings
- ΔkWh_{gross}/unit = Gross energy savings per measure (unit savings)
- Units = number of measures (count)
- NTG = net to gross ratio[4]
- EUL = effective useful life[5]

[3] For more information on different types of algorithms and methodologies used in calculating gross and net savings, refer to (Zhou, Romankiewicz, Vine, Khanna, & Fridley, 2012).
[4] The net-to-gross (NTG) ratio determines the actual energy savings attributable to a particular program, as distinct from energy efficiency improvements that would have occurred without the program. The NTG ratio equals the program's net energy savings divided by the program's gross energy savings. Factors such as free ridership and market effects contribute to NTG ratios according to definitions which vary widely throughout the world.
[5] The effective useful life (EUL) refers of a measure (i.e., measure lifetime) is the period of time that the measure is expected to perform its intended function in a typical installation. (Vine, Hall, Keating, Kushler, & Prahl, 2012)

This algorithm is a standard form algorithm for net energy savings that is widely accepted in the U.S., though definitions vary on how NTG ratios should be calculated (Zhou, Romankiewicz, Vine, Khanna, & Fridley, 2012). The number of measures (count), gross energy savings per measure (unit savings), and net to gross ratio all have uncertainty. This uncertainty is typically calculated by establishing high and low-end values for each data point. For instance, if 10,000 CFL rebates were cashed in, the range of estimates for CFL's installed and replacing incandescent bulbs (number of measures) could be between 6,000 (low-end) and 7,000 (high-end). The EUL could be between 8 and 10 years depending on the different types of CFL's purchased. A simple convention assumes that the range between those two values represents a 90% confidence interval or that there is less than a 5% chance that the value lies outside of that range (Kiefer, 1993). Figure 4 shows how each of those data points contribute to overall uncertainty for each of the programs previously referenced in Figure 3. Unit savings tends to have the largest contribution to uncertainty, followed by NTG ratios with count contributing the least uncertainty.

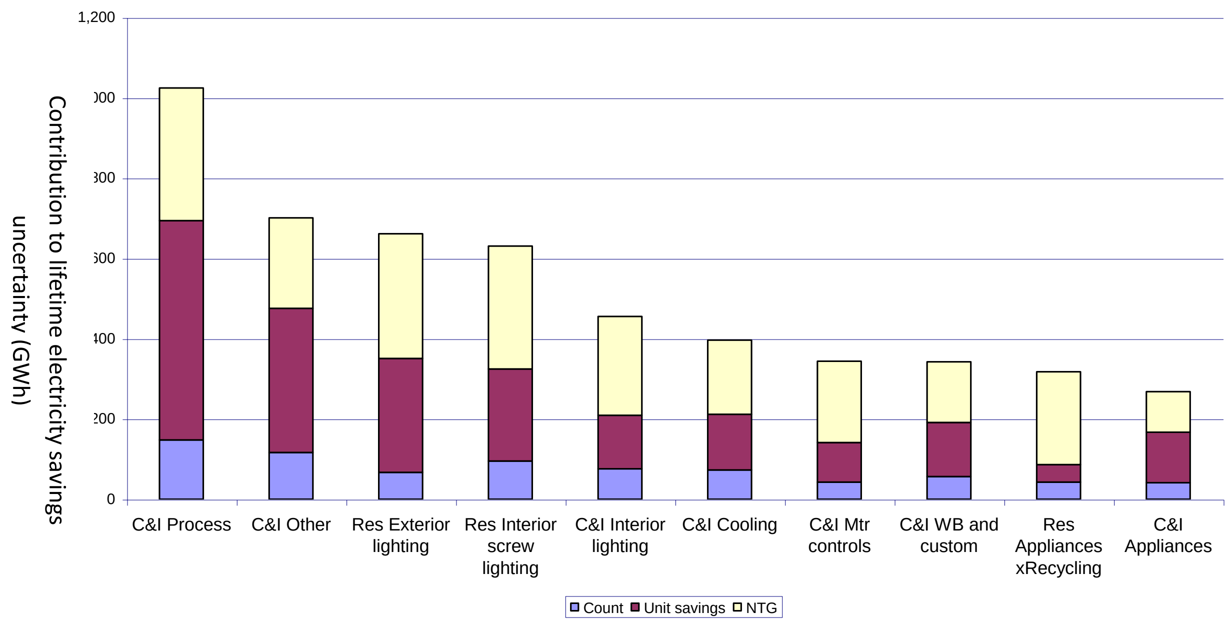

Figure 4. Contribution to lifetime electricity savings uncertainty by risk category (Jacobs, Kromer, & Hall, 2006) Note: The uncertainty bars in Figure 3 for each program correspond to those in this figure. The values are not perfectly matched due to uncertainty contributions from EUL and other factors.

In California, there is an impetus for the utilities to ensure that their energy savings are credible and reliable (not just estimated) given energy efficiency targets and cost effectiveness mandates (getting the most efficiency with the least amount of money). In China, there are energy efficiency targets, but explicit cost effectiveness mandates do not yet exist and estimated savings are acceptable. Given the limited budgets of organizations like CNIS, however, it is essential that only the most critical issues influencing the estimated savings be investigated – for example, the data issues that bring the most uncertainty. The previous example – and Figure 4 in particular – demonstrates how a risk analysis may help prioritize measures to improve data quality and reduce uncertainty. Besides uncertainty, other program parameters may also have an impact on the evaluation result, such as the types of measures or implementation method for those measures. As energy efficiency targets become harder and harder to meet (with potentially the added pressure of tighter budgets), this type of risk and uncertainty analyses will help CNIS focus on conducting the most accurate and credible evaluation studies possible. Accurate and credible evaluations will aid future policy decisions and enable cost-effective energy savings. (Kiefer, 1993)

Table 1 on the following page outlines the common primary and secondary data sources used in standards development and program evaluation. Often, there is an overlap between the data required for development and evaluation. Additionally, some governments carry out national energy surveys, such as the Residential Energy Consumption Survey (RECS), which gather socio-demographic data and energy usage in homes across the U.S. Sections 3 (standards development) and 4 (program evaluation) will review these data sources in detail, provide an initial assessment of data availability in China, and offer methodology recommendations for China and options to increase data collection over time.

Table 1. Primary and secondary data sources for standards development and program evaluation, adapted from (Kiefer, 1993)

		Primary data sources							**Secondary data sources**		
		Direct measurement (metering)	**Laboratory test data**	**Manufacturer surveys**	**Program participant surveys**	**Retailer surveys**	**National energy surveys (e.g. RECS)**	**Government statistics**	**Energy/evaluation literature**	**Deemed values (technical resource manuals)**	**Simulations and modeling**
Standards development only	**Electricity prices (real or forecasts)**							X	X		
	Product/component costs			X							X
	Retail prices & mark-up					X			X		
	Discount rates							X	X		
	Costs of use (installation, repair)			X	X						
	Technology options, product classes		X	X							
Standards development and program evaluation	**Usage**	X					X		X		
	UEC	X	X	X			X		X	X	
	Existing stock, saturation						X	X	X		X
	Lifetime	X		X	X		X		X		

	Sales/shipments (real or forecasts)			X		X		X			
	Site-to-source energy conversion factors							X	X		
	Emission factors							X	X		
Program evaluation only	**Free ridership**				X				X	X	
	Compliance rate		X								
	Number of participants and non-participants				X				X		
	Participant spillover				X				X		
	Market effects				X				X		
	Heating and cooling degree days							X	X		

3.Standards development options

3.1 Options for standards development

On the basis of the LBNL review of international standard-setting frameworks (Zhou et al. 2012a), possible options for standards development can be considered from two different perspectives: two options in terms of the procedural process for conducting standards development and four groups of possible analyses in terms of the content of standards development. The procedural options frames the type of process through which standards development occurs, while the analytical options shapes the scope and content of the standards development analyses.

3.1.1 Procedural Options for Standards Development: Legislative versus Analytical Process

From a procedural point of view, there have been two main tracks of standards development: a legislative or rulemaking-based process and an analytical process.

The U.S. and Japan both follow a specific administrative process that is guided by legislation for setting its standards, with the processes and timeline for standard-setting analyses clearly defined by legislation. As part of this government-driven process, most of the standard-setting analysis is conducted and managed in-house by the government (by either technical experts or sub-contractors). There are also specified requirements that must be met in the review process for proposed standards, such as strict timeframes for public reviews and comments. Because the processes and timeline are laid out in specific legislation, the administrative approach to setting standards provides a standardized and uniform framework with guaranteed stakeholder participation for developing new or revised standards regardless of the product.

In contrast, the EU and Australia tend to follow a more analytical approach to standard development that is shaped by comprehensive studies to evaluate if new or revised standards are needed and where the standard should be set. Both the EU's Ecodesign Preparatory Studies and Australia's Regulatory Impact Studies follow a predefined set of analyses, although there is flexibility in the process for conducting these analyses. In both cases, the majority of the analyses is conducted by a consulting firm hired through a competitive process, rather than conducted by government agencies. The results of the studies are then reviewed with stakeholders through discussions and meetings. This approach can result in a faster and more flexible process for developing new or revised standards, while still providing standardization in terms of the core sets of analyses conducted as part of the process.

China currently has a standards development framework that is not procedurally-based. Although the products chosen for a new or revised standard are determined by the central government, the subsequent analyses to set the standard level are not standardized and do not follow either a legislative or analytical process. In the absence of new legislation that specifically dictates the process for setting a standard, China may consider adopting the analytical approach of using a standardized comprehensive study to frame their standard-setting analyses.

3.1.2 Content-based Options for Standards Development: Sets of Analyses for Informing Standard-Setting

From a content perspective, standards development can be divided into four main groups of analyses:

1. ***Market and Technology Assessment:*** characterizes the existing product market and sheds light on the existing and prospective technology landscape; the assessment serves as the basis for further technological analysis and to justify policy action such as the implementation of MEPS [6]
2. ***Technological Analyses:*** assesses all possible options for efficiency improvements and includes engineering analysis conducted to determine specific gains and costs of efficiency options
3. ***Economic or Life-cycle Cost Analyses:*** evaluates the life-cycle cost to consumers (including sub-groups of consumers in the case of the U.S.) of potential standard levels taking into consideration products' up-front costs, usage costs, repair and maintenance costs, and savings associated with water and energy savings
4. ***Specific Impacts Analyses:*** evaluates specific economic or environmental impacts using underlying data on existing stock and forecasted future sales; include national impact (financial costs and savings), manufacturer impact, and environmental impact (energy, CO_2, water) analysis.

These four groups of analyses are rarely conducted in isolation, and outputs from one analysis often serve as inputs into another analysis. For example, in the major standards programs reviewed, the standard-setting processes in use internationally utilized several analyses from each of the four groups with many of the analyses closely integrated and interlinked across the groups. In the U.S., for instance, the market and technology assessment identified technology and design options that could be considered in the screening analysis, with the viable options that survived the screening analysis being evaluated as potential standard levels based on their life-cycle cost, national, and environmental impacts.

China currently conducts a combination of limited analyses from the four groups and the linkages between the four groups of analyses are less evident. In the specific impacts analysis, for instance, CNIS only conducts a simplified environmental impact analysis on energy and emissions and does not delve into national financial impact analysis or manufacturer impact analysis. Depending on data and resource

[6] MEPS, or minimum energy performance standard, legislates the maximum amount of energy that may be consumed by a product of a certain type.

availability, China may consider adopting a standard-setting framework based on all of the analyses across the four groups.

3.2 Key data requirements

Table 2 shows the key data requirements for the major analyses in both the procedural and content-based options and highlights the major types of data that are crucial inputs to the standard-setting process. As shown in

Table 2, the key data requirements are incorporated into the standard-setting analyses in major standards programs of the U.S., Australia, and the European Union. However, some data requirements are program-specific and reflect an additional layer of complexity added to the major analyses, such as the use of household demographic data, household energy usage data, and government administrative costs data.

Table 2. Key Data Requirements for Standards Development

Data Needs	Data Availability and Use Internationally	International Data Sources (U.S. unless indicated otherwise)
	Market and Technology Assessment	
Historical Shipments	U.S., Australia, EU	Trade associations, market research companies, customs data (Australia)
Market Segmentation (e.g., technology subgroups)	U.S., Australia, EU	
Non-regulatory programs	U.S., Australia, EU	Review of past and existing efforts by manufacturers, utilities and other parties
	Screening Analyses	
Product classes	U.S., Australia, EU	Discussions with manufacturers, trade associations, etc.
Technology options	U.S., Australia, EU	Trade organizations, manufacturers, consultant experts
Product prototypes	U.S., Australia, EU	
	Engineering Analysis	
Design options or efficiency levels	U.S., Australia, EU	Manufacturer data, engineering models, market analysis, stakeholder consultation
Maximum technologically achievable efficiency	U.S., Australia, EU	Same as above, existing product prototypes
Manufacturing cost (by design option)	U.S., Australia, EU	Tear-down analysis; bottom-up manufacturing cost assessment based on detailed bill of materials
	Life-cycle Cost and Payback Period Analysis	

Retail prices for products	U.S., Australia, EU	Survey of existing price on market, survey of manufacturers and market experts, applying mark-up to manufacturer costs
Mark-up for products	U.S., Australia, EU	Financial statements for major manufacturers from Securities & Exchange Commission
Annual energy use per unit	U.S., Australia, EU	Calculated using formula and assumed usage
Annual water use per unit	U.S., Australia, EU	Calculated using formula and assumed usage
Average energy prices	U.S., Australia, EU	Published EIA data from utilities
Average water prices	U.S., Australia, EU	Consulting/research company's survey of water utilities
Installation costs	U.S., Australia, EU	Historical Producer Price Index data (from Bureau of Labor Statistics)
Maintenance and repair costs	U.S., Australia, EU	No specific source; often assumed to be same for different efficiency levels

Data Needs	Data Availability and Use Internationally	International Data Sources (U.S. unless indicated otherwise)
Shipments Analysis		
Existing stock	U.S., Australia, EU	Calculated from shipments
Market saturation (ownership, market shares)	U.S., Australia, EU	Residential household and national housing surveys; trade association data on market shares
Lifetime or retirement function	U.S., Australia, EU	Residential household and national housing surveys
National Impact Analysis		
Future shipment forecasts	U.S., Australia, EU	Regression analysis, analysis of relative price elasticity of demand from Shipments Analysis
Total annual average energy use	U.S., Australia, EU	Calculated using other inputs
Energy price forecasts	U.S., Australia, EU	EIA Forecasts
Site-to-source energy conversion factors	U.S.	Values in National Energy Modeling System corresponding to EIA Annual Energy Outlook
Discount rates	U.S., Australia, EU	Derived from estimates of cost to finance purchases
Rebound effect considerations	U.S.	Literature review on rebound effects for particular product
Costs to government for administering program	Australia	Australia Business Cost Calculator
Manufacturer Impact Analysis		
Manufacturer prices	U.S., Australia, EU	Corporate annual reports to government
Manufacturer interviews on technology trends	U.S., Australia, EU	Manufacturer interviews
Additional investment needs	U.S., Australia, EU	Manufacturer interviews
Production cost changes (materials, labor, overhead)	U.S., Australia, EU	Engineering analysis
Revenue impacts (e.g., due to higher prices but lower sales)	U.S., Australia	Cash flow analysis
Additional non-production industry cost (e.g., training, record keeping, etc.)	Australia	Australia Business Cost Calculator
Life-cycle cost Subgroup Analysis		
Demographic data for representative consumer sample (usage, energy prices)	U.S.	EIA residential household energy consumption survey
Household energy use data for representative consumer sample	U.S.	EIA residential household energy consumption survey
Environmental Assessment		
Emission factors	U.S., Australia, EU	NEMS model; standard emission factors for fuel combustion

Overall, there are three major categories of key data requirements that feed into the main standard-setting analyses: market data, technical data, and cost data.

3.2.1 Market data

Market data on product sales and groupings and data on the product's energy-related characteristics (e.g., usage patterns, per unit energy consumption and lifetime) are both needed to characterize the scale for potential energy savings and justify the development of a new or revised standard. The market data are often collected through trade associations, market research companies and government statistics while the product usage data can be obtained from household user surveys, manufacturer surveys and metering studies. While the level of detail of the market and usage-related data may vary by country – some countries such as the U.S. may have detailed usage data broken out by consumer subgroups while others only have aggregate national averages – estimates of sales and average values for hours of use, per unit energy consumption and lifetime are needed at a minimum to characterize baseline energy consumption against which to measure the potential gains by a standard.

3.2.2 Technical data

Technical data on whole product or component design options are a second category of required data in the standard-setting process. Data related to a product's technical design options and performance are needed to identify the potential efficiency levels at which to set the standard and to evaluate the potential energy savings at each level. The necessary technical data include data on technology and design options associated with different efficiency levels, including the current maximum technically feasible efficiency levels. These data are often obtained through engineering studies and models, manufacturer interviews, and studies by technical experts. In the case of the U.S., tear-down analysis[7] is also undertaken by consulting firms as a primary source of detailed technical data. The scope of technical data that are available and accessible to standard-setting regulators directly impacts the effectiveness and stringency of the resulting standard. If a regulator only has limited technical data and cannot thoroughly grasp the technological landscape of the product being regulated, the level at which the standard is set may not be stringent enough to produce significant savings.

3.2.3 Cost data

A third category of key data requirement is cost data – particularly data related to consumer and manufacturer costs – which serve as inputs to life-cycle cost analysis that is conducted to evaluate the economic impact of proposed standard levels and justify the setting of a new or revised standard. Key consumer cost data typically include retail prices and retail mark-up margins and energy and water prices, while manufacturer data include manufacturing and production costs and prices. For design options that are not yet commercially available and for which retail prices are not available, engineering analysis and manufacturer interviews can be used to estimate consumer costs. Consumer cost data can generally be obtained from market and manufacturer surveys, consulting companies and utilities, while manufacturer data are often collected from interviews with manufacturers, and engineering and financial analyses. In most programs (including the U.S, Australia and European Union), cost data are essential as life-cycle cost effectiveness is one of the criteria that must be met by a new standard.

3.3 China data availability

Table 3 outlines the key data requirements for standard-setting analyses and their availability in China. The bolded items indicate data types that China is already collecting to some extent through surveying and other data collection methods and which are incorporated into the standard-setting process (CNIS, 2012) (Liu, 2011). Most of the data that China is already collecting is not readily accessible to the public, although some information (e.g., historical sales, average energy use per unit, product retail price) is publicly available through the annual white papers published by CNIS and on related program websites

[7] Tear-down analysis is an analysis performed in the U.S. standard-setting process where a product is purchased from the retail market and dismantled in order to itemize individual parts and estimate costs.

such as TopTen China website. Most of the other data is compiled through the China Energy Label product registration database. For each data type, its associated uncertainty and its impact on the results of the standard-setting analyses (i.e., sensitivity) are indicated. For the purposes of this preliminary study, the levels of uncertainty and sensitivity are estimated based on the authors' knowledge of evaluation literature and familiarity with Chinese data availability. These hypothesized levels can later be tested once formal standards development studies for China are underway.

Table 3. Standard-setting Data Requirements with China's current data sources and associated uncertainty and sensitivity, Note: bolded data types are for data that China has already collected to some extent

Data input	Data source	Uncertainty (High, medium, low)	Sensitivity (High, medium, low)
Historical Sales	CELC white paper, market research	Low	High
Forecasted Sales	Assumptions, calculation	Low	High
Market Segmentation (Product classes)	Market survey, research	Low	Medium
Technology Options	Manufacturers, literature, market survey	Low	Medium
Design and Efficiency Options	Manufacturers, engineering models, market survey	Medium	High
Maximum Technologically Feasible Efficiency	Literature, manufacturers	Low	Low
Manufacturing Cost by Design Option	Calculations	Medium	High
Product Retail Price	Market survey, manufacturers survey, Calculations	Medium	High
Product Mark-Up	Calculation, assumptions	High	Medium
Annual Energy Use Per Unit	CELC white paper, calculations	Medium	High
Average Energy Prices	Utilities, assumptions	Low	High
Lifetime or retirement function	Literature, assumptions	Medium	Medium
Installation Costs	Market survey, manufacturers	Medium	Medium
Maintenance and Repair Costs	Market survey, manufacturers, assumptions	Medium	Low
Forecasted Energy Prices	Assumptions, calculations	Medium	High
Production Cost Changes	Calculations	High	Medium
Household Demographic Data	Provincial and city-level statistical yearbooks	Low	Low
Household Energy Usage Data Sample	Household surveys	High	Medium
Site-to-source Energy Conversion Factor	Literature, assumptions	Medium	Low
Discount Rate	Calculation, assumptions	Low	Low
Emission Factors	Literature, assumptions	Medium	Low

From Table 3, it can be seen that the basic market and technical data needed for the market and technology assessment, and technological and economic analyses in the standard-setting processes are available in China, although the rigor of the available data and the reliability of the data sources are uncertain. In some instances, such as data on shipments – both historical and forecasted – and on product groupings and technology options, existing data in China are relatively sound because they are either collected directly by the China Energy Label Center or can be ascertained from simple market or manufacturer surveys.

In other instances including data on the annual average unit energy consumption, energy prices, retail prices and design and efficiency options, there is more room for improvement in data collection. These data points all have high impacts on the analytical results but face medium uncertainty. Improving the data on design and efficiency options – a key input to several techno-economic analyses - is especially important for China given the high number of manufacturers that exist for some products and the greater likelihood of overlooking some efficiency options.

Lastly, cost data are a major area of weakness in China's standard-setting process. While CNIS has collected data on manufacturing costs by design options, the rigor of the data is low as the data are often based on interviews with manufacturers rather than from the more comprehensive engineering models and tear-down analyses used to independently derive manufacturing costs outside of China. Likewise, data on product mark-ups, an important factor in determining the true retail prices of products, are not incorporated into the standard-setting process in China. Other cost data related to manufacturer-side production changes and the product's use-phase are also not readily available, and consequently, limit the potential for China to conduct comprehensive manufacturer and consumer impact analyses when considering different proposed standard efficiency levels.

3.4 Identifying standards development options for China

Given China's current data availability and analytical capacities, three options are possible for bolstering the standard-setting framework in the immediate future, in the short-term (1-3 years) and in the long-term (5-10 years). These options can be applicable to either the legislative or analytical approach to setting standards (which are not currently conducted in China), but would require China to adopt a standardized approach of conducting an expanded set of analyses rather than selected analyses completed in an ad hoc manner. As shown in Table 4, the first two options require data for a combination of the four main groups of analyses, while the long-term option requires data for all of the analyses in the four groups. Each of the three options, therefore, requires – to varying degrees – additional data collection and input to expand the scope of the standards-setting analyses and to strengthen the technical basis for standards development. Specifically, the range in depth of data collection needs (from low to high) include the following sources:

1. Existing data or methodology: use China-specific data that have already been collected; or if no Chinese data are available, use international data that have been collected and can be used as suitable proxies. For forecasts, use existing forecasting methodology without further refinements
2. Literature: collect data from Chinese or international literature or similar secondary sources
3. Datasets: collect data from existing datasets that may not have been previously used, or from new datasets that China should have in the near future
4. Survey: conduct a comprehensive survey to gather new data

Table 4. Data sources for standards development options, Note: Exclude means that data point is not needed for the specified analysis in this option. Derive means that additional analyses or reviews are needed to gather the necessary data for the specified analysis in this option.

Data input	Option 1 Implement now	Option 2 Short-term (1-3 years)	Option 3 Long-term (5-10 years)
Historical Sales	Existing data	Existing data	Existing data
Forecasted Sales	Existing methodology	Estimate with literature	Survey
Market Segmentation (Product classes)	Existing data	Existing data	Survey
Technology Options	Existing data	Existing data	Survey
Design and Efficiency Options	Existing data	Survey	Survey
Maximum Technologically Feasible Efficiency	Existing data (international)	Existing data (international)	Derive (international review)
Manufacturing Cost by Design Option	Existing data	Datasets, compare with international data	Derive (analyses)
Product Retail Price	Existing data	Survey	Survey
Product Mark-Up	Literature	Datasets, survey	Derive (analysis)
Annual Energy Use Per Unit	Existing data	Survey	Survey
Average Energy Prices	Existing data	Datasets	Survey
Lifetime or retirement function	Literature	Datasets	Survey
Installation Costs	Literature	Datasets	Survey
Maintenance and Repair Costs	Exclude	Literature	Survey
Forecasted Energy Prices	Existing methodology	Literature	Datasets
Production Cost Changes	Exclude	Literature	Survey, datasets
Household Demographic Data	Exclude	Existing data	Survey, datasets
Household Energy Usage Data Sample	Exclude	Exclude	Survey
Site-to-source Energy Conversion Factor	Literature	Datasets	Datasets
Discount Rate	Existing data	Literature	Datasets
Emission Factors	Literature	Datasets	Datasets

Option 1 builds on China's current data availability and analytical capacities by adding in a few cost analyses that can be conducted using existing data or data estimates from literature review. This option can be realized through a simplified analytical framework such as the Policy Analysis Modeling System, which provides a tool and default international data for conducting the major techno-economic analyses. Under this framework, China's technological analysis can be improved with international data on maximum technologically feasible efficiency levels. Cost data and cost-related analyses in China's standards development will also be bolstered, with the addition of a first-order estimate of product mark-up and use of existing data on installation costs.

Option 2 would enable China to both expand its standards development data collection for key analyses but also widen the scope of standard-setting analyses to include new areas such as manufacturer impact analysis and more comprehensive technological analysis. This option would require increased budget to support expanded technical and market surveys in areas such as design and efficiency options, product retail prices, and annual energy use per unit. It would also require more literature review to gather cost data that have not been collected before, including maintenance and repair costs and production cost changes. For data that were previously estimated using only literature, these estimates will be refined and better tailored to China by using actual datasets that are already available or can be made available over the short-term.

Option 3 reflects a path in which China adopts a comprehensive framework for standards development featuring all of the standard-setting analyses being conducted internationally. Adopting this framework will enable China to develop extensive sets of technical, economic (particularly cost) and environmental data as well as build a complete set of increasingly complex analyses across the four main groups. Building up this framework in China will require significant budget increases to conduct expanded surveys to collect specific data that had previously only been estimated using existing datasets and to refine existing data through expanded data collection. This option will, however, enable China to develop and consistently use new analytical capabilities in areas such as product mark-up analysis and manufacturing cost by design option analysis to derive data tailored specific to Chinese products. Under this option, China can establish a standards development framework that is on par with – if not better – than international standards programs.

4. Program evaluation options (standards, labeling, and incentives)

4.1 Options for program evaluation (impact evaluation)

The basic research question in impact evaluation is: What were the true effects produced by a program or intervention in terms of energy savings (as well as other impacts, such as changes in electricity demand and carbon emissions), separated out from what would have otherwise occurred absent the program or intervention? The recent international evaluation framework review prepared by LBNL referenced over 60 evaluation studies from the U.S., E.U., Australia, and other countries, and it looked at 30 studies in depth for unique evaluation methodologies. In general, evaluation of standards, labeling, and incentives follow a similar methodology, which is summarized in Figure 5 and can be broken down into three main parts: A) stock model (steps 1-3), B) baseline setting (step 1), and C) *ex-post* evaluation options (steps 4-7).

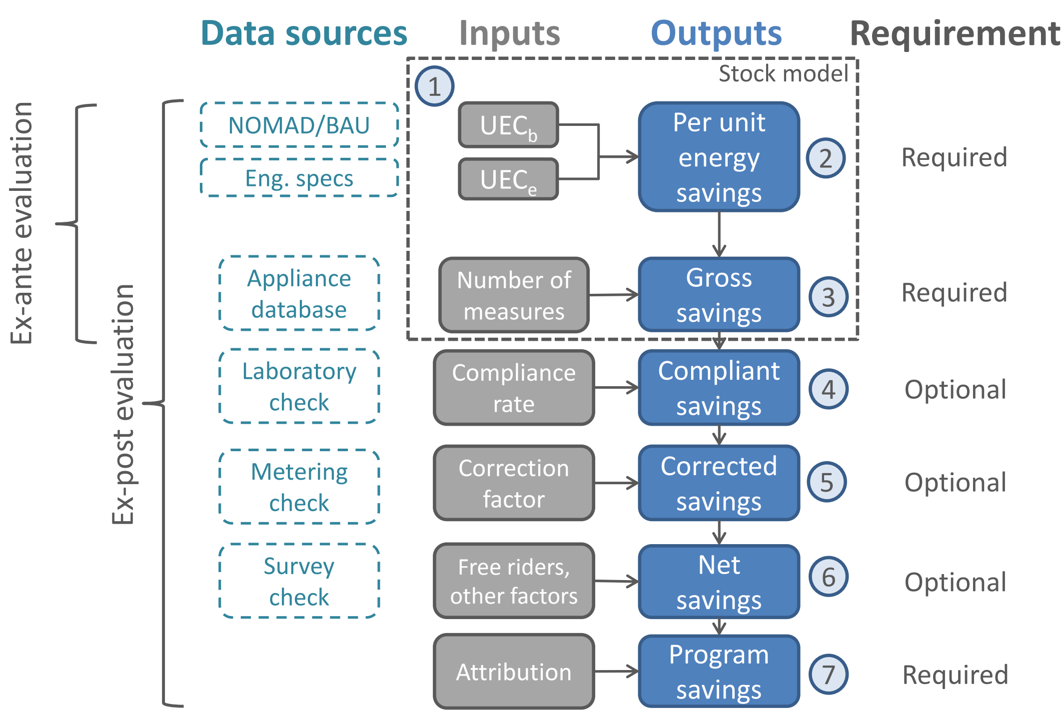

Figure 5. ***Ex-ante*** **and** ***ex-post*** **evaluation frameworks for standards, labeling, and incentives**

Ex-ante evaluation of appliance standard programs plays a large role in a number of countries' standards development process, whereby the impact on national energy demand can be estimated for different levels of proposed standards – essentially, different unit energy consumption (UEC) levels against the same baseline. Shipment projections are used to estimate the "number of measures", the number of products that will be sold under the new standard or labeling program. In the case of incentive programs, the "number of measures" typically mirrors the number of rebates given out for more efficient appliances. Baselines can be set to estimate some initial level of market penetration for high efficiency products in the absence of the standards or labeling. These elements (UEC and number of measures) compose what is typically called an engineering-based model or a stock model in the literature. *Ex-post* evaluation can use the same estimates as *ex-ante* evaluation or it can update them based on collected data rather than projections, for example, using actual shipments or sales and prices as opposed to projected shipments and prices. Additionally, *ex-post* evaluation has the option of using a number of correction factors to get a more accurate estimate of total energy saved from the standard in the cases where the performance of a product when installed is different from the performance as declared by the manufacturer. Finally, *ex-post* evaluation of incentives offers the option to include free ridership, reflecting the fact that some consumers would have purchased a more efficient appliance in the absence of an incentive.

4.2 Key data requirements

Table 5 below outlines the required and optional data requirements for *ex-post* and *ex-ante* analysis of standards and labeling.

Table 5. Required and optional data requirements for ex-post/ex-ante analysis of standards and labeling

Data type	Used in *ex-ante* or *ex-post*	Required or optional
Annual energy savings per unit product	*Ex-ante, ex-post*	Required
Existing stock	*Ex-ante, ex-post*	Required
Market saturation (ownership, market shares)	*Ex-ante, ex-post*	Required
Lifetime or retirement function	*Ex-ante, ex-post*	Required
Usage	*Ex-ante, ex-post*	Required
Future shipment forecasts	*Ex-ante*	Required
Usage adjustment factor (UAF)[8]	*Ex-ante, ex-post*	Optional
Naturally occurring market adoption (NOMAD)	*Ex-ante, ex-post*	Optional
Compliance rate	*Ex-post*	Optional
Real shipments/sales	*Ex-post*	Optional/required
Site-to-source energy conversion factors	*Ex-ante, ex-post*	Optional
Emission factors	*Ex-ante, ex-post*	Optional

A stock model is key for most standards and labeling evaluations; it keeps track of the efficiency distribution and energy consumption of a fleet of appliances based on engineering specifications, usage, lifetimes, and unit energy consumption. If future shipment forecasts are not available, then a stock model will be able to determine a forecast using estimates for market saturation and lifetime. Future shipments are the key for determining the savings of any standards and labeling program. When performing an evaluation of a labeling program, as opposed to a standards program, the stock and shipments data (whether projected or real) will need to be separated by efficiency class in order to calculate energy savings. For an ex-post evaluation of labeling, real shipments data would be required in the case of labeling because there needs to be some indication of sales by efficiency (even if it is only a

[8] The usage adjustment factor (UAF), also known as the correction factor, accounts for the difference between a manufacturer's claimed performance for a product and how that product actually performs when installed in the field. This is different from a compliance rate, which is the performance differential between and a manufacturer's claim for a certain product and that product's laboratory claim when submitted for verification testing.

sampling of the market). This is in contrast to standards evaluation where there is an implicit assumption that all new products simply meet the new standard efficiency level (which can be corrected via compliance rate or usage adjustment factor).

Baselines help define how the efficiency of the appliance fleet would have improved without the standard, often referred to as naturally occurring market adoption (NOMAD). The main options here are as follows:

1. Frozen baseline: the efficiency of new products remains constant in the base case. (This is the simplest approach, but often inaccurate.)
2. Improvement baseline: where historic unit energy consumption (UEC) data exist, the efficiency of new products improves at a similar rate of historic autonomous efficiency improvement, or declines into the future
3. Market share baseline: where data on market share for models of different efficiencies exist, a baseline efficiency can be estimated for future years
4. Bass model baseline: the most advanced curve fitting of market adoption of energy-efficient products to predict NOMAD

Data requirements for an impact evaluation of incentive programs are shown in Table 6. Once unit energy savings have been determined, then that quantity can be multiplied by the number of participants, or number of consumers receiving an incentive or rebate, in order to calculate gross savings. Net savings is calculated by taking into account free ridership, participant spillover, and market effects. Free ridership reflects the fact that some consumers would have purchased a more efficient appliance in the absence of the incentive or rebate being offered. Participant spillover is the savings from program participants who, as a result of the program, installed additional energy efficiency measures, but who did not obtain a program incentive for those additional measures. Market effects are the market level savings that resulted from program influences on the market and the operations of that market (sometimes referred to as nonparticipant spillover, since these end users did not participate in the program and did not obtain a program incentive for those measures), but the market for energy efficiency was affected by the program.

Table 6. Required and optional data requirements and sources for impact evaluation of incentive programs

Data type	Required or optional for gross energy savings	Required or optional for net energy savings
Annual energy savings per unit product	Required	Required
Number of participants and non-participants	Required	Required
Free ridership	Optional	Required
Participant spillover	Optional	Required
Market effects (participant & nonparticipant spillover)	Optional	Required
Site-to-source energy conversion factors	Optional	Optional
Emission factors	Optional	Optional

Finally, there are site-to-source energy conversion factors to help differentiate between primary and secondary energy use and efficiency. Primary energy is the raw fuel that is burned to create heat or electricity. For example, natural gas may be burned to produce electricity at a natural gas-fired power plant, or it may be directly burned in a building's heating boiler or an apartment's instant gas-fired water heater. Secondary energy refers to that electricity or heat that was produced. Units of primary and secondary energy are not directly comparable because there is energy lost in the conversion of primary energy into useful secondary energy (heat, electricity). The U.S. Environmental Protection Agency has a standardized definition for these conversions: "When primary energy is consumed on site, the conversion to source energy must account for losses that are incurred in the storage, transport and delivery of fuel to the building. When secondary energy is consumed on site, the conversion must account for losses incurred in the production, transmission, and delivery to the site. The factors used to restate primary and secondary energy in terms of the total equivalent source energy units are called the source-site ratios." (U.S. EPA, 2012)[9] Source-site ratios are also known as site-to-source energy conversation factors. Once the quantity and type (coal, natural gas, oil, etc.) of primary energy or source energy units is known, then a carbon dioxide emissions factor can be used to find the associated carbon emission savings for the energy saved. Similar emissions factors could be used to determine the amount of emissions avoided for other common air pollutants included sulfur dioxide, nitrogen oxides, and ozone.

4.3 China data availability

Before outlining the data availability for China's evaluations, the relationship of the relative sensitivity and uncertainty of those data need to be established so as to set data collection priorities. Table 7 outlines the high, medium, and low priorities for data collection based on sensitivity and uncertainty parameters. For instance, if an input has a high sensitivity on the evaluation algorithm but a very low uncertainty based on the current data source, then it is a low priority for improved data collection methods.

Table 7. Relationship of sensitivity and uncertainty of data inputs to data collection priority

Sensitivity \ Uncertainty	High	Medium	Low
High	High priority	High priority	Low priority
Medium	Medium priority	Medium priority	Low priority
Low	Medium priority	Low priority	Low priority

Table 8 outlines the data types needed for evaluations of standards, labeling, and incentive programs. The bolded items indicate which data types that China is already collecting through surveys and other types of data collection (CNIS, 2012). The shaded items are those data types that are optional in some cases. For instance, in the evaluation of incentives, free ridership is not needed for a gross energy savings evaluation but is required for a net energy savings evaluation. As in the section on China data

[9] The EPA ENERGY STAR program uses source-site ratios of ~3.34 for electricity, 1.047 for natural gas, and 1.01 for propane and fuel oil. See: http://www.energystar.gov/ia/business/evaluate_performance/site_source.pdf?3890-de20

availability for standards development, the levels of uncertainty and sensitivity have been estimated for this preliminary study based on the authors' knowledge of evaluation literature and familiarity with Chinese data availability. These hypothesized levels can later be tested once formal program evaluations are underway in China.

Table 8. Data types required for standards, labeling, and incentives evaluation with China's current data sources and associated uncertainty and sensitivity. Note: bolded data types are for data that China has already collected, shaded data types are optional in some cases.

Data type	Standards, labeling, incentives? (S, L, I)	Data source	Sensitivity (high, medium, low)	Uncertainty (high, medium, low)
Annual energy savings per unit	S, L, I	CELC white paper, assumptions	High	Medium
Existing stock	S, L	Calculation	Low	Low
Market saturation (ownership, market shares)	S, L	China Statistical Yearbook	Medium	Low
Lifetime or retirement function	S, L, I	Literature, assumptions	Medium	Medium
Future shipment forecasts	S, L	Assumptions, calculation	High	Medium
Usage adjustment factor (UAF)	S, L, I	Literature, assumptions	Medium	Medium
Naturally occurring market adoption (NOMAD) or other baseline	S, L	Estimation	Medium	Medium
Compliance rate	S, L, I	Literature, assumptions	Medium	Medium
Real shipments/sales	S, L	CELC white paper	High	Low
Site-to-source energy conversion factors	S, L, I	Literature, assumptions	Medium	Low
Emission factors	S, L, I	Literature, assumptions	Medium	Low
Number of measures (e.g. rebates)	I	Ministry data (e.g. Ministry of Finance)	High	Low
Free ridership	I	Survey needed	High	High
Participant spillover	I	Survey needed	Medium	Medium
Market effects	I	Survey needed	Medium	Medium

Associated with each data type is an assigned sensitivity based on that data's impact on the evaluation algorithm. In section 2, it was described how in the case of an ex-post net savings evaluation for an incentives program, the number of measures, unit energy savings, and net to gross ratio had the highest impact on the end result. In Table 8, those options have all been flagged as high sensitivity.

Each data type also has an assigned uncertainty based on the currently used data source. For example, the annual unit energy consumption data currently collected by the China Energy Label Center (CELC) are a combination of the unit's efficiency (since energy labeling is now mandatory for many consumer products, these data must be reported by the manufacturers to CELC) multiplied by some assumption on usage. Although there is low uncertainty about the declared unit efficiencies, there could be higher uncertainty on the usage of the product (air conditioner usage varies widely, for instance).

As mentioned in Figure 1, once the sensitivity has been determined, this can assist in deciding priorities for enhanced data collection. Here, we add the uncertainty as an additional parameter to help determine data collection priorities. Table 8 shows that real sales for products have a high sensitivity for standards and labeling evaluation but a low uncertainty. Sales has perhaps the largest impact on the evaluation algorithm (thus the high sensitivity), but China has had data collection in place for about three years now via a retail sales survey company, so this would be a relatively low priority for any new or improved data collection. As another example, lifetimes of products can help in projecting sales or in determining useful life of a product and thus carry a medium sensitivity. The data are currently estimated based on international literature giving them a medium uncertainty since product lifetimes within China may vary from those in the U.S. or E.U.

In the following section, we will highlight the evaluation options China can undertake now given its current data availability. Its evaluation options in the short-term will be dictated based on the high priority items. Based on Table 7, that would be any data type with high sensitivity and a high or medium uncertainty: annual energy savings per unit and free ridership. Its evaluation options in the long-term will be expanded based on the medium priority items: lifetime, UAF, baselines, compliance rates, participant spillover, and market effects.

4.4 Identifying evaluation options for China

Identifying where China currently collects data will aid in recommending options for China for evaluations that it can currently carry out with little or no additional data collection. Assigning sensitivity and uncertainty to all other data types will aid in recommending options for evaluation that it can carry out in the short-term (1-3 years) and long-term (5-10 years) based on expanded and improved data collection. In this section, we identify evaluation options for China for standards and labeling as well as incentive programs. We outline three options based on: 1) current capabilities, 2) an increase of select data collection over the next few years, and 3) a more expansive increase in data collection over the next five to ten years.

4.4.1 Standards and labeling options

The three options for standards and labeling evaluation are as follows:

1. Ex-ante or simple ex-post evaluation based on existing capabilities (to implement now)
2. Ex-post evaluation with improved data on product usage, compliance, and site-to-source factors (to implement on a 1-3 year time scale)
3. Full ex-post evaluation with expanded surveying, testing, and/or metering (to implement on a 5-10 year time scale)

Within each option (outlined in Table 9), there is a specific direction for each data source:

1. Literature: Pull data from domestic or international literature
2. Existing data: Use data sources that China has already collected, including the China Energy Label Center's annual white paper
3. Datasets: Use a dataset that China should have at a future date (e.g. emissions factors)
4. Survey/metering: Employ new survey, laboratory testing, or on-site metering to gather new and improved data

Table 9. Data sources for standards and labeling evaluation options, Note: Exclude means that data point is not needed for the specified analysis in this option.

Data type	Option 1 Implement now	Option 2 Short-term (1-3 years)	Option 3 Long-term (5-10 years)
Annual energy savings per unit product	Existing data	Survey/metering	Survey/metering
Market saturation	Existing data	Existing data	Existing data
Lifetime	Literature	Literature	Survey/metering
Usage adjustment factor (UAF)	Exclude	Exclude	Survey/metering
NOMAD or other baseline	Exclude	Exclude	Datasets
Compliance rate	Exclude	Datasets	Survey/metering
Real shipments/sales (or forecasts if *ex-ante*)	Existing data	Existing data	Use existing data
Site-to-source energy conversion factors	Literature	Datasets	Datasets
Emission factors	Literature	Datasets	Datasets

Option 1 is roughly based on China's current evaluation capabilities and data availability and, therefore, does not involve any expanded surveys. Unit savings, market saturation, and sales can be taken from CELC's existing databases and retail surveys. All ex-post evaluation adjustments (such as usage adjustment factor and compliance rates) are excluded. Site-to-source energy conversion factors and emission factors can be estimated with existing domestic literature such as project design documents reported under the Clean Development Mechanism.

Option 2 There are two major changes in moving from Option 1 to Option 2. First, there would be a set amount of surveying or on-site metering to determine hours of usage for various products (e.g. lighting, televisions, air conditioners, etc.). Data collection should characterize not only average values, but also the range. This would help decrease the uncertainty surrounding the high sensitivity data point on unit energy savings but would also carry an associated need for budget to finance that surveying or metering. Second, new estimates could be made on compliance rate and emission factors based on expanded datasets that China is likely to have on a one to three year time scale. For example, China is currently expanding its product verification testing (within CNIS's Energy Efficiency Laboratory Division) and may have an expanded dataset on product compliance in the coming years. Additionally, with China's expanding reporting for carbon intensity targets and growing carbon market programs, data on emission factors will likely improve over the coming years.

Option 3 suggests a full set of evaluation capabilities, the extent of which is even beyond current evaluation scopes in the U.S., E.U., and elsewhere. The main difference in going from option 2 to option 3 is in using more sophisticated ex-post evaluation adjustment factors, including usage adjustment factors, compliance rates, and baselines. Through a mixture of laboratory testing and on-site metering, China can get a more accurate energy savings estimate by comparing how products perform in the field versus how they are declared to perform by manufacturers. It should be noted, however, that extensive testing like this is not yet regularly performed for ex-post evaluation in major developed countries due to the high budget required. There will always be a balance between budget and certainty, and China should make appropriate decisions based on their evaluation goals. For baselines, China will have enough data to perform a more sophisticated baseline analysis given that it will have collected 10-15 years of retail sales data broken down by energy efficiency labeling level.

4.4.2 Incentives options

Unique data points for an incentives evaluation (as compared to a standards or labeling evaluation) are highlighted in Table 10. The options for evaluation follow a similar categorization.

1. Simple ex-post gross savings evaluation based on existing capabilities (to implement now)
2. Ex-post net savings evaluation with improved data on free ridership and unit energy savings (to implement on a 1-3 year time scale)
3. Full ex-post net savings evaluation with expanded surveying on free ridership, participant spillover, and market effects (to implement on a 5-10 year time scale)

Table 10. Data sources for incentives evaluation options, Note: Exclude means that data point is not needed for the specified analysis in this option.

Data type	Option 1 Implement now	Option 2 Short-term (1-3 years)	Option 3 Long-term (5-10 years)
Annual energy savings per unit product	Existing data	Survey/metering	Survey/metering
Number of measures	Existing data	Existing data	Existing data
Free ridership	Exclude	Survey/metering	Survey/metering
Participant spillover	Exclude	Exclude	Survey/metering
Market effects	Exclude	Exclude	Survey/metering

China has implemented a number of rebate programs for energy efficient products in the past few years including Rural Area Household Appliance Subsidy Program (also known as “Appliances to the countryside”) and Energy Efficient Products Discount Program (also known as “Huimin gongcheng”). Evaluations for gross energy savings can be performed on these programs by simply multiplying the number of rebates by the unit energy savings and expected useful life of the product. This assumes, however, that all of the measures or efficient products were successfully installed which may not be the case. Inspections and surveying could be performed on a selection of sites to determine the proportion of measures installed to rebates granted. For a net energy savings evaluation, a survey will be needed to evaluate the amount of free ridership. Further down the line, a more sophisticated survey can be administered to estimate other values including participant spillover and market effects. The degree to which these values are used in estimating net-to-gross ratios in ex-post evaluation varies widely in the U.S. Additionally, there is a concern that with a proliferation of energy efficiency policies (standards, labeling, incentives, other awareness campaigns) impacting any one consumer’s purchase, it may become increasingly difficult to attribute the savings from an energy-efficient product to any one incentive program (Vine, Hall, Keating, Kushler, & Prahl, 2012).

5. Conclusion

As China ramps up its efforts in energy efficiency in the 12^{th} Five Year Plan and beyond, it will seek to maximize the achievable savings from all of its programs. Program evaluation will become an increasingly important tool for administrators to both improve program design and better understand remaining market potential for energy efficiency. Data availability for appliance efficiency standards development and program evaluation is becoming an increasingly relevant question. In this study, we have recommended options for China to improve its capabilities in appliance standard development and evaluation. In order to establish the foundation of these recommendations, we first outlined the relationships between program goals and budget, choice of algorithm and methodology, data gathering, and data sensitivity and uncertainty. Improvement to data inputs should be prioritized based on how much those inputs impact the end result of the analysis (sensitivity) and an assessment of data quality and collection method (uncertainty). Given the scope of China's appliance standards, labeling, and incentive programs and the scale of energy consumption growth in this sector, improvements to practices in standards development and program evaluation should continue to be recommended.

Increased involvement of and data collection from manufacturers may aid in improving and increasing the transparency of standards development. China currently lacks sufficient data and data collection resources on product design and efficiency options as well as their related cost data. These are all crucial inputs to sound techno-economic analyses at the core of standard-setting frameworks adopted internationally. Given the absence of legislation and administrative requirements for developing standards, China may consider following an analytical approach of using a standardized comprehensive study encompassing all four groups of key analyses (market and technology assessment, technological, economic or life-cycle cost and specific impacts analysis) as a general framework for standards development. The steps that China could take to realize this framework could range from expanding its technical and cost data collection and widening the analytical scope in the medium-term to developing extensive sets of technical, economic (particularly cost) and environmental data to support additional and deeper analyses.

In program evaluation for standards, labeling, and incentives, we find that data types with high sensitivity and high or medium uncertainty should be prioritized for increased or improved data collection in the medium term (1-3 years). Both annual energy savings per unit and free ridership fit this description, and China can plan increased surveying and metering to reduce the uncertainty of these values. Any increased data collection will need extra budget and human resources, which may be difficult for CNIS to manage. A small-scale surveying and metering pilot program could be introduced as a near-term, low budget option for China to test out how China might improve its evaluations and decrease the uncertainty of energy savings achieved by its expansive standards, labeling, and incentive programs.

Acknowledgments

We would like to express our deep appreciation to the China Sustainable Energy Program of the Energy Foundation and the Collaborative Labeling and Appliance Standard Program (CLASP) for providing funding for this project. We would especially like to thank Kevin Mo and Steven Zeng. This work was supported by the China Sustainable Energy Program of the Energy Foundation and the Collaborative Labeling and Appliance Standard Program (CLASP) through the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

CNIS. (2012). *White paper for the energy efficiency status of China energy-use products*. Beijing: China Standards Press.

Dent, S., & Enterline, S. (2012). *Energy Efficiency Programs: Deemed Savings and Technical Reference Manuals.* Washington, DC: EPA Energy Efficiency EM&V Webinar Series.

Jacobs, P., Kromer, S., & Hall, N. (2006). Risk Analysis Overview. *California Evaluation Protocols Public Workshop.* San Francisco, CA: California Public Utilities Commission.

Kiefer, K. (1993). A Framework for Strategic Evaluation Planning. *International Energy Program Evaluation Conference (IEPEC).* Chicago, IL: IEPEC.

Liu, M. (2011). Case Study: Development and Implementation of Refrigerator ES&L in China. *Second Annual Asian Energy Efficiency Standard and Labeling Forum.* Guilin.

U.S. EPA. (2012). *Understanding Source and Site Energy*. Retrieved January 9, 2012, from U.S. Environmental Protection Agency: http://www.energystar.gov/index.cfm?c=evaluate_performance.bus_benchmark_comm_bldgs

Vine, E., Hall, N., Keating, K., Kushler, M., & Prahl, R. (2012). Emerging issues in the evaluation of energy-efficiency programs: the US experience. *Energy Efficiency*, 5:5-17.

Zhou, N., Khanna, N., Fridley, D., & Romankiewicz, J. (2012). *International Review of Frameworks for Standard Setting and Labeling Development.* Berkeley, CA: Lawrence Berkeley National Laboratory.

Zhou, N., Romankiewicz, J., Vine, E., Khanna, N., & Fridley, D. (2012). *International Review of Frameworks for Impact Evaluation of Appliance Standards, Labeling, and Incentives.* Berkeley, CA: Lawrence Berkeley National Laboratory: LBNL-6003E.