Establishing Performance Metrics for Oil Spill Response, Recovery and Restoration

A Final Report Submitted to
The Coastal Response Research Center

Submitted by

Drs. Seth Tuler, Dr. Thomas Webler, and Thomas P. Seager
Social and Environmental Research Institute
278 Main Street, Room 404
Greenfield, MA 01301
-and-
Igor Linkov
Intertox, Inc.
83 Winchester Street, Suite 1
Brookline, MA 02446

May 11th, 2007

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

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

Abstract

Establishing a systematic approach to assessing the effectiveness of a spill response is extremely challenging. Comparing responses to different spills can be even more daunting. 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. 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-government groups outside the command structure, including the media. 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. However, the question of how best to assess response successes and shortcomings *after actual events* has not received the same level of systematic attention among planners, although some frameworks have been proposed.

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, and provide information for legal disputes about responsibilities for impacts from the spill and response activities. 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 primary goal of this project was to develop a framework to guide the selection of metrics for assessing spill responses, prior to a spill response during response planning, by a broad range of potentially affected and interested parties.

To achieve this over-arching project goal we organized the project around a series of literature review and empirical research activities to achieve four objectives. The objectives were to:

1. characterize the objectives that stakeholders desire to achieve by oil spill response and upon which performance assessments should be based;
2. characterize the features of good performance metrics;
3. identify the kinds of performance metrics that have been proposed to assess oil spill response performance and the challenges of applying them in assessments of particular spill responses;
4. develop a framework by which performance metrics may be compared and selected.

The results of this project extend prior understandings and fill important gaps in information and guidance for spill response planning and assessment by a) providing empirical findings from multiple case studies about spill response objectives and performance metrics and b) developing a framework for identifying, discussing, and selecting performance metrics for oil spill response assessment that involves a wide range of interested and affected parties.

Keywords:
Oil spill response, Oil spill planning, performance metrics, response assessment, evaluation

Acknowledgments

Funding for this project was provided by the Coastal Response and Research Center, University of New Hampshire through NOAA Grant Number: NA04NOS4190063 and project Number: 05-983.

We benefited from the input of many individuals. We would like to thank the NOAA SSCs who talked about oil spill response with us, as well as the many individuals we interviewed in regard to the Bouchard-120 spill, the Chalk Point spill, and San Francisco Bay response planning. We especially would like to thank the SERI graduate student research assistant, Rebecca Kay, for her willingness to help out on all aspects of the project, good cheer, and quick learning. F. Kyle Satterstrom provided support to Igor Linkov.

Of course, the work is ours and any mistakes, mis-interpretations, and confusions are our responsibility.

Table of Contents

List of Figures and Tables

1.0 Introduction

As part of ongoing efforts to improve national, state, and local oil spill preparedness, extensive planning and coordinating efforts have been undertaken since the 1980s, culminating in the creation of the National Response Plan (DHS 2004, DHS No date, Caudle 2005). 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 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).

Establishing a baseline context to measure the effectiveness of a response is extremely challenging (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 (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-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 (Chess et al. 2005, Lindstedt-Siva 1999, Harrald 1994).

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.

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,

- justifying 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 2001, Abordaif et al. 1995, Harrald and Mazzuchi 1993). However, the question of how best to assess response successes and shortcomings *after actual events* has not received the same level of systematic attention among planners, although some frameworks have been proposed (Kuchin and Hereth 1999, Lindstedt-Siva 1999).

2.0 Objectives

The primary goal of this project was to develop a framework to guide the selection of metrics for assessing spill responses, prior to a spill response during response planning, 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:
1. the objectives that should guide spill response and
2. the criteria or metrics that should be used to assess performance with respect to the objectives.

To achieve this over-arching project goal we organized the project around a series of literature review and empirical research activities to achieve four objectives. As part of each objective we have prepared reports, book chapters, and articles. The objectives were to:
5. characterize the objectives that stakeholders desire to achieve by oil spill response and upon which performance assessments should be based (Tuler et al. 2006a; Tuler and Kay 2007, Tuler et al. 2007);
6. characterize the features of good performance metrics (Seager et al. 2006, Seager et al. Forthcoming);
7. identify the kinds of performance metrics that have been proposed to assess oil spill response performance and the challenges of applying them in assessments of particular spill responses (Tuler et al. 2006a, Tuler et al. 2007);
8. develop a framework by which performance metrics may be compared and selected (Linkov et al. 2007a, Linkov et al. Forthcoming, Tuler et al. 2007).

In the following sections we discuss the methods and results related to each of these activities.

3.0 Methods

A variety of methods were used to achieve the research objectives, including interviews, archival research, and Q method. Relevant literature was reviewed on various topics, as discussed below. Using these standard social science research methods several cases studies were conducted to explore objectives for oil spill response and performance metrics proposed to assess oil spill response performance and the challenges of applying them in assessments of particular spill responses. In this report we provide brief overviews of the research methods used to investigate each objective. Further details can be found in the technical reports and published book chapters and papers.

3.1 Approach to characterizing objectives and performance metrics (objectives 1, 2, an 3)

We conducted case studies of two recent oil spill responses and completed a literature review to address the objectives of 1) characterizing objectives that stakeholders desire to achieve by oil spill response, 2) characterizing good performance metrics, and 3) the types of performance metrics that have been proposed to assess oil spill response performance and the challenges of applying them in assessments of particular spill responses.

Our first two case studies were of:

1. the *Bouchard-120* spill response that began on April 27th, 2003 as the tugboat *Evening Tide* ran its tanker aground and released No. 6 home heating fuel just at the entrance to Buzzards Bay, Massachusetts and
2. the Chalk Point spill response that began on April 7, 2000 when an intrastate pipeline that transports oil from the Potomac Electric Power Company's (PEPCO) Chalk Point electrical generating facility to residents in Prince George's County, released No. 2 and No. 6 home heating fuel oil into Swanson Creek and subsequently the Patuxent River.

The purpose of the case studies was to better understand how people assess or judge performance of spill response efforts in particular cases. Understanding how people assess spill response efforts requires us to also understand their goals and objectives in oil spill response. To explore these issues in the context of the Bouchard-120 and Chalk Point spills and responses we conducted a series of interviews and investigated published reports to gather information about:

- roles and experiences of key responders and other interested and affected parties,
- their concerns about spill impacts and response,
- their views about the response, and
- their views about response performance metrics.

We conducted 12 interviews for the Bouchard-120 case study. In one instance two individuals were interviewed together. All interviews were conducted by either Seth Tuler or Rebecca Kay and lasted between 1 to 1.5 hours. Those interviewed represented a diverse group involved in the spill response, including federal, state, and local government officials, NGO representatives, and local community members. The Federal on-scene coordinator and State on-scene coordinators were included in the interview sample. Specific organizational affiliations included:

- US Coast Guard
- NOAA
- FWS
- Buzzards Bay National Estuary Program
- MA DEP
- Town fire departments
- Town Harbor Masters
- Town Environmental Affairs Coordinator

- Contractor for the responsible party

We conducted 18 interviews for the Chalk Point case study. In one instance two individuals were interviewed together. All interviews were conducted by Rebecca Kay and lasted between 1 to 1.5 hours. All except one were recorded digitally. Those interviewed represented a diverse group involved in the spill response, including federal, state, and local government officials, NGO representatives, and local community members. The Federal on-scene coordinator and State on-scene coordinator were included in the interview sample. Specific organizational affiliations included:

- EPA
- MDE
- MD DNR
- DOT
- Homeowner
- MD Board of Public Works administrator
- MD Department of Housing and Community Development
- FWS
- NOAA

Our approach to identifying the interviewees was a snowball sampling technique. In snowball sampling participants are not selected to be representative of a random selection of a population but instead are meant to represent cogent perspectives. We endeavored to select people that represent different points of view, have well-formed opinions based on their experiences with the spill, and were willing to be interviewed. We identified these people through our review of documents about the spill (i.e., official reports, newspaper articles) and suggestions by others. During each interview (or preliminary conversation with a potential interviewee) we also asked: "who else should we be talking with?" In this way, our interview sample was built as we went – this is the "snowball" approach.

A summary of each interview was written. Then, each digitally recorded interview was listened to (sometimes multiple times) where we then coded the interviewee's responses. We did not prepare verbatim transcripts of each interview in its entirety, but some limited transcription was made of key parts of selected interviews.

With many hours of recorded interviews, it was necessary to identify the information relevant to our guiding questions and interests. Toward this end, we used a grounded theory approach (Glaser and Strauss 1967). Grounded theory is a qualitative research tool. Moore (1996) and Tuler and Webler (1999) used this approach in their studies of participants' definitions of successful outcomes in risk and environmental assessment and decision-making. In grounded theory, important concepts emerge *inductively* during the data analysis rather than in advance of the investigation. Data are categorized with respect to relevant similar characteristics in a process called "coding." At first, a relatively large number of categories are developed. Then, through iteration these categories are grouped into more abstract categories of conceptual relevance to the analysis; data and categories are grouped according to their relationships with each other. For example, all statements related to "reducing bird injuries" or "protecting nesting habitat" can be grouped into a category named "response should protect bird populations." The primary reason for choosing this methodology is that it seeks to minimize researcher influence and instead let stakeholder concerns and possible performance metrics emerge directly out of the data themselves.

After coding the interviews we grouped the identified spill response objectives into more abstract categories; this is referred to as *axial coding* in a grounded theory framework. In this way common themes among the coded objectives are identified. We then extracted all performance metrics expressed by the interviewees that were related to each of the objectives. These performance metrics were then coded into the following categories (Seager et al. 2006, Seager et al., Forthcoming, Tuler et al. 2005, Tuler et al. 2007a):

- resource-based metrics: measures of the resources applied operationally to perform the activities,
- process-based metrics: measures of the quality of activities performed to achieve the desired outcomes, and
- endpoint-based metrics: measures of the state variable relating to a desired outcome or objective.

In addition, we compiled all interviewee comments regarding:

- appropriate uses of performance metrics for assessing oil spill response efforts and
- characteristics of "good" performance metrics.

This information was combined with insights gained from a literature review about performance metrics to develop an understanding of characteristics of good metrics for oil spill response (Seager and Theis 2004, Seager et al. 2006, Seager et al., Forthcoming).

We further considered the way that stakeholders *prioritize* objectives for spill response. Toward this end we used Q methodology (Brown 1986, 1996; McKeown and Thomas 1988; Niemeyer et al. 2005; Tuler et al. 2005, Tuler and Webler 2006). This is a type of discourse analysis that integrates quantitative and qualitative analyses to understand, in depth, the points of view on a subject. By inquiring of people with unique points of view, Q researchers can reveal patterns in how elements of perspectives are related.

To learn more about the ways that stakeholders prioritize different objectives we wanted to return to one of the initial case studies, as well as gather data related to a new case. Thus, we first returned to conduct a Q study in Buzzards Bay and we met with many of the same people that were interviewed initially as part of the Bouchard-120 case study. For our second case we wanted to gather information from people in a new region. This would allow us to test the robustness of the protocol that we developed for the Q study, including the applicability of the set of Q statements (see below). We approached selection of a new case in two ways. First, we identified those regions for which ecological risk assessments had been recently completed (i.e., within the last several years). Second, we asked CRRC staff for assistance to identify a point of contact that could help us identify research participants efficiently. Ultimately, this led us to San Francisco Bay. An ERA was completed for this region in 2000.

We selected individuals to participate in our research who have been actively involved in spill response planning and implementation; represented different institutional affiliations; and, most importantly, were likely to have different views about spill response objectives.

In the Buzzards Bay case we were familiar with government officials and regional and local stakeholders from our earlier work and the masters thesis work of our graduate student research assistant. 16 people completed Q sorts; twelve we had interviewed earlier as part of our initial case study effort (Table 1). Data for this case were gathered during August – September, 2006.

In the case of San Francisco Bay we discussed our needs with Jordan Stout, NOAA Scientific Support Coordinator (SSC) for the region. He helped us by providing background information about the region and spill response planning and by identifying a diverse group of people to include in our Q study. Thirteen people completed Q sorts, as shown in Table 2. Data for this case were gathered during November 2006.

Then, the identified individuals were approached via telephone and introduced to the project and told how they were selected. We described our data collection procedures and what we wanted them to do. We told people we would visit them at a time and place convenient to them and that the entire process would take about one and one-half hours.

Table 1. Affiliations of participants in the Buzzards Bay Q study

- Environmental educator/advocate
- Harbor master
- Local environmental planner
- Local fire chiefs (2)
- Local residents (3)
- Marine Spill Response Corporation
- NOAA
- Regional environmental planning organization (2)
- State agency staff / first responder
- USCG
- USFWS (2)

Table 2. Affiliations of participants in the San Francisco Bay Q study

- BLM's California Coastal National Monument
- CA Coastal Commission
- CA Office of Spill Prevention & Response
- Chevron Energy Technology Company
- Fish & Game NRDA Specialist, OSPR
- Fish & Game NRDA Specialist, OSPR
- Gulf of the Farrallones, National Marine Sanctuaries

- Marine Spill Response Corporation
- NMFS, NOAA Damage Assessment Center
- NOAA
- NOAA's National Marine Fisheries Service
- Salmon & Crab Fisherman
- USCG, District 11 - Response

In Q methodology, the researchers gain access to various perspectives on a subject – what Q practitioners often call "social discourses" – by having a small number of people with different, but well-formed opinions sort a group of statements according to their personal opinions. Participants in our case study were handed a set of small cards (about the size of a normal business card). Each card had a statement printed upon it that described a single objective that might be important to a spill response.

The statements sorted by the participants were chosen by the research team to represent the fullest possible extent of content relative to the topic.[1] As part of our case studies about the Bouchard-120 and Chalk Point spills we identified objectives that research subjects cared about in those spill responses (see Tuler et al. 2006a, Tuler et al. 2007a for more details). In addition, we completed a review of literature about spill response planning (e.g., Baker 1999, Kuchin and Hereth 1999, Ornitz and Champ 2002, Ott 2005, Pond et al. 2000, USCG 2005). We then created statements describing these objectives, by sampling from the quotes we extracted from the interviews and literature. Ultimately, we ended up with 42 statements, as shown in Table 3. It was essential that these statements capture the full range of objectives that might be important in any spill response. In other words, we were seeking a set of statements that could be used to study perspectives about spill response objectives in any region.

A condition of instruction specified the context under which the participant was to interpret and react to the Q statements. In both cases the condition of instruction was:

> When you think about past oil spills, what do you think should be the objectives that guide responses to future oil spills in this area? Sort the statements to indicate what <u>you</u> would be *most unlikely to emphasize (-4)* to *most likely to emphasize (+4)* in a future response.

This condition of instruction was designed to focus the participants' thinking on the emergency phases of spill response; we asked them not to consider these objectives in latter response efforts, including restoration and damage assessment. We wanted to draw on each participants' experiences to-date and at the same time get his or her ideas of what would be the most important objectives to guide a response in the future. We did not ask people to evaluate, for example, the Bouchard-120 spill response, although we expected, of course, that their experiences would inform their ideas about a future effort.

The analysis that is part of Q method reveals both the content of the social discourses present in the group of participants and the extent to which particular individuals believe or subscribe to the

[1] It is important to note that in a Q study the sample is *not* the people who sort the statements; rather, the sample in a Q study is the set of Q statements, the population is the "concourse" of utterances that have been made on the topic, and the completed Q sorts are the variables. This is just the opposite of standard survey techniques.

different discourses. The assumption is that these social discourses exist partially in the subjectivity of individuals, but they are also a product of social interaction. In addition, while perspectives are held subjectively, similarities among individual views make it possible to articulate a small number of social discourses on a topic.

We arrive at the meaning of each of the social discourses that emerges from the analysis by using three approaches. First of all, Q sort data were entered into a computer program called MQMethod.[2] This program computes the factor analysis that is used to identify the basic features to the perspectives expressed by the research participants.. Second, we recorded the conversation we had with the participant during the sort. We asked the participant to interpret their sort and to explain how he or she interpreted specific Q statements. We used their comments to help interpret the statistical output when composing the perspective narratives. Third, we mailed a narrative description of each social discourse to participants whose individual sorts were most strongly correlated with it. That is, we endeavored to find participants who were most representative of the perspective represented by the social discourse and then asked each of them to verify its clarity, content, and emphasis.

[2] This freeware program is available through http://www.qmethod.org. Readers interested in learning more about Q method will find this website informative.

Table 3. List of 42 statements used in the Q sorts.

1. Economic impacts to towns from costs of clean-up should be mitigated.
2. Consumption of contaminated seafood should be prevented.
3. Get on with response efforts early for areas that have been pre-identified as sensitive areas.
4. Even if a species is not native to this area, mitigate impacts to the local population.
5. When faced with a spill, it is most important to protect the adults of a species at risk because the adults can come back next year and reproduce.
6. Give priority to protecting those areas that have multiple resource values, like those that are undeveloped, pristine, and that provide for recreation.
7. Attention should be focused on protecting species that are especially critical for the functioning of an impacted ecosystem.
8. Health and ecological impacts from clean-up activities should be mitigated.
9. The response should 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.
10. Damage to cultural artifacts (e.g., shipwrecks) from oil and its clean-up should be prevented.
11. Economic impacts from lost recreation should be mitigated.
12. Inconveniences to local residents and tourists should be mitigated.
13. The economic impacts to local commercial fishermen should be reduced, including impacts that might arise from people's perceptions (for example, about shellfish tainting).
14. There should be no situations that threaten human health whatsoever during the response.
15. Costs to the responsible party resulting from the response should be minimized.
16. Getting clean-up contractors on-scene should ramp up quickly, even if there is uncertainty about how many gallons have been spilled.
17. The clean-up should address aesthetic concerns – like oil stains on rocks.
18. Get a good estimate of the amount of oil spilled.
19. Local responders/leaders should be integrated quickly into response planning because of their knowledge of local conditions, resources, etc.
20. Coordination among participating government agencies, contractors, etc. should be established rapidly.
21. Establish meaningful ways of involving volunteers in the response.
22. A well-organized unified command with a clear chain of command should be established.
23. A well-coordinated expert scientific effort should drive the gathering of decision-relevant information, not public concerns and perceptions.
24. Implement the contingency plan.
25. There should be no residual oil or buried oil that is going to show up later.
26. Restoration planning should be tightly integrated with the response effort so that decisions are based on future restoration needs.
27. As much on-water recovery and removal of oil as possible should be achieved.
28. Conduct monitoring of response activities, such as booming, to actually see whether things are working.
29. Clear definitions of what counts as "clean" should be used so that there is a clear end-point.
30. Tell members of the public about the things they want to know about.
31. Responders should listen to the publics' concerns, even if they cannot be addressed to their complete satisfaction.
32. Unified Command should gain public support for the response effort.
33. Unified Command should develop and maintain trust with members of the public.
34. Efforts to communicate with and engage with the community should be proactive and timely.
35. Consistent and accurate information should be provided to the public.
36. Response efforts should direct oil to a "sacrificial area" – such as a sandy cove that will be easier to clean-up than other, more rocky areas.
37. Unified Command should reconcile the preferences and points of views of all parties about what impacts are important to avoid.
38. Unified Command should manage expectations about the clean-up so that they are reasonable.
39. Responsible authorities should assign flexible and experienced decision makers – who can implement contingency plans right away and then step back and ask "what do we need?"
40. Response efforts need to avoid disrupting the integrity and culture of local communities.
41. Subsistence fishing and shellfishing areas should be protected.
42. Make determinations of "clean" with relevant stakeholders, including local residents.

3.2 Approach to proposing a collaborative process by which performance metrics may be compared and selected (objective 4)

The goal of this activity was is to propose a process design that can be used to 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:

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

The design of the process was informed by:

- a review of prior research on collaborative decision-making about environmental risk assessment and management (Renn et al. 1993, Webler and Tuler 1999, Webler 1995, National Research Council 1996),
- consideration of the role that can be played by multi-criteria decision analysis (MCDA) in the selection of performance metrics (Lahdelma et al. 2000, Linkov et al. 2006, Figueira et al. 2005), and
- outcomes from other research activities conducted as part of this research project (discussed above).

4.0 Results

In this section we summarize the results from our four research activities. We discuss them in regard to:

- objectives for oil spill response, including differences in priorities among stakeholders' objectives,
- performance metrics proposed to assess oil spill response performance,
- characteristics of good performance metrics and the challenges of applying them in assessments of particular spill responses, and
- framework by which performance metrics may be compared and selected.

4.1 Objectives for oil spill response

Several objectives for oil spill response emerged from the Bouchard-120 and Chalk Point case studies, as shown in Table 4 (see Tuler et al. 2006a, 2007a for further details). Many of the same objectives were identified among the respondents in both cases. They follow directly from the goals of oil spill response as expressed in federal policy:

- 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 (from http://response.restoration.noaa.gov/book_shelf/910_response.pdf, pg. 9).

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

In addition to similarities on the broader categories of objectives, many of the more specific objectives are shared. For example, interviewees in both cases wanted to protect sensitive habitats and populations of threatened and endangered species. They were concerned with mitigating impacts from clean-up actions. They want to mitigate economic impacts related to lost recreation, tourism, fisheries, and to towns from their efforts during the response. Furthermore, they were concerned about the timely gathering and use of relevant, accurate, and credible information for decision making. Strong and flexible leadership that can learn from past experiences was important to them so that the response could be well-planned and implemented. And, they shared objectives related to addressing public concerns (e.g., providing accurate information to the public).

At the same time, there are a few interesting differences, which are reflective of the particular contexts of each spill.

First, mitigating impacts to cultural resources was identified in the Chalk Point case and not the Bouchard-120 case. This reflects the presence of significant artifacts in the region affected by the Chalk Point spill. It may also be a reflection of whom we interviewed – or more accurately, did not interview in the Bouchard-120 case. However, it is noteworthy that no one in the Bouchard-120 case made mention of historical or other culturally significant artifacts.

Second, the mitigation of social nuisance impacts was not a category of objectives that emerged from our analysis of the Chalk Point interviews. However, some factors that can be related to social nuisance impacts were raised in the context of other objectives. For example, addressing the potential for an "increase in crime rate [in southern MD due to influx of people from cleanup crews]" might be considered by some to be a social nuisance, but the interviewee was clearly talking about this in the context of economic impacts.

Third, while people in both cases expressed objectives (and performance metrics) related to establishing a coordinated and effective response framework, there were some differences in emphasis. These differences suggest the importance of the particulars of experience that inform people's views. Critical comments emerged about the integration of non-federal officials and responders into the response effort. In the Bouchard-120 case this concern was raised about local officials and local first responders, but not in regard to state-level responders. In the Chalk Point case we found the opposite.

In the Bouchard-120 interviews the roles and participation of local officials and local residents was a concern. In fact, the way in which local responders along Buzzards Bay were contacted and integrated into the response was a subject of contention and criticism among those we interviewed. Location of the unified command was an example of ways that local officials thought they were not included to the extent that they thought desirable. In addition, there was a concern with how local volunteers were brought into the response effort in the Bouchard case.

On the other hand, among the Chalk Point interviewees neither of these two concerns were strongly articulated; volunteers were not even mentioned in these interviews. Instead, there was a concern with whether expert input (e.g., about cultural resources) was obtained and how well federal and state agencies' were coordinated (e.g., clarity about jurisdictions). In the Chalk Point case a federal official spoke of the importance of involving local stakeholders in decisions about clean-up goals, but the metrics suggested were about one-way communication and outreach to the public (e.g., number of meetings). There was also a concern expressed from *state* officials about their involvement in the response effort. In the Chalk Point case there was much said about the quality of leadership (FOSC) and the coordination and communication among responding federal and state agencies. There were some suggestions that involvement of local officials would have been useful in Chalk Point in terms of understanding local nuances, such as currents.

Another difference among objectives relates to the kind of people that expressed a concern with gaining public support for the response. Only state and federal officials spoke of this as an objective. This is a very instrumental perspective that one often finds from officials involved with hazard management.

Finally, it is apparent that objectives can be in tension. Responders can have a difficult time balancing the ways that they strive to achieve goals and objectives. This was expressed clearly by a federal official involved with the Bouchard-120 response. Using a "best response" framework helps the Unified Command simultaneously focus on all the critical elements so that none are neglected: "As part of the best response we don't get stuck on one area or issue."

One interviewee stated that directing oil to "sacrificial areas" is more important to do well than preventing oiling anywhere:

> Direction I think is more important than stopping. If you can direct it into a "sacrificial area" send it into a cove that yes will have an environmental impact but you will save this amount of other shoreline. I think there's an indication of a more approachable loop than having this visual image of "I'm going to stop it all."

However, trade-offs can be made differently. For example, another interviewee privileged marshes and other sensitive areas saying:

> Sandy recreational beaches, it's hard to compare say to a salt marsh because it depends what perspective you are coming from, from an environmentalist perspective… I would say the salt marsh is much more important to protect… its also important to protect those sandy recreational beach areas in part because that's what connects people to the resource

> and imbues them with more of a stewardship ethic and it ties into the public trust doctrine we all commonly own the ocean's resources…at least our public beaches and I think its really important to maintain that connection.

As another example, an interviewee from the Bouchard-120 case expressed the objective of minimizing the costs to the responsible party from the response: "be fair to the responsible party. Don't over-burden responsible party."

Another person spoke of this being an objective that might conflict with other important goals. When gallons spilled is used to determine size of initial response, it can result in too slow of a response:

> Because they were unsure about how many gallons had been spilled, the spin-up getting the clean-up contractors there on-scene started out pretty slow. Again, the CG is pretty responsive to the RP and they don't want to mobilize an entire army if it turns out that the spill isn't as big then you've spent all this money when you didn't have to…[later] the CG admitted that, that the response spin-up time was actually slower than it should have been.

While these examples show how there can be tensions among the objectives, we also found that objectives can be differently weighted. Based on our interviews it appears that some people believe there is a "pecking order" of objectives. Protection of historical underwater sites may not be the most important – there is need for trade-offs and making choices about priorities:

> The whole point of a contingency plan is to prioritize; if they say, 'ok, we need a point from which to deploy boom and this one is most logical and there's a site there, is it sufficiently significant to ask them to move or can you let that one go.'

One arena in which the tension among objectives is most readily apparent relates to objectives of mitigating costs of the response, making informed decisions based on relevant information, and addressing public concerns and need. In some cases, such concerns were dealt with as requested, but in others not:

- we had a rule of thumb that in rocky areas we would be better off letting some natural cleaning to go on, so we were washing rocks but not high powered steam cleaning – but our rule was to let some natural processing go in certain areas. Well, there is rock in a bay and the community said that is the swim test rock. We were not going to clean it…all our children who learn how to swim have to swim to that rock…oh so you need that rock cleaned and we will do it. [Another example is that]…we were trying to communicate with some communities that natural processing is the best alternative, and this man wrote that I am 75 and I don't have a lot of years left to watch natural processing and I waited all my life to buy this waterfront property – I did the best I could to listen to every person. We did not do anything to clean up this man's property any better. We just cleaned it up [as we did other locations]. We just cant clean up someone's beach better than others.
- The rumors were that there were competing interests for these cleanup crews…to some degree CG wants to be responsive to the complaints of the public… clean-up crews were so concerned to not damaging the habitat on Ram island… that when they first sent their

> crews in they were hand-scrubbing the rocks…we finally told them that its more important that you get it clean cause they were afraid [of damaging the ecosystem].

These objectives identified as important during the interviews are all related to the protection and promotion of what people value, such as protection of critical habitats and promotion of decisions based on the best information available. *Why* they value certain things can differ; they may value certain outcomes for intrinsic reasons (e.g., value of species for their own sake) or instrumental reasons (e.g., they allow other things to be accomplished) and the relative weights given to each may also differ.

To explore whether objectives could be prioritized differently by different stakeholders or in different regions, we conducted the Q studies in Buzzards Bay and San Francisco Bay.

4.1.1 Priorities among spill response objectives in Buzzards Bay, Massachusetts

In our further study of Buzzards Bay stakeholders three distinct and coherent perspectives emerged from the analysis; we call these Perspectives A, B, and C. Each of the three perspectives represents a distinct view about the relative importance of objectives that should guide oil spill response in Buzzards Bay. Of course, these perspectives share some features, while still having some important differences.[3] Detailed presentation of the three perspectives can be found in Tuler and Kay 2007.

Perspective A highlights the need for a well-organized response system. It should ensure good coordination (20), integrate local responders (19), have a clear chain of command (22), get clean-up crews on-site rapidly (16), and initiate efforts rapidly (3). On the other hand, objectives related to economic costs (1, 11), aesthetic concerns (17), and inconveniencing local residents (12, 40) were ranked low. Objectives related to public health and ecological impacts were important (2, 8, 7, 6, 36, 14) but not as important as objectives related to response organization – nor did they receive the same relative emphasis as they did in Perspectives B or C.

Objectives that are articulated in policy (e.g., National Response Plan) were rated high in Perspective B (statements 14, 2, 7, 8, 41); these are the statements regarding protection of public and worker health and that special attention should be focused on protecting species that are especially critical for the functioning of an impacted ecosystem. In addition, those loading highly on this factor placed a high priority on implementing the contingency plan (24); Perspective A, instead, rates more highly the need for people that can be flexible (39). This perspective also shares the concern that there be a clear chain of command (22), that as much oil as possible be recovered off-shore (27), and the need for the ramping up response quickly (16). Perspective B does not represent a perspective that places a high premium on science to guide the response effort (23), relative to Perspective A or C. The importance of flexible leadership (39), coordination (20), and integration of local responders (19) are all ranked much lower than for Perspective A.

---

[3] In the following discussion, numbers in parentheses refer to statement numbers shown in Table 3.

The perspective represented by Perspective C rates some issues with a different priority than the other two factors. Monitoring of activities (28), providing accurate and consistent information to the public (35), and integrating restoration planning with clean-up activities (26) are much more important in this factor. However, Perspective C shares with Perspective A the objective for good local responder integration into the effort (19); this may be a result of conflicts that arose among federal, state, and local responders about the notification and integration of local first responders immediately after the spill occurred (see Tuler et al. 2006a). Their concerns about impacts are mostly focused on ecological impacts (7, 6). They are unconcerned about threats to public health (14), as well as aesthetic concerns (17), costs to the responsible party (15), inconveniences to local residents (12), and costs to local towns (1). Science as a guide to decisions is most important in this perspective (23).

This brief overview has focused on the relative rankings that particular statements received in each of the three perspectives. In Table 5 we compare the relative emphasis that is given by each of the three perspectives to the different types of objectives; these objectives were found to be important in our case studies of the Bouchard-120 and Chalk Point spills and they were the primary basis for the selection of the Q statements. In addition, we have included two other statements for comparison:

- Implement the contingency plan.
- A well-coordinated expert scientific effort should drive the gathering of decision-relevant information, not public concerns and perceptions.

For each type of objective the z-scores of the statements associated with that objective were summed. Z-scores are a measure of the relative importance of a statement in a factor emerging from the factor analysis. These data, then, reveal the relative importance of a particular objective across the three perspectives. The data also provide some information about the relative importance of different types of objectives within a particular perspective. However, these data should be used cautiously, as a) there are different numbers of statements that define each objective and b) z-scores can be both positive and negative, so that sums of z-scores can mask how some statements are ranked relative to others (e.g., statement #35 is ranked highly in Perspective C, although all of the other statements relating to the objective of addressing public concerns are emphasized more weakly).

Table 5. Summed z-scores for statements defining types of objectives in the Buzzards Bay case.

Type of objective	Persp. A	Persp .B	Persp .C
Protect worker and public health and safety	1.555	4.753	-1.678
Protect environment and mitigate environmental impacts	2.412	4.274	2.968
Mitigate economic impacts	-5	-2.198	-4.474
Protect cultural resources	-1.874	-1.468	-0.43
Mitigate social nuisance impacts	-2.954	-1.987	-3.011
Gain public support for the response	0.071	-2.048	-0.172
Address needs and concerns of the affected public/communities	-0.638	-4.813	-0.602
Establish a coordinated and effective response framework	6.53	5.253	2.968
Implement an effective and timely response	6.135	4.157	0.904
Implement the contingency plan.	0.268	1.394	-0.602
A well-coordinated expert scientific effort should drive the gathering of decision-relevant information, not public concerns and perceptions.	0.832	-0.943	0.99

Inspite of these caveats, the data in Table 5 reveal important similarities and differences among the perspectives:

- Perspectives A and B place much more emphasis on the coordination and timeliness of the response effort, relative to Perspective C;
- mitigating economic impacts are relatively unimportant compared to mitigating ecological impacts;
- addressing public concerns or gaining public support are much less of a priority in Perspective B then in the other two perspectives;
- Perspective B places more emphasis on following the contingency plan than Perspectives A or C; and
- having a well-coordinated and expert scientific effort that drives the gathering of decision-relevant information is important in Perspectives A and C, but not in Perspective B.

These differences cannot be simply attributed to the affiliations of the individuals who load highly on the various factors:

- almost all of the local responders included in our sample of research subjects loaded highly on Perspective A;
- both federal officials *and* local residents loaded highly on Perspective B; and
- Perspective C represents a perspective that includes a federal official and an NGO environmental advocate.

4.1.2 Priorities among spill response objectives in Buzzards Bay, Massachusetts

Three distinct and coherent perspectives emerged from the analysis; we call these Perspectives D, E, and F. Each of the three perspectives represents a distinct view about the relative importance of objectives that should guide oil spill response in San Francisco Bay. Of course, these perspectives share some features, while still having some important differences.[4] Detailed presentation of the three perspectives can be found in Tuler and Kay 2007.

Perspective D emphasizes very strongly (relative to the other two factors) the need for a well-organized response system. It should ensure good coordination (20), integrate local responders (19), have a clear chain of command (22), get clean-up crews on-site rapidly (16), and initiate efforts rapidly (3). On the other hand, objectives related to mitigating impacts to human systems were ranked low (1, 11, 12, 13, 17, 40). Objectives related to ensuring public and worker health and safety and to mitigating ecological impacts were ranked as important in this Perspective; most emphasis is given to protecting key species and sensitive areas that have been identified during pre-spill planning (3, 24, 7).

The concern for mitigating impacts to ecological systems is shared by Perspectives E and F as well. But this general statement can mask some important differences. In particular, Perspective E represents a view that mitigation of impacts is not only a function of the amount of oil that is removed or remains. It is also a function of *how* the oil is removed. Thus, for example, it might be better to avoid using dispertants or *insitu* burning. Furthermore, there is additional emphasis on the need to protect adults of a species that can continue to reproduce in the following years (5). Interestingly, this also appears to have an influence on the relative importance of worker and public safety in the overall response effort: those loading on Perspective E believe that oil spill response is inherently risky so that avoiding *all* situations that threaten human health whatsoever during the response (14) is not a reasonable objective.

In addition, Perspectives E and F also place a strong emphasis on the need to protect the human components of the impacted region (e.g., statement 13). This issue does not emerge as a strong concern in Perspective D. Those ascribing to Perspective E express concern with long-term impacts that may result from "hidden" changes; such as those to lower levels in the food chain that may cause negative impacts to resources that are important ecologically and economically (commercial fisheries). That is why Perspective E does not emphasize that there should be as much on-water recovery and removal, as this can require use of dispersants that harm plankton (27) or that there should be no residual oil or buried oil (25). This is also suggested by Perspective E's emphasis on reducing economic impacts to local fishermen (13).

Those who load on Perspectives E and F give more emphasis to the objective that the response should 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 (9). Again, this reflects their concern with the impacts to human components of the impacted region (i.e., local economies,

[4] In the following discussion, numbers in parentheses refer to statement numbers shown in Table 3.

culture). Perspective F appears to place somewhat more emphasis on avoiding potential impacts to recreation (11, 12) – which can be important to the local economy – and to subsistence fishing (41), relative to Perspective E. This may be because those individuals who load on Perspective E have a different sense of the importance of these than do those who load highly on Perspective F.

Finally, the way that decisions about response actions are made is another area of significant disagreement between Perspectives D and E. Those ascribing to Perspective D believe that the response should closely follow the contingency plan (24), while at the same time having leadership that is flexible and experienced (39). Those ascribing to Perspective E suggest that the contingency plan should not be implemented blindly. Case specific contingencies need to be accounted for, and the best way to do this is with a well coordinated expert scientific effort (23) and clear definitions of what end-points are desired (29); these end-points may not be clearly spelled out in the contingency plan. It is better to have responders who ask: what do we need here? Now?

This brief overview has focused on the relative rankings that particular statements received in each of the three perspectives. In Table 6 we compare the relative emphasis that is given by each of the three perspectives to the different types of objectives, using the same approach as described above.

Table 6. Summed z-scores for statements defining types of objectives in the San Francisco Bay case.

Type of objective	Persp. D	Persp. E	Persp. F
Protect worker and public health and safety	1.879	-1.803	3.154
Protect environment and mitigate environmental impacts	3.312	3.604	4.055
Mitigate economic impacts	-4.82	-0.001	0.901
Protect cultural resources	-1.031	-0.451	-0.451
Mitigate social nuisance impacts	-3.407	-2.253	-0.901
Gain public support for the response	-0.639	-1.352	-0.902
Address needs and concerns of the affected public/communities	-1.501	-0.45	-4.506
Establish a coordinated and effective response framework	7.681	1.353	2.254
Implement an effective and timely response	5.48	1.803	1.353
Implement the contingency plan.	1.657	-1.802	0.451
A well-coordinated expert scientific effort should drive the gathering of decision-relevant information, not public concerns and perceptions.	0.087	1.802	-0.451

Inspite of these caveats, the data in Table 6 reveal important similarities and differences among the perspectives:

- Perspective D places much more emphasis on the coordination and timeliness of the response effort, relative to Perspectives E and F;
- Perspective D places more emphasis on following the contingency plan than Perspectives E or F;
- Mitigating economic impacts are relatively unimportant compared to mitigating ecological impacts in Perspectives D. Perspective E gives significant weight to the objective that the economic impacts to local commercial fishermen should be reduced, including impacts that might arise from people's perceptions (for example, about shellfish tainting) (13), while Perspective F gives significant weight to this objective as well as the objective that economic impacts from lost recreation should be mitigated (11);
- Addressing public concerns or gaining public support receive little emphasis in all three perspectives, but especially so in Perspective F. Objectives related to public communication, trust, support, etc. are ranked as mid to low level of importance in all perspectives. These are statements 30, 31, 32, 33, 34, 35, 37, 38, 42. As one person completing the Q sort stated, "why do responders care about trust from the public? Does it help clean up the oil? No…the public is not so critical to getting the job done;'" and
- Having a well-coordinated and expert scientific effort that drives the gathering of decision-relevant information is important in Perspective E, but not in Perspectives D or F.

These differences cannot be simply attributed to the affiliations of the individuals who load highly on the various perspectives. 11 out of 13 of the individuals completing the Q sorts about San Francisco Bay loaded on Perspective D. Only one person each loaded on Perspectives E (a local fisherman) and F (a state NRDA specialist). At the same time, we cannot discount Factors E and F as unimportant or marginal, although each of these factors is only defined by a single individual. However, because we used Q method we do not know anything about the relative frequencies of the perspectives represented by the factors in the population.

It is interesting to speculate on why this may be so. At the start it is important to note that this may be an artifact of the individuals we had complete the Q sorts. We may not have chosen a very good sample to represent the diversity of perspectives about relative importance of objectives that should inform response in this region. While we asked the regional NOAA SSC for input about whom to contact, we should not expect that he would have knowledge of people's specific points of view.

On the other hand, the degree of consensus may in fact be real. A high level of consensus about objectives may be a result of recent planning efforts in this region, such as the 2000 ecological risk assessment (Pond et al. 2000). However, this claim must be balanced against the information that only *two* of the participants in our research attended workshops for the ecological risk assessment (see Pond et al., 2000 Appendix A); both of these were high loaders on Factor D. We do not know how the fact that many of the others involved in our study work at the agencies and organizations whose representatives participated in the ecological risk assessment affects the perspectives expressed to us via the Q sorts.

In addition, there has been no recent large-scale oil spill in this region that would raise questions about the contingency plan and general agreement about objectives. Such disjunctures have been discussed in the literature on disaster planning and hazard management. For example, Clarke (1999, pg. 2) has written, with one of his case studies about the Exxon Valdez response, that "organizations and experts use plans as forms of rhetoric, tools designed to convince audiences that they ought to believe what an organization says. In particular, some plans have so little instrumental utility in them that they warrant the label 'fantasy document.'" This is, in fact, the critique of the high loader on Perspective E. Without making too strong a claim about the adequacy of the contingency plan for San Francisco Bay (which we did not attempt to evaluate in any way), we also note that in a region like southern New England where there have been two more recent large spills (Naragansett and Buzzards Bays), any consensus that might have emerged from the planning process (contingency planning) may have been ruffled by actual experiences – such as integration of local responders into the response effort (as reflected in factor loadings for the Buzzards Bay case).

Overall, the Q study reveals 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 – even while they can all agree with higher order goals as expressed in policy and statute. Some general observations about the two cases are that:

- While protection of health and safety was important to all perspectives represented by the perspectives, in each case there is one that emphasizes health to a much greater extent than the others. In addition, in each case there was a perspective that suggests that protection of health and safety must be balanced with other objectives – oil spill response is dangerous work and to achieve other objectives may entail putting people in some risky situations (Buzzards Bay Perspective B and San Francisco Bay Perspective F);
- Mitigating ecological impacts is emphasized in each case, but not equally by all perspectives in each case;
- Mitigating impacts to cultural resources was not very salient in either case – perhaps because they were not so important as in the Chalk Point spill. On the other hand, mitigation of socio-cultural impacts, such as those that could occur to subsistence fishing, was ranked important in some perspectives;
- Mitigating economic impacts was not a high priority to any perspective – but it was especially de-emphasized in Buzzards Bay. On the other hand, in two of the San Francisco Bay Perspectives (E and F) concern was raised about impacts to local fishermen (statements #11). While tourism and recreation are arguably important in both areas studied, impacts to local recreation was only emphasized in one factor (San Francisco Bay Perspective F);
- There was no support for minimizing costs to the responsible party resulting from the response in any of the perspectives; even though this was expressed as a concern because costs must be justified. Similarly, a requirement of OPA 90 is that sites affected by the spill must also be returned to their conditions prior to the spill; support for this objective (statement #9) was weak in all but two of the perspectives (San Francisco Perspectives E and F);
- Objectives related to addressing public concerns were almost universally ranked by respondents as objectives they would be unlikely to emphasize. There were two exceptions. In Buzzards Bay Perspective C it was very important to ensure that

consistent and accurate information be provided to the public (32) and in San Francisco Bay Perspective E it was important that Unified Command manage expectations about the clean-up so that they are reasonable (38). The relative lack of emphasis on objectives related public opinion and public satisfaction stands in contrast to the emphasis that they were given in our case study interviews (Tuler et al. 2006a) and in the literature (e.g., Lindstedt-Siva 1999); and

- In most factors the role of scientific analysis (statement #23) was emphasized when following the contingency plan (statement #24) was not – and vice versa. There was a negative relationship between the two in five of the six perspectives.

There are several limitations to Q studies in general and to our studies in particular. First a limitation of Q method is that it cannot tell us:

- that these are the only perspectives that exist within the populations of interested and affected parties in the Buzzards Bay and San Francisco Bay regions (i.e., we may not have included people in the study with other points of view); and
- the frequency of these perspectives within a population. This is why the two perspectives that are defined by single Q sorts (individuals) in the San Francisco Bay case should not be discounted. They may be important – and the perspectives they define may be ascribed to by additional people that did not participate in our study.

Second, during the initial Q sorts we felt that respondents did not always understand our interest in the relative priority of objectives during the emergency phase of spill response. We discovered that our condition of instruction and explanations were not always adequate to ensure that we created this context for their responses. Thus, in some Q sorts it is difficult to disentangle what aspect of response individuals were thinking about. However, our notes about respondent comments during the Q sorts were helpful to some extent to make sense of this issue.

Third, we the two case studies provide only initial evidence about the adequacy of the 42 Q statements for describing the full range of objectives that interested and affected parties might believe are important for the emergency phases of oil spill response. To test the adequacy of this set, after all Q sorts we also asked individuals if there were any important objectives missing from the set of statements that they thought should have been included. In the San Francisco Bay case none were suggested. In the Buzzards Bay case three gaps in the statements were suggested: a) the importance of meeting policy and regulatory requirements, b) addressing political aspect of response, and c) getting an estimate or a number (during the response) of the amount of oil that hit the shoreline so that this could inform damage assessment later.

4.2 Performance metrics proposed to assess oil spill response performance

Often a distinction is made between goals, objectives, and performance 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.[5]

[5] 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 2001,

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; in other words, be assessed relative to different performance metrics. For example, whether or not a sensitive species has been protected may be characterized by the performance metrics of:

- 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.

In an actual spill these metrics will have certain values, or measures.

Many federal agencies use performance metrics 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.[6] 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).

Kuchin and Hereth 1999). Despite some terminological differences, the challenges of defining, selecting, and measuring them are similar.

[6] 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).

While many federal agencies have developed ecological performance metrics, few specifically address oil spill *response* or *impacts*. On the other hand, 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 2001, 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 particular format, rather than their capacity to ensure that desired outcomes are achieved or tasks accomplished (Abordaif et al. 1995, Haynes and Ott 2001).

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).”

Our case studies of the Bouchard-120 and Chalk Point spill responses reveal a diversity of performance metrics that may be used to assess spill response. While some of them would satisfy Kuchin and Hereth’s desire for metrics that are “reflective of processes and activities being carried out in the response” and of overall outcomes, there are also many that would not.

We grouped the performance metrics identified from the two case studies into end-point, process, and resource based metrics, as shown in Tables 7-15 (see Seager et al. 2006). Also, performance metrics are best defined with respect to specific objectives, which is why we have organized the elicited performance metrics with respect to each objective as presented above in Table 4. By coding performance metrics in the context of particular objectives it is apparent that many metrics were discussed for some objectives (e.g., mitigating ecological impacts) and few in relation to other objectives (e.g., protection of public and worker health and safety). This should not be taken to mean that those objectives associated with more performance metrics are more important than those with fewer performance metrics; this conclusion is not justified based on the evidence available. Instead, some objectives may be easier to gauge than other. For example, protecting health and safety can be assessed by counting OSHA-reportable injuries and fatalities and work-hours without accidents, whereas degradation of cultural resources may be difficult to capture quantitatively. Another reason may be an artifact of how we grouped objectives into larger categories. Furthermore, one might argue that many of the metrics suggested for assessing the prevention and mitigation of ecological impacts and economic impacts could also be measures for whether or not public concerns were addressed. However, they were not always expressed as such. Instead, addressing public concerns was often measured in terms of what effort was made for public outreach and involvement: How many meetings? How many leaflets distributed? Our inability to systematically identify all performance metrics with respect to each objective is a shortcoming of using interviews.

Table 7: Metrics related to protecting public health and safety.

End-Point	Number people killed or injured Number of mishaps during hours worked Presence of contaminants (e.g. PAHs) in water samples? Concentrations of oil in fish tissues. Number life threatening situations to human health Toxins in smoke plume if do in situ burning
Process	(None identified).
Resource	Number of IRAC team members OSHA/HAZMAT trained

Table 8: Metrics related to protection of cultural resources (Chalk Point).

End-Point	Number of critical sites protected Soil concentrations, smell or residual presence of oil on artifacts.
Process	Did trench digging affect sites? Experts contacted early for input about sites potentially at risk? Command responsive to requests for protection of sites? Were less destructive response actions chosen (e.g., sorbents and booms rather than burning)?
Resource	Number of GIS, hard maps, laptops and accuracy of location information. Amount of boom deployed

Table 9: Metrics related to mitigating social nuisances (Bouchard).

End-Point	Presence of stained rocks, oil on beach and complaints?
Process	(None identified).
Resource	(None identified).

Table 10: Metrics related to protection of environment.

End-Point	Number of oiled birds, eggs or other wildlife. Number of miles of shoreline impacted or cleaned. Amount of oil or globules on shore. No re-oiling or residual oil causing chronic toxicity to something. Number of fish, birds or other wildlife killed or injured (per unit search area). Number of “appropriate” (not exotics) animals rehabilitated and released. How far sheen at surface extended out [miles]? How long oil stayed? Presence of odors of oil. How much grass was destroyed or acres of marsh were impacted. Dead and stressed organisms found (rather than estimated). Time to achieve background levels/concentrations of contaminant or clean-up standards and recover from clean-up related damage. Areas protected (e.g., by redirecting or containing oil). Observe water blowing over booms. Degree of change to beaches and sandbars from clean-up actions. Types of animals and vegetation present after spill cleanup.
Process	Did getting required permits delay response action? Time for wildlife rehabilitation efforts to set up operations. Rate of bird handling at rehabilitation center. Gallons of oil and pounds of contaminated debris recovered and disposed of. Accuracy of cataloguing and enumeration of findings. Monitoring stations established. Time to deploy booming – and double-booming in sensitive areas. Immediacy of rehabilitator organization’s response to call for assistance. Change of helicopter flight patterns in response to requests from biologists to not disturb nesting birds. Oil direct to sacrificial (rather than sensitive) areas. Oil being captured in open water before it hits the beach. Area covered in search and recovery.
Resource	Amount of oil containment boom deployed. Number of volunteers. Number of floating resources to pick up oil in open water oil. Number of bodies to manage different aspects of response, inc. SCATs. Is there a ‘bird searcher’ on each team? Number sandbags deployed. Number people on cleanup crews to deal with oiled beaches

Table 11: Metrics related to economic impacts.

End-Point	Lost rental income, tourism dollars, property values, wages to fisher/watermen Duration of beaches closures. Recreational opportunities that were lost that are now back to what they were? Acres and duration of shellfish areas closed or acres of closures reopened. Acres of shellfish beds lost and number of lost fishing days. Costs of laboratory work, other research studies and money spent on response. Increase in crime rate [in southern MD due to influx of people from cleanup crews] Number of dead fish, ducks & geese.
Process	Kept track of all costs. Local municipalities reimbursed by responsible party?
Resource	(None identified).

Table 12: Metrics related to coordinated and effective response framework.

End-Point	Clear chain of command established and Incident Command System used? Are pre-identified areas potentially being affected? Did we have information to keep Governor's office and other state Senators and Reps abreast of what was going on? Understanding of whether oil is still stored offshore re-contaminating cleaned up beaches? Accurate accounting of volume oil spilled and on shore? Number areas cleaned as of today? Number of miles of shoreline impacted right now? Bad feelings among locals responders toward the Unified Command staff?
Process	RP is responsive? Number of hours to set-up stable incident command center. Location easily accessible? Key people became involved early? Local and federal responders notified quickly? Clear understanding of rights the state trustees have as a state agency? Is Incident Commander able to reach key people, does their phone number work, did they respond? Clear communication protocols and reliable technology working? How quickly decisions made? Are decisions correct (in hindsight)? Is there conflict or chaos in command center? Chauvinistic behavior? Cooperation? Did Unified Command resist information that did not conform to their expectations? Number and frequency of meetings, daily reports for morning meetings? Informed of meetings in advance (i.e., lead time)? Presence of watchdog to see what's going on? Modelers able to get 24 hours ahead of spill with accurate projections? Experts consulted for input on response strategies? Are there clear protocols and schedules? Plans communicated day in advance? Time it takes to implement tasks, such as boom deployment? Are players familiar with each other? Frequency of resource and personnel changes. Time taken to re-staffing response people after contractor fired. Pick-up and shipping schedule for waste generated by clean-up organized? Resources placed in the proper locations? Response organized by segments? System established to track progress? Are crews visiting hard hit areas every day and recording information in a unified way? Is all pertinent data gathered and recorded?
Resource	Number of supervisors, Spanish speaking supervisors assigned per section. Number of radios; availability of GIS and computers; phones available. Number of teams of trained observers walk coastline and make observations of extent and coverage area of oil Number crews trained, hours worked. Type of oil.

Table 13: Metrics related to effective and timely response

End-Point	Number of days company (PEPCO) shut down. Presence/absence of sheen, oil in water, tarballs, oil on shoreline, oil in sediments. Areas impacted to be cleaned up to the ecological state the environment was before the spill happened. Number of areas to be signed-off compared to total. Number of days until endpoints achieved? Is oil being contained? Patchiness of oil? Solid sheet of heavy oil. What it looks like a year later. Number of public meetings organized? Members of the public voice support? Good working relationships with all parties involved? Are response actions having the desired effect?
Process	How often beach is searched – both oiled and non-oiled? Digging holes to look for oil on shoreline? Quarterly checks to see if oil is present or not on beaches Breach of water over boom? Protection equipment put in the right place at the right time? Quality of contractor work. Number of times a orders were given but still not done. Number of public meetings, newsletters published by RP. Time it takes to get response in order. Lead resources mobilized? Number/frequency of fly-overs for real time aerial photography. Clear chain of command established? Communication to appropriate people? Accurate reporting and counting of crews in field. How quickly SCATs out in field? Are clean-up crews assigned effectively to do a good job? Is all pertinent data gathered and recorded? Data sheets available and sufficient? Local officials set up task force for spill response? Muster all forces in town? Basic training for volunteers on bird collection and rehabilitation conducted? Cleanup of impacted areas organized to be manageable and able to monitor? Access established for recovery and clean-up crews through private property? Do efforts correspond to tides? Recovery or rehabilitation of wildlife conducted? How soon have experts been called and set up triage and rehab centers?
Resource	Number of teams/people/supervisors in the field? Number of volunteers? Number of people working at one time. Hours worked. Number of monitors in field to give direction and warning to clean up crews Resources adequate for planned tasks? Cost of response. Types of skills represented on team? A 'bird searcher' on each SCAT? Pounds of sorbent material. Number of packets of baby oil for oil removal. Amount of oil containment boom.

Table 14: Metrics related to achieving legal regulatory requirements.

End-Point	Achieve termination endpoints? Shoreline back to conditions prior to spill? No oil should come off to the touch. Evaluate response with respect to endpoints achieved. For sandy beach no visible oil, no odor of oil. For marshes no sheen. For groin (jetties between properties), riprap no sheen or no oil available when touched.
Process	Number of days until endpoints achieved. All procedures followed (e.g., NIMS).
Resource	(None identified).

Table 15: Metrics related to addressing public concerns, needs and support.

End-Point	Public reimbursements for private property losses. Number calls from public. Members of the public voice support? Public comments from critics and local residents (re: response effort, not outcomes)? Level of staining Complaints about stained rocks? Residual oil on shore (tar balls)? Was spirit of state regulations for public involvement met? Level of public trust?
Process	Relationships and trust with local officials developed? Immediacy of public meetings (number days after spill). Number of public meetings. People given examples of what was impacted and what kinds of cleanup was going on? Number of fliers and informational packets delivered door to door, visuals for media. Public provided the kinds of information it wants? Incorrect information disclosed? A timeframe for ending the cleanup established? Ongoing monitoring and addressing of issues post-spill? People have a place or someone to go to with concerns? Public receives assurances that beaches will be cleaned up to the level of their expectation? Is a forum provided to public so they can hear what's going on and give their feedback? Feelings: level of conflict/anger or happiness. Establish and keep up to date website for public information. Quality of questions from media Able to 'stay on message' during public meetings, press conferences, etc.? Unified Command accessible for public questions and comments? Amount of oil removed manually from shoreline.
Resource	Number hours agencies spent on public outreach (meetings). Number of pamphlets distributed to inform public of hazards. Number of stakeholders involved in setting clean-up standards. Number of dispatch teams arranged to reach-out to various stakeholders. Frequency of information postings on website.

4.3 Characteristics of good performance metrics and the challenges of applying them in assessments of particular spill responses

The existing literature characterizing indicators and performance metrics emphasizes the *way* the metric is expressed (mathematical), the *purpose* of the metric (within the organization), and the cause-effect relationships between different indicators (Seager and Theis 2004). However, more recent attention has shifted to indicators as an expression of the *values* of an organization and as a method of facilitating communication both within the organization and with outside or stakeholder groups (e.g., Chess et al. 2005). In this regard it is helpful to create a taxonomy that classifies different indicators according to their qualitative, value-based characteristics (Seager & Theis 2004). As part of this project, we have developed a better understanding of a) types of metrics for oil spill response and b) characteristics of good metrics, through a combination of literature review and empirical case study research as discussed in this section.

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 (for further detail see Seager & Theis 2004).

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). 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 ecological 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).

Table 16 classifies a number of example oil spill metrics with regard to the multiple value-based dimensions described in the previous section.

Table 16. Example oil spill metrics illustrating different types of performance metrics.

Economic	Thermodynamic	Environmental	Ecological	Human Health	Socio-Political
Clean up costs Property & ecosystem damage Ecosystem damages or lost services Lost marginal profits Volunteer opportunity costs	Volume of oil spilled, recovered, destroyed, or contained Slick area and thickness Mass of clean up wastes generated Volume cleaning agent deployed	Chemical concentration and toxicity Habitat suitability, e.g., acres shellfish bed Length of oiled shoreline Degradation rates	Wildlife deaths, pop-ulations, fecundity and recovery rates Biodiversity Catch sizes Plantings, seedings	Quality-adjusted-life-years (QALYS) Disability-adjusted-life-years (DALYS) Life expect-ancy Lost time injuries	Newspaper column inches, minutes TV coverage, web hits. Volunteerism. Public meeting attendance. Direct messages (email, letters, phone calls).

In addition, performance metrics 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); performance metrics should be:

- *scientifically verifiable*; two independent assessments would yield similar results.
- *cost-effective*, using technology that is economically feasible and does not require an intensive deployment of labor to track.
- *easy to communicate to a wide audience*; the public should understand the scale and context and be able to interpret the metric with little additional explanation.
- *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 because metrics that are independent of human action cannot inform a management, policy-making, or design processes.
- *credible*, so that they are perceived by most stakeholders as accurately measuring what they are intended to measure.
- *scalable over an appropriate time period and geographic region*; they should be indicative of short, medium, and/or long term effects as appropriate (e.g., 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).
- *relevant*, with respect to reflecting 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.

- *sensitive* enough to capture the minimum meaningful level of change or make the smallest distinctions that are still significant, and they should have uncertainty bounds that are easy to communicate.

It may be difficult, if not impossible, to find metrics that satisfy all of these conditions for all stakeholders. In other words, defining and selecting "good" performance metrics is challenging. The literature review and case studies of the Bouchard-120 and Chalk Point spill responses revealed a number of important challenges in the context of oil spill response.

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. 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 are 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 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. Performance metrics that may 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 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.

4.4 Framework by which performance metrics may be compared and selected

The final objective of this project was 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, which can be implemented prior to a spill event (or, during spill response *planning*). 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, Webler and Tuler 2001), 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 addition, the proposed process design is informed by findings that emerged from our work on the other project objectives. In particular, the empirical research findings and literature reviews conducted as part of the project, discussed above, suggest that:

- The objectives that drive oil spill response and inform oil spill response assessment can vary among regions and among stakeholders. In particular, even if there is agreement about objectives in principle, there may be disagreements about their relative priorities within a particular context. This suggests a need for broad and structured deliberation among interested and affected parties to identify objectives.
- Selection of objectives and performance metrics requires input about technical factors as well as input related to social values and preferences. Different stakeholders have different kinds of knowledge and expertise that can inform decisions. A process should be structured to ensure that appropriate kinds of expertise are sought at the right time, and that appropriate expertise informs deliberations and decisions.
- Metrics chosen to evaluate performance can measure a variety of dimensions that may be of interest. A process should seek input about appropriate metrics from a wide range of interested and affected parties. On the other hand, there are important qualities of metrics that should be considered. Selection of metrics for assessing performance on oil spill response objectives may be informed by multi-criteria decision analysis (Linkov et al. 2007; Linkov et al. Forthcoming).
- Objectives and metrics for assessment of oil spill response can be context dependent. In addition, the needs of stakeholders can vary in regard to their ability to participate. Thus, a process should be designed that can be replicated in different regions, while being sensitive to the environmental, economic, and social context that may affect choices of objectives and metrics and to the factors that can ensure meaningful participation by interested and affected parties. In other words, the process design should be adaptable.

The process we propose is guided by two overarching principles: fairness and competence. 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 2001). Efficiency is another principle all processes should aspire to attain. Obviously, it is better when these objectives are met efficiently, but efficiency is not a principle on par with fairness or competence. 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 incremental progression. 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 learning 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, iteration, derives from the idea that the process should be adaptable and promote learning. 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 integration of analysis and deliberation. 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, Dietz and Stern 1998, 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.[7]

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.

[7] 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.

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.

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 17; details about each of these steps are provided in Tuler et al. 2007.

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.[8] 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. At the same time, we have proposed the use of innovative tools, such as Q method and multi-criteria decision analysis, as part of the process. The potential use of Q method was discussed above, with further detail available in Tuler and Kay 2007. We have also provided an illustrative example of a multi-criteria decision analysis application for oil spill response assessment (Linkov et al. 2007, Linkov et al. Forthcoming). Specifically, the example considers the problem of prioritizing performance metrics for oil spill response according to specified management objectives using 14 alternative metrics. These metrics represent different phases of the response (response, recovery, restoration) and different types of metrics (economic, thermodynamic, environmental, human health, ecological, sociopolitical).

Table 17: 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.

5.0 Discussion and importance to oil spill response/restoration

The results of this project extend prior understandings and fill important gaps in information and guidance for spill response planning and assessment by a) providing empirical findings from multiple case studies about spill response objectives and performance metrics and b) developing a framework for identifying, discussing, and selecting performance metrics for oil spill response assessment that involves a wide range of interested and affected parties.

[8] 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.

First, prior work has highlighted the challenge of selecting performance metrics for oil spill response. It has also argued for the need to closely couple performance metrics to the assessment of particular objectives. However, there has not always been a clear distinction made between objectives and performance metrics. In addition, there is a lack of empirical research to identify the performance metrics utilized by different stakeholders in spill response. For example, Lindstedt-Siva (1999) has proposed a long list of performance metrics that may be used to assess spill response. They were generated from her long work experience in the field. While there are some overlaps among those she proposes (and attributes to particular kinds of stakeholders), our work elaborates this list – and it provides an empirical basis for the claims she makes based on her experience. Thus, our work both supports and extends understandings on this issue.

In addition, our works provides additional empirical support for the claim made by others that "that the elements of a contingency plan are of very unequal importance" (Abordaif et al. 1995). They go on to write that "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." This conclusion relates clearly to the discussion in the previous section about what makes for a "good" performance metric.

Second, prior work has suggested the need for multi-stakeholder processes. In brief, a variety of models or frameworks for evaluating oil spill contingency plans and response efforts have been proposed (see Tuler et al. 2007 for additional discussion; Haynes and Ott 2001, Abordaif et al. 1995, Harrald and Mazzuchi 1993, Kuchin and Hereth 1999, Ornitz and Champ 2002). Many of them are based on the concepts of best response and critical success factors. According to Haynes and Ott (2001) "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). Lindstedt-Siva (1999) briefly writes about a process for identifying stakeholder performance criteria (pg. 31), embedded within a much more elaborated *system* of spill response assessment that includes contingency planning and conducting exercises and response assessment studies.

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. For example, Haynes and Ott (2001) write that "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*" (*emphasis* added).

Such observations strongly suggest that consideration of a broad range of perspectives about objectives for a response and assessments of the quality of a response is critical (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:

- taking into account the interests of the public in the management of a spill response and
- 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* 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 framework we have proposed provides a specific approach to answering these 'up-front' questions as well as achieving the final goal of developing a clear set of defensible performance metrics for oil spill response assessment.

6.0 Technology transfer

The results of the project have been documented in journal articles, book chapters, and technical reports. In addition, conference presentations and posters have been delivered. In addition, these materials are available on the CRRC website and the SERI website (xx-need to finish uploading).

The results have been shared with:

- The oil spill response community, including NOAA SSCs. For example, Gary Ott has used our reports to inform his recent work on spill response assessment.
- Federal and state disaster management agencies, including NOAA, US Coast Guard, and the Army Corps of Engineers.
- Researchers and consultants concerned with oil spill response specifically and disaster/hazard management planning in general (e.g., Society for Risk Analysis).

7.0 Achievement and dissemination

Journal articles and book chapters

Linkov, I., Tkachuk, A., Levchenko, A., Seager, T., Figueira, J., Tervonen, T, in press. "An Multi-Criteria Decision Analysis Approach for Establishing Performance Metrics under

Government Performance and Results Act: Example of Oil Spill Response." In: Linkov,I., Wenning, R., Kiker, G., eds. *Risk Assessment and Decision Support for Ports and Critical Infrastructure*, Springer, 2007.

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*.

Tuler, S., Seager, T., Kay, R., and Linkov, I. 2007a. 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.

Presentations and posters at conferences

- Linkov, I. 2007. Rapid Environmental Planning: Examples of Applications of Risk Assessment, Multi-Criteria Decision Analysis, and Adaptive Management to Better Serve Project Needs. US Army Corps of Engineers, Seattle, March.
- Linkov, I., Kiker, G., Bridges, T. 2005. Linking Estuarine Science and Management: Comparative Risk Assessment. Delaware Estuary Science Conference May 10-11, University of Delaware, Newark, DE
- Linkov, I., Seager, T. Multi-criteria Decision Analysis: Performance Metrics Example. Training Course. US EPA (Cincinnati), April 28, 2006
- Linkov, I., Seager, T., Satterstrom, K., Linkov, I., Tuler, S., and Kay, R. 2006. Environmental Performance Metrics For Oil Spill Response. NATO Workshop on Ports Security, 16-19 March, Venice, Italy.
- Linkov, I., Seager, T., Tuler, S, and Kay, R. 2005. Poster presentation prepared for 2005 EPA and Tri-Services "Toxicology and Risk Assessment Conference," April 25-28, 2005, Fairborn, Ohio.
- Satterstrom, K., Linkov, I., Seager, T., Linkov, I., Tuler, S., and Kay, R. 2005. Oil Spill Response, Recovery & Restoration: Multi Criteria Decision Analysis Approach. NATO Workshop on Ports Security, 20-24 April, Thessalonica, Greece.
- Seager, T., Tuler, S, Kay, R., and Linkov, I. 2005. Performance Metrics for Oil Spill Response, Recovery & Restoration: Understanding Stakeholder & Expert Perspectives. Poster presentation at the Society for Environmental Toxicology and Chemistry Annual Meeting, November 13-17.
- Tuler, S., Webler, T., Seager, T., Kay, R., and Linkov, R. 2006. Bringing clarity to competing objectives and performance metrics in marine oil spill response: A deliberative approach to planning and assessing response. Paper presented at the Society for Risk Analysis Meeting, Baltimore, MD, 6 Dec.

Reports (unpublished)

Linkov, I., Satterstrom, F. K., Tkachuk, A., Levchenko, A., Seager, T., Figueira, J. R., Tervonen, T. (2007a). MCDA Approach For Selecting Performance Metrics: Applications To Government Performance And Results Act (GPRA) And 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, 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.

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.

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., Webler, and Kay, R., 2007b. Defining and selecting objectives and performance metrics for oil spill response assessment: A process design integrating analysis and deliberation. Technical report submitted to the Coastal Response Research Center, NOAA Grant # NA04NOS4190063. Project Number: 05-983. Greenfield, MA: Social and Environmental Research Institute, Inc.

8.0 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.

Assessments: Environmental Protection in Oil Spill Response Planning. A Guidebook. United States Coast Guard. Washington, D.C. 148p.

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.

Aurand, D., Walko, L., and Pond, R. 2001. *Developing Consensus Ecological Risk*

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

Clarke, L. 1999. Mission improbable: Using fantasy documents to tame disaster. Chicago: University of Chicago Press.

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.

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.

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

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.

Glaser B. 1992. *Basics of Grounded Theory.* Mill Valley, CA: Sociology Press.

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.

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 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. 2001. 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.

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.
Hofstetter P, Hammit JK. 2002. Selecting human health metrics for environmental decision-support tools. *Risk Analysis.* 22(5):965-983.
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. 2000. Using multicriteria methods in environmental planning and management. Environmental Management, 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. R., Tervonen, T. (2007a). MCDA Approach For Selecting Performance Metrics: Applications To Government Performance And Results Act (GPRA) And 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.
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.
Loureiro ML, Ribas A, López E, Ojea E. 2006. Estimated costs and admissible claims linked to the Prestige oil spill. *Ecological Economics*. In press.
McKeown, B. and Thomas, D. 1988. *Q Methodology*. Sage University Paper Series on Quantitative Applications in the Social Sciences 07-066 Beverly Hills, CA: Sage.
Moore, S. A. 1996. Defining 'successful' environmental dispute resolution: case studies from public land planning in the United States and Australia, *Environmental Impact Assesment Review* 16:151-169.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.
Niemeyer, S., Petts, J., and Hobson, K. 2005. Rapid climate change and society: Assessing responses and thresholds, *Risk Analysis* 25(6):1443-1456.
Odum HT. 1996. *Environmental Accounting*. John Wiley & Sons: New York NY. pp127-132.

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. 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.
Rapport DJ. 1999. Gaining respectability: Development of quantitative methods in ecosystem health. *Ecosystem Health* 5(1)1-2.
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 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.
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.
Shaw DG. 1992. The Exxon-Valdez oil spill – ecological and social consequences. *Environmental Conservation* 19(3):253-258.
Tuler, S. and Kay, R. 2007c. 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.
Tuler, S. and Webler, T. 1999. Voices from the forest: What participants expect of a public participation process. *Society and Natural Resources* 12:437-453.
Tuler, S. and Webler, T. (2006). Competing perspectives on a process for making remediation and stewardship decisions at the Rocky Flats Environmental Technology Site, *Research in Social Problems and Public Policy,* volume 13.
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. 2007a. 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., Webler, and Kay, R., 2007b. Defining and selecting objectives and performance metrics for oil spill response assessment: A process design integrating analysis and deliberation. 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., Webler, T., and Finson, R. 2005. Competing perspectives on public involvement: Planning for risk characterization and risk communication about radiological contamination from a national laboratory, *Health, Risk, and Society* 7(3):247-266.
USCG 2005. Long Island Sound Area Contingency Plan DRAFT. Available at www.uscg.mil/d1/units/seclis/acp/draft-acp.html
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. 1999. Integrating Technical Analysis with Deliberation in Regional Watershed Management Planning: Applying the National Research Council Approach. *Policy Studies Journal* 27(3):530-543.
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