Survey of Benchmarking Methodologies:
Executive Summary

March 1, 2006

Prepared By

Sanford V. Berg
Director of Water Studies
PURC (www.purc.ufl.edu)
University of Florida

The focus of this Survey is on *Metric Benchmarking* that quantifies the relative performance of organizations or divisions. The study identifies the strengths and limitations of several techniques that yield *Performance Scores based on Production or Cost Estimates* ("total" methods). Other methodologies are briefly summarized. One utilizes *Core Indicators* which can be combined into a Summary or *Overall Performance Indicator* (partial metric method). Another is an engineering approach to benchmarking that determines *Performance Relative to a Model Company* (an artificial firm that has optimized its network and minimized its operating costs). *Process Benchmarking,* a fourth methodology, requires detailed analysis of utility operating characteristics at different stages of the production process. Finally, *Customer Survey Benchmarking* identifies customer perceptions regarding utility responsiveness, reliability, and other service dimensions. Studies might also focus on financial and/or water resource sustainability since both outcomes affect the long term water sector performance.

The production and cost applications are to the water sector, but the principles of Data Envelopment Analysis and econometric analysis can be applied to any industry. Using well-established empirical procedures, the analyst can measure cost or productivity performance and identify performance gaps. Such metric benchmarking is essential for those developing and implementing water policy. The tools are important for documenting past performance, establishing baselines for gauging productivity improvements, and making comparisons across service providers. Rankings can inform policymakers, those providing investment funds (multilateral organizations and private investors), and customers regarding the cost effectiveness of different water utilities. If decision-makers do not know where they have been or where they are, they cannot set reasonable targets for future performance

Funded by

World Bank

Prepared by

PURC Studies in Energy Telecommunications Water
"Leadership in Infrastructure Policy"

Survey of Benchmarking Methodologies: Executive Summary

By Sanford V. Berg[1]
Director of Water Studies, Public Utility Research Center (www.purc.org)
University of Florida
March 1, 2006

Metric benchmarking quantifies the relative performance of organizations or divisions, controlling for external conditions. Using well-established empirical procedures, the analyst can measure performance and identify performance gaps. Such benchmarking is essential for those developing and implementing water policy. The tools are important for documenting past performance, establishing baselines for gauging productivity improvements, and making comparisons across service providers. Rankings can inform policymakers, those providing investment funds (multilateral organizations and private investors), and customers regarding the cost effectiveness of different water utilities. In addition, if managers do not know how well their organization (or division) has performed (or is performing), they cannot set reasonable targets for future performance.

The first step in benchmarking involves collecting information on water/sewerage system operations, network capacity, financial flows, and outputs. Consistent data that facilitate comparisons are essential for good management and for public policy oversight; such data are becoming available via the *Water & Sanitation International Benchmarking Network* (IBNET, funded by the UK Department for International Development and the World Bank).

When conducting benchmarking analyses, water professionals must understand the strengths and limitations of different metric methodologies. This Survey is directed to regulators and utility managers to help them make performance comparisons over time, across water utilities, and across countries. Although consultants and academic researchers have published over thirty empirical studies of water utilities (summarized in an Appendix), few regulators and companies are using metric benchmarking on a regular basis. This Survey and its supporting Technical Appendices are designed to bridge the gap between technical researchers and those practitioners currently conducting studies for government agencies and water utilities.

We believe that the application of more sophisticated quantitative tools is necessary (but not sufficient) for promoting policies that can improve company (and sector) performance. The introduction of greater rigor allows stakeholders to quantify utility progress towards meeting policy objectives, helps water specialists identify high performing utilities (whose production processes might be adopted by others), and enables regulators to fine tune targets and incentives for utilities.

First, consider a caveat. A single index of utility performance has the same problems of any indicator: it will be neither comprehensive nor fully diagnostic. A physician can collect information on a patient's temperature, pulse, height and weight. Those four indicators help the physician determine whether the person has a dangerous fever and/or is overweight. The indicators point to potential or existing health problems. A fever is a short term problem that can be addressed with specific medications; weight is a longer-term health issue with implications for the heart attacks and other problems—diet and exercise programs might be prescribed. However, a set of blood tests will provide more detailed information that can aid in diagnosing the physical problems that are only partly reflected in the two health indicators. Similarly, diagnosing and treating mental health issues would require other diagnostics . . . and treatments.

[1] The Project Director gratefully acknowledges the assistance of Maria Luisa Corton, Chen Lin, and Guillermo Sabbioni (PURC Research Associates), and Liangliang Jiang and Aaron Jones (PURC Research Assistants). The project received funding from the World Bank, but the views presented here are those of the authors.

Five Methodologies

In recognition of the wide range of issues that might be addressed when evaluating water utility performance (as in evaluating a patient's health), analysts have developed a variety of methodologies for addressing specific issues:

- Core Indicators and a Summary or Overall Performance Indicator (partial metric method),
- Performance Scores based on Production or Cost Estimates ("total" methods),
- Performance Relative to a Model Company (engineering approach),
- Process Benchmarking (involving detailed analysis of operating characteristics), and
- Customer Survey Benchmarking (identifying customer perceptions).

These techniques are briefly described below, with the Survey focusing on the second methodology.

Core Overall Performance Indicators include a number of *Specific Core Indices*, such as volume billed per worker, quality of service (continuity, water quality, complaints), unaccounted for water, coverage, and key financial data (operating expenses relative to total revenues, collections). Usually these indicators are presented in ratio form to control for the scale of operations. These partial measures are generally available, and provide the simplest way to perform comparisons: trends direct attention to potential problem areas. Policymakers often combine the specific core indices to create an *Overall Performance Indicator (OPI)*, generally using a weighted average of core indices. Thus, an OPI provides a summary index that can be used to communicate relative performance to a wide audience. Although its components are easily understandable, in practice the weights used to compute the OPI are not determined through a process that prioritizes the different indicators. For example, the OPI used by SUNASS (the Peruvian water regulator) is the sum of nine specific indices. In addition, many factors will affect the specific indices, including population density, ability to pay (income levels), topography, and distance from bulk water sources. Finally, an OPI fails to account for the relationships among the different factors. A firm that performs well on one measure may do poorly on another, while one company doing reasonably well on all measures may not be viewed as the "most efficient" company.

Performance Scores based on Production or Cost Estimates are used to identify the best performers and the weakest performers in a group of utilities. The metric approach allows quantitative measurement of relative performance (cost efficiency, technical/engineering efficiency, scale efficiency, allocative efficiency, and efficiency change). Performance can be compared with other utilities at a point of time and over time, using statistical and/or nonparametric frontier methods. Analysts apply these quantitative techniques to determine relationships among variables: for example, utilities that produce far less output than other utilities (who are using the same input levels) are deemed to be relatively inefficient. Similarly, a utility might have much higher costs than expected (based on observations of others producing the same output level but having lower costs). A finding of excessively high costs would trigger more in-depth studies to determine the source of such poor performance. Thus, performance scores and relative rankings identify under-performing and high-performing utilities. Rankings can be based on the analysis of production patterns and/or cost structures. Production function studies (requiring data on inputs and outputs) show how inputs affect utility outputs (such as volume of water delivered, number of customers, and service quality). Similarly, cost functions show how outputs, inputs and input prices affect costs; such models have heavy data collection and analysis requirements. One advantage of cost models is the ability to analyze components of total cost; for example, Ofwat (England and Wales) has examined how different types of operating expenses depend on various cost-drivers, such as length of pipe, volume of water delivered, and customer density. Using data from a group of utilities at a point in time (or over a long time period) allows analysts to incorporate cost-drivers beyond management's control (such as population density or topology). In both types of studies, estimated parameters can give an indication of economies of scale and/or economies derived from the joint supply of water service and wastewater collection and treatment. Studies have also examined the relative performance of privately-owned and publicly-owned water utilities. Data availability and the issue under investigation dictate whether production or cost functions are utilized and influence the choice of analytic technique (statistical estimation or data envelopment analysis).

Engineering/Model Company approach has been used to establish baseline performance. This methodology requires the development of an optimized economic and engineering model: based on creating an idealized benchmark specific to each utility—incorporating the topology, demand patterns, and population density of the service territory. The use of an "artificial" firm that has optimized its network design and minimizes its operating costs can provide insight into what is possible if a firm is starting as a Greenfield Project. As with any methodology, this approach also has its limitations. The engineering models that support it can be very complicated, and the structure of the underlying production relationships can be obscured through a set of assumed coefficients used in the optimization process. Chile and Argentina have used this approach for establishing infrastructure performance targets.

Process Benchmarking focuses on individual production processes in the vertical production chain. One advantage of this approach is the ability to identify specific stages of the production process that warrant attention. For example, to obtain finished drinking water involves the following steps: pumping up, intake, transport, clarification and filtration of groundwater as well as the purification and treatment of raw surface water. Detailed examination of production facilities and their operations would be the starting point for process benchmarking. Similar studies would be performed for distribution processes (network design, pipeline construction and maintenance), sales processes (including meter reading, data processing, billing, collections, and customer relations), and general processes (like planning, staff recruitment and retention, and public relations). Many water associations focus on process benchmarking as a mechanism for identifying potential benchmarking partners, preparing for and undertaking benchmarking visits, and implementing best practices.[2] Thus, water utility managers recognize that information sharing and coordination is a significant performance driver across companies. From the standpoint of public policy, there must clear delineation of utility obligations and regulatory responsibilities so that the process benchmarking activity does not involve undue interference with managerial decision making.

Customer Survey Benchmarking focuses on the perceptions of customers as a key element for performance evaluation. Unlike the other approaches, this technique can shed light on consumer concerns, reflected in complaints or captured in customer surveys. Customer perceptions regarding service quality are central to evaluating water utility performance. One widely-used model[3] identifies five dimensions of service quality as perceived by customers: external characteristics (tidy workplace, employee appearances), reliability (meeting deadlines, consistency in interactions), responsiveness (providing service promptly), consideration (personnel who are courteous, friendly, and helpful), and empathy (giving individual care and attention). Surveys can reveal performance gaps and identify areas of concern. Disaggregating complaints by type of customer, location, and type of complaint can help managers identify problem areas. In addition, trends over time can be used by regulators and policy-makers to evaluate utility performance. Nevertheless, many other factors are relevant for evaluating the efficient provision of water services.

---

[2] Mats Larsson, Renato Parena, Ed Smeets and Ingrid Troquet, (2002). *Process Benchmarking in the Water Industry: Towards a Worldwide Approach*, International Water Association.

[3] Parasuraman, Zeithaml and Berry (1985) "A Conceptual Model of Service Quality and its implications for future research," Journal of Marketing 49 (4), Fall, 41-50. See also Alegre, Helena, Wolfram Hirnir, Jamie Melo Baptista and Renato Parena (2000). *Performance Indicators for Water Supply Services*. IWA Manual of Best Practices. London: IWA Publishing.

The Chart on the next page shows how input prices, input levels, and external circumstances enter into the production process. Some variables are under current management's control (like variable inputs), while others are the result of past managerial decisions, like the network (reflecting inherited assets and past maintenance outlays). The cost of capital and the prices of variable inputs determine total economics costs. Due to data difficulties for the cost of capital, analysts sometimes only can identify the determinants of Operating Expenses. Of course, many factors affecting the production process and associated costs are determined external to the utility (population density, topology of the service territory, customer ability to pay, and access to water resources). Performance scores based on production or cost models need to take such factors into account, so that analysts are comparing apples to apples.

The Chart includes a box labeled Process Benchmarking. The present Survey does not attempt to dig deeply into the various sub-processes that link inputs to outputs. Rather, efficiency and productivity are emphasized, using cost and production models to gauge relative performance. In addition, the bottom of the Chart contains boxes reflecting three other aspects of water sector performance: financial sustainability, customer satisfaction, and water resource sustainability.

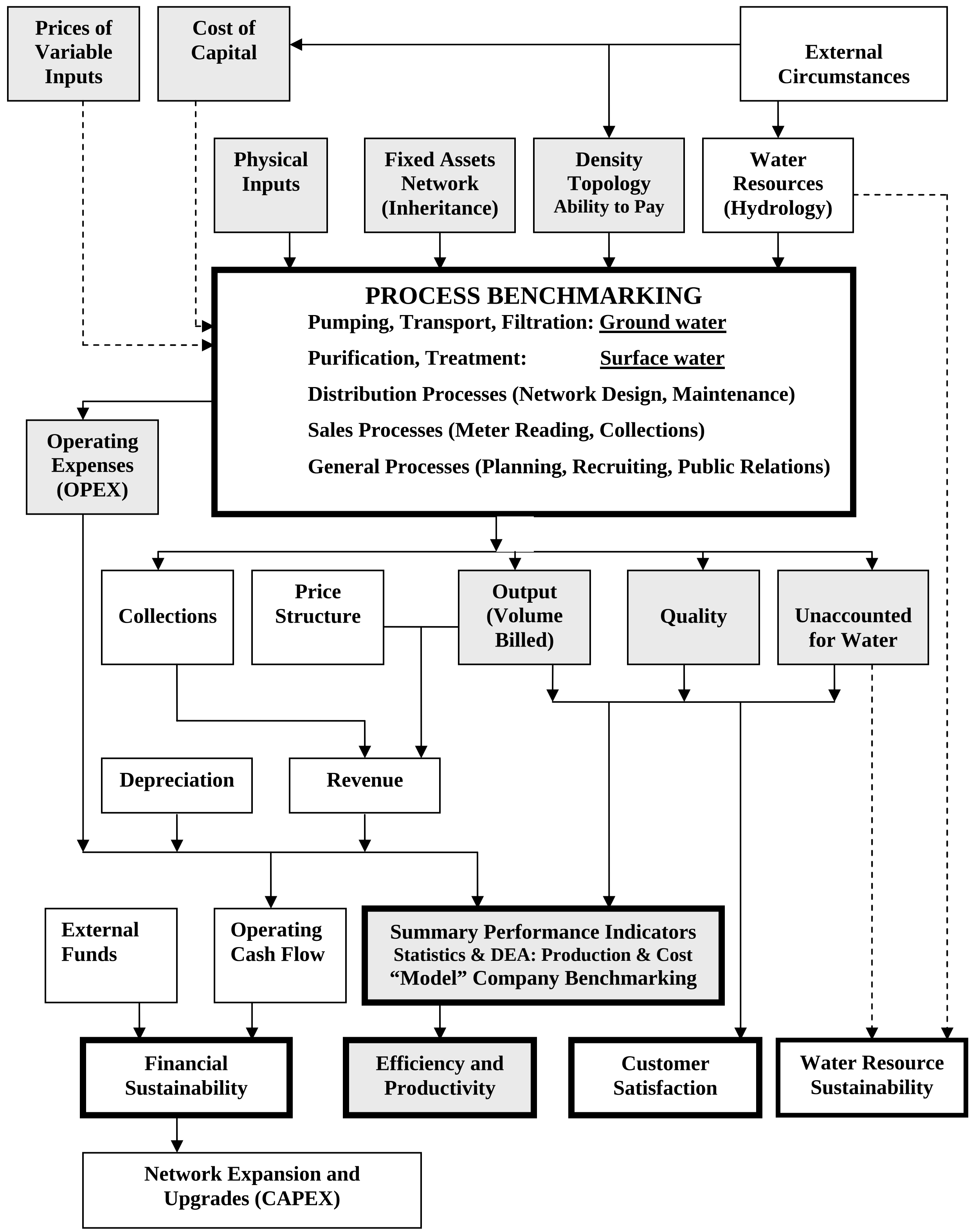
Financial Sustainability Benchmarking considers the role of collections, revenues (price times volume billed), and operating expenses (OPEX). The price structure includes hook-up fees, monthly fixed fees, and price per unit consumed (which can involve inclining or declining block prices). Key financial ratios serve as indicators of long term performance. Revenues (less operating expenses) provide the cash flows that facilitate future capacity investments: for both network expansion and upgrades (CAPEX—capital expenditures). Obtaining external funding (either through the issuance of bonds to private investors or to government agencies or development banks) can be contingent on current cash flows more than covering OPEX. Clearly, this dimension of utility performance warrants attention.

Customer Satisfaction Benchmarking has already been identified as one of the five most-used methodologies: from the standpoint of Customer Survey Benchmarking. Nevertheless, survey benchmarking only gives a "rough" picture of how customers perceive utility service offerings. For example, studies could compare these perceptions with the utility's own record of day-to-day customer complaints; such studies would reveal if there is consistency in identified areas of weakness. While customer survey benchmarking enables utility managers to attain information of customers' feelings from a large sample, some responses reflect emotional attitudes that may not capture verifiable features of service delivery. Utility rankings based on survey scores might simply be reflect customer sentiments rather than technical features of service quality. Of course, such sentiments warrant attention whether correct or not: "Believing is seeing." More attention might be given to identifying cost-effective methods of capturing customer sentiments and designing programs that address (subjective) consumer perceptions.

Water Resource Sustainability is one issue that is given inadequate attention in the analysis of water sector performance. It is basically a public policy issue related to resource management, and not addressed here.

The focus here is on performance scores based on comprehensive production and cost studies. The Chart categorizes this type of analysis as one of the three that emphasize efficiency and productivity. The other two (summary indicators and the "model" company—engineering approach) are not covered in detail here. The performance scores discussed here are based on statistical studies and Data Envelopment Analysis (DEA). After reviewing the Chart, the reader will find a Checklist providing an overview of steps in performance benchmarking; these steps yield relative performance scores.

Chart: Inputs, Processes, Outcomes, and Performance Benchmarking



Checklist for Conducting Benchmarking Studies that Yield Performance Scores

Decision-makers understand that no single methodology captures all the elements that are relevant for ranking utilities (or nations) in terms of water sector performance. Because Core Overall Performance Indictors are widely utilized and Engineering/Model Company, Process Benchmarking, and Customer Survey Benchmarking are well documented elsewhere, this study focuses on performance scores based on production or cost estimates. The Survey does not attempt to identify specific processes that fall short of best practice nor does it show how to utilize indicators in the development of incentive plans, although both activities are important. Rather, it shows how analysts can *measure* water utility *operations* (costs, physical inputs and outputs) to *perform company comparisons* in the context of *infrastructure reform*.

The benchmarking process can be divided into five steps: identify objectives, select methodology and gather data; screen and analyze data; utilize specific analytic techniques; conduct consistency/sensitivity tests; and develop policy implications. Each step includes a set of sub-steps (depicted in the Figures at the end of this Executive Summary).

Step 1. Identify Objectives, Select Methodology and Gather Data. Choices must be made regarding the issues to be addressed, the time period to be analyzed, and the types of comparisons to be made. These choices will reflect current analytic capabilities, an initial understanding of data availability, and preliminary methodological choices. The objectives of any benchmarking study will depend on the most important policy issues under consideration. Clearly, staff members with finance and accounting backgrounds are required to monitor financial statements and check the financial status of the firm. Some team members will need backgrounds in econometrics and statistics to specify and test empirical models. In addition, professionals who understand regulatory and organizational issues will be needed to help interpret the results and develop policy implications.

Step 2. Screen and Analyze Data. Conducting a benchmarking study is an iterative process. Detailed screening of the data will result in greater refinement in the timeframe, sample size, and statistical techniques. The sub-steps of this second stage require more technical skills than the initial framing of the issues to be investigated. However, feedback should be elicited from those who participated in Step 1 so that there is agreement regarding decisions taken at this stage. After assembling the raw data, the benchmarking team should screen the data carefully to ensure the quality and quantity of information being gathered meet the requirements to allow for a successful project. This process is crucial to the study because poor data quality (inconsistent definitions, missing data or extreme data values) may lead to biased results. Insufficient data could result in the use of more limited models that might constrain functional forms so as to yield biased or skewed results. Four broad metric methodologies might be utilized:

- *Core Overall Performance Indicators* (OPIs combine partial indicators of operating or financial performance; these are summary indices),
- *Total Factor Productivity Indices* (an index number approach that considers output per unit input—where multiple inputs are taken into consideration to gauge efficiency levels and changes),
- *Relative Performance using Data Envelopment Analysis* (a non-parametric technique that makes no assumptions about the functional form of production or cost functions),
- *Relative Performance using Statistical Techniques* (parametric approaches that involve assumptions about functional relationships)

Each of the approaches has strengths and limitations.

Step 3. Utilize Specific Analytic Techniques. The process of model specification and technique selection process depends on benchmarking objectives, data availability, and the team's willingness to adopt specific assumptions for each type of model. Hence, the benchmarking team may need to draw upon professional consultants or specialists at research institutions before moving to more sophisticated models. Two broad approaches to quantifying performance are outlined here:

Data Envelopment Analysis: DEA is a method in which linear programming techniques are applied to a selected set of variables to calculate an efficiency coefficient for each water utility. The framework has been adapted from the multi-input, multi-output production and functions. The production function is a basic concept in economics used to determine how much output can be produced with a given basket of inputs. A cost function relates total cost to input use and input prices. DEA develops a function whose form is determined by the most efficient producers. This method differs from OLS that would base comparisons with respect to an average producer. DEA benchmarks firms only against the best producers (so it is a non-parametric frontier analysis). It can be characterized as an extreme point method that assumes that if a firm can produce a certain level of output utilizing specific input levels, another firm of equal scale should be capable of doing the same. The most efficient producers can form a "composite producer," allowing an efficient solution for every level of input or output. Where there is no actual corresponding firm, "virtual producers" are identified to make comparisons. DEA is a non-parametric approach to evaluating performance: functional forms are not specified, although the type of returns to scale is incorporated into the model.

DEA has been used for both production and cost data. Utilizing the selected variables, such as unit cost and output, DEA software searches for the points with the lowest unit cost for any given output, connecting those points to form the efficiency frontier. Any company not on the frontier is considered inefficient. A numerical coefficient is given to each firm, defining its relative efficiency. Different variables that could be used to establish the efficiency frontier are: number of employees, service quality, environmental safety, and fuel consumption. A recent survey of studies of electricity distribution companies identified more than thirty DEA analyses—indicating widespread application of this technique to that network industry. Less than five such studies have been published for water utilities. The main advantage to this method is its ability to accommodate a multiplicity of inputs and outputs. It is also useful because it takes into consideration returns to scale in calculating efficiency, allowing for the concept of increasing or decreasing efficiency based on size and output levels. A drawback of this technique is that model specification and inclusion/exclusion of variables can affect the results.

Statistical Analysis: Linear regression analysis seeks to derive a linear relationship between firm performance and market conditions and characteristics of the production processes. Statistical analysis of a production relationship can isolate the effects of specific conditions or input levels on the level of output—so the roles of multiple independent variables can be determined. Data from the firms being compared can then be used to calculate expected output, given the model coefficients and values of variables for each firm. The process is similar for a cost function. A major advantage of this method is its ability to reveal information about cost structures and distinguish between different variables' roles in determining cost. Predicted versus actual cost can provide a measure of relative performance. The quality of these results can then be statistically evaluated to provide the policy-maker with a framework for evaluating firms.

Econometric analyses have two major disadvantages. First, a large data set is necessary for reliable results. Obtaining the number of observations needed to derive an efficient and unbiased estimate of cost (or production) structures can often prove to be a difficult task. The second disadvantage is a statistical one. The regression results are sensitive to model specification (for example, a linear vs. a non-linear functional form). In addition, for some models, the interpretation of the error term becomes important.

Nevertheless, these techniques are widely used to analyze other network industries. The Annotated Bibliography of Water Benchmarking Studies identifies over thirty statistical studies for the water sector.

The early studies tended to utilize Ordinary Least Squares (OLS) to estimate cost functions for firms. Due to data limitations, most of these studies were cross-sectional in nature. Besides using data from only a single year researchers utilized data from England and Wales or from the United States. These academic studies often focused on the relative performance of private vs. publicly-owned water and sewerage utilities. In addition, they investigated the extent of scale economies and economies of joint production (providing both water and sewerage services). In some cases, they considered the impacts of residential vs. industrial/commercial customers.

As data from Brazil, Peru, and other emerging nations became available, additional country studies were published—often using more advanced econometric (parametric) or non-parametric data analysis techniques. Studies of utilities in France, Italy, and other nations began to appear in the academic literature. Techniques associated with Stochastic Frontier Analysis began to be applied to both production functions and cost functions. Panel data facilitated the incorporation of customer density, topology, and other variables.

A number of cross country studies have been conducted by researchers at the World Bank. These studies characterize production processes in a region and to draw conclusions about the impacts of different institutional features (such as the presence of good governance procedures, the regulatory environment, and public/private ownership). In addition, quality of service has begun to be incorporated in empirical studies, through the use of hedonic output measures (adjusting the output for quality) or as separate outputs altogether (in multi-output models).

In many cases, research raises more questions than it answers. Public policy-makers would like empirical studies to answer questions that have decision-relevance. The publications and studies summarized in the Technical Appendices have primarily involved academic researchers, sometimes in the context of rate cases—or as a retrospective look at how methodologies were used by regulators. Consulting firms have entered the fray, as well. Water utility managers are likely to fear the misuse of empirical studies: they tend to prefer process benchmarking studies that can identify changes in specific practices that would improve performance.

One categorization of the many alternative models yields eleven analytic techniques, shown below.

Partial Indicators
(1) Summary Indices (Core Indicators)

Total Factor Productivity
(2) Tornqvist Index (requires data on prices and quantities of inputs and outputs)
(3) Malmquist Index (only uses data on inputs and outputs)

Nonparametric (Data Envelopment Analysis—no specification of functional form)
(4) DEA Production Scale (can handle multiple inputs and outputs and returns to scale)
(5) DEA Cost Efficiency (with price information, determines technical and production efficiency)
(6) DEA Super Efficiency (estimates the frontier using all the <u>other</u> utilities)

Parametric (Stochastic Frontier Analyses and other Statistical/Econometric Models)
(7) SFA Production Function (uses inputs and output; functional form specified)
(8) SFA Cost Function (uses inputs, input prices, and output; functional form specified)
(9) SFA Distance Function (handles multiple outputs and inputs; requires other assumptions)
(10) Corrected Ordinary Least Squares, COLS (similar to simple OLS regression)
(11) Fixed and Random Effects Models (involving different assumptions on error terms)

These methods are basically arrayed in terms of the technical quantitative skills required for implementing the different approaches. Availability of software and specialized skills can be as important as data availability in determining the types of models utilized. Summary indices allow analysts to review trends of core indictors over time. Total factor productivity indices provide more comprehensive characterizations of trends over time and of relative performance across a set of water utilities. However, these (and parametric and nonparametric methods) may not incorporate a number of important dimensions of performance noted above (e.g., financial sustainability or some aspects of customer service or water quality) due to difficulties in incorporating these factors into DEA, SFA, or COLS models. Thus, more sophisticated techniques are not necessarily the most useful in the context of benchmarking. Sometimes simple partial indicators can provide useful comparisons that are intuitive. For example, if the analyst is not concerned with classifying individual utilities but rather with determining broad trends for the group, the results from using partial indicators may be credible. However, rankings and performance comparisons that have financial impacts require that the results be convincing; in that case, the use of partial indicators alone might not suffice.

Step 4. Conduct Consistency/Sensitivity Tests. In making performance comparisons through benchmarking analyses, the analyst is interested in obtaining a measure of firms' relative efficiency. Such information can be used to develop X-factors in price cap regulation, to reward (or punish) companies. Or the regulator might want to publish the rankings or efficiency scores to provide the public with information, putting pressure on managers of poor performing utilities to improve the performance of their firms. In both cases, the accuracy and robustness of inefficiency estimates are very important because they may have significant financial or social impacts. In particular, if the estimated inefficiency scores or rankings are sensitive to the benchmarking method, a more detailed analysis is required to justify the adopted model. Tests for mutual consistency are becoming standard.

Nevertheless, in most cases there is no ''ideal'' model among the set of potential models. Issues include model specification (cost vs. production and functional form), alternative specification of inputs or outputs (e.g., network length vs. fixed assets as an input), assumptions about error terms, and alternative methodologies (e.g., DEA vs. SFA). Following the work of others, we suggest three levels of sensitivity tests.

Level 1: Sensitivity tests of efficiency scores. The analyst can find the correlation of efficiency scores between pairs of techniques to determine whether scores are highly sensitive to model specification.

Level 2: Sensitivity tests of efficiency ranking: If the efficiency scores are not consistent across the different methods, it is still possible that these approaches will generate similar rankings of firms by their efficiency score. A ranking determined in this manner can help a regulator establish the X-factor to be used in setting prices for the firms in the sector.

Level 3: Sensitivity tests of efficiency ranking (best and worst performers): Even if the efficiency level and rankings are not consistent, it is still possible that several methodologies can identify the best and worst performers. This result can be especially helpful for rewarding the best performers and punishing the worst performers. Therefore, the regulator (or manager—in the case of comparing several divisions) can compare rankings yielded under the different techniques and summarize from that comparison the overlapping rate of the best and worst performers.

Thus, to check for the robustness of performance rankings, researchers have begun to compare results from different methodologies: using correlation matrices or verifying whether models identified the same set of utilities as the most efficient and least efficient firms. Clearly, if efficiency scores are to have any use for managerial incentive or as elements in regulatory mechanisms, stakeholders need to be confident that the scores reflect reality, and are not just artifacts of model specification, sample selection, treatment of outliers, or other steps in the analytic process. Thus, analysts are performing sensitivity tests.

Step 5. Develop Policy Implications. If the results pass at least one of the sensitivity tests, the benchmarking team can start to analyze scores and rankings and explore in greater detail the potential determinants of inefficiencies across firms and over time. The utilities can be divided into different groups by various factors, such as region, population density, regulatory environment, ownership structure, and vintage to compare their efficiency scores. Firms should not necessarily be ranked as poor performers if they operate under conditions that differ from those of the other firms. As noted earlier, population density, topology, distance from raw water sources, and political constraints on prices (affecting the financial sustainability of operations) affect relative performance. Thus, it is useful to seek public comments that might augment the benchmarking analysis effort.

Audiences for Benchmarking Studies:

There are at least six audiences for yardstick comparisons: benchmarking specialists, the press, the general public, regulators, national and international policymakers, and utility managers.

Benchmarking specialists produce and critique studies that utilize various methodologies. Rankings can be manipulated by choice of variables, model specification, sample size, time frame, and treatment of outliers. Because the stakes are high, affected parties have an interest in the relative and absolute performance evaluations prepared by analysts, and studies can be controversial.

The *press* filters and highlights reports, using executive summaries and interviews. Although technical reports are not amenable to sound bites, most newspaper and television journalists seek the clear message that emerges from a benchmarking study. However, some seek sensational factoids that support their own ideological predilections and some lack the expertise to interpret technical studies.

The *general public* is not well-positioned to evaluate conflicting claims. Long before a benchmark comparison is released, the responsible agency should be engaging in an information-dissemination campaign, informing political leaders and the press about the purpose of the forthcoming report. NGOs and formal citizen advisory committees that can be established by regulatory commissions provide opportunities for input and feedback for citizens.

The *regulator* reviews studies and creates performance incentives to achieve policy objectives. Productivity measures and other measures of technical efficiency provide valuable information for regulators. However, excessive simplicity can result in a distorted analysis. Sector performance also depends on efficient price signals, benefits from quality improvements, incorporation of environmental impacts into decisions, recognition of transition costs, and ability to meet agreed-upon social obligations.

National policymakers (elected representatives and appointed officials) react to and utilize technical studies in setting priorities and interacting with international organizations. Solid data regarding the actual performance of public and privately owned utilities can be used to counter the politicization of infrastructure pricing.

Water utility managers are sensitive to comparisons as they have much to lose (and something to gain) when information is made public. It is extremely difficult for outsiders to evaluate managerial performance. Inadequate reports and the selective presentation of information mean that only insiders know whether the organization is managed well or poorly. Benchmarking reduces the extent of this information asymmetry. For this reason, utility managers might delay or block serious benchmarking initiatives.

Concluding Observations

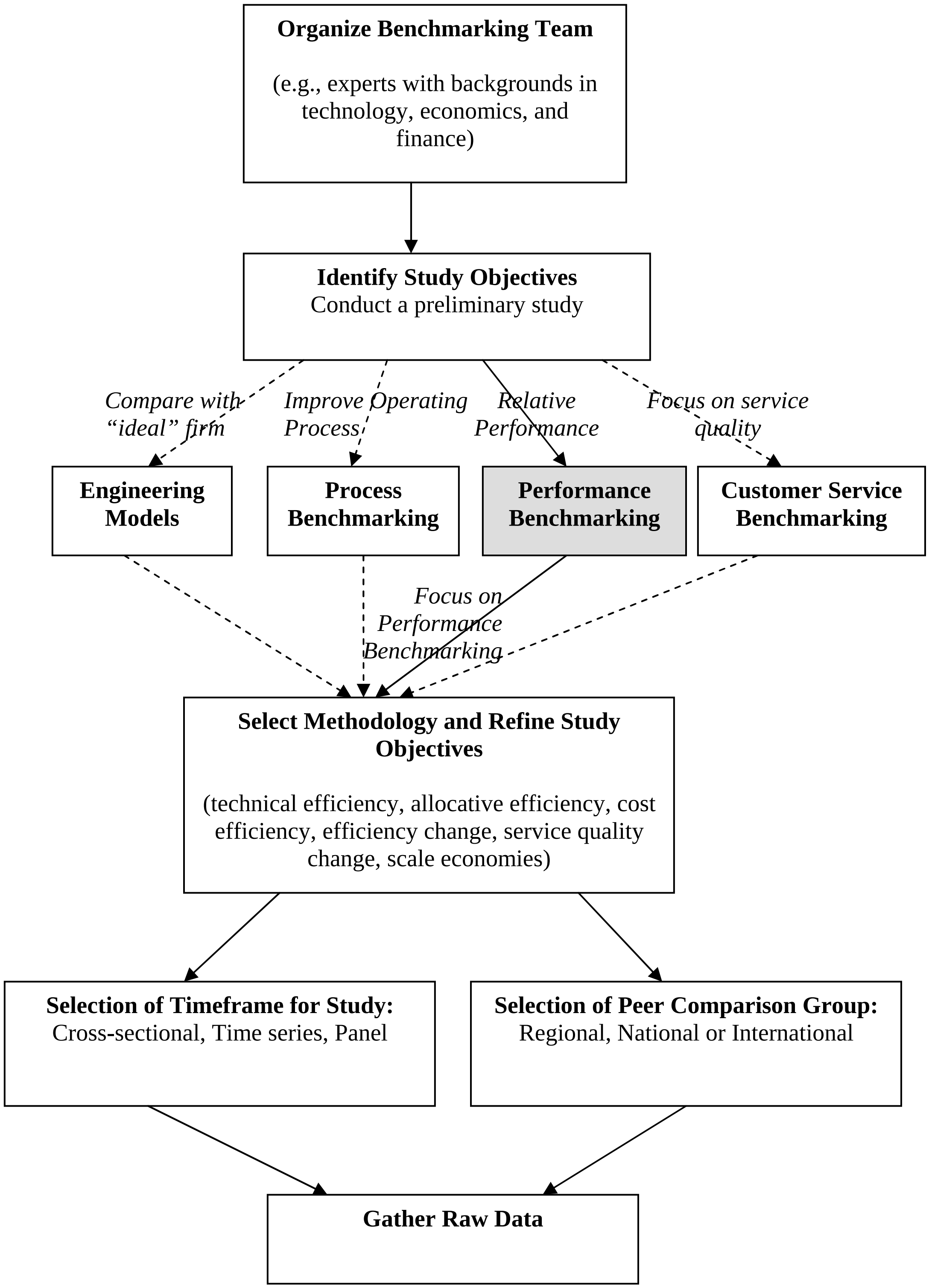
Benchmarking studies should target specific audiences with appropriate levels of detail. Each of the five benchmarking approaches can provide insights on aspects of water utility performance. *Overall Performance Indicators* combine partial metrics and provide information time trends and patterns across firms. This Survey focused on *Performance Scores based on Production or Cost Estimates* ("total" methods); these techniques can utilize the data sets that are becoming more widely available. We identified thirty-two empirical studies of water utilities, with research on production and cost functions continuing at a rapid pace. Techniques that are being applied vary in terms of technical skills required and ease of interpretation, but all can provide important information for analysts who assess industry and firm performance. The Survey outlines in greater detail the strengths and limitations of these approaches.

In addition, the Survey draws attention to three other Benchmarking Methodologies: the *Model Company* (or engineering approach), *Process Benchmarking* (involving detailed analysis of operating characteristics), and *Customer Survey Benchmarking* (identifying customer perceptions). These methodologies can supplement (but not supplant) the statistical models and Data Envelopment Analysis. Other areas of sector performance warranting attention include financial sustainability and water resource sustainability.

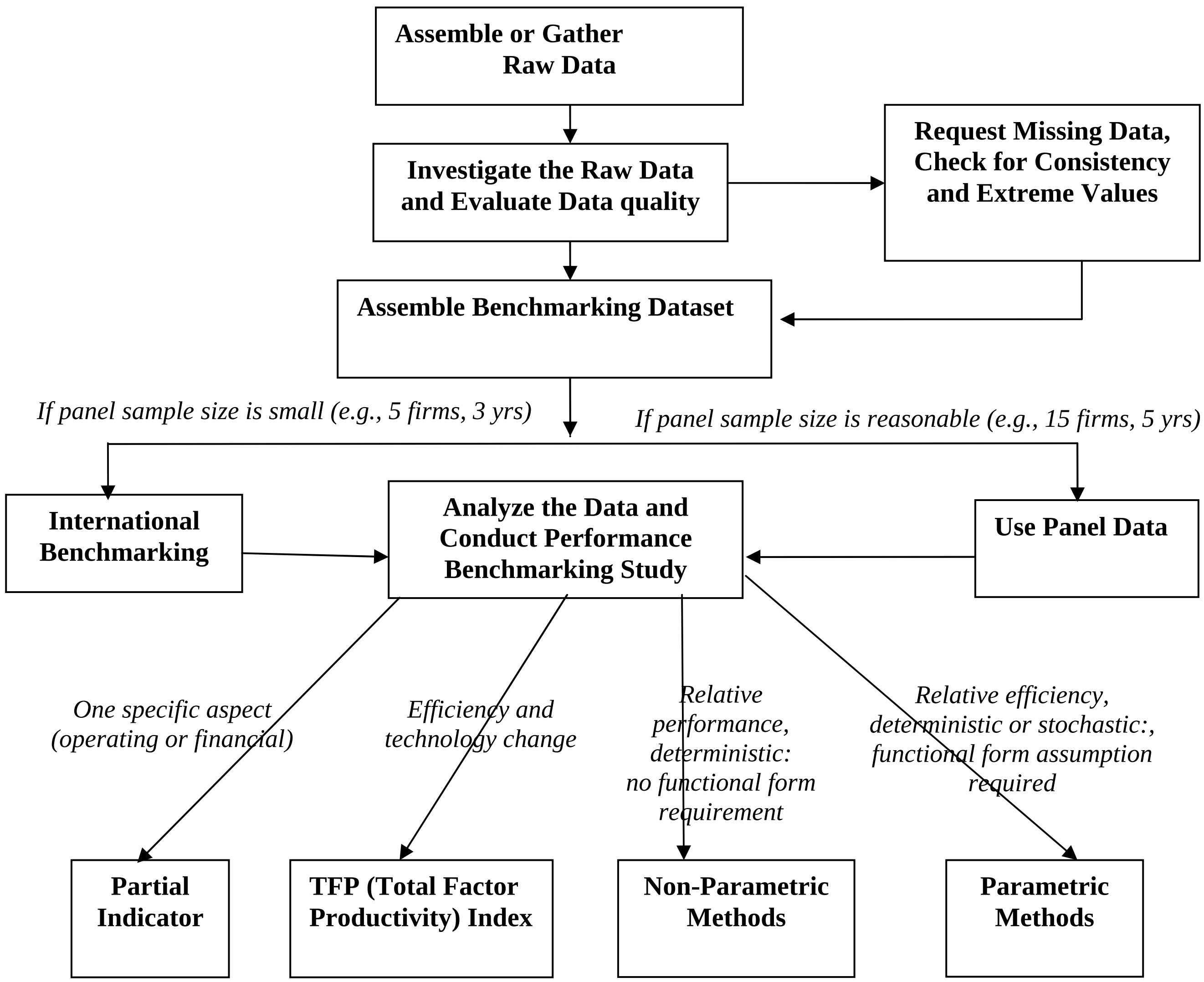
To even come close to achieving the Millennium Development Goals for Water, nations will need to include benchmarking as a tool for documenting the relative performance of utilities —or public and private investments are less likely to be forthcoming. The application of the techniques summarized here can improve service quality, expand networks, and optimize utility operations. Any benchmarking study will have limitations, but sound studies can be used to place the burden of proof on other parties who might argue that the analysis is incomplete or incorrect. Over time, data availability will improve and studies will be strengthened as professionals gain experience with these quantitative techniques. In the process, governance procedures within companies can incorporate this information into managerial incentive packages. Thus, rankings can serve as catalysts for better stewardship of water and other resources. Still, care must be taken to use comprehensive indicators, lest those being evaluated "game" the system. If only a subset is used, performance may improve for some dimensions of a firm's operations but may diminish for others.

The concluding point is that like the physician's injunction to "Do no harm," a benchmarking specialist needs to avoid doing harm. It may be that the quantitative analysis of cost functions and production functions are under-utilized by regulators and managers for fear that results will be misinterpreted—leading to the misuse of rankings or efficiency scores. Others are skeptical of simplistic performance comparisons. To the skeptics, one can say, "If not now, when? If not here, then where?" Benchmarking is a fundamental requirement of good management. If managers do not have the data required for such comparisons, then one must question what they are actually managing. If regulators cannot identify historical trends, determine today's baseline performance, and quantify relative performance across utilities, then as an Indian regulator said, they may as well be "writing pretty poetry".

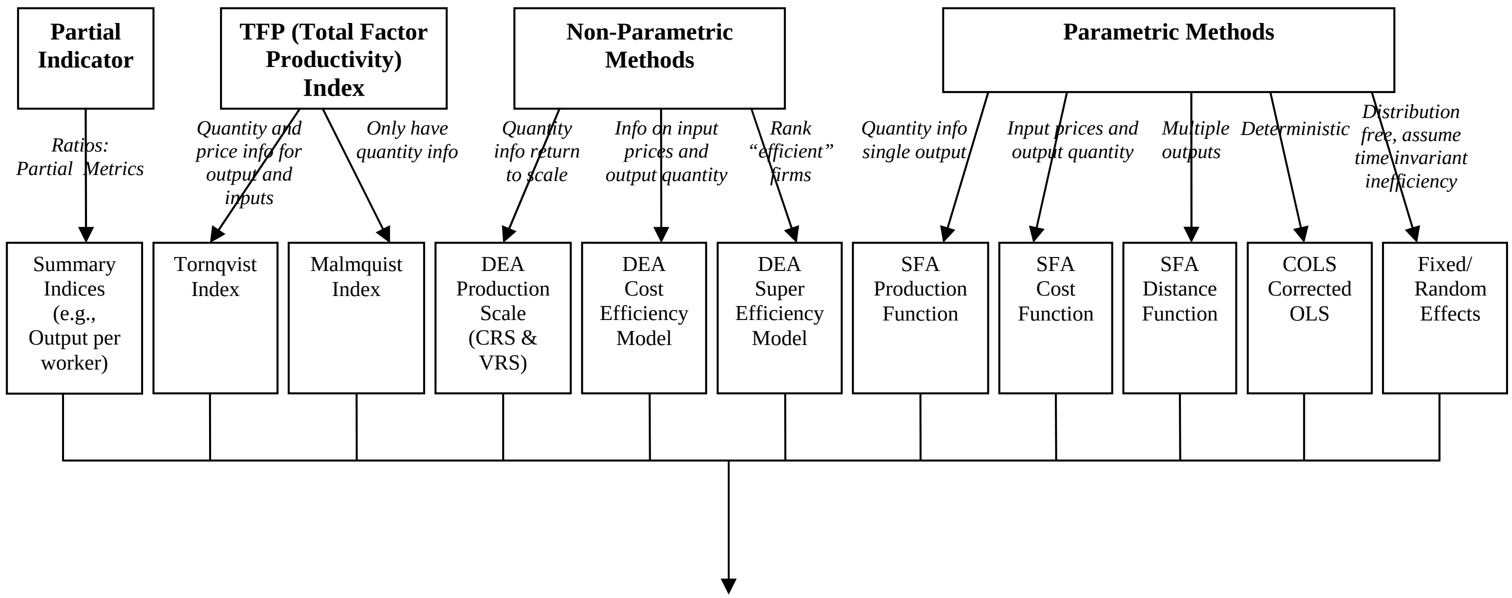
Step 1: Identify Objectives, Select Methodology, and Gather Data



Step 2: Screen and Analyze Data



Step 3: Utilize Specific Analytic Techniques



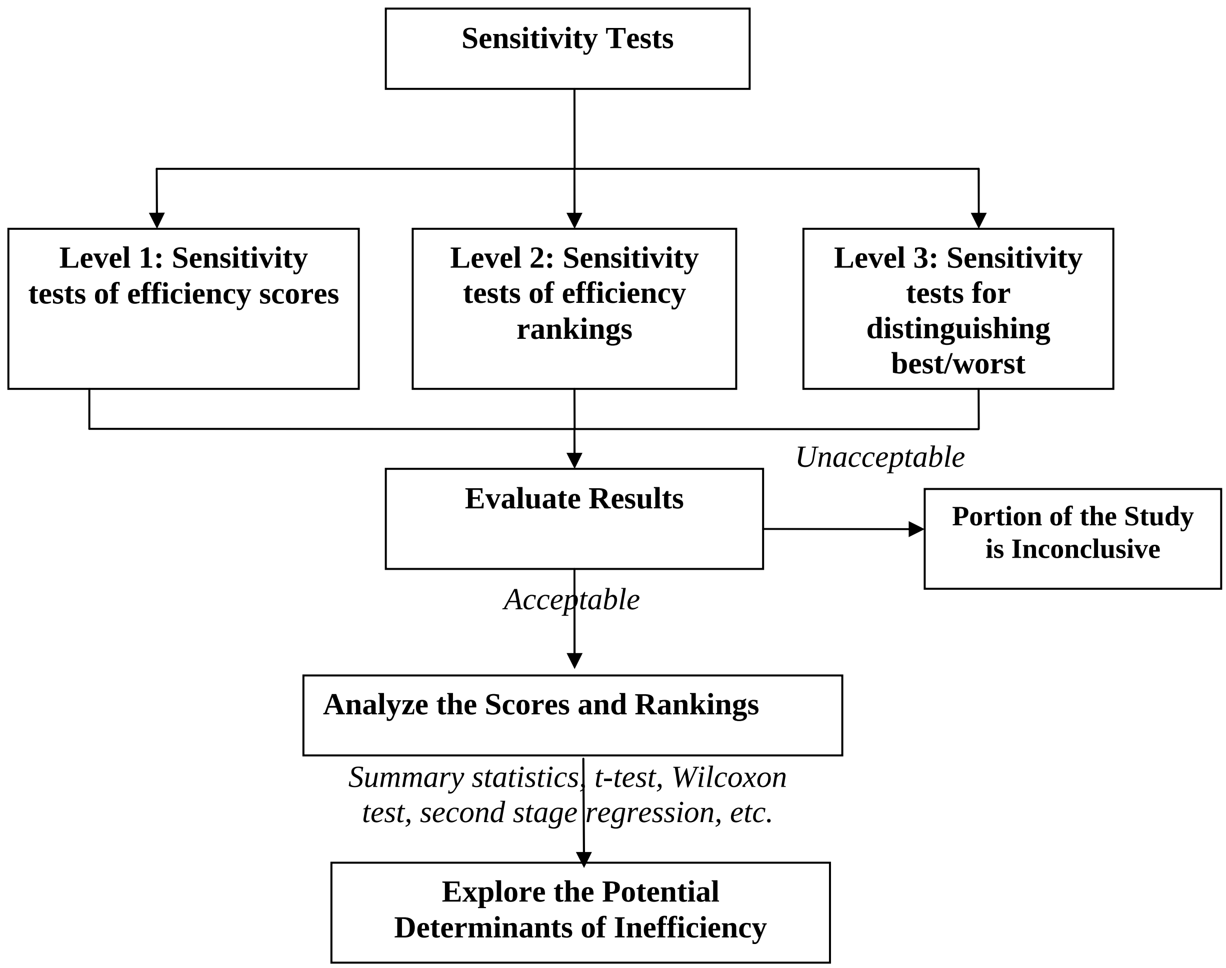
Sensitivity Tests:

Model Specification (e.g., translog vs. log-linear)

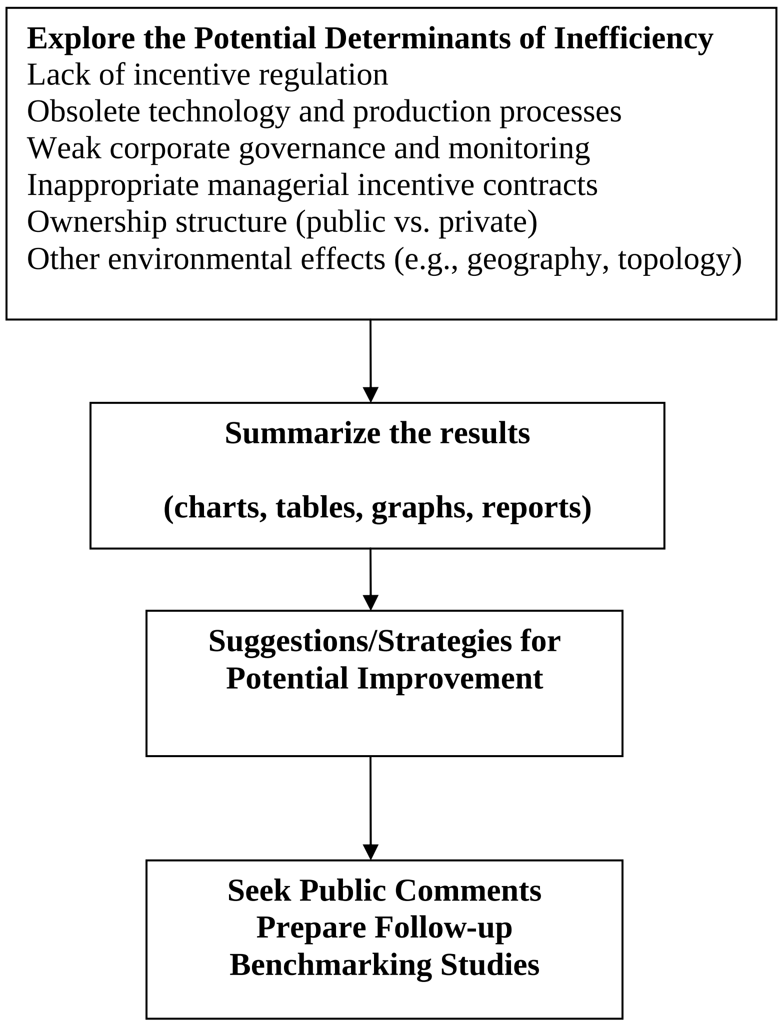
Alternative Inputs and Outputs (e.g., network length vs. fixed assets; quality as an output)

Alternative Methodologies (e.g., DEA vs. SFA)

Step 4: Sensitivity Tests



Step 5: Develop Policy Implications



\