

**Developing Performance Measures for Complex Evaluations:
An Introduction and an Application to Upgrading Infrastructure Systems**

Detlof von Winterfeldt
School of Policy, Planning, and Development
University of Southern California

May 1, 2000

D R A F T

Preparation of this paper was supported by the Institute for Civil Infrastructure Systems (ICIS) at New York University (in partnership with Cornell University, Polytechnic University of New York, and the University of Southern California). This paper is based on activities supported by the National Science Foundation under Cooperative Agreement CMS-9728805. Any opinions, findings, and conclusions or recommendations expressed in this document are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Abstract

Performance measures define how decision alternatives should be measured to determine how well they achieve decision making objectives. For example, one performance measure for decisions on upgrading an electrical power line is the improvement in service reliability, measured as the annual number of customer interruption hours. Developing and using performance measures is one of the most important and difficult tasks in the evaluation of complex decision alternatives. This paper describes some concepts, suggests some guidelines, and illustrates their use for performance measurement. Throughout the paper, the concepts and guidelines are illustrated with an example of the decision of whether or not to underground a residential electrical distribution line. The example illustrates that the choice of a performance measure can have a powerful effect on the analysis results. In some cases, simply changing the performance measure can switch the preferred alternative from undergrounding to leaving the powerline above ground. The paper concludes by suggesting that the development and choice of appropriate performance measures for complex decisions such as infrastructure upgrade decisions should be a deliberate step, involving technical experts, decision makers, and external stakeholders.

Introduction

An important step in the evaluation of complex decision alternatives is the development of measures that characterize the performance of the alternatives in terms of the decision-makers' objectives. Performance measures, as they will be called throughout this paper, are important, because they

1. provide the necessary bridge between the qualitative concerns of decision makers and the quantitative assessments necessary for the comparison and evaluation of alternatives
2. define, what data and judgments should be collected and what calculations need to be performed, when assessing the performance of alternatives
3. Determine how the alternatives are evaluated.

For example, when evaluating a proposal to underground a stretch of electrical distribution lines in a residential neighborhood, a city planner may be concerned with the following objectives: increase service reliability, improve aesthetics, and reduce cost. Each of these objectives can be measured in many different ways. Even a simple objective like cost can have many measures, for example, as construction cost vs. life cycle cost, discounted vs. undiscounted cost, including vs. not including financing expenses.

The development of performance measures is one of the most difficult tasks of evaluating alternatives. The main reason for this difficulty is that there usually are many different ways to define a performance measure for the same objective and that tradeoffs

need to be made, for example between the availability of data for a given performance measure and how meaningful the performance measure is to the decision makers and stakeholders.

This paper describes concepts and principles for developing performance measures and it illustrates them with an example of performance measures for an infrastructure upgrading decision. The concepts and principles build on a report by the National Research Council (1995), which discusses performance measures for infrastructure systems. Section 2 introduces the concepts of decisions, objectives, and performance measures. Section 3 describes how performance measures can be developed and how to select among alternative measures. Section 4 illustrates these issues in the context of performance measures for five objectives for the undergrounding decision: Property values, aesthetics, and health effects, service reliability, and cost. Section 5 provides conclusions and guidelines on how to develop performance measures.

Decisions, Objectives, and Performance Measures.

Decisions. The reason for developing performance measure is to improve the clarity of the comparison and the evaluation of decision alternatives. This paper uses a decision analysis framework for performance measurement (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986; Clemen, 1990). In this framework decision alternatives and objectives are defined first. Objectives are subsequently quantified as performance measures in order to allow the decision-maker to make better choices. The ultimate concern with good performance measures and performance measurement is to improve decision-making.

Objectives. Objectives are operational definitions of values in a specific decision problem. Values are the underlying principles that guide a decision-maker's actions. An individual, for example, may have values that include a concern with family, health, and personal well being. An organization may have values concerned with quality, growth, profit, and good citizenship.

An objective is a specific formulation of a value that includes an object of value and a direction of preference. The object of value can be attractive to the decision-maker (e.g., health), unattractive (e.g., cost). To indicate this, it is often useful to state the direction of preference (more or less of the valuable object) explicitly. For example, by stating to "increase health" or to "reduce cost."

To illustrate the concept of objectives, consider the decision on whether or not to underground a stretch of electrical distribution lines in a residential neighborhood. The decision-maker is the city's planning department. The decision alternatives are to leave the line as is or to underground it. Among the objectives are:

- Improve service reliability
- Increase property values
- Improve aesthetics
- Reduce potential health effects
- Reduce cost

Objectives tend to conflict. In the above example, undergrounding will increase service reliability, property values and aesthetics, but it will do so at a substantial cost.

Ultimately, the decision-maker will have to make tradeoffs: It is worth spending the required amount on undergrounding to receive the benefits?

It is useful to distinguish between means, ends, and process objectives (see also von Winterfeldt, 1999; Keeney, 1990). Ends objectives are the ones that a decision-maker truly cares about. Ends objectives can be discovered with a simple test. Ask, why the decision-maker cares about a stated objective. If the answer is “That is self evident”, it is an ends objective. If the answer is “Because achieving this objectives contributes to achieving another objective,” it is a means objective. For example, when asked why the objective “Increase property value” is important, the decision-maker in the city planning department might state that it is important because this increase leads to an increase in property taxes, which benefit the city. Thus “Increase property tax revenues” may be a better ends objective, while “Increase property values” may be a means objectives. On the other hand, if the decision-maker is a property owner, “Increase property values” may indeed be an ends objective.

Process objectives are concerned with the quality of the decision process. For example, stakeholders may be concerned with the public involvement, fairness, and accountability of a decision process. Process objectives do not differentiate among the alternatives, but they differentiate among decision processes. In the undergrounding example, a process objective may be to involve residents in the decision making process of the city planning department. Clearly, this involvement could be done better or worse, independently of the alternative that is ultimately chosen.

The remainder of this paper will be concerned primarily with performance measures for means and ends objectives.

Performance measures. Objectives express values from the perspective of a decision-maker. Often, they are general and qualitative statements that leave a fair amount of room for interpretation. To make them valuable for evaluation and decision making, one needs to define more quantitative measures that capture how well the alternatives perform on the objectives. These are called performance measures. Ideally, they have a one-to-one relationship with objectives that expresses, in quantitative terms, how an objective is to be measured.

This is not a trivial task. Generally, there are many different definitions of performance measures for a given objective. Often the available data for a performance measure do not capture the decision-maker's concerns precisely. Sometimes several performance measures are required to define a single objective. At other times, the same performance measure can serve to measure several objectives.

Consider the objective "Improve service reliability" in the undergrounding example. There are many different definitions for service reliability, for example, in terms of outage frequency, average outage duration per event, or the product of frequency and duration. The definitions of outages can be system-wide, per circuit, or per mile. They may include only unplanned outages (accidents) or also planned ones (for service, for example). The affected population can be customers or individual users of electricity. Which alternative (overhead lines or underground lines) is better, depends on the choice of the performance measure. For example, when only considering unplanned outages, underground distribution lines generally perform better. However, when adding planned outages, underground lines perform slightly worse than overhead lines,

Developing Performance Measures

A taxonomy of performance measures. Performance measures can be classified as either natural or constructed, direct or indirect (see von Winterfeldt and Edwards, 1986; Keeney and von Winterfeldt, 1981). Natural measures have a common quantitative interpretation and use. Examples are the percent change of property values in case of undergrounding a distribution line, or the direct dollar cost of construction for undergrounding one mile of distribution line.

Not all objectives have obvious natural measures. For example, there is no accepted quantitative scale for the objective “Improve Aesthetics.” Constructed measures provide specific guidance on how to score the consequences of alternative with respect to an objective. When scoring the negative aesthetic impacts of alternative powerline configurations, for example, it may be possible to develop a scale from 0 (no negative impact, clear view) to 5 (impact comparable to multiple lines mixing transmission and distribution on large towers).

In some cases, a constructed measure can be developed by combining several natural measures. The Dow Jones Industrial Index is a constructed measure of the value of Blue Chip stocks based on the value of selected stocks. Keeney and Nair (1977) developed a constructed measure for biological impact of an industrial facility that combines the area of land lost, the type of use of the land, and the loss of wetlands and endangered species (see Table 1).

Level	Impact Description
0	Loss of 1.0 square mile of entirely agricultural or urban “habitat) with no loss of any native communities.
1	Loss of 1.0 square miles of farmed (50%) and disturbed (i.e. logged or new second growth forest) (50%) habitat; no measurable loss of wetlands or endangered species habitat.
2	Loss of 1.0 square miles of farmed or disturbed (logged, plowed) habitat (50%) and mature second growth forest or other undisturbed community (50%); 15% loss of wetlands or endangered species.
3	Loss of 1.0 square mile of mature second growth (but not virgin) forest community; 50% loss of big game and upland game birds; 50% loss of local wetlands and local endangered species habitat.
4	Complete loss of 1.0 square mile of mature virgin forest; 100% loss of local wetlands and local endangered species.

Table 1: Constructed Measure for Biological Impacts of an Industrial Facility (adapted from Keeney and Nair, 1977)

When constructing a measure based on several natural measures, one can choose to use some simple aggregation model (e.g., averaging or weighted averaging) or to translate combinations of the natural measures into levels of the constructed measure. The Dow Jones Industrial Index is an example of the former technique, Keeney and Nair’s biological scale is an example of the latter.

If a constructed measure is based on several natural measures, many different points in the n-dimensional space of the natural measures are mapped into one point of the constructed scale. Thus, the constructed measure loses some of the detail and specificity of the natural measures. Furthermore, in the process of establishing the mapping from the many n-dimensional points into a few one-dimensional categories, tradeoffs have to be made. For example, the biological impact measure equated levels of impact, based on tradeoffs between the seriousness of losing agricultural land vs. wetland vs. endangered species.

Natural measures are preferable over constructed measures, since they prescribe more precisely how the measurement is to be done and leave little ambiguity or judgment in assigning consequences to alternatives. Constructed measures should only be developed, when natural measures are not available, or when they cannot capture the meaning of an ends objective. Constructing measures is an art that requires substantial experience. Keeney and von Winterfeldt (1981) and Keeney (1990) provide some guidelines.

Direct performance measures measure an ends objective directly. For example, a direct measure of service reliability is the expected outage duration per year per customer. Indirect performance measures measure a means objectives. An indirect measure of service reliability is the frequency of outage repair service calls per mile of a distribution line. In a different decision domain, a direct measure of the health effects of a pollutant may be the number of cancers it causes per year. An indirect measure may be the amount of pollutants emitted per year.

There usually is a probabilistic relationship between indirect and direct measures. For example, knowing the amount of a pollutants, measured in tons/year that is emitted from a smokestack is linked probabilistically to the health effects that the pollution will produce (see Figure 1). Even if we knew precisely the amount of a pollutant emitted, we would not know the exact number of health effects, because of limited knowledge about meteorological conditions, population exposed, and dose-response functions, to name only a few uncertainties. Figure 1 illustrates this by indicating that a precise knowledge of 50 tons of emitted pollutants per year could lead anywhere from about 30 to 180 additional asthma cases per year.

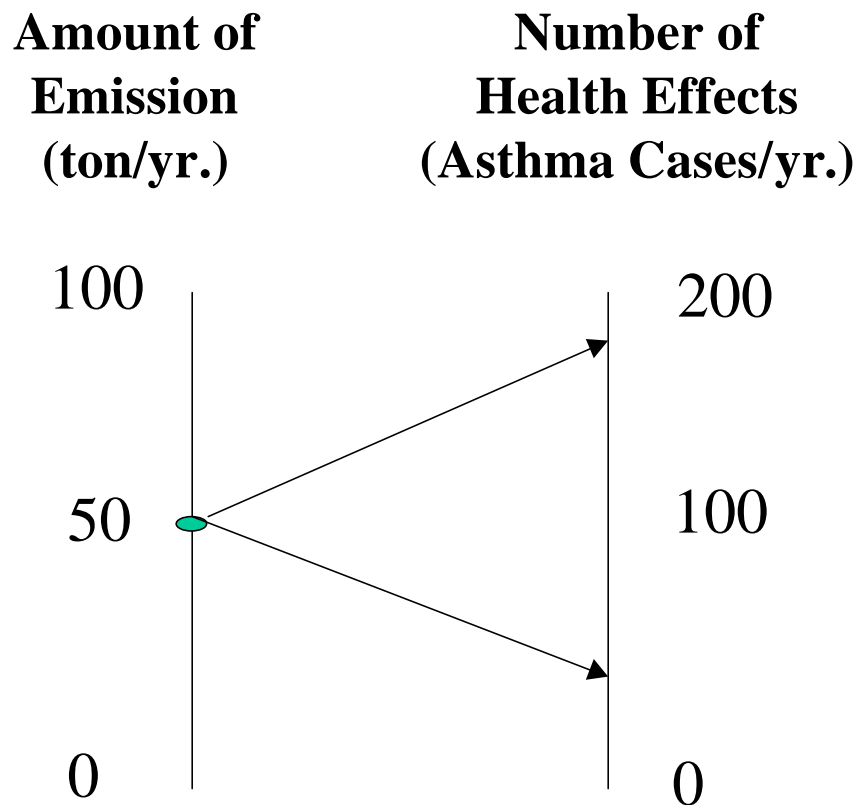


Figure 1: Illustration of the Probabilistic Relationship between an Indirect and a Direct Performance Measure

In general, direct measures are preferable over indirect ones, because they are more linked to the ends objectives and therefore to the preferences of the decision-maker. However, tradeoffs have to be made here too, for example, when indirect measures are available as natural measures, while no direct measures have natural measures. For example, a natural measure for “Aesthetics” may be the number of lines on a pole that disturb the view. This captures one aspect of the objective, but the overall concern with aesthetics includes more, including the height and location of the poles, the surrounding environment, etc. A constructed measure, based on judgments of pictures, may be more appropriate in this case.

Table 2 shows the taxonomy of performance measures with some examples for measuring the potential health effects due to exposure to electromagnetic fields (EMF) from electric power lines. Ideally, one wants to use *direct, natural* measure like the number of health effects per year. However, since there is much uncertainty about whether or not there is a health effect, some researcher prefer to use an *indirect, natural* measure like the average exposure at some distance from the powerline. Wertheimer and Leeper (1979) created an interesting *indirect, constructed* measure, called the “wire code.” They visually scored the appearance of powerlines and the distance of the residences to the lines in what they called “Current Configurations.” A very high current configuration (VHCC) was characterized by line with a thick set of conductors (indicating high loads) and a short proximity from the residence (indicating high exposure). A very low current configuration (VLCC) was characterized by no powerlines in the vicinity of the residence. Intermediate levels were low current

configurations (LCC) and high current configurations (HCC). In this example, it is hard to think of a direct, constructed measure. Perhaps a direct judgment by experts on a rating scale of the seriousness of the health effects could serve that purpose.

	Natural	Constructed
Direct	Number of Cancers per Year	Judged Seriousness of Health Effects
Indirect	Average Field Strength at 50 Feet from Line	Wertheimer and Leeper's "Wire Code"

Table 2: Taxonomy of Performance Measures with Examples

Criteria for selecting performance measures. In most cases several performance measures exist for the same ends objective. Ultimately, the decision-maker ultimately has to make a choice. When making this choice, one needs to consider the following criteria:

1. The degree to which the measure reflects the full content of the objective
2. The availability of data and expertise for assessing alternatives on the measure
3. The degree of certainty with which alternatives can be assessed on the measure
4. The meaningfulness and acceptability of the measure to experts
5. The meaningfulness and acceptability of the measure to laypeople
6. The ease of making tradeoffs between units of this measure vs. other measures.

No measure meets all criteria perfectly. Consider, for example, the measure “Number of Health Effects (e.g., cancers) per year from EMF exposure.” This measure scores well on criteria 1, 4, 5, and 6. However, it scores very poorly on criteria 2 and 3, because there is little data and expertise to assess the alternatives (overhead vs. underground lines) on this

measure, and because there is substantial uncertainty about the extent of a possible health effect, if any. On the other hand, the measure “Average field strength at a 50 foot distance from the power line” scores well on criteria 2, 3, and 4, but poorly on 1, 5, and 6.

Process Considerations when selecting performance measures. When selecting performance measures, it is useful to involve experts, decision-makers, and possibly other stakeholders. The process should begin with a clear definition of the ends or means objective that is to be measured. If the experts, decision-makers, and stakeholders cannot agree on this definition, it is pointless to pursue the performance measures any further.

Next, the experts should propose a possible set of performance measures and discuss their pros and cons in terms of the criteria listed above. However, the ultimate choice of a performance measure is not made by the experts. Decision-makers have a major stake in this choice, since they are affected in several ways. First, they will have to communicate the measures to other stakeholders, most of whom will be laypeople, not experts. Second, they will have to be able to understand the measure and make tradeoffs using them. Other stakeholders have an interest in using measures that are meaningful and that reflect their interests.

Because there are tradeoffs between the criteria for good performance measures, reasonable people can disagree about which one is best. In the end, the decision-maker has to choose, but is it wise to listen to both the experts and the external stakeholders, when making this choice.

It is also possible to carry multiple alternative performance measures through an analysis, if a particular measure is highly controversial. In the undergrounding example a

major choice of the performance measure for cost was the extent of financing chosen and the discount rate – both influenced the ultimate decision substantially. By keeping variants of the cost measure as part of the decision analysis, valuable insights could be gained about the preferred alternative as seen from the perspective of different stakeholders.

Use of performance measures. There are two closely related uses of performance measures: To allow an accurate, quantitative comparison of the alternatives; and to facilitate making tradeoffs when evaluating alternatives. The first use is best illustrated by the tables of Consumer Report magazine. These tables use many measures, some natural, some constructed, some direct, some indirect, to compare the performance of consumer products. The data are presented in a compact table that provides the user with all the comparative information they need. The final evaluation is, however, still up to the user.

Often this evaluation involved tough tradeoffs. For example, should I spend \$200 more on a computer to increase its hard disk space from 5 GB to 10 GB? These tradeoffs are made easier with good performance measures, especially if they measure the ends objective of the decision-maker directly. In addition, performance measures can often be converted directly into equivalent dollar costs or benefits. For example, if the performance measure is the number of cancers per year, it is possible to assign an equivalent dollar value to the elimination of one cancer per year and then apply this to the range of the measure. For example, if avoiding one cancer per year is valued at \$2 million, the avoiding ten cancers per year should be valued at \$20 million.

Illustration of performance measurement when deciding whether to underground an electrical distribution line.

This section uses material from a multi-year effort by the author and his colleagues on developing policies for addressing the concern with health effects from electromagnetic fields exposure from electrical power lines (see von Winterfeldt et al., 1996; von Winterfeldt et al., 1999). As this investigation unfolded, it became clear that a major issue in the EMF debate had to do with the costs and benefits of undergrounding powerlines. Undergrounding is an effective strategy for reducing EMFs and it also improves the aesthetics and possibly increases property values. Thus, those concerned with EMF exposure forged a coalition with those interested in undergrounding powerlines for aesthetic or property value reasons.

The scenario that will be considered in this section considers the decision of whether or not to underground a four-mile stretch of a distribution line. The decision-maker is a consortium of city planners, utility representatives, and homeowners.

The decision alternatives are clear: To leave the power lines on poles above ground to underground them. Many objectives were considered (see von Winterfeldt, 1999), but only the following ends objectives made a difference to the decision:

- Improve service reliability
- Increase property values
- Improve aesthetics
- Reduce potential health effects

- Reduce cost
-

Each of these ends objectives has different possible performance measures. Some of these are debated among the stakeholders. Some lead to different preferred alternatives. These issues will be discussed separately for the five performance measures below.

Improve service reliability. There are numerous measures of the reliability of electrical service, including:

1. SAIFI: System average interruption frequency index measured as the average number of interruptions per customer per year,
2. SAIDI: System average interruption duration index measured as the average time a customer is without electricity per year.
3. MAIFI: Momentary average interruption frequency index measured as the average number of interruptions of less than a few seconds per customer per year,
4. CAIDI: Customer average interruption duration index measured as the average length of electricity interruption per event.

Other measures include the frequency and duration of equipment failures and of failures of circuits. These failures do not always lead to customer interruptions, because of the redundancy in the electrical service network.

From a customer perspective, SAIDI is the best measure of reliability, because it measures the impact of systems failures most directly. However, there are a few

problems with SAIDI. First, it is measured on the customer level (i.e., on the level of an individual meter). Yet, some customers are single-family households, others are businesses with many employees. Clearly, one hour of annual electricity interruptions for a single-family household is not as severe as a one-hour interruption of a business with 50 employees. Furthermore, SAIDI does not distinguish between day and nighttime interruptions or between short term and long term interruptions, which may have non-linear consequences. Finally, SAIDI is not always conveniently available separately for the decision alternatives that are considered. For example, when studying undergrounding in California, it took a special data collection effort to determine SAIDIs separately for overhead and underground lines, and even then this led to satisfactory results only for distribution lines, not for transmission lines (von Winterfeldt et al., 1999).

Equipment outage data (frequency and duration) are more readily available, but they are only loosely linked to customer impacts. For example, most high-voltage transmission lines are in networks that create sufficient redundancy to provide full customer service, even if one of the lines has an outage. This is less true for distribution lines, which are often designed in a “radial” fashion emanating from a substation. In any case, it is not trivial to correlate equipment outages with the more relevant customer impacts.

An additional complication comes from the distinction between unplanned and planned outages. Unplanned outages are events of nature, e.g., due to storms that knock over power poles or to corrosion of underground cables. Planned outages occur during services or upgrades. Many utilities allow “live-line” service and upgrades for overhead lines, but underground lines are typically shut down for these activities, because of the

restricted space and the possible danger of electrocution. As a result, planned outages can be substantial for underground lines (about the same time as unplanned outages), but they are typically negligible for overhead lines. While planned outages do occur more frequently and for longer times for underground lines, they also occur in “of-peak” hours and they are typically announced to the customers before they occur. Thus, they should not be considered to be as serious as unplanned outages.

Stakeholders tend to prefer measures that align with their own positions and preferences. Utilities like SAIDI and MAIFI, because they are used to it and because it correlates with their responsibilities to maintain a reliable flow of electricity to customers. They also like to consider both planned and unplanned outages, because they are both results of systems effects. Resident like SAIDI, but would prefer a more people oriented measure, for example, how many person hours are interrupted, rather than customer or meter hours. They would like to focus on unplanned outages and ignore planned ones, primarily, because planned outages can be managed more effectively from the customer point of view. Some engineers in utility companies and systems operators like equipment outage data, because this data tells them where equipment problems occur.

As in the previous cases, the analyst has a tough choice. In the analysis by von Winterfeldt et al. (1999), we chose to use SAIDIs split by overhead and underground distribution lines and equipment outage data for transmission lines, which were coupled with a model that translated equipment outages into customer interruptions. We gave users of the analysis an option to either use unplanned outages only or to use both planned and unplanned ones.

Increase property values. Properties located near electrical powerlines typically have slightly lower values than those further away (see, for example, Hamilton and Schwann, 19989; Gregory and von Winterfeldt, 1986). Estimates of the impact of powerlines on property values range from no decrease to about 10% decrease, depending on the type of line (transmission vs. distribution), the aesthetic impact of the line, closeness of the property to the line, and other considerations. When contemplating the decision to underground a powerline, the impact on property values thus becomes an important consideration. But how should this impact be measured?

According to economic cost-benefit thinking this impact is evaluated by looking from the decision point into the future, suggesting that *undergrounding produces an increase* in existing property values. A reasonable natural measure would be the percent increase in the value of all homes in the vicinity of the line. This percentage measure can be easily translated into a total dollar figure, once the property values are known.

Residents living near powerlines strongly object to measuring the impact of undergrounding on property values as an increase, and instead want to measure it as a decrease in property values for the “leave powerlines above ground” alternative. The reason given is that the action to underground is simply restitution for a negative impact that homeowners were burdened with some time ago. In this world view, the measure would be the percent decrease in property values due to the overhead line and undergrounding would simply be judged to have a zero percent decrease. The net results of a cost-benefit analysis using these two measures would be identical, but the

forward looking view of the economist and the restitution oriented view of the residents clash on moral and ethical grounds.

The situation is complicated by the fact that not all homeowners lived in the home when the powerline was built. Those homeowners who moved in later presumably bought the home at a lower price to compensate them for the powerline impacts. For them, the property value benefit of undergrounding is not restitution, but a windfall. As a practical matter, tracking home ownership in the calculation of who “wins” and who receives “just restitution” can easily become an analyst’s nightmare.

Ultimately, the decision-maker and analyst has to decide on the appropriate measure. In this case, it is not too difficult to side with the economists’ point of view, acknowledging that the benefits that accrue to homeowners would be judged differently by them, depending on the time that they moved into their homes. Nevertheless in the project by von Winterfeldt et al. (1999), the residents’ point of view was so strongly advocated that the computer model that was developed to support the decision now includes a toggle that lets users switch between the “loss” and the “gain” framing” of the property value impacts of undergrounding.

Improve aesthetics. Everyone agrees that undergrounding electrical powerlines improves the aesthetics of the environment. By aesthetic impact we mean the impact on people other than the homeowners, for example, people who drive through an otherwise beautiful canyon road that is lined with fairly ugly powerlines. Unfortunately, agreed upon measure of aesthetics, nor is there a metric that would allow one to translate such a measure into economic consequences. von Winterfeldt et al. (1999) proposed a simple aesthetics measure that combines the following aspects:

1. Whether it is a transmission line or a distribution line
2. Whether the line is on a right-of-way vs. street side
3. The type of structure (pole or tower)
4. The number of wires on the structure

The idea was that transmission lines on the street side on towers with multiple wires have the largest aesthetic impact, while distribution lines on right-of-ways (i.e., in backyards) have the smallest aesthetic impact.

Some critics of this approach have argued that whatever the impact of aesthetics of powerlines may be, it is already incorporated in property values. The argument goes further by challenging that including an additional aesthetic value is, in a sense, double counting the aesthetic impacts of power lines. There is clearly the potential for double counting. But one has to be clear that there is a non-property value of aesthetics, for example, on the well being of people who don't own property, but have a reduced enjoyment of the scenery and environment. In the end, it is probably reasonable to use two measures, but to be careful what is counted under each. For example, the non-property values impact of aesthetics needs to be evaluated carefully by only considering the potential benefits of undergrounding for those who visit and drive by the area.

Reduce health effects. The possibility of health effects due to exposure to electromagnetic fields from powerlines have been debated ever since Wertheimer and Leeper (1979) published an article reporting a positive association between a constructed measure of exposure (the “wire code” described above) and the rate of childhood

leukemia. Research is still inconclusive, with some epidemiological studies reporting an association, but with no clearly established physical and biological mechanism that could support the epidemiology. Research opinions are well entrenched ranging from expression that there could not possibly be a link of EMFs to health effects because of the low energy levels, to an acknowledgment that there is some evidence of biological effects, to serious allegations that there are serious ongoing health concerns.

Overhead lines produce fields that are (roughly) inversely proportional to the square of the distance from a powerline and proportional to the current on the line. Undergrounding is an effective means of reducing EMF exposure, provided that the cables are placed in close proximity to produce a field cancellation effect.

In this case it is clear what one wants to measure: The number of health effects, separately for several implicated health endpoint (e.g., leukemia, brain cancer, Alzheimer's disease, spontaneous abortions), and affected age groups. The age concern arises because some health effects that are possibly associated with EMF exposure primarily affect children (leukemia), while others almost exclusively affect old people (Alzheimer's disease). The problem is that there are numerous uncertainties that make it virtually impossible to come up with good estimates of the health effects of exposure to EMFs.

The next best solution is to select an indirect measure, for example, the average field strength measured in homes near powerlines. The advantage of this indirect measure is that it can be taken without much uncertainty, using standard and readily available instruments. The disadvantage is the very uncertain relationship of this indirect measure with health effects. In fact, there is surprisingly little evidence of an

association between measured or estimated average field strength and health effects. In addition, there are many possible alternative measures of exposure, for example, peak fields, time spent in high fields, rapid field strength changes, etc. Each of these alternative measures may be more closely linked to possible health effects.

The third route is to use a constructed measure like Wertheimer and Leeper's wire code. This constructed measure has shown the most consistent association with health effects, even in studies that showed no association with measured fields. The disadvantage of this measure is that it may be correlated with some other impact (e.g., noise or pollution from cars), and thus create a spurious correlation with health effects.

The approach taken by some analysts (e.g., von Winterfeldt et al., 1999) is to discard the constructed measures, explore alternative exposure metrics and develop probabilistic models that link these measures and to possible effects (including none at all). Others rely more on the use of indirect measures (Morgan et al., 1994).

Reduce Cost. On surface, one would expect that developing a measure for this objective should be straightforward. In practice, it turned out to be a very difficult task because stakeholders disagreed on

1. The appropriate detail of cost measurement and cost components
2. How to treat cost streams and discounting
3. How to deal with financing and the cost of money

In the project by von Winterfeldt et al. (1999), most stakeholders agreed early on that the guiding principle should be to determine total life-cycle cost (TLCC) of the decision of whether or not to underground a line. TLCC would include total project cost (engineering, design, construction, construction management, etc.), operation and

maintenance (service, repairs, tree trimming, capital upgrades, etc.) and line losses (losing power due to line impedances and the associated opportunity cost of selling electricity). Obtaining total life cycle cost is not a trivial step in itself. Often, project costs are simply measured by their up-front construction expenses.

However, the agreement ended there. Utility representatives wanted cost estimates at the level of final engineering design, while residents and most analysts involved in the project were comfortable with cost estimates at the level of a conceptual design. To some degree this is just a matter of appropriate level of detail. However, it also became a focal point of controversy, since, as more cost detail was added, more criticisms were leveled against both the component cost estimates and the other details that were left out of the cost equation. For example, reasonably good data is available on the average cost of undergrounding a mile of distribution line. Once one begins to consider different terrains, soil conditions, number of required manholes, underpasses, cost for taking down poles and their salvage value and so forth, the comfortable agreement about the rule of thumb per-mile estimates is overshadowed by minute arguments of the salvage value of a power pole, for example.

Another issue had to do with the proper treatment of streams of cost. Standard economic theory suggests to calculate the constant dollar cost stream over the life-cycle of the project and then to discount this stream to determine net present value. The argument then is typically about the appropriate discount rate. Those who favor projects with a large cost stream in later project stages favor discounting, usually at a high rate. Those who favor large up-front investments with low costs later in the project favor no discounting or a low discount rate. For government project, the federal Office of

Management and Budget (OMB) provides guidelines for what discount rate to use. Yet, in practice, stakeholders want to see an analysis performed with no, OMB, and a high discount rate (see, for example, von Winterfeldt and Schweitzer, 1998).

While discounting at some rate is the appropriate economist's solution, there is another concern, not frequently mentioned. Undiscounted costs are usually easy to trace and recalculate – they are simply constant dollar costs spread over the life time of the project. Discounted costs are much more difficult to trace. For example, it is not obvious what impact a mid-life cycle capital upgrade will have on discounted costs. Furthermore, the net discount rate adjusted for inflation is usually fairly small (2-3%). Thus, as a benchmark it is often not unreasonable to include an undiscounted case in the analysis.

A third controversy surrounded the issue of how to deal with financing an undergrounding project. If this project is privately financed or financed through an assessment district, the results are very clear and essentially translate into a cost stream over n years. However, if the costs are absorbed into an increased utility rate, the results can be very different. For example, it may be possible, at least in principle, to pay for a staggered undergrounding program in a large city through staggered increases in utility rates without financing the actual project costs. Another issue had to do with the cost of money for the equity contribution of a utility company in the case that the project had to be financed by the utility. Some utility financial programs assigned a very high cost of money to their equity portion, while others considered this essentially part of their operating budget.

The stakeholder positions on these issues were clear. Residents favoring undergrounding argued that the analysis should consider an unfinanced, highly

discounted case with quite general per-mile cost numbers. Utilities wanted very specific, component-wise cost numbers, financing, and moderate discounting.

Again, the analysts had to make compromises. We included a financed (with a user specified x percent equity) and a non-financed case, and a discounted (with a user specified y percent discount rate) and non-discounted case. We also included a highly controversial detailed cost estimation model and a user specified per-mile input. As we will discuss below, the choice between these different options for measure strongly influenced the results.

The consequences of using different performance measures. In this subsection, we will demonstrate how choices among performance measures influence the results of the decision on whether or not to underground a stretch of distribution lines. Principally, these changes came from cost and property issues. Aesthetics, health effects, and service reliability also had impacts, but they did not change the decision nearly as much.

Impact of assumptions about discounting and financing. Table 3 shows the equivalent costs of leaving the line overhead vs. undergrounding it for four different assumptions about discounting and financing, split up by the major objectives. The total equivalent costs tell the main story here. For the two financed cases, undergrounding loses convincingly, because the high up-front costs need to be financed at a 10% interest rate, which is much higher than the discount rate of 3%. For the two unfinanced cases, undergrounding wins by about \$1.4 million and \$900,000 respectively. However, this assumes a \$3 million property values benefit, which some stakeholders question. We turn to this issue next.

Table 3: Equivalent Costs of the Undergrounding Decision Using Different Assumptions about Financing and Discounting
(in thousands of 1999 dollars)

Case 1: No discounting, total project cost not financed

	Service Reliability	Property Values	Health-EMF	Cost	Other	Total	Result
Overhead	\$567	-\$187	\$1,800	\$1,069	\$164	\$2,846	
Underground	\$415	-\$3,000	\$69	\$4,400	-\$57	\$1,412	UG

Case 2: No discounting, 80% of total project cost financed at 10% interest rate

Overhead	\$567	-\$187	\$1,800	\$1,069	\$164	\$2,846	OH
Underground	\$415	-\$3,000	\$69	\$11,020	-\$57	\$8,032	

Case 3: Discounting at 3%, total project not financed

Overhead	\$349	-\$124	\$1,100	\$656	\$101	\$1,733	
Underground	\$254	-\$3,000	\$42	\$3,900	-\$60	\$882	UG

Case 4: Discounting at 3%, 80% of total project cost financed at 10% interest rate

Overhead	\$349	-\$124	\$1,100	\$656	\$101	\$1,733	OH
Underground	\$254	-\$3,000	\$42	\$7,000	-\$60	\$3,982	

Impact of property values frame and amounts. Table 4 shows the results of the analysis using either a “loss” frame for property values or a “gain” frame and using either 2.5% or 5% property values impacts. In all cases, we assumed that the discount rate is 3%, but that the projects are not financed (i.e., they are paid out of an increased rate base).

As expected, the loss vs. gain frame has no impact on the ultimate result, since in the loss frame one simply adds a constant amount to the equivalent cost of property values. However, by halving the impact of undergrounding on property values, the preferred alternative switches back again to leaving the line overhead.

Table 4: Equivalent Cost of Undergrounding Using Different Assumptions about the Impact of Property Values

Case 1: Gain Frame and 5% Property Impacts

Overhead	\$349	-\$124	\$1,100	\$656	\$101	\$1,733	
Underground	\$254	-\$3,000	\$42	\$3,900	-\$60	\$882	UG

Case 2: Loss Frame and 5% Property Impacts

Overhead	\$349	\$2,876	\$1,100	\$656	\$101	\$4,733	
Underground	\$254	\$0	\$42	\$3,900	-\$60	\$3,882	UG

Case 3: Gain Frame and 2.5% Property Impacts

Overhead	\$349	-\$108	\$1,100	\$656	\$101	\$1,749	OH
Underground	\$254	-\$1,500	\$42	\$3,900	-\$60	\$2,382	

Case 4: Loss Frame and 2.5% Property Impacts

Overhead	\$349	\$1,392	\$1,100	\$656	\$101	\$3,249	OH
Underground	\$254	\$0	\$42	\$3,900	-\$60	\$3,882	

Conclusions

This paper has described the concepts and principles for developing performance measures and it has illustrated the powerful effects of alternative definitions of performance measures on the possible results of a decision analysis. The main conclusions and recommendations are:

1. Before working on performance measures, one should develop means and ends objectives, because they provide the basis for what is to be measured. Stakeholders should be involved in the development of objectives.
2. Performance measures should be derived from the objectives. Performance measure should not be used merely because data are readily or because others

have used them before. Experts and stakeholders should be involved in identifying and selecting performance measures that are appropriate for the objectives.

3. Natural and direct measures are preferred over constructed and indirect ones
4. When it is impossible to use natural and direct measures, carefully the tradeoffs between using constructed direct measures vs. using natural indirect measures should be explored carefully. The criteria for good performance measures listed in section 3 should help.
5. When there are several contending performance measures, an analysis should be conducted with alternative measures to determine the impact on the results.

Perhaps the best advice is to make a deliberate effort to develop performance measures in complex evaluations, involving decision-makers, experts, and stakeholders. As the introductory sections of this report indicate, this is a very complex and difficult task. As the example of the distribution line upgrading case demonstrates, the results can vary dramatically, depending on the performance measure chose.

References

- Gregory, R. and von Winterfeldt, D. The effects of electromagnetic fields from transmission lines on public fears and property values. *Journal of Environmental Management*, 48, 1996, 201-214.
- Hamilton, S. and Schwann, G. Do high voltage transmission lines affect property values? *Land Economics*, 71, 1995, 436-444.
- Keeney, R.L. *Value focused thinking*. Cambridge, MA: Harvard Business School Press, 1990.
- Keeney, R.L. and Raiffa, H. *Decisions with multiple objectives*. New York: Wiley, 1986.
- Keeney, R.L. and von Winterfeldt, D. Operational procedures to evaluate decisions with multiple objectives. EPRI Report No. RP 2141-10, Palo Alto: Electric Power Research Institute, 1981.
- Keeney, R.L. and Nair, K. Selecting nuclear power plant sites in the Pacific Northwest using decision analysis. In. D. Bell, R.L. Keeney, and H. Raiffa (Eds.) *Conflicting objectives in decisions*. New York: Wiley, 1977.
- Morgan, G. and Nair, I. Alternative functional relationships between ELF field exposure and possible health effects: Report on an expert workshop. *Bioelectromagnetics*, 13, 1995, 335-343.
- National Research Council. *Measuring and improving infrastructure performance*. Board on Infrastructure and the Constructed Environment. Washington, DC: National Academy Press, 1995.
- von Winterfeldt, D. and Edwards, W. *Decision analysis and behavioral research*. New York: Cambridge University Press, 1986.
- von Winterfeldt, D. and Keeney, R.L. Managing magnetic fields: Insights from decision analysis. EPRI Report No. 2560-03. Palo Alto: Electric Power Research Institute, 1996.
- von Winterfeldt, D. and Schweitzer, E. An assessment of tritium supply alternatives to support the nuclear weapons stockpile. *Interfaces*, 28, 1998.

von Winterfeldt, D., Adams, J., Eppel, T., and Nair, I. Power Grid and Land Use Policy Analysis. Final report submitted to the California Department of Health Services. CDHS: Oakland, 1999.

Wertheimer, N. and Leeper, E. Electrical wiring configurations and childhood leukemia in Rhode Island. *American Journal of Epidemiology*, 109, 273-284, 1979.