# Defining and selecting objectives and performance metrics for oil spill response assessment: A process design integrating analysis and deliberation

**A technical report submitted to the Coastal Response Research Center by**


**Dr. Seth Tuler, Dr. Thomas P. Seager, Rebecca Kay, and Dr. Thomas Webler**

Social and Environmental Research Institute, Inc.
278 Main Street, Room 404, Greenfield, MA 01301
413.773.9955 www.seri-us.org

**-and-**

**Dr. Igor Linkov**

58 Charles St., Cambridge MA 02141


**30 March, 2007**


Funding provided by the
Coastal Response Research Center
www.crrc.unh.edu

Coastal Response Research Center

NOAA Grant Number: NA04NOS4190063. Project Number: 05-983.

# Defining and selecting objectives and performance metrics for oil spill response assessment: A process design integrating analysis and deliberation

SP TULER, T WEBLER, TP SEAGER, R KAY.
Social and Environmental Research Institute
278 Main Street, Room 404, Greenfield, MA 01301, USA

I LINKOV
Intertox, Inc.
83 Winchester Street, Brookline, MA 02446, USA

**Abstract**
This document lays out a framework by which defensible performance metrics for oil spill response can be developed in an analytic-deliberative process that includes a broad range of stakeholders and is informed by high quality scientific knowledge. Performance metrics can be used to: improve oil spill response planning, promote institutional learning post-response, support public communication during and after spills, etc. The framework takes as a starting point the existing “business as usual” approach to response planning and proposes specific activities that can be undertaken to integrate selection of performance metrics into contingency planning efforts. This paper builds on theories and empirical research on public participation in hazard management, typologies of performance metrics, and decision analysis. A core principle of this framework is that performance metrics for oil spill response must be tightly coupled to objectives for oil spill response. Objectives, performance metrics, and methods for measuring need to be selected so that they represent both best available science and appropriate democratic agreement. Using the best available science means that metrics should be measurable, meaningful, and are consistent with previous scientific work. It also means that the measures of metrics (and objectives) selected are justified with evidence, models, or theories. Using the appropriate democratic agreement means that the objectives, metrics, and measures selected adequately represent the interests at risk. It also means that the objectives, metrics, and measures chosen reflect a consensus among interested and affected parties. The integration of performance metrics into oil spill response planning will help ensure plans better reflect public concerns and values, promote learning, and improve spill managers’ communications with the public.

## 1. Introduction

Oil spill response is defined as encompassing all activities involved in containing and cleaning up oil to[1]:

- maintain safety of human life;
- stabilize a situation to preclude it from worsening, and
- minimize adverse environmental and socioeconomic impacts by coordinating all containment and removal activities to carry out a timely, effective response.

Achieving these goals is a complex matter. Spills are surprises, and response activities must ramp-up quickly with responsibilities and personnel distributed among multiple agencies and organizations. Often important information is unknown or uncertain. Experienced spill responders often say that "all spills are different" because the conditions that define the spill can be highly variable, including the type of oil, weather conditions, and resources at risk (due to seasonal changes, for example).

As part of ongoing efforts to improve national, state, and local spill preparedness, extensive planning and coordinating efforts have been undertaken during the last fifteen years, culminating in the creation of the National Response Plan (DHS 2004). In the case of oil spills, the National Response Plan (NRP) "describes the lead coordination roles, the division and specification of responsibilities among Federal agencies under anticipated crisis scenarios and the national, regional, and onsite response organizations, personnel, and resources that may be used to support response actions" (DHS 2004 pg. ESF#10-1). Responsible federal agencies may include the US Coast Guard and US Environmental Protection Agency. Although the frequency of domestic spills has declined precipitously since passage of the Oil Pollution Act of 1990, the effectiveness of response efforts is more difficult to gauge (partly because spills are rarer than ever; see Kim 2002). More recently, attention has shifted within the Department of Homeland Security to *capabilities based planning* (DHS No date, Caudle 2005).

Establishing a baseline context to measure the effectiveness of a response is extremely challenging. The fact that incident-specific strategies must be identified early and on a case-by-case basis increases the complexity and challenge faced by spill managers (Grabowski et al. 1997). While there may be general agreement about the over-arching goals for spill response, the objectives that define these goals in specific incidents may vary and their relative priority may vary – from spill to spill and among those with a stake in the spill response (e.g., Abordaif et al. 1995, Lindstedt-Siva 1999). Consequently, setting objectives, tracking progress and communicating or determining success can be an *ad hoc* process depending upon the experience of the On-Scene Coordinator and the level of interaction with state, local or other non-federal government groups outside the command structure, including the media. Even in the case that the response is closely coordinated among agencies and planning documents are scrupulously adhered to, public perceptions may be that the response has failed – partly because it is not apparent what normative standards of success should be applied or how the measures of success employed by decision-makers will be interpreted by the public or intermediaries (such as journalists or non-government organizations; see Chess et al. 2005, Lindstedt-Siva 1999, Harrald 1994).

Because of the complexity of spill response and the significant variation that can occur among spills and responses, the question of how best to assess response, both in general and in

1. http://response.restoration.noaa.gov/book_shelf/910_response.pdf, pg. 9. See also www.uscg.mil/d1/staff/m/rrt/

regard to specific spills, is not easily resolved. After a spill event occurs and the spill response is completed (or largely complete), there are several reasons why they may be evaluated, including to:

- facilitate organizational learning to inform future spill response efforts,
- determine adequacy of area contingency plans.
- provide information for legal disputes about responsibilities for impacts from the spill and response activities,
- help determine responsible party liabilities,

In addition, it can be useful to select criteria on which to judge the performance of spill response during contingency planning; in other words, prior to a spill event to inform preparedness planning. The advantages include:

- clarifying objectives that will further inform area contingency planning,
- facilitating better communications with interested and affected parties (e.g., local residents),
- establishing expectations and help to align objectives and perspectives,
- input into institutional planning, such as GPRA and capabilities-based planning requirements,
- informing organizational planning to develop procedures and capabilities (e.g., obtain and position appropriate equipment, improve training), and
- establishing procedures to gather relevant data, including baseline data and during a response that will inform both response-related decision making and post-response assessments and evaluations (e.g., natural resource damage assessment).

In general, learning can be facilitated by assessment that is systematically integrated into the overall response planning and response management system (Lindstedt-Siva 1999). Over the last decade various approaches have been proposed to assess the quality of contingency *plans*, using expert input about the appropriate criteria and measures to use (Haynes and Ott No date, Abordaif et al. 1995, Harrald and Mazzuchi 1993). However, the question of how best to assess response successes and shortcomings has not received the same level of systematic attention among planners, although some frameworks have been proposed (Kuchin and Hereth 1999, Lindstedt-Siva 1999).

The goal of this report is to propose a process design that will facilitate the selection of metrics for assessing spill responses by a broad range of potentially affected and interested parties. Our concern here is with the process by which those who have a stake in spill response can collectively discuss:

- the objectives that should guide spill response and
- the criteria or metrics that should be used to assess performance with respect to the objectives.

First, we discuss the challenges of defining clear objectives and performance metrics for spill response. Second, we provide a brief overview of the use of performance metrics for assessing oil spill response and models for how that can be done. Then, we outline a process that can be implemented to guide stakeholders through the activity of clarifying objectives and performance metrics that can be used to a) guide area contingency planning and b) assess performance of response efforts. As part of this discussion, we suggest how the proposed process can be integrated into current oil spill response planning requirements and approaches, such as national response plans, area contingency planning, and ecological risk assessments.

## 2. Challenges to defining objectives, metrics, and measures

Often a distinction is made between goals, objectives, and metrics. Goals are over-arching state of affairs that are desired. In regard to oil spill response these are protection of human life and mitigation of adverse ecological and socio-economic impacts. Objectives define the ways that goals are achieved, by identifying the priorities associated with each goal. Metrics are specific and measurable ways of assessing whether objectives are being achieved. Measures are the specific values of the metrics in a given situation.[2]

As an example, consider the goal of mitigating ecological impacts from an oil spill. This goal can be made more concrete by defining a set of associated objectives. These might include:

- ensure that areas or species pre-identified as sensitive are quickly protected,
- protect the adults of a species at risk,
- protect areas that have multiple resource values, like those that are undeveloped, pristine, and that provide for recreation.
- protect species that are especially critical for the functioning of an impacted ecosystem,
- ensure that impacts from clean-up activities are minimized, and
- remove enough oil so that impacted species, habitats, and local communities can return to the way they were before the spill in a reasonable amount of time.

Each of these objectives may be measured in a variety of ways. For example, whether or not a sensitive species has been protected may be measured by:

- the number of individuals found dead or injured,
- the number of recovered and rehabilitated individuals,
- the number of impacted individuals that were capable of reproducing,
- the degree to which their habitat has been oiled, and
- whether or not the next breeding season is successful.

While it is possible to develop long lists of possible objectives and metrics, there are a number of challenges to defining objectives and using metrics for assessing oil spill response.

The first challenge involves defining the objectives. Spill response is organized to achieve the three over-arching goals as defined by OPA 90, etc. However, what it means to achieve these goals in particular situations can vary.

In order to achieve these goals, a number of participants are required to take part in a response. They may include federal, state, and local officials (i.e., USCG, EPA, state environmental protection agencies, fire chiefs, harbormasters); the Responsible Party (RP) and its contractors; non-local clean-up crews hired by private contractors; environmental and community advocates at the national, state, and local levels; and community residents who have a vested interest in the response (i.e., business- and homeowners, beach associations). However, because so many different and interested parties are affected, there is the potential for conflicting ideas about how clean-up goals should be achieved and how a response should be organized and implemented.

[2] Others in the spill response community have used a different set of terms to describe similar concepts, including key business drivers, critical success factors, key elements, and performance indicators (e.g., Haynes and Ott No date, Kuchin and Hereth 1999). Despite some terminological differences (that will be discussed further below), the challenges of defining, selecting, and measuring them are similar.

Differences of opinion may arise over the broadly applied response goals in specific contexts, what objectives are associated with each goal, and the relative weightings given to different objectives (e.g., Abordaif et al. 1995, Lindstedt-Siva 1999). For example, response strategies are based on the characteristics of the oil released, the location and timing of the release, and the kinds of resources in harms way. Objectives may vary depending on the specifics of a spill event. For example, seasonal changes in tourism and local economies (e.g., shell-fishing) may affect the relative priority of preventing oiling of some areas because there would be time to clean oil off the area before tourists return. The timing of a spill (and response) may also affect how important will be efforts to protect sensitive species – some of which may not be in the area (i.e., migration) or may be nesting. Similarly, the type of oil spilled may affect the relative importance of some objectives, such as the need to protect cultural resources or sensitive areas.

In addition, there may be varying preferences for some objectives, even within the same context. Some people may care more about mitigating impacts to local fishermen, which would highlight the need to protect some species more than others. In short, one individual's goal(s) may create tension with another. In addition, members of the public and local stakeholders can have differing opinions from experts about when and to what extent clean-up goals should be met, mistrust the decisions made about how to meet these goals, and question why particular goals were set.

As a result of these various goals and perspectives, choices about clean-up priorities may become a political and/or social decision based on competing values. As one federal official noted, termination endpoints depend "on whose ox is being gored" (from Tuler et al. 2006). For instance, will shellfish beds or frequented tourist beaches have more stringent endpoints than other less valued resources or beaches? Will limited response resources be devoted to saving a much used recreational area or a sensitive marsh habitat even if the endpoints are unattainable or the law of diminishing returns applies?

In addition to the potential conflict about specific objectives in any particular situation, there are also challenges in determining how to assess whether objectives are achieved in a spill response. This is the issue that requires defining metrics.

The first challenge associated with selecting metrics for specific objectives is that there are an enormous number of potential metrics that can be applied to assessing the response effort. It is not realistic to expect that all possible metrics should be used. Choices must be made about which metrics are most relevant and meaningful. But, relevance is not easily determined. Just because something can be measured does not mean it is relevant to understanding the success of a spill response or important to many stakeholders (Abordaif et al. 1995). For example, the amount of boom deployed in a spill response can be used as a metric to assess how well a shoreline was protected from oil contamination. However, while it is easy to measure, it may not be a good indicator of whether the shoreline is actually impacted. Deployed boom may not always effective when currents or winds are strong. Furthermore, deploying booms may be a good decision at the time it was made, but weather conditions can shift and cause them to fail. That is, the metric does not have a causal relationship between the state of the system and the variables that are under a decision-maker's control. Similarly, the amount of money spent – for numbers of crew, manhours, and amount of boom deployed – may be very appealing politically as a measure of performance, but it may not have a direct relationship with the quality of a response in a particular situation. As a third example, the amount of oil removed from beaches might be a relevant metric for assessing the response. In one sense this value might be easy to measure. But, the measurement may not meaningful: it is very difficult to quantify the actual amount of oil on the beach because it is mixed with sand and rocks and in sorbent material. Finally, bird mortality and rehabilitation rates as an often used metric, but one that might not always be that meaningful. For example, the number of impacted swans is not a good measure because they are not native species. A more relevant issue is the *effort* that was made to recover all the wildlife, not the rehabilitation of those that have greater importance in the public's eyes.

A second challenge associated with selecting performance metrics is that they are often indirect measures of something that the people really care about. They are not direct measures of, for example, economic impacts, public satisfaction and support, public and worker health, or ecological impacts. These challenges are certainly not unique to the assessment of oil spill response. For example, problems associated with complex, dynamic systems for which longterm outcomes are of interest but very difficult to assess have been discussed in the context of international conflict resolution (National Research Council 2000a) and longterm stewardship (National Research Council 2000b):

> We have noted that evaluation is complicated by the fact that short-and long-term definitions of success may be quite different. This difficulty can be addressed in part by focusing on particular outcomes rather than overall "success." Evaluations can be separated according to time horizon, with outcomes at different times analyzed separately. As noted, it is important to have short-term indicators of progress even for interventions intended mainly to have long-term effects. This is so partly to provide interim indications of progress and also to allow for meaningful evaluation even in cases in which intervening events not brought about by the intervention throw the process of conflict resolution off its intended course (National Research Council 2000a, pg. 63).

A third, and related, challenge suggested by the quote from the NRC report is that metrics may assess intermediate features of a system, such the state of an ecological system, habitat, or populations, rather than on the actual end-points desired. For example:

- How many pounds of debris, oil contaminated debris, has been recovered and sent off for disposal?
- Number of miles of shoreline impacted depends on the degree that its impacted, the type of shoreline, the environmental sensitivity of that shoreline, and the socio-use of that shoreline
- How many birds were oiled, how many recovered, how many released, and how many released and survived?
- How many acres of shellfish areas were closed and then re-opened?

A final challenge is that thresholds that separate 'good' versus 'bad' may not be easily defined. For example, an endpoint for marshes could be no sheen or no oil available when touched. However, oil coating may be left on peat or leaves because to address that might result in salt marsh destruction or further injury. Thus, the question of *how much?* or *how clean is clean?* is not so easy to define.

In addition to the challenges of defining objectives and metrics, complexities of measurement are also important. For example, what is most relevant may also be very difficult to measure. Many of the performance metrics that could be easy to communicate to a wide audience, relate to something that is important to many stakeholders, changed by human intervention or would be relevant to human action, credible, and relevant are not easy to measure. Measures that might at first glance seem easy to measure may not be, such as number of lost fishing days: are fishing days "lost" due to impacts of the oil spill or because of foul weather or other reasons?

Similarly, when counting injured or dead wildlife how can natural mortality be differentiated from mortality or injury due to a spill? The problem can be exacerbated when if there are significant uncertainties associated with their measurement. Again, these challenges are certainly not unique to the assessment of oil spill response. Problems associated with measurement and the lack of baseline data are widely documented.

In addition, there are several other challenges associated with obtaining accurate measures of metrics for spill response. They are:

- the reliability of data recording,
- there is little accurate baseline data on which to base comparisons or assessments,
- inability to quantify events or changes can leave the meaning of measurements open to dispute, and
- measurements may vary depending on *when* they are made.

## 3. Approaches to assessing oil spill response performance

Many federal agencies use performance measures to assess their progress toward achieving ecological or environmental goals. The primary regulatory driver of performance measurement has been the Government Performance and Results Act of 1993.[3] Different agencies are in differing stages of performance metric adoption. The development of performance metrics is not always straightforward. Problems commonly arise when attempting to measure the performance of programs that:

- have outcomes that are extremely difficult to measure,
- have many contributors to a desired outcome,
- have results that will not be achieved for many years,
- are characterized by causal relationships or feedbacks are not well understood,
- relate to inherently uncertain or stochastic systems,
- operate at multiple temporal and spatial scales,
- relate to deterrence or prevention of specific behaviors,
- have multiple purposes and funding that can be used for a range of activities, and
- are administrative or process oriented, relating to bureaucratic effectiveness, rather than outcomes-based (OMB, 2005c).

While many federal agencies have developed ecological performance metrics, few specifically address oil spill *response* or *impacts*. For example, the USEPA Strategic Plan 2003-2008 performance measures related to oil spills are predicated on the number of spill responses and the percentage of oil storage facilities inspected, rather than the severity of the spills or damage that results (EPA 2003). The US Department of Transportation has targeted a 20% reduction in the volume of oil spilled by maritime shipping sources and by piped sources between the years 2001 and 2006 (DOT 2002). While a laudable goal, this performance metric is unaffected by spill response and is independent of the ecological damage caused. There have been efforts to develop models for assessing the quality of area contingency plans for spill response (Abordaif et al. 1995, Haynes and Ott No date, Kuchin and Hereth 1999). A criticism of area contingency planning for oil spills is that they have been evaluated with respect to how well they conform to a

[3] GPRA requires federal agencies to prepare performance reports that are then reviewed by the Office of Management and Budget (OMB). GPRA was enacted to "provide for the establishment of strategic planning and performance measurement in the Federal Government" (OMB, 2005a). It embodied a push for better planning, greater accountability, and straightforward performance evaluation in government by requiring a federal program to have an overall strategic plan and to prepare annual performance plans and reports (OMB, 2005b).

particular format, rather than their capacity to ensure that desired outcomes are achived or tasks accomplished (Haynes and Ott No date).

Kuchin and Hereth (1999; see also Ornitz and Champ 2002, pgs. 58-61 and Appendix 3) “found no comprehensive system, agreed upon by the response community, that systematically evaluates the success of the response effort.” Furthermore, they note that “the historical focus has typically been on measuring activities such as: speed in responding; feet of boom deployed; and gallons spilled and recovered rather than the actual impact of those activities. While traditional metrics are important matters in the response, they are largely reflective of processes and activities being carried out in the response and do not always directly relate to the overall outcomes”. Instead, they believe that the intent should be to “measure outcomes that directly relate to minimizing consequences to people, the environment, property, and the economy. Ideally, we want specific information that will relate to the value provided by out response efforts (i.e., through reduced consequences).”

A variety of models or frameworks for evaluating oil spill contingency plans and response efforts have been proposed (Haynes and Ott No date, Abordaif et al. 1995, Harrald and Mazzuchi 1993, Kuchin and Hereth 1999, Ornitz and Champ 2002). All of them are based on the concepts of best response and critical success factors. According to Haynes and Ott (No date) “the term best response means that the oil spill response organization will effectively, efficiently, and safely respond to oil spills or hazardous material releases and, having done so, that the response will be perceived as a success.” Critical success factors are things “that must go well or be done right in order for the key business driver to be protected or receive some benefit” (Kuchin and Hereth 1999).

For example, Abordaif et al. (1995) proposed a process for eliciting expert input about critical success factors for oil spill response. These factors were grouped according to whether they were under the control of the response organization, not under the control of the response organization, or specific to the spill scenario that should be accounted for by the contingency plan. Experts assigned weights to critical success factors, and existing plans were compared to a “standard plan” defined by the experts’ weightings using the analytic hierarchy process. They conclude their analysis by stating that “the fundamental conclusions of this project support the premise that the elements of a contingency plan are of very unequal importance. Unfortunately, the easiest things to provide and document are those typically evaluated by the experts as the least important, and the important elements are difficult to thoroughly develop. An emphasis by reviewers [of area contingency plans] will therefore lead to quantity, not quality.”

Haynes and Ott (No date, Ott 2005) have proposed a method for evaluating area contingency plans using critical success factors that uses a scorecard approach: “the task of measuring an area contingency plan is to assess how well it enables the response community to achieve this success.” The propose a set of critical success factors that must be accomplished to achieve key business drivers, which include public health, environmental protection, public communications, economic well-being, and stakeholder support. Key business drivers are the activities that should be achieved to ensure that a *best response* is accomplished. The critical success factors they propose are based on the same set of expert input described by Abordaif et al. (1995). For example, the key business driver of environmental protection can be measured by the critical success factors of: source of discharge minimized, source contained, sensitive areas protected, and resource damage minimized. A scorecard is used to indicate the degree to which various critical elements may be achieved by following the contingency plan.

Kuchin and Hereth (1999) present an evaluation model for assessing spill responses, as opposed to contingency plans. Their model is tightly coupled to the prevailing best response model and National Incident Management System that defines the US approach to oil spill response planning (Ornitz and Champ 2002). It includes several criteria and is also based on assessments of key business drivers, which are understood to be critical to obtaining a successful outcome. They propose six key business drivers that include operational outcomes (#1-3), customer service outcomes (#4-5), and organizational outcomes (#6), as shown in Table 1.

For each key business driver, critical success factors were developed. Examples of critical success factors for each key business driver are provided in Table 1 (from Kuchin and Hereth 1999). The critical success factors for oil spill response were developed based on the ability to measure them and the ability to discriminate between good and bad outcomes. Input to their selection came from: participant experience, ICS implementation experience, incident specific pollution reports (ISPRs), lessons learned database, job-task analysis, and response management job aids.

**Table 1. Key business drivers and critical success factors for oil spill response**

| Key business drivers | Example critical success factors |
|---|---|
| human health and safety | No spill related public injuries or death |
| natural environment | Sensitive areas protected |
| economic impact | Spill effectively contained/controlled |
| public communication, | Positive media coverage of response |
| stakeholder service and support | Positive meetings with stakeholders |
| quality and effectiveness of the response organization. | Clarity in leadership and responsibility at all levels |

Measurement of the critical success factors was to be obtained though the use of surveys to capture qualitative assessments of those involved in a response. After each criterion is evaluated, they are integrated as part of a “balanced response scorecard” for measuring the success of spill response.

In comparison to Kuchin and Hereth (1999), a more recent model being developed by NOAAs Office of Response and Restoration (ORR) segregates the importance of an effective response organization as an administrative rather than outcomes-based metric, and it lumps “communication” and “stakeholder service and support” into a single, broader category. The ORR approach also recognizes the importance of “leading indicator” metrics (such as hours worked) that are indicative of progress toward the eventual goal (also see Kuchin and Hereth 1999). Although they are not direct measures of the goal itself, such as safeguarding human or ecological health, the leading indicator metrics can be used to gauge the intensity of efforts or forecast potential outcomes. ORR identifies four “key elements” to a successful response: objectives, organization, resources and mobilization (see Haynes and Ott No date).

These approaches are informing spill response planning in federal agencies. However, criticisms of them have also been raised. For example, Ornitz and Champ (2002) report that some stakeholders are concerned about the degree to which “success” may be dependent on public opinion, at the expense of more scientifically-based and technical criteria. On the other hand, the importance of considering a broad range of perspectives about objectives for a response and assessments of the quality of a response is often highlighted (e.g., see our case study report Tuler et al. 2006, Tuler et al. Forthcoming, Baker 1999). For example, there may be multiple definitions of "clean" (Baker 1999). Even in this more limited discussion (in contrast to assessing the *entire* response effort) "decisions regarding definitions of clean and assessment of net environmental benefit for clean-up methods need to be reached through consensus, involving all interested parties during the contingency planning process. Decisions need to be well-informed, and there is a continuing need for education, better communication, and constructive involvement of the media" (Baker 1999, pg. 144).

Ornitz and Champ (2002) elaborate on this point; in their discussion of the Kuchin and Hereth model they note how that model is intended to promote:

a) taking into account the interests of the public in the management of a spill response and
b) consensus building about what defines "success."

In earlier work by John Harrald (Abordaif et al. 1995, Harrald and Conway 1981, Harrald and Mazzuchi 1993) that informs these approaches by Kuchin, Hereth, Ott, and Haynes there was also a recognition that the *process* also matters:

> Contingency planning is a process, not a product. The true and lasting value of the planning process is the understanding of the problem domain gained through the planning process and the relationships established in resolving interorganizational issues" (Abordaif et al. 1995).

However, fulfilling the intent of their model, as well as subsequent efforts, is no easy task. There may be competing perspectives that are not easily integrated (Lindstedt-Siva 1999). There may also not even be agreement about *who* should participate and how.

The goal of this report is to propose a process design that will facilitate the selection of metrics for assessing spill responses by a broad range of potentially affected and interested parties. To achieve this goal we build on the insights of those in the spill response community as well as lessons learned from hazard management in a variety of domains that rest on the idea that diverse interested and affected parties -- people with a stake in the process and outcomes – should be involved in planning and decision-making that involves complex (and uncertain) scientific information. How to best do this is a complex challenge. Our proposed process is informed by our own experience and a vast– and growing – literature on these issues that touch on a wide range of policy arenas (NRC 1996, Beierle and Cayford 2002, Creighton 2005, Frewer and Rowe 2005, Rowe and Frewer 2002, Kasperson et al. 1992), including natural resource management (Stevens and Walker 2001, Leach et al. 2002), clean-up of contaminated areas (Bradbury et al. 2003, Ashford and Rest 1999, Carnes et al. 1998), disaster planning (Burby 2003, Wood et al. 2002, Godschalk et al. 2003), and public health (Henry S. Cole Associates 1996). In the following sections we turn to the principles that inform our process design and a description of the proposed process.

## 4. Principles for process design

The complexities associated with oil spill response overall, and with defining clear, and widely accepted objectives and metrics in particular, have implications for how stakeholders can best work through a process that will lead to meaningful input for contingency planning and organizational learning.

The process we propose is guided by two overarching principles: fairness and competence, described in more detail in the accompanying text boxes). Fairness and competence ensure that the right people are participating and that they use the best available information to reach agreement or decisions (Webler 1995, Webler and Tuler 2000).

Every decision-making process needs to be fair and competent. Efficiency is another principle all processes should aspire to attain. But efficiency is a quality that the process achieves while in pursuit of other objectives. What is important is that the process be fair and competent. Obviously, it is better when these objectives are met efficiently, but efficiency is not a principle on par with fairness or competence. In other words, no one would trade fairness or competence for efficiency. We learn to make a process more efficient by repetition, evaluation, reflection, and innovation.

The principles of fairness and competence are ensured through several structural features upon which the process rests.

The first feature is <u>incremental progression</u>. As the process moves along in a step-wise manner, it involves citizens, stakeholders, regulators, industry, and experts – or what we call the potentially interested and affected parties. In any single step, the process can draw upon any of numerous techniques and participants in order to achieve its purposes. Incremental progression promotes both competence and fairness.

The second feature of the process is adaptability. This means that while the process is on-going all decisions should be viewed as preliminary. We prefer to think of this in terms of designing a <u>learning</u> process. Participants in the process should continuously learn from new information, other participants, and past experiences. However, this learning and experimentation happens within the context of adhering to the central guiding principles, seeking to achieve the goals of fairness and competence.

The third feature derives from the idea that the process should be adaptable and promote learning. The third feature is <u>iteration</u>. Although the process is defined in terms of steps, the process will cycles through these steps as needs arise. Furthermore, depending on the situation at hand, the sequence of steps may be adjusted.

When the process is viewed as adaptive and iterative, then decisions and recommendations made early in the process need to be considered as very rough drafts, open to revision and clearer definition. But to preserve a sense of progression along the process, such products need to gather more and more inertia. In other words, participants should converge on a more and more stable set of decisions and recommendations – in this case on definitions of spill response objectives, performance metrics, and measures.

The fourth feature is the <u>integration of analysis and deliberation</u>. Doing policy making well means finding the right combination and interplay of analysis and deliberation at each and every step of the policymaking process (NRC 1996, Stern and Dietz 1996, Gregory and Failing 2002, Apostolakis and Pickett 1998); the challenge and importance of doing so is highlighted by the kinds of criticisms leveled at the model Kuchin and Hereth propose for assessing spill response success (1999; see also Lindstedt-Siva 1999 and Ornitz ands Champ 2002, pg. 60). Seeing a policy making process as a combination of analysis and deliberation is an important advancement, mainly because it corrects shortcomings with the traditional view that policy making is an uneasy combination of science and politics.[4] Following the NRC's report we use analysis and deliberation to mean specific things ways of knowing or generating knowledge. *Analysis* means the use of systematic, rigorous, and replicable methods to formulate and evaluate knowledge claims. Methods include the natural, social, or decision sciences, mathematics, logic, and law. *Deliberation* is any formal or informal process for communication and collective consideration of issues. In deliberation, participants ponder, exchange observations and opinions, reflect upon information and judgments, and practice persuasion.

Analysis and deliberation are both understood as ways of making sense out of the world. In analysis people use systematic ways of gathering and interpreting data. The overarching principle of analysis is that results can be validated through systematic studies. Note that lay people as well as scientists can do analysis. For instance, lay monitoring teams collect data on water quality, compile it, and plot trends in the data.

Deliberation is a different way that people make sense of the world. Here people "confer, ponder, exchange views, consider evidence, reflect on matters of mutual interest, negotiate, and attempt to persuade each other" (NRC 1996, pg. 73). People do not deliberate only over values. The courtroom is a clear example of where facts are contested. As with analysis, deliberation is *not* performed only by decision makers and interested and affected parties. In scientific circles, deliberation takes place about how to best do analysis. Scientists also deliberate with publics, with decision makers, with stakeholders. Deliberation refers to all contexts of discussion that occur in the policy process.

Technical analysis does not happen in a vacuum. It is not a value-free activity. Values clearly inform how analyses are done, who does them, and when they are done. Thus, analysis needs to be integrated with a process in which judgments are made. Such a process clearly involves people talking with each other and pondering the situation, hence the word *deliberation*. Clearly there are different kinds of judgments to make. Some require input from all interested and affected parties. Others are best made by limited groups of certain individuals. Consequently, deliberation is *not* synonymous with open public participation. It includes that, certainly, but also describes processes with many fewer participants.

The selection of metrics to assess spill response performance requires knowledge of complex systems about which there is uncertain information. In addition, the selection of performance metrics should factor into consideration human behavior. Spills and responses to them have multi-faceted consequences for human individuals, groups, and communities that may be displaced across time and space. As such, their assessment involves making complex value-laden judgments under conditions of imperfect knowledge in the context of a democratic, highly litigious society. Technical knowledge is of obvious importance. But, because of the uncertainties involved, dialogue and collaboration among all knowledgeable, interested, and affected parties is also essential.

[4] The main shortcoming of this view is that it fails to acknowledge that the two are really highly interconnected and it advances the misconception that science is objective and free of human values or interests.

*Fairness*
Fairness includes two basic components – inclusivity and opportunity. Inclusivity has to do with who is invited or allowed to be part of the process. Opportunity concerns what people are allowed to do as part of the process. Clearly these two questions cannot be completely separated from each other. The question of who to involve is closely linked to how to involve them. Once the affected population is identified, there are several different ways to go about achieving inclusivity. One option is to make the process open to anyone who wants to participate. This approach relies on the principle of *self-selection*. Self-selection is legitimate if everyone in the potential population was made aware of the reason for the process, the opportunity to participate, and the stakes involved. This requires an outreach effort to inform the identified population. Another approach is to seek *representativeness*. We distinguish among three styles of this approach. The first we call *satisficing with efficiency.* In this style, organizers of the process pull together a group that can work together well. This approach works well if the community has a history of successful collaborative decision-making and the people who have been involved are widely seen as representative of the population's interests. The second we call *satisficing with diversity.* In this style, the focus is on bringing as many different perspectives to the table as possible. It is a way of respecting every perspective in the community, no matter how small it may be. This approach can be effective for re-building trust, as it works to equalize power relations, but it is not necessarily at all efficient. The third style for achieving representativeness is to use *elections*. This approach has rarely been use, because it is highly inefficient, however, when done correctly, it may be the best way to acquire legitimacy. Any of the above-mentioned approaches can be used to establish inclusivity.

Deciding upon opportunities for those who participate is the second major challenge in building a fair process. Opportunities do not necessarily need to be equal or identical. Doing so may not be logistically feasible, or it may be highly inefficient. It also may not be what the participants or the population want. Opportunity is identical to power or the ability to influence a decision. A fair process does not merely replicate existing power relations in society, it develops a structure for the exercise of power in a manner that respects both the need for fairness and competence.

Although competence is discussed below, we recognize here that there is an unresolved tension about the role of power in producing competent outcomes. Some argue that putting power be in the hands of the experts may produce better decisions, but this will certainly be seen as unfair among many. Others point to the wisdom of "local knowledge" and argue that non-experts use commonsense reasoning that can produce more competent outcomes. Balancing the influence of knowledge on decision-making means striking a balance between how power is realized in the process.

We identify four activities that must be undertaken to ensure fair opportunity. First, all participants need to have opportunities to influence the *definition of the problem*. How the problem is defined is directly linked to the items on the agenda, the timetable of the process, and the desired goals. It specifies what brands of expert knowledge are needed and is also associated with defining the relevant population. Second, all participants also need to be able to participate in *every discussion* on the agenda. Third, for a process to be fair, it must provide opportunities for all participants to *influence decisions* made as part of the process. Without this quality, participation runs the risk of being symbolic or superficial. Fourth, an important consideration is *providing multiple means for people to participate.* Experience has shown that, when processes offer many different ways of participating, more people participate. This is simply because people have different preferences. Some like smaller one-on-one meetings, while others are more comfortable in a large meeting. Some enjoy interacting on-line, while others want to see people face-to-face.

*Competence*
Competence is an idea that has relevance in a procedural sense and in a substantive sense. *Procedural competence* has to do with how the process is carried out – how information is distributed, how discussion and learning happens, how closure is reached, and so on. *Substantive competence* has to do with the actual problem and the actions taken to remedy it. We argue that a process that is procedurally competent will also produce better decisions (e.g. be substantively competent). Thus, below, we speak only generally of competence.

Competence is promoted by having participants be committed to the process. This means that, when agencies appoint individuals, those people should remain involved for the life of the process and not replaced with other appointees. Transparency of information and decision-making is a second major quality associated with a competent process. People need to be informed in order to make the best decisions possible and they need to understand the opportunities to act in the process. Having accurate, or high quality information is obviously critical to competence. Finally, a competent process requires adequate resources be made available to the participants. They need resources to obtain or interpret information. They also need resources to conduct the process itself (arrange meetings, print minutes, summarize findings, etc.). Competence is not about maximizing the amount of information, but acquiring satisfactory information. A parallel to Herbert Simon's idea of satisficing is appropriate here. Simon argued that decision makers do not attempt to optimize decisions, they satisfice. That is, rather that thoroughly evaluating every decision option on every decision criterion, they use rules of thumb to reduce options until they come to a decision that is "satisfactory enough." Our view of competence is similar to this. Efficiency enters when we contemplate which solution is satisfactory "enough."

Another important aspect of competence has to do with actual decision-making, in particular, how uncertainty is handled. One of the most challenging decisions to make is what to do when there are large uncertainties about what the outcomes of the decision options are. The most competent way to move forward through a time of uncertainty is with an incremental and iterative approach. Like walking through a dense fog, each step should be tentative and short. One must always leave open the opportunity to retreat and advance again in another direction. Thus, elements such as flexibility, adaptability, iterativeness, incrementalness are all aspects of a competent process. Incrementalness can be seen as antithetical to efficiency. It may seem more appropriate to push forward through periods of uncertainty in the interest of reaching an endpoint expediently. However, as any driver knows, driving headlong into a dense fog can result in a collision, which will impede or prevent you from reaching your goal. Moving slowly may seem less efficient, but it may well end up being the more efficient path. It is very difficult to predict which path will be the most efficient.

## 5. Process design for selecting objectives and metrics to assess spill response

The process we propose is to be conducted in three phases that combine six steps. As we discussed in the previous section these steps can be iterative, although it is more simple to describe them in a generally sequential series. The set of six steps are shown in Table 2.

An iterative, analytic-deliberative process has the potential to take a large amount of resources and time to complete. Therefore, it is very important to consider ways that a process to clarify objectives and performance metrics can be conducted in a reasonable amount of time with reasonable demands on agency staff and resources. We have proposed a process design that can be conducted as part of existing contingency planning and spill monitoring activities in a region.[5] Ideally, it would be integrated into existing planning processes, such as ecological risk assessments, capabilities-based planning (as promoted by DHS), or existing stakeholder involvement processes, such as the Prince Williams Sound Regional Citizens' Advisory Council (www.pwsrcac.org). Ecological risk assessments and the stakeholder involvement processes already exemplify the kind of broad involvement of interested and affected parties that are envisioned as part of the proposed process.

**Table 2: Steps of the analytic-deliberative process for selecting objectives and performance metrics**

*Phase 1. Preparatory activities*

- Step 1: Problem formulation
- Step 2: Process design -- the participants, the agenda, and defining opportunities and means to participate.
- Step 3. Gather information about oil spill response scenarios, including preferences for objectives, performance metrics, and measures.

*Phase 2. Deliberate about options*

- Step 4: Assess objectives, performance metrics, and measures.

*Phase 3: Reach closure and evaluate outcomes*

- Step 5: Select objectives, performance metrics, and measures and attempt to reach closure.
- Step 6: Evaluate performance and gauge the effects of the actions.

In the following sections we elaborate in some detail the six steps of the proposed process. We also highlight illustrative points at which iteration may be desirable.

**<u>Phase 1. Preparatory activities</u>**

The very first action is for the team responsible for conducting the process to identify themselves in an unambiguous manner. This membership of this group can expand or change, but clear identification ensures that communication about the project is accountable. We will call this group "the organizing team." In oil spill response planning the organizing team may consist of staff from relevant federal agencies (e.g., NOAA, Coast Guard), but it may also include contractors hired to design and implement the process.

[5] The process we propose elaborates the first step of the proposed response performance assessment process proposed by Lindstedt-Siva (1999); the first step of her process is to identify stakeholder performance criteria.

The organizing team is responsible, at this first step, for drafting preliminary definitions of the problem, proposing a design for the process, and gathering initial information.

*Step 1: Problem formulation*
The problem definition specifies the origin of the motivation for the process as well as the desired end state. The problem statement should also be clear about the desired outcome of the process: agreement about a set of objectives, performance metrics, and sources of measures (i.e., data) for assessing oil spill responses. If the goal of the process includes integrating this information into existing contingency planning or for developing outreach and educational activities, then these should also be specified.

In the case of oil spill response planning, desired end states may take the form of a general statement about the goals of oil spill response. They may also be defined in the context of a set of scenarios around which oil spill response planning is based, which is the approach taken in ecological risk assessments (Pond et al. 2000, Aurand 2003).

Analysis should aid the activity of problem formulation by characterizing key challenges for oil spill response in the area under consideration: types and sources of oil that may be spilled, shoreline characteristics, presence of sensitive species and habitats, important economic features, potential seasonal variations, etc. These data help inform assumptions and preconceptions. Deliberation gives people from different interest positions an opportunity to posit their definitions of the problem (e.g., relevant scenarios) in a space which guarantees they are heard. Meaningful involvement of interested and affected parties early on in a process is a widely recognized principle of good public participation (Bradbury et al. 2003, Kasperson 1986, Creighton 2005, DOE 2003, EPA 2003).

The organizing team has the responsibility to ensure that information and analyses are brought into these deliberations. It schedules presentations by different groups, for example. It should also pay attention to the discourse space it creates -- for example: location of meetings, time of meetings, physical setting of the room, groundrules for discussion, whether or not a facilitator is used.

The importance of approaching the initial problem definition as being preliminary cannot be overemphasized. Any definition of a problem is loaded with normative judgments, uncertainties, and understandings, which may be appropriate or inappropriate for the given context. All these aspects need to be vetted through a process that engages many types of involved people in a discussion of how best to define the problem. For example, assumptions may be made in a problem definition about the most appropriate scenarios to guide discussions. There may not be universal agreement about these. Furthermore, there may not be consensus that the outcomes of this process should be input into existing contingency plans. These issues should be made explicit and agreement reached about them before the process begins. Just how the problem formulation is officially changed is a topic for the process design step.

*Step 2: Process design -- the participants, the agenda, and defining opportunities and means to participate*
Who participates and in what role or capacity, are the questions tackled during the process design step. In pursuing these objectives, the funding organization and the convening organization adopt a view of process design characterized by the following principles:

- produce a clear and explicit design, stipulating how all relevant parties are to be included,
- document the design and make it available for all to consult,

- explicitly note where policy judgments need to be made and clearly specify how interested and affected parties can participate in making these judgments, and
- build-in opportunities for iteration and explain how decisions to revisit previous judgments will be made.

With a tentative problem definition in hand, the organizing team can move on to identify a potential population of interest. In a democratic society, anyone with an interest in a collective decision or who is affected by a collective decision ought to have a say in making that decision. However, what amounts to “an interest and affected party” can be highly subjective. To some extent this matter is unambiguous. People whose health, way of life, or economic welfare is affected are clearly interested parties. The family whose property may be contaminated with oil is clearly affected. The same can be said for people whose economic interest is linked to the perception of negative stigma associated with having a marine oil spill in their community. Communities neighboring the affected community may also be interested. To some extent, interest is a matter of self-determination.

We suggest that the organizing team prepare a preliminary list of the potentially interested and affected population. This should begin with a list of all the key interest groups with a definite interest in the problem. They may be active already or they may need to be motivated. Around this core group can be defined a secondary group that is possibly interested. These should be groups that are reasonably likely to participate. A third circle of possibly interested parties should also be drafted. This population analysis report should be summarized and used as the basis for making first contact.

Before beginning the participatory aspects of the process, there is one more step to take. This is to define a preliminary agenda. At this stage the agenda is merely a list of things to do, not a specific outline for a given meeting. The list may distinguish between information gathering, analytic activities (e.g., characterize what is known about sensitive areas and bird populations in the area), and deliberative activities (e.g., decide on performance metrics for cleanup). The purpose here is begin to link the definition of the problem to the process design.

With the preliminary problem definition, population analysis, and draft agenda in hand, the organizing team now contacts the groups and individuals listed in the core group. They explain the steps taken to date, the principles of the process, and invite their involvement in the next step. They may then expand their effort by contacting people in the secondary and tertiary groups.

A critical component of successful collaborative and participatory planning processes is the creation of a “safe” setting for dialogue. In part this requires that participants are “co-convenors” of the process. They need to think of it as “their process.” It is the responsibility of the facilitators, however, to ensure that a respectful, open, collaborative setting for dialogue is established and maintained. One means of achieving this outcome is to pay close attention to the structure of deliberations. For example, facilitators can have participants agree to basic groundrules. This model has been implemented by some organizations in controversial environmental policy arenas and proven effective for establishing on-going, constructive dialogue. Expectations about iteration, learning, and incrementalism should be established among all of the participants.

With the preliminary decision made about who to include in the process, it is now necessary to define what they will do. There are two aspects to the question of how people may best contribute to the process: 1) how they will participate and 2) what information and experience they will bring to inform deliberations and decisions.

*How they will participate?*
The organizing team should sketch out activities that the participants can undertake and spell out what is needed to take part. The product of this effort is a preliminary plan for the full process. Analysis and deliberation are integrated in process design activities when, for example, the organizing team members draw upon their own experiences and the expertise of others. Consultants and experts can provide advice on how to establish groundrules and operating procedures. Public participation practitioners and social scientists have assembled a wealth of knowledge on techniques for public participation (Bleiker and Bleiker 1995, Connor 1994, Webler 1997, Creighton 2005).

In the process of clarifying objectives and selecting performance metrics for oil spill response, activities are coordinated and focused. Meetings are not held for the sake of merely holding a meeting, but to review a specific segment of information. Press conferences and public meetings may be planned to correspond with releases of reports.

The outcome of the process design phase is a *working draft process plan*, prepared for review by the relevant stakeholders involved with oil spill response. It is at this point that the organizing team takes an initial cut at sketching out how many meetings there will be, where they will be held, what the informational needs are, how the sequence of learning, discussion, and decision will take place, and how many resources will be needed to carry out the process. This is the place where the four structural features of the process that we discussed above – incremental progression, iteration, learning, and integration of analysis and deliberation – are intentionally considered. In addition, the proposed process should include a statement of the intended products or outcomes. Defining specific products along the course of the process is a good way to ensure that the criteria of good process – fairness and competence – are achieved.

As with the preliminary problem formulation, this draft process design is considered a work in progress. As more experience is gained, unanticipated problems (e.g. opposition from groups or dropout) may emerge and external conditions may change (e.g. new legislation). Any of these are valid reasons to re-visit the process design and problem formulation.

Legitimacy begins with the very first steps in the process. People need to feel that their input was desired, that they were listened to, and that they made a difference. A good process is one in which a wide variety of groups feel ownership over the process. For individuals involved, this often is associated with a high sense of solidarity and group cohesiveness. When people believe this is “their process” and these are “their recommendations,” they are much more likely to support implementation and follow-up.

*What information and experience they will bring to inform deliberations and decisions:*
One way to determine what people will do in the process is to be clear about what different groups of people can contribute. There are two dimensions to this: 1) what information can they provide and 2) what discussions should they participate in, because their views and knowledge are important. For each of these, there are three possibilities:

- *What do people care about?* For example, are residents in potentially affected areas concerned about loss or property values? Opportunities for recreation? Impacts to the local economy by disruptions of tourism and fishing? Are they concerned about how water quality affects fish habitat and recreational opportunities? Or are their concerned with habitat damage and mortality of endangered species? Objectives and performance metrics can be fully considered when people have opportunities to raise their concerns, propose ideas, and identify outcomes they care about. This may happen in a variety of deliberative venues. Analysis informs these deliberations through the use of survey or interview techniques to measure, for example, how widely held viewpoints are in a community and whether experts believe given management options have high or low probability of success.

- *What should people care about?* If people are not informed about existing conditions, they may not give adequate attention to a significant outcome. Analysis and deliberation can help clarify potentially significant issues, such as the presence of, for example, important cultural resources, sensitive habitats, and ecological resources. Similarly, what are the risks of dispersants or *in* situ burning to things that people care about?

- *What are ways to assess response success?* For every concern or objective, there may be a variety of metrics that can be used to assess how well a response effort addresses them. But, not all metrics may be as equally informative. Or, while a metric may be relevant and meaningful, there may be no way to ensure that the necessary data to measure it will be available. Some metrics may be indirect measures of what people care about, while others may be more directly connected the expressed concern/objective. In other words, some metrics may be better than others, in particular situations and under particular conditions.

Different people may be more capable of providing answers to these questions or participating in meaningful discussions about them. For example, a wide range of interested and affected parties should provide input about what people care about. However, it may be more appropriate – and competent – to gather information about the relative utility of different performance metrics from people with relevant technical expertise. The organizing team needs to carefully consider who can best provide relevant information and who should discuss different aspects of the issue, for all steps of the process. Table 3 illustrates the information that can be obtained from different stakeholders. This table can be used to structure the information gathering effort (Step 3), which is discussed next, and characterization of options (Step 4).

**Table 3. Roles in each step: Examples of what information should be provided by different stakeholders**

| Participants | Objectives | Performance metrics | Measures |
|---|---|---|---|
| **Organizing team** | Legal requirements | | |
| **Federal responders (NOAA, Coast Guard, EPA)** | Best response model and critical success factors | Critical success factors from past spill response experience | |
| **State and local responders** | State specific objectives related to local economies, closure of fisheries, etc. | Critical success factors from past spill response experience; number of constituent complaints | Media stories |
| **Responsible parties** | | The kinds of data they routinely keep track of about their activities (e.g., man-hours worked, equipment deployed) | |
| **Contractors** | | The kinds of data they routinely keep track of about their activities (e.g., man-hours worked, equipment deployed) | |
| **Community members (.g., business interests, recreational interests, local residents)** | What are people most concerned about regarding future oil spills? | Degree of interference to local business from spills and spill responses | |
| **Researchers** | | Critical success factors; What is known about ecological responses to oils of different types and different seasons? | What is known about effects of oil on particular bird species? (e.g., dose-response relationship) |

*Step 3. Gather information about oil spill response scenarios, including preferences for objectives, performance metrics, and measures*

This step explicitly focuses on gathering all of the informational inputs the process requires and assembling the information in a way that facilitates further deliberation.

During this step the organizing team is responsible for compiling and summarizing information about three topics:[6]

1. objectives: what do people care about and believe should be addressed in order to achieve the three over-arching goals of spill response?
2. performance metrics: how can judgments about the success or failure of the response in meeting the objectives be made?
3. measures: what data can be used to measure performance metrics, and what are appropriate ways to make differentiate between 'success' and 'failure' (e.g., thresholds)?

There are three general sources for information about these topics: scientific knowledge, legal knowledge, and experiential knowledge. To obtain information about the topics from these sources, the organizing team must tap the knowledge of different stakeholders who "specialize" in that type of knowledge, as discussed in Step 2. We have tried to highlight the diversity of forms that information or input can take. Input can be scientific data, governmental reports, private records, public discussions, expert testimony, or private opinions, attitudes, or beliefs. It can take the form of raw data or it can be interpreted in some form. For example, Kuchin and Hereth (1999) gathered relevant information about potential performance metrics from: participant (i.e., first responders') experience, ICS implementation experience, incident specific pollution reports (ISPRs), lessons learned database, job-task analysis, and response management job aids. Lindstedt-Siva (1999) used case studies of major spills, technical and social science literature on spill preparedness, management, and response, and her personal experience and observations during her long career in the field.

The initial identification of objectives and performance metrics is best accomplished by asking relevant stakeholder groups (i.e., interested and affected parties, or socially organized groups that are or perceive themselves as being affected by the decision) to reveal their values and criteria for preferring particular objectives. It is important that all relevant stakeholder groups are represented and that a variety of value clusters, including economic, political, social, cultural, and religious values, is integrated into the analysis. Although strategic reasoning and hidden agendas may influence the responses of stakeholder groups, the mere listing of concerns as expressed in values and, subsequently, the deduction of criteria helps to expose inconsistencies and to avoid hidden agendas.

An effective approach for the organizing team to gather initial information about stakeholders' views and ideas is to meet separately with individual or small groups of stakeholders. A variety of techniques may be used for this purpose, including interviews, surveys, and review of documents, reports, etc.; our case study report exemplifies how interviews can be used to gather information about objectives and candidate performance metrics. Different stakeholders should be asked for their views about the kinds of information identified in Table 3, above.

[6] We have adopted the terminology of goals, objectives, performance metrics, and measures in the proposed process. However, other options and models may be used. For example, Kuchin and Hereth (1999) and Haynes and Ott (No date) have proposed a model using the concepts of key business indicators and critical success factors. While these are not identical to objectives and performance metrics, they ultimately will provide the same kinds of information that will support assessments of oil spill response performance.

During the initial stages of the process we recommend that the organizing team frame its gathering of information in regard to the region of interest, rather than specific scenarios (as is done for ecological risk assessments; see Pond et al. 2000, Aurand 2003, USCG 2001). Early in the process it is more useful to get all the concerns of interested and affected parties "on the table." This is also true for their ideas about performance metrics. This can be thought of as a creative brainstorming exercise. This approach has the advantage of allowing stakeholders to express their concerns and ideas more fully, without having to orient themselves to what people think is likely (i.e., a particular scenario). Furthermore, it can help participants feel that their input is really desired, without having decisions (e.g., about scenarios) pre-determined. If scenarios frame the initial gathering of ideas about objectives and performance metrics, the risk is increased that some important information may be missed or neglected because it is not expected or cannot easily be put into the context of a scenario.

The organizing team should then compile all of the information into a form that stakeholders can review and use to *compare* different options – a form of analysis. This requires a method for organizing objectives and metrics into a form that encourages stakeholder deliberation about alternative sets of objectives and performance metrics that can be used to guide contingency planning and to assess oil spill responses in a way that facilitates learning. One option for accomplishing this data organization task is to group suggested objectives and performance metrics into a small set of categories. Performance metrics that were proposed for different objectives can be listed with them, as another option. The point is to compile the information that helps to increase the understanding of stakeholders about a) what the options are and b) what others think. Consequently, it is very important that there be definitions that clearly distinguish between objectives and performance metrics; the organizing team should promote conceptual clarity.[7]

For example, many different objectives can be sought in oil spill response efforts, and it is helpful to group them into meaningful categories. Kuchin and Hereth (1999) have identified several, which they call key business drivers: protection of human health and safety, protection of the natural environment, mitigation of economic impact, public communication, stakeholder service and support, and quality and effectiveness of the response organization. Lindstedt-Siva (following a model from Brown (1996)) suggests sorting performance criteria according to five performance elements of inputs, process, outputs, outcomes, and goals. Tuler et al. (2006, Forthcoming) identified objectives that people involved with two spills (i.e., Bouchard-120 and Chalk Point spills) cared about in those spill responses. Tuler et al. found that many different objectives for the response to these two oil spills were important to those interviewed (although they were not all shared among research subjects in each of the cases), and they fall into several categories as Shown in Table 4.

[7] This is one weakness of some prior efforts to list and classify metrics. For example, Lindstedt-Siva (1999, Table 1) lists items that appear to be objectives (e.g., protection of marine mammals, responsive to constituent's needs) as response performance criteria.

**Table 4. Types of objectives for oil spill response derived from two case studies**

- Address needs and concerns of the affected public/communities
- Establish a coordinated and effective response framework
- Gain public support for the response
- Implement an effective and timely response
- Meet legal and regulatory requirements
- Mitigate economic impacts
- Mitigate social nuisance impacts
- Protect cultural resources
- Protect environment and mitigate environmental impacts
- Protect worker and public health and safety

At this point, the organizing team should facilitate the gathering of additional information. It may also benefit from shifting to a scenario-based approach. For example, objectives and performance metrics can be evaluated for their relevance to particular scenarios for future spills and response (e.g., see Lindstedt-Siva 1999, pg. 32). Again, there are a variety of techniques that can be used to gather information and characterize the ideas expressed by the participants. Different kinds of stakeholders should be engaged in this effort in a way that builds on their areas of expertise and experience.

As a first step, however, it will be important to assemble information about possible scenarios. This effort will require the input of people closely involved with the transport of oil in a particular region, as well as first responders, scientists, and others that have appropriate technical expertise and experience with spills and spill responses in the region. Judgments must be made about the likely causes of oil spills, the dynamics of the oil once it is in the marine environment, etc. The generation of scenarios can be a well-defined process, such as used for ecological risk assessments.

These scenarios can be helpful for further characterizing the information needed to assess objectives, performance metrics, and measures. Here we want to illustrate two ways that information can be characterized that will facilitate learning and deliberations.

Characterizing preferences for different objectives

As noted previously, there can be many different objectives that inform people's views about what ought to happen in a spill response. However, it is probably incorrect to assume that there will be universal agreement about the objectives, either in general, in regard to specific scenarios, or in regard to specific spills (Abordaif et al. 1995, Lindstadt-Siva 1999, Tuler et al. 2007). Interested and affected parties may disagree about the *relative* importance of specific objectives that underlie the over-arching goals of spill response. This kind of information is important to reveal as part of this process – and such information should be used to inform planning and to inform selection of performance metrics.

As is often the case, a variety of techniques may be used to gather information about the relative importance placed on various objectives. At this point it can be helpful to ask about relative importance in the context of specific scenarios, and to analyze how objectives are weighted or ranked among people and of how preferences for objectives vary among stakeholders. One option for gathering the needed information about objectives and developing this sort of information about relative priorities is to use Q method; this is the technique employed by Tuler et al. (2007). We asked people in two regions to tell us what were the most important objectives during the emergency phase oil spill response to meet goals (and not during other phases, such as restoration, damage assessment, compensation). Most participants told us that all of the objectives were important. However, Q method pushes people to express relative priorities. In doing this, the participants in the research reveal that they attempt to realize the over-arching goals of spill response by emphasizing different objectives. Lack of emphasis did *not* necessarily mean that an objective was rejected. What these results indicate is that, indeed, people that have experience with oil spills and responses in a particular region can agree about the relative importance of some objectives and disagree about the relative importance of others.

Characterizing the appropriateness of different performance metrics

It is unlikely that all stakeholders (or agency managers) will agree on which metrics should be employed to determine the relative success of any response. In most situations, there can be a large number of potential metrics that can be defined (Tuler et al. 2006, Forthcoming). Baker (1999) provides an illustrative example in the context of assessing "clean" by measuring petroleum hydrocarbon concentrations, in which he concludes that "there are many different definitions of clean, drawing upon both ecological and socio-economic criteria. There is no 'best' definition" (pg. 144; also see Lindstedt-Siva pg. 32). Possible metrics he suggests that may be appropriate include (pg. 135):

- do not exceed normal background levels for a particular location;
- do not exceed statutory limits;
- are not lethal to specified organisms;
- do not cause deleterious sublethal effects to specified organisms;
- do not cause tainting of food organisms;
- have no detectable impact on the function of an ecosystem;
- do not impair the human use of an area;
- are not visible to the human eye; and,
- cannot be reduced by enhanced clean-up actions without causing an overall retardation of recovery.

Participants in the process can be helped to assess the relative quality of different performance metrics if a) they are grouped systematically in relation to specific contexts (e.g., scenarios) and b) if they are evaluated on a clear set of criteria.

To start, it may be worthwhile to use multiple taxonomies for categorizing and evaluating metrics to highlight where there may be gaps in suggested performance metrics (and if the gaps remain, subsequently in the selection of performance metrics as part of the process. One way of grouping objectives and performance metrics is according to the phase of the response, such as emergency response, project management, and investigative phases (Lindstedt-Siva 1999, Ott et al. 1993). Others define the phases of spill response differently. For example, there are at least three phases within response that have a spatial dimension:

1. containment and removal of oil at the scene of the spill or while it is still on the open water, in order to reduce or eliminate the impact on shorelines or sensitive habitats.
2. interception, containment, and removal of oil in the near-shore area. The intent of removal or containment near-shore is the same as containment and removal in the open water: remove the spilled oil before it impacts sensitive environments.
3. protection of sensitive areas in the path of the oil or to minimize that impact to the maximum extent practical. [From http://www.dec.state.ak.us/spar/perp/grs/faq.htm].

Two additional options for grouping performance metrics are provided in Text Boxes.

In addition, each performance metric that is proposed can be assessed according to several criteria about quality. An ideal metric should have several characteristics (Graedel & Allenby 2002, Seager & Theis 2004, Seager et al. 2006, Forthcoming):

- It would be *scientifically verifiable*; two independent assessments would yield similar results.
- It would be *cost-effective*; it would use technology that is economically feasible and does not require an intensive deployment of labor to track.
- It would be *easy to communicate to a wide audience*. The public would understand the scale and context and be able to interpret the metric with little additional explanation.
- It could be *changed by human intervention*. The metric would have a causal relationship between the state of the system and the variables that are under a decision-maker's control. Metrics that are independent of human action do not inform a management, policy-making, or design process.
- It would be *credible*; it would be perceived by most of the stakeholders as accurately measuring what it is intended to measure.
- It would be scalable *over an appropriate time period and geographic region*. It would be indicative of short, medium, and/or long term effects as appropriate. For example, it would not be meaningful to attempt to measure the effects of chronic low-level toxic dosages over a period of weeks or months, just as it would not be appropriate to average local environmental conditions over a widely varying region.
- It would be *relevant*. It would reflect the priorities of the public and other stakeholders and enhance the ability of spill managers and/or regulators to faithfully execute their stewardship responsibilities. There is no point assembling a metric no one cares about.
- It would be *sensitive* enough to capture the minimum meaningful level of change or make the smallest distinctions that are still significant, and it would have uncertainty bounds that are easy to communicate.

Finally, candidate performance metrics should be assessed in terms of their *measurability*. As we discussed earlier (Section 2), there can be many challenges to using performance metrics because of the challenges associated with finding direct measures of the quality of interest or data for making comparisons. Thus, for example, performance metrics should be assessed for (but not limited to):

- the reliability of data recording,
- the availability of accurate baseline data on which to base comparisons or assessments,
- the types of uncertainties that may affect measurements, and how they may leave the meaning of measurements open to dispute, and
- whether the meaning of measured values may vary depending on *when* measurements are made.

The purpose of such assessment is to *inform* stakeholders about the strengths and weaknesses of different metrics (Step 4). The purpose is *not*, at this step, to choose metrics. One option for assessing proposed performance and measures and developing this sort of evaluative information about relative strengths and weaknesses of particular options is to use group delphi (see Linstone and Turoff 1975, Renn et al. 1993). It is similar to the original Delphi exercise but based on group interactions instead of individual written responses. The objective is to reconcile conflicts about factual evidence and reach an expert consensus via direct confrontation among a heterogeneous, preferably representative, sample of experts in the field. Evaluation of the proposed metrics should be made by people with the expertise to evaluate them. This may mean that different types of expertise are called for. For example, assessment of economic metrics should be done by those with expertise in economics. Those who assess the quality and relevance of metrics related to organizational capabilities and activities, should have a thorough understanding of the organizational environment in which spill responses occur as well as administrative and legal requirements.  The public may also have a role in the evaluation of some metrics using criteria such as ease of communicating and credibility.

The outcome of this step should be a profile for each objective and performance metric. The profile for performance metrics should specify the range of scientifically, socially, and institutionally legitimate and defensible expert judgments for each metric, illustrate the distribution of these opinions among the expert community, and include justifications for opinions that deviate from the median viewpoint. As input to further deliberations about the metrics, the experts should 'rank' each metric's ability to provide a useful measure of spill response performance with respect to a particular objective. They can do this by, for example, color-coding the metrics (red = bad, yellow = OK, green = good; Haynes and Ott (No date) use this approach as well). Alternatively, each metric can be characterized by where in the organization the metrics applies, how (and how well) the information is expressed, and what type of information is collected (Table 5) according to the typology of environmental performance metrics described in previous work (Seager et al. 2006, Forthcoming).

**Table 5: Characterization of Performance Metrics**

| HOW | | WHERE | | | WHAT |
|---|---|---|---|---|---|
| MATHEMATICAL | QUALITY | DECISION LEVEL | CAUSAL CHAIN | | TYPE |
| quantitative<br>semi-quantitative<br>qualitative | verifiability<br>cost<br>communicability<br>sensitivity<br>credibility<br>scalability<br>relevancy | strategic<br>tactical<br>operational | inputs<br>processes<br>outputs<br>outcomes | planning<br>response<br>recovery<br>restoration | economic<br>environmental<br>ecological<br>socio-political<br>human health<br>thermodynamic |

The *how* column in Table 5 characterizes the mathematical form and quality of the metric. Mathematically, all metrics could be classified as either quantitative (e.g., cardinal measurement), semi-quantitative, or qualitative. In the GPRA, a clear preference is expressed for quantitative measures -- although they may not always be available. Regarding quality, different metrics may have multiple attributes to different degrees. Therefore, a judgment regarding the quality of any metric may depend upon the relative importance of each attribute such as cost or verifiability to the decision maker.

The *where* column describes the relationship of a metric to both organizational structure and sequence events in crisis response. A different suite of metrics may be applicable at the strategic level of thinking than at the tactical or operational. Similarly, different metrics are applicable for an accounting of resources, processes or outcomes depending upon whether the concern is emergency response, project management, investigative phases, or systems recovery and restoration. In some cases, outcome metrics for response (e.g., rescued birds) may be viewed as input metrics for recovery (e.g., viable breeding population) or restoration. For example, an overarching strategy may be to reduce the severity of oil spills (as measured by wildlife deaths, for example).  One tactical approach may be to contain and remove slicks. In pre-spill planning, a resource metric might be to measure capital equipment expenditures for purchase and pre-positioning of additional equipment. In the event of an oil spill, the on-scene coordinator might track tactical measures such as the time required to deploy the prepositioned equipment during the response. At the operational level, the effectiveness of the deployed equipment must be tracked to ensure success of the overall strategy. The importance of tracking resource and process (or output) measures must be emphasized in cases where outcomes are significantly delayed or disconnected in time from the actual decisions. In many instances, mid-course corrections must be made before final outcomes can be tallied.

**A taxonomy of metrics based on phases of spill response**

Different metrics may be relevant to different stages of oil spill response, recovery and restoration. Figure 1 depicts a typology of oil spill metrics that shows both how oil spill management progresses in time (from left to right) and how metrics may be characterized as resource, process, or end-point based. An effective response will mitigate the damaging economic, thermodynamic, environmental, ecological, and socio/political effects. Because these are manifested on different time scales, leading indicator metrics are essential to provide feedback to the response organization and allow adaptation to changing or unexpected events. As a response efforts progress, the locus of concern moves from the lower left hand corner of the graph towards the upper right. That is, resources are required to drive processes which are directed at endpoints. However, response phase endpoints soon become recovery phase resources, and so on.

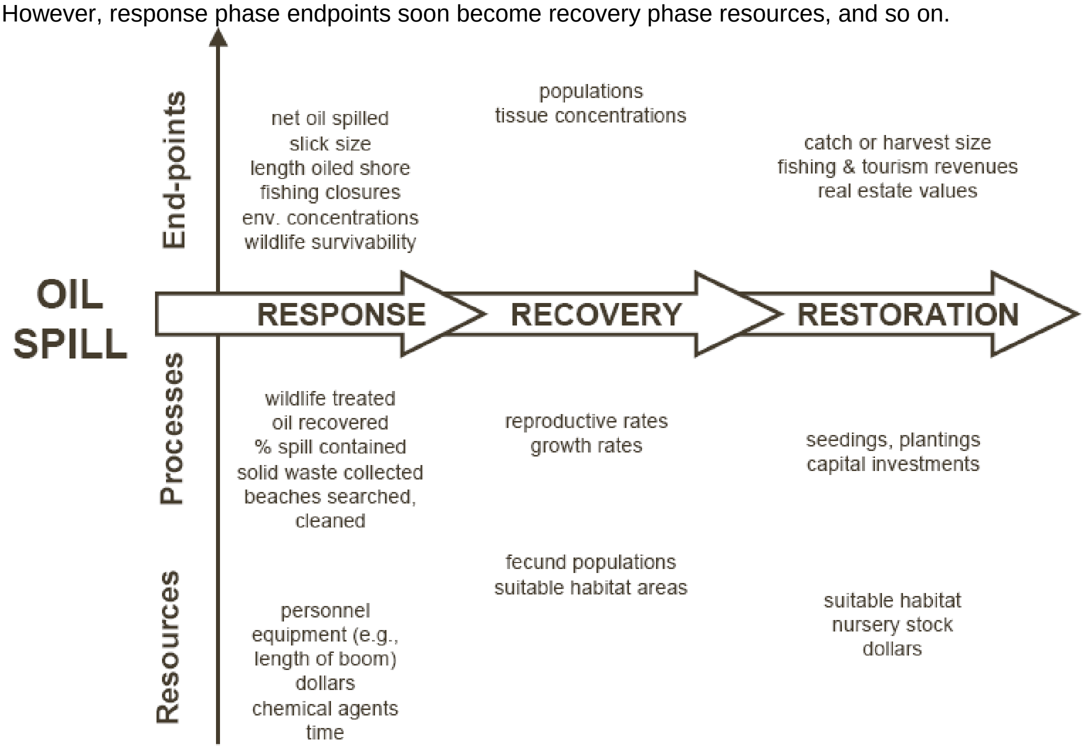


**Figure 1: Typology of Oil Spill Response & Assessment Metrics.** Management of oil spills may be characterized in three stages that generally blend together: response, recovery, and restoration. Each stage is overlapping in time, read from left to right in the direction of the block arrows. Within each stage, progress is from the lower edge of the figure towards the top as resources are applied in processes that are intended to improve endpoints. Decisions made in early stages may have a direct impact upon the resources, alternatives, and endpoints in other stages. Mapping these concerns provides a visual depiction of potential opportunities for better communication and coordination, as well as potential conflicts.

As different organizations become engaged in managing the spill, the metrics relevant to those organizations efforts can be plotted in the metrics map depicted by Figure 5. For example, mechanical recovery contractors will be primarily concerned with resource availability and operational effectiveness in recovering oil and emulsion from the environment. The metrics relevant to their work will be found in the lower left-hand region. A wildlife biologist, on the other hand, may be more concerned with habitat restoration and population recovery – concerns better plotted closer to the upper right-hand corner. Considering the distance between these two perspectives, it may be challenging to communicate to contractors how their efforts or decisions influence the measures of concern to the biologist. Opportunities for closer coordination can be identified where two different agencies identify measurable objectives that are plotted at some distance apart on the map. These agencies (or stakeholder or public groups) may disagree entirely on the perceived success of spill management efforts simply because they are examining different aspects of those efforts.

**A taxonomy for classifying metrics**

Virtually all metrics relevant to chemical release management may be characterized into five broad dimensions: economic, thermodynamic, environmental, ecological and/or human health, and socio-political. As they relate to oil spills, the broader categories include (see also Seager & Theis 2004, Seager et al. 2006, Forthcoming):

**Economic.** In addition to direct and indirect costs, economic metrics convert non-market resources or impacts into monetary values to allow comparison with monetary transactions or industrial accounts. Economic estimates of non-market impacts are required by benefit-cost analysis for estimating the value of damages caused by an oil spill in terms of fish catch, property damage, clean up costs, or for prioritizing new investments. Broader economic analysis could include estimates of lost tourism revenues, decreased property values, or opportunity costs (Loureiro et al. 2006). In theory, proper pricing of environmental goods and services could allow market forces to optimally allocate resources between ecological and industrial activities. However, in practice both the calculation methods and the validity of the concept of pricing the environment are recognized as controversial. Because there are no markets for most environmental goods, such as pollution attenuation, external or social costs are highly uncertain, as are the methods and figures reported for the value of ecosystem services. Moreover, monetization may lead to the erroneous assumption that environmental exploitation can be revocable in a manner analogous to pecuniary transactions, although in some cases ecological systems may be damaged beyond recovery.

**Thermodynamic** metrics such as total pollutant loading or release are indicative of environmental *pressure* (e.g., pollution to be attenuated), whereas measures such as energy use are more indicative of resource consumption or scarcity. Sometimes, thermodynamic metrics are normalized to intensive units such as kg/person or oil equivalents of energy/product, which attempt to capture the *eco-efficiency* of a process. However, in the case of oil spills, extensive measures such as total barrels lost or recovered are appropriate. Usually thermodynamic metrics do not indicate the specific environmental response associated with resource consumption or loss. For example, the severity of an oil spill may be determined on the basis of total volume spilled. Nonetheless, ecological effects are dependent upon a number of other factors such as the type of oil and the location, mobility, and timing of the spill. On the basis of a thermodynamic measure called *emergy,* which measures energy consumption in terms of the equivalent solar energy required to replace the consumption, Odum (1996) criticized the extensive clean up efforts that followed the grounding of the *Exxon Valdez* as an unproductive deployment of energy resources. His study claimed that more diesel fuel was expended on clean up efforts than barrels of oil were lost in the spill. Nonetheless, thermodynamic metrics are only indirectly related to the human and ecological health objectives that guide oil spill response -- conservation of diesel fuel is not the primary objective of any large spill response.

**Environmental** metrics estimate the extent of chemical change or hazard in the environment. Environmental metrics often use physical or chemical units such as pH, temperature, or concentration. Concentration measures – especially for toxic oil components such as polycyclic aromatic hydrocarbons (PAHs) – are difficult to put in an appropriate context unless they are tied to some ecological or human manifestation such as carcinogenicity, mutagenicity, or even non-health based endpoints such as beach or fisheries closures. Environmental metrics are generally measures of the residuals released by industrial processes that *pressure* the environment or are indicative of the environmental *state* (e.g., chemical contamination). Similarly, *response* (such as clean-up or remediation) actions are often motivated by measurable environmental objectives, such as reducing contaminant concentrations.

**Ecological** metrics attempt to estimate the effects of human intervention on natural systems in ways that are related to living things and ecosystem functions. The rates of species extinction and loss of biodiversity are good examples, and they are incorporated into the concept of ecosystem health (Rapport 1999, Rapport et al. 1998). Oiled bird counts, marine mammal death counts, and time to ecological recovery are all examples of ecological metrics that are typically applied to oil spill damage assessments and restoration efforts. Although there may be good agreement among experts on the importance and relevancy of ecological outcomes, there may still be considerable disagreement about the response alternatives (such as mechanical removal, in-situ burning, dispersion, or natural attenuation) that will best achieve the ecological objectives. When ecological systems are unable to recover naturally from the effects of a spill, responsible parties are typically required to sponsor restoration efforts that return ecological systems to an approximation of their undamaged state.

**Human health** metrics are indicative of the state of the human population just as ecological metrics indicate the state of natural systems. Human health includes worker and public safety. For example, worker injuries per total response hours worked is representative of one type of anthropocentric oil spill effect. There may also be increased health risks from inhalation exposure to toxic chemicals released to the environment. These risks are difficult to aggregate due to the breadth of different end-points entailed (Hofstetter and Hammit 2002). Protection of human health is a primary goal for federal agencies engaged in oil spill response, reflecting the primacy of anthropocentric concerns. Nevertheless, the measures devised (e.g., injuries, deaths, treated patients) may be directly analogs –if not identical -- to those used to track animal effects.

**Socio-political** metrics evaluate whether industrial activities are consistent with political goals like energy independence or eco-justice, or whether collaborative relationships exist that foster social solutions to shared problems. Major oil spills undoubtedly have far-reaching social and political impacts (e.g., Shaw 1992). However, these are difficult to gauge quantitatively. In some cases, the political and social dimensions are translated or communicated primarily through the media. That is, although spill responders may understand the importance of public perceptions, they may have no basis for measuring improvement or deterioration of public sentiment, except through the tone of media coverage – which they may feel powerless to influence (Harrold 1994).

**Phase 2. Deliberate about options**

*Step 4: Assess objectives, performance metrics, and measures.*

The purpose of this step is for the stakeholders to learn from the experts about the strengths and weaknesses of the various metrics to allow for a good assessment of the objectives. This whole step is oriented toward stakeholder learning; the emphasis at this Step is on deliberation rather than analysis. The challenge here – and the reason this step is required – is that no single performance metric is going to be adequate to measure any particular objectives. Rather, stakeholders must think in terms of *sets* of performance metrics for particular objectives. That means that not all metrics must be 'perfect' – some can work well on some criteria, others on other criteria. The goal should be to encourage thinking about how to assemble satisfactory 'packages' or 'bundles' of metrics that a) do a good job of characterizing the objective, b) assess the various aspects of the objective that a broad range of stakeholders cares about and c) that are practical to use.

To promote learning a workshop or meeting (or a series of them) should be scheduled, in which stakeholders meet with the experts involved in assessing the metrics in Step 3. All of the information compiled and summarized as part of Step 3 should be shared. For example, scientist's traffic light assessments should be shared with stakeholders. During the meetings, there should be ample time for stakeholders to ask questions of the experts.

Of course, there are many ways to organize such workshops, and we will not review them here. However, the organizing team will have to make a number of decisions about how to best organize them. For example, they should assess the need for iterations between small group discussions and full group discussions, the utility of having multiple meetings, and the utility of facilitating "keyboard-to-keyboard" discussions to supplement face-to-face discussions at meetings (e.g., Hamlett 2002, Hamlett and Cobb 2006). Experiences with ecological risk assessments may prove useful in their efforts to decide how to best structure these deliberations, as well as the large literature on structuring effective public participation in environmental decision making (e.g., Pond et al. 2000, Aurand 2003).

**Phase 3: Reach closure and evaluate outcomes**

During phase 3 of the process, the focus is on reaching agreement about the objectives and performance metrics for oil spill response in the region. Performance metrics should be defined in close connection to specific objectives.

*Step 5: Select objectives, performance metrics, and measures and attempt to reach closure*

As part of the process design (steps 1-2) the organizing team should have made clear to participants what are the desired outcomes of this deliberative planning process. There are two options that should be considered: to identify a set of objectives that can be used to a) guide spill response planning and b) assess performance in a spill event. In fact, these are tightly coupled, and can be thought of as two essential elements for organizational planning and learning. For example, Lindstedt-Siva (1999) outlines a response performance assessment process based on four steps that are linked in a feedback loop to facilitate learning: 1) identify stakeholder performance criteria, develop contingency plan, conduct exercises and responses, and 4) assess response performance. The set of objectives should be responsive to the breadth of concerns expressed by interested and affected parties. For each objective a set of performance should be selected that can be used to assess performance on the particular objective.

In order to reach agreement on alternative objectives and performance metrics these must be ranked using the same criteria that were used to assess them. The interested and affected parties that are participating in the process should do this together, in a well-structured deliberation based on a satisficing approach. This can occur in one or more meetings, and paralleling the approach taken in Step 4 there may be opportunities for online discussions to supplement face-to-face meetings.

Many of the challenges posed by selection of objectives and performance metrics are amenable to multi-criteria decision analytic approaches such as those that have been adopted in some instances for social problems, such as environmental decision-making (Lahdelma et al. 2000, Kiker et al. 2005). Multi-criteria decision analysis refers to a group of methods used to impart structure to the decision-making process. Generally, these decision analysis methods consist of four steps: (1) creating a set of criteria relevant to the decision at hand, for use in evaluating the decision alternatives, (2) weighting the relative importance of the criteria, (3) scoring how well each alternative performs on each criteria, and (4) combining scores across criteria to produce an aggregate score for each alternative or pair-wise comparisons of alternatives. The goal of the process is often to select a single best alternative, but a ranking of alternatives is very useful.

Ideally, in Steps 3 and 4 the organizing team (with the assistance of relevant experts) will have provided an overall assessment of the proposed objectives and performance metrics. If the typology proposed in Table 5 is used, the organizing team can populate the entire *where* and *what* dimensions of the typology with suggested alternatives of high quality metrics that are applicable to every level of decision-making within the organization. The deliberations of Step 4 should have allowed the identification of gaps that may concern managers or stakeholder groups. For example, they may discover a paucity of strategic or quantitative metrics, which may cause them to return to the step of information gathering and analysis to fill the gaps they consider important (Step 3 and 4). In addition, the organizing team must reorganize the information in multiple directions to discover whether a suitable cross-section of different types of feasible, high-quality metrics are available for each organizational objective (strategic, tactical, operational) or at each step in the agency's chain of influence (inputs, process outputs, and outcomes) or for each type (economic, environmental, ecological, socio-political, human health and thermodynamic). Once again, significant deficiencies may be corrected by further information gathering and analysis (i.e., iteration of Steps 3 and 4).

The typical approach to an environmental MCDA involves identifying feasible alternatives and the criteria by which they should be judged and assessing the performance of each alternative relative to the salient criteria. While alternatives may be generated by experts, stakeholders and/or public groups, the decision *criteria* and their relative importance may justifiably be the purview of affected parties (such as stakeholder groups), who may or may not also be experts. The performance assessments are typically performed by experts. For example, in considering the problem of managing oil spills, stakeholder groups may emphasize the importance of environmental quality, preservation of ecological habitat, or economic development. All of the alternatives generated by expert and stakeholder groups must then be assessed relative to measurable criteria that capture each of these performance criteria – although different groups may value each criterion differently. The results are called a *performance table* that is subsequently analyzed with different mathematical approaches including multi-attribute utility theory, outranking, pair-wise comparison or others. The advantage of MCDA is that alternatives can be ranked or prioritized for multiple decision criteria, trade-offs can be elucidated, different types of information including semi-quantitative or uncertainty information can be handled in a structured way and conflicts or opportunities for compromise can be observed that may not otherwise have been discovered (Lahdelma et al. 2000, Figueira et al. 2005). The principal disadvantage of MCDA approaches is that they can be time-intensive, especially when multiple stakeholder perspectives are important.

In this context, we propose a unique approach for the purposes of the oil spill response planning and assessment (Linkov et al. Forthcoming). In this case, the *metrics* are the alternatives. Usually, metrics are associated with assessment criteria and relate to the underlying objectives or goals of the organization. However, spill response assessment requires agencies to select a limited number of metrics from a theoretically infinite universe of potential outcomes metrics. Not all metrics would be cost-effective for the agency to track, nor can any single metric perfectly fulfill the requirements for an assessment of spill response. As in a typical environmental MCDA, some agency customers may value certain metrics within the suite more highly than others. (Some may even place zero weighting on certain metrics). Consequently, despite a common set of assessment metrics it is possible that agency performance could be viewed positively by one group and negatively by others – a situation that may be all too familiar to government managers. The critical difference between the process we propose for spill response planning and assessment and a typical MCDA is that the decision alternatives are *which metrics to use*. The criteria that represent the other half of the performance table are characteristics that agency and key stakeholder feel the selected metrics should have. Thus, the performance table is a guide to determining which metrics are better than others for any particular program.

The product of this step is a report detailing the selected objectives and their performance metrics, as well as any special observations about the measurement of performance metrics. The report should summarize the information gathered and interprets it in the context of the policy options, explaining uncertainties and assumptions. Quantitative information can be summarized through a variety of techniques including statistics, representations of uncertainty, and simulations. Qualitative information can be summarized in map overlays and narratives. Generally, each technique is limited and can only provide one viewpoint on a complex picture. Thus, more than one method of summarizing information should be used to facilitate understanding of an issue. In writing the report, the organizing team oversees a synthesis that considers preferences of interested and affected parties and produces recommendations.

In this context, it is *not* essential for all participants to agree about each objective and performance metric. To be sure, *opposition* to an objective or performance metric must be addressed. However, it will also be appropriate to include all objectives and performance metrics in the selected set to the extent that they are not opposed or inconsistent. For example, some stakeholders may think that protection of economic resources is critical. Others may think that the protection of cultural resources are critical. To the degree that these do not create conflict, it is appropriate to select *both* as objectives by which to assess a spill response. Similarly, it can be appropriate to select a set of performance metrics for a particular objective. Different stakeholders may want to assess performance on the objective in different ways.

Of course, the stakeholders and organizing team must be cognizant of the potential for objectives and performance metrics to be inconsistent or conflict. At this point, those responsible for organizing the process (i.e., organizing team) may need to make a final decision. To the extent that the objectives or performance metrics desired by a stakeholder are rejected, a clear rationale must be provided; the decision must be transparent.

*<u>Step 6: Evaluate performance and gauge the effects of the actions.</u>*
Building on the idea of learning and iteration, the process should incorporate plans to assess the utility and practicality of

- using the selected objectives and performance metrics and
- the process used to develop them.

*Evaluation of the utility and practicality of selected objectives and performance metrics*
The goal of the process is, at the very least, to develop a set of objectives and performance metrics that can be used to assess spill response in a particular region. Their utility should be assessed: do they support a meaningful assessment of spill response? Can they help to differentiate among those things that went well and those things that did not in a response? This can be tested, of course, after any future spill. Assessments may also be conducted after training exercises.

The objectives and performance metrics may also be used as input into contingency planning. The effects of integrating this information into contingency planning can also be assessed. As discussed earlier (Section 3) frameworks for assessing the quality of spill response preparedness and plans have been proposed. They (or others) can be used.

The critical importance of evaluation should not be underestimated. Lindstedt-Siva (1999) argues that spill response performance is a multi-phase process that includes:

1. Identification of stakeholder performance criteria
2. Developing contingency plans
3. Conduct of exercises and actual responses
4. Assessment of response performance

She writes that: "this process encourages stakeholder participation in the establishment of specific performance criteria (Phase 1) for spill response. These criteria form the reference points from which goals and strategies are developed in contingency plans (Phase 2). Development of specific, measurable, achievable, result-oriented goals will enable a response organization to better manage spill response and measure improvement (Phase 3). Additionally, goals enable a response organization and stakeholders to track performance during exercises and response to adjust performance criteria and strategies in contingency plans (Phase 4). This continuous process should result in improved relations and increased preparation for response" (pg. 41; note that *phases* in this quote refer to her phases for a systematic approach to assess performance, not the phases we outline for the objective and metric selection process). Without the fourth step, organizational learning and improvements to contingency planning and spill response is constrained. Thus, as part of the process for selecting performance metrics, effort should also be made to plan for the ways that the information developed as part of the process (Step 5) will be integrated into current spill response planning and spill response assessment procedures.

*Evaluation of the process*
It will also be useful to evaluate the process by which the objectives and metrics are identified, analyzed, deliberated, and selected (Steps 1-5, above). While members of the spill response community have called for deliberative, consensus building processes to be implemented, experience is sparse. While there is increasing understanding of the instrumental and substantive benefits of public involvement in environmental and risk assessment and decision-making, understandings of what works and what does not in particular situations is still limited. Thus, building in a systematic evaluation component to the process can further improve

organizational learning that can support better contingency planning in the future and in other regions.

Evaluation should not be considered an activity that is only conducted at the end of the process, although evaluation is often thought of as an end-of-the-project report card. Evaluation at the end of a process is not useful for improving the *process* by correcting problems mid-course. "Real time" evaluations aimed at improving programs in progress are a means of *formative evaluation* that provides stakeholders with feedback during program development and implementation (Posavac 1991). To improve programs as they evolve, formative evaluation considers relationships among stakeholders, perceptions of agency communication, effectiveness of various outreach or involvement tools, etc. This feedback can be used during the entire process for selecting objectives and performance metrics in order to improve both the analytic and deliberative efforts and positively influence spill response.

## 6. Conclusion

Effective response to oil spills is very challenging. Judgments about whether they are successful can be contentious. There are many criteria on which judgments about the success or failure of a response can be made. Moreover, there may be conflicting ideas about what are the objectives that should be sought in a response effort. A variety of challenges to defining clear objectives and measurable performance metrics were discussed in Section 2.

As if these did not create enough challenges, oil spill response has additional features that make them particularly hard to evaluate. In particular,

- Effective oil spill response requires close coordination between multiple public and private agencies (such as the responsible party and response contractors). The perceived success or failure of a response is a function of the collective efforts of the wide-ranging *ad hoc* administrative structure. Disentangling the contribution of each of these agents to the eventual outcome may be impossible, if not explicitly counter-productive to the extent that organizations may fail to cooperate effectively if they become preoccupied with their own performance assessment at the expense of overall system effectiveness.
- Every oil spill is unique. Although planning efforts are typically intense with regard to pre-positioning of equipment, establishing communication and leadership protocols and personnel training, the response *tactics* are virtually impossible to plan ahead of time. Optimal alternatives are highly dependent upon the circumstances of individual spills. Therefore, assessment of performance and comparison with other spill responses – in which strategic planning is an essential aspect – is complicated by the inability to establish a universal vision of how to achieve a successful response.
- Even with generalizable goals for oil spill response in place, performance measurement requires a benchmark to determine the sensitivity of outcomes to agency efforts. In the case of oil spills (and other crisis response situations), it may be impossible to say what the level of measurable outcomes *would* be under alternative response scenarios. The unique nature of the crisis precludes comparative assessments to control or 'normal' circumstances. Moreover, the time-sensitive nature of the crisis typically prohibits dynamic modeling of alternative scenarios in real time. Consequently, decision outcomes must be judged without the benefit of context.

The goal of this report is to propose a process design that will facilitate the selection of metrics for assessing spill responses by a broad range of potentially affected and interested parties. Our concern here is with the process by which those who have a stake in spill response can collectively discuss:

- the objectives that should guide spill response and
- the criteria or metrics that should be used to assess performance with respect to the objectives.

In this report we have described a process design using six steps to take the first step toward systematically assessing oil spill response performance. What comes next is the actual integration of the outcomes from the process into a response performance assessment system. We have proposed a process design that will result in a set of objectives and performance metrics that can be used to assess responses in a particular region. They may also be used to inform contingency planning in the region. The process promotes integration of analysis and deliberation, in a way that is iterative, promotes learning, and incremental. It rests on the principles of fairness and competence to ensure that decisions are based on broad input of (and discussion among) interested and affected parties and are based on the best available science.

## 7. References

Abordaif, F., Harrald, J., and Mazzuchi, T. 1995. Evaluating oil spill contingency and prevention plans using measures of quality, not quantity. Proceedings of the 1995 International Oil Spill Conference. Washington, DC: American Petroleum Institute.

Apostolakis, G. E., and S. E. Pickett. "Deliberation: Integrating Analytical Results into Environmental Decisions Involving Multiple Stakeholders." Risk Analysis . 18, no. 5 (1998): 621-34.

Ashford, N. and Rest, K. 1999. Public participation in contaminated communities. Cambridge, MA: Center for Technology, Policy, and Industrial Development, Massachusetts Institute of Technology.

Aurand, D. (Compiler). 2003. Ecological Risk Assessment: Consensus Workshop. Environmental Tradeoffs Associated With Oil Spill Response Technologies. Maryland Eastern Shore. A report to Commander, US Coast Guard Activities, Baltimore. Ecosystem Management & Associates, Inc., Lusby, MD 20657. Report 02-02, 36 pages.

Baker, J. 1999. Ecological effectiveness of oil spill countermeasures: How clean is clean?, *Pure Applied Chemistry* 71(1):135-151.

Beierle, T. and Cayford, J. 2002. Democracy in practice: Public participation in environmental decisions. Washington, DC: Resources for the Future.

Bradbury, J., Branch, K., and Malone, E. 2003. *An evaluation of DOE-EM public participation programs. Report prepared for the DOE Environmental Management Program*. Report PNNL-14200. Washington, DC: Pacific Northwest National Laboratory.

Brown, M. 1996. Keeping score. New York: Quality Resources, Krause Organization, Limited.

Burby, R. 2003. Making Plans That Matter: Citizen Involvement and Government Action, Journal of the American Planning Association, Vol. 69.

Carnes, S. A., Schweitzer, M., Peelle, E. B., Wolfe, A. K., Munro, J. F. (1998). Measuring the success of public participation on environmental restoration and waste management activities in the US Department of Energy, *Technology in society*, 20(4):385-406.

Caudle, S. 2005. Homeland security capabilities-based planning: Lessons from the defense community, *Homeland Security Affairs* 1(2). Available online at www.hasj.org/hsa

Chess C, Johnson BB, Gibson G. 2005. Communicating about environmental indicators. *Journal of Risk Research* 8(1):63-75

Creighton, J. 2005. *The public participation handbook: Making better decisions through citizen involvement.* San Francisco: Jossey-Bass.

Daniels, S. and Walker, G. 2001. Working through environmental conflict: The collaborative learning approach. New York: Praeger.

Department of Energy 2003. Public participation and community relations. Policy document DOE-P-141.2, approved 2 May, 2003. Washington, DC: Office of Congressional and Intergovernmental Affairs, Department of Energy.

DHS No date. Capabilities-based planning: Overview. DHS/SLGCP/OPIA/Policy and Planning Branch. Washington, DC: Department of Homeland Security. Accessed 30 March, 2007. Available at: www.scd.state.hi.us/grant_docs/Capabilities_Based_Planning_Overview_12_17.pdf

Dietz, T. and Stern, P. 1998. Science, values, and biodiversity, *Bioscience* 48(6):441-444.

DOT. 2002. DOT Performance Plan – FY 2003 and Performance Report – FY 2001 (2002). http://www.dot.gov/performance/pdf/environment.pdf, accessed 24 June 2005.

Environmental Protection Agency 2003. *Public involvement policy of the US Environmental Protection Agency*. EPA-233-B-03-002. Available on the web at: www.epa.gov/policy2003.Henry S. Cole Associates 1996. *Learning from success: health agency efforts to improve community involvement in communities affected by hazardous waste sites*. Atlanta, GA: Agency for Toxic Substances and Disease Registry.

Figueira J, Greco S, Ehrgott M. (eds.) 2005. Multiple criteria decision analysis: state of the art surveys. New York: Springer.

Frewer, Lynn and Gene Rowe. 2005. Evaluating public participation exercises: Strategic and practical issues. In: Joanne Caddy (ed.), Evaluating public participation in policy making. OECD: Paris, pp. 85-106.

Godschalk, D., Brody, S., Burby, R. 2003. Public Participation in Natural Hazard Mitigation Policy Formation: Challenges for Comprehensive Planning, Journal of Environmental Planning and Management 46 (5):733-754.

Grabowski MR, Harrald JH, Roberts KH 1997. Decision support and organizational forms in a high velocity environment: responses to catastrophic oil spills. In: Grabowski MR, Wallace WA, editors. Advances in expert systems for management, vol. 2. Greenwich: JAI Press.

Graedel TE, Allenby BR. 2002. Hierarchical metrics for sustainability. *Environmental Quality Management* 12(2):21-30.

Gregory, R. and Failing, L. 2002. Using decision analysis to encourage sound deliberation: water use planning in British Columbia, Canada, Journal of Policy Analysis and Management 21(3):492-499.

Hamlett, P. and Cobb, M. 2006. Potential solutions to public deliberation problems: Structured deliberations and polarization cascades, Policy Sciences Journal 34(4):629-648.

Hamlett, P. W. 2002. Adapting the Internet to citizen deliberations: Lessons learned. In Proceedings: Social Implications of Information and Communication Technology, IEEE International Symposium on Technology and Society (pp. 213-218). Raleigh, North Carolina: Institute of Electrical and Electronics Engineers.

Harrald JR. 1994. Preparing for success: A systems approach to pollution response. Paper presented to the *Five Years After the* Exxon Valdez *Oil Spill Conference*, Anchorage Alaska.

Harrald, J. and Conway, S. 1981. The use of decision support aids in the development of an action oriented contingency plan. Proceedings of the 1981 Oil Spill Conference, Washington, DC: American Petroleum Conference.

Harrald, J. and Mazzuchi, T. 1993. Planning for success: A scenario-based approach to contingency planning using expert judgment, Journal of Contingencies and Crisis Management 1(4):189-198.

Haynes, D. and Ott, G. No date. Area contingency plan scorecard: determining a plan's ability to achieve critical success factors. Paper presented at the International Oil Spill Conference. Washington, DC: American Petroleum Institute.

Kasperson, R. E. 1986. Six propositions for public participation and their relevance for risk communication, *Risk Analysis* 6(3):275-281.

Kasperson, R. E., Golding, D., and Tuler, S. 1992. Siting hazardous facilities and communicating risks under conditions of high social distrust. *Journal of Social Issues* 48(4):161-187.

Kim I. 2002. Ten years after the enactment of the Oil Pollution Act of 1990: A success or a failure. *Marine Policy* 26(3):197-207

Kuchin, J. and Hereth, L. 1999. Measuring response: A balanced response scorecard for evaluating success. Proceedings of the 1999 International Oil Spill Conference, Seattle, WA, pg. 685-690. Washington, DC: American Petroleum Institute.

Lahdelma R, Salminen P, Hokkanen J. Using multicriteria methods in environmental planning and management. Environmental Management. 2000; 26(6):595-605.Leach, William, Neil Pelkey and Paul Sabatier 2002. Stakeholder Partnerships as Collaborative Policymaking: Evaluation Criteria Applied to Watershed Management in California and Washington. *Journal of Policy Analysis and Management* 21(4): 645-670

Lindstedt-Siva, J. 1999. Judging oil spill response performance: The challenge of competing perspectives. Issue paper presented at the 1999 International Oil Spill Conference, Seattle, WA. Technical Report IOSC-008. Washington, DC: American Petroleum Institute.

Linkov, I., Satterstrom, F.K., Tkachuk, A., Levchenko, A., Seager, T., Figueira, J., and Tervonen, T. Forthcoming. An MCDA approach for selecting performance metrics: applications to Government Performance and Results Act (GPRA) and oil spill response. In Linkov, I., Wenning, R. (eds.), *Managing Critical Infrastructure Risks: Decision Tools and Applications for Port Security*. Springer, Amsterdam.

National Research Council 1996. Understanding risk: Informing decisions in a democratic society. Washington, DC: National Academy Press.

National Research Council 2000a. *International Conflict Resolution After the Cold War. Washington,* DC: National Academy Press

National Research Council 2000b. *Long-term institutional management of US Department of Energy legacy waste sites*. Washington, DC: National Academy Press.

National Research Council. 1996. *Understanding risk: Informing decisions in a democratic society*. Washington, DC: National Academy Press.

OMB. 2005a. Management/GPRA -- Government Performance Results Act of 1993. http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.html, accessed 17 May 2005.

OMB. 2005b. OMB Circular No. A-11. http://www.whitehouse.gov/omb/circulars/a11/current_year/s26.pdf, accessed 17 May 2005.

OMB. 2005c. Performance Measurement Challenges and Strategies – Draft. http://www.whitehouse.gov/omb/part/challenges_strategies.html, accessed 17 May 2005.

OPA90. Oil Pollution Act of 1990. Title 33 US Code 2702-2762. http://www.epa.gov/Region5/defs/html/opa.htm

Ornitz, B. and Champ, M. 2002. *Oil spill first principles: Prevention and best response*. NY: Elsevier.

Ott, G., Lindstedt-Siva, J., and Walker, A. 1993. Evolving spill management systems under OPA 90 could reduce response effectiveness. Proceedings of the 1993 Oil Spill Conference, Washington, DC: American Petroleum Conference.

Ott. G. 2005. Personal communication regarding efforts at the Coast Guard Training Center to measure how well an Area Contingency Plan might help an organization achieve success. Norfolk, VA: Office of Response and Restoration, NOAA.

Pond, R., Aurand, D., and Kraly, J. 2000. *Ecological risk assessment principles applied to oil spill response planning in the San Francisco Bay Area*. Sacramento, CA: California Office of Spill Prevention and Response.

Posavac, E. J. (1994). Misusing program evaluation by asking the wrong question. In C. J. Stevens & M. Deal (Eds.), *Preventing the misuse of evaluation* (pp. 69-78). San Francisco: Jossey-Bass.

Renn, O., Webler, T., Rakel, H., Dienel, P., and Johnson, B. 1993. Public participation in decision making: A three step procedure, *Policy Sciences* 26:189-214.Rowe, Gene and Lynn Frewer, 2000. Public participation methods: A framework for evaluation. *Science, Technology, and Human Values* 25(1):3-29.

Seager, T., Satterstrom, K., Linkov, I., Tuler, S., and Kay, R., Forthcoming. Typological Review of Environmental Performance Metrics (with Illustrative Examples for Oil Spill Response), *Integrated Environmental Assessment and Management*.

Seager, T., Tuler, S., Kay, R., Linkov, I. ,and Satterstrom, K. 2006. Environmental performance metrics for oil spill response. Technical report submitted to the Coastal Response Research Center, NOAA Grant # NA04NOS4190063. Project Number: 05-983. Greenfield, MA: Social and Environmental Research Institute, Inc.

Seager TP, Theis TL. 2004. A taxonomy of metrics for testing the industrial ecology hypotheses and application to design of freezer insulation. *Journal of Cleaner Production* 12(8-10):865-875.

Tuler, S., Kay, R., Seager, T., and Linkov, I. 2006a. *Objectives and performance metrics in oil spill response: The Bouchard-120 and Chalk Point spill responses*. Technical report submitted to the Coastal Response Research Center, NOAA Grant # NA04NOS4190063. Project Number: 05-983. Greenfield, MA: Social and Environmental Research Institute, Inc.

Tuler, S., Seager, T., Kay, R., and Linkov, R. Forthcoming. The Bouchard-120 and Chalk Point oil spill responses: Objectives and performance metrics. In I. Linkov, G. Kiker, and R. Wenning (eds.), *Environmental security in harbors and coastal areas*, pp. 209-228. Netherlands: Springer.

Tuler, S. and Kay, R. 2007. Comparing stakeholders' objectives for oil spill response: A Q study of Buzzards Bay and San Francisco Bay. Technical report submitted to the Coastal Response Research Center, NOAA Grant # NA04NOS4190063. Project Number: 05-983. Greenfield, MA: Social and Environmental Research Institute, Inc.

Aurand, D., Walko, L., and Pond, R. 2001*. Developing Consensus Ecological Risk Assessments: Environmental Protection in Oil Spill Response Planning. A Guidebook*. United States Coast Guard. Washington, D.C. 148p.

Webler, T. 1997. Organizing Public Participation: A Review of Three Handbooks. *Human Ecology Review* 3(1): 245-254.

Webler, T. 1995. 'Right' Discourse in Citizen Participation: An Evaluative Yardstick. In: O. Renn, T. Webler, and P. Wiedemann (Eds), *Fairness and Competence in Citizen Participation: Evaluating Models for Environmental Discourse* (Dordrecht: Kluwer Academic Press). pp. 35-86.

Webler, T. and Tuler, S. 2000. Fairness and Competence in Citizen Participation: Theoretical Reflections from a Case Study. *Administration & Society* 32(5): 566-595.

Wood, N., Good, J., and Goodwin, R. 2002. Vulnerability Assessment of a Port and Harbor Community to Earthquake and Tsunami Hazards: Integrating Technical Expert and Stakeholder Input, *Natural Hazards Review* 3(4):148-15