9-27-2002

Evaluation of Monitoring Programs For the NHEP Monitoring Plan

Phil Trowbridge
NH Department of Environmental Services

New Hampshire Estuaries Project

Evaluation of Monitoring Programs For the NHEP Monitoring Plan

Final

September 27, 2002

Prepared by:

Phil Trowbridge
NHEP Coastal Scientist
New Hampshire Department of Environmental Services

This document was funded in part by the U.S. Environmental Protection Agency through a cooperative agreement with the New Hampshire Estuaries Project.

ACKNOWLEDGMENTS

This report was peer-reviewed by the NHEP Technical Advisory Committee. The members of this committee deserve thanks for their time and thoughtful input.

NHEP Technical Advisory Committee

Name	**Organization**
Steve Jones, Chair	UNH-JEL
Tom Ballestero	UNH
Gregg Comstock	NHDES
Paul Currier	NHDES
Ted Diers	OSP-NHCP
Jennifer Hunter	OSP-NHEP
Natalie Landry	NHDES
Richard Langan	UNH-CICEET
Stephen Mirick	ASNH
Chris Nash	NHDES
Chris Neefus	UNH
Fay Rubin	UNH-CSRC
Fred Short	UNH-JEL
Brian Smith	NHF&G
Sally Soule	OSP-NHCP

INTRODUCTION

The objective of this report is to evaluate the monitoring programs for indicators in the NHEP Monitoring Plan (NHEP, 2002) to determine if sufficient data are being collected for each indicator and to identify any datagaps. The outcome of this report is a list of recommended changes to monitoring programs and a list of follow-up research projects. These recommendations will be considered to develop the workplans for NHEP monitoring funds in 2003.

METHODS

The monitoring programs for each indicator were evaluated relative to three basic questions:

- Are the correct parameters being measured?
- Are the measurements being made in the correct locations?
- Are the measurements being made frequently enough to provide good statistical power?

The first two questions were addressed through qualitative assessments, because quantitative assessments are not possible. Statistical power calculations have been conducted to answer the third question. Details of the statistical methods used are summarized in Appendix A.

ASSESSMENTS OF NHEP INDICATORS

Bacteria Indicators

BAC1: Acre-Days of Shellfish Harvesting Opportunities in Estuarine Waters

No assessment needed. All monitoring for this indicator is covered by the DES Shellfish Program. There are no datagaps in the monitoring program and the data quality objectives of National Shellfish Sanitation Program are being met. The NHEP supports the DES Shellfish Program through programmatic funds. No additional NHEP monitoring funds are needed.

BAC2: Trends in Dry-Weather Bacteria Indicators Concentrations

- Are the correct parameters being measured?

Yes. Fecal coliforms, enterococci, and *E. coli* are being monitored at the estuarine stations. *E. coli* is being monitored at the freshwater tributary stations.

- Are the measurements being made in the correct locations?

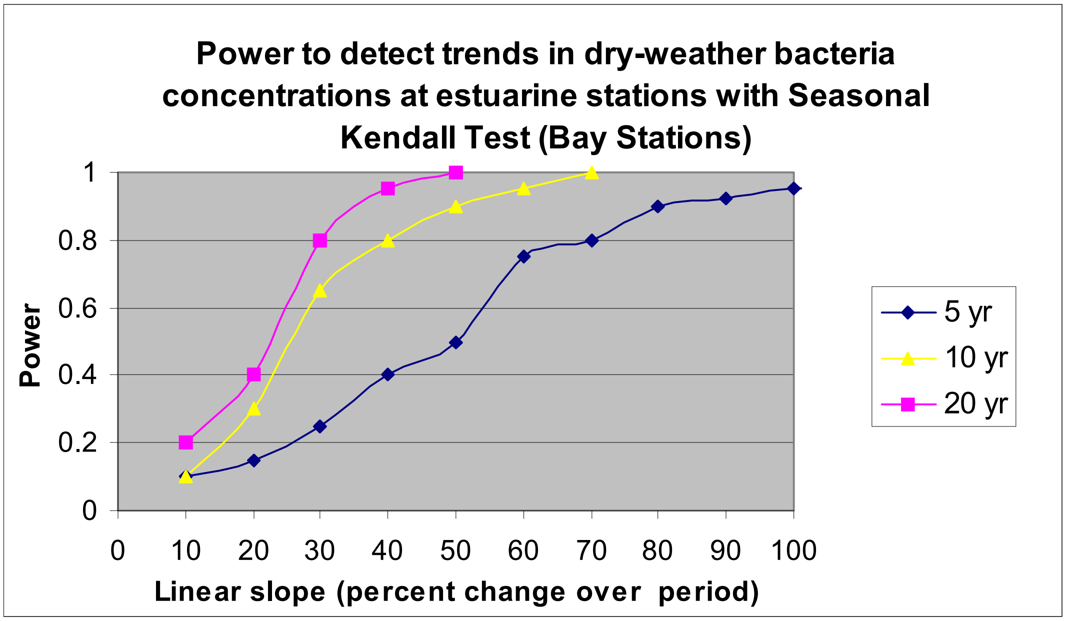
Yes, there is good coverage throughout the estuaries and all tributaries to the Great Bay. It may be possible to reduce the number of stations in the future if there are consistent trends between stations in the same area/mixing zone and the spatial coverage is not needed for other purposes (e.g., 305b assessments).

- Are the measurements being made frequently enough to provide good statistical power?

No. The Monitoring Plan set a performance criterion of “0.8 power for detecting a 10% change from 2000 concentrations over a 5 year period using 0.10 as the level of the test” for this indicator. This criterion is not being met by the existing monthly monitoring programs. The minimum detectable trends for a monthly monitoring frequency are:

Minimum detectable trend with power of 0.8

	Bay stations (e.g., Adams Point)	Tributary Stations
After 5 years	70% (0.4 cfu/100ml/yr)	>100% (8 cfu/100ml/yr)
After 10 years	40% (0.1 cfu/100ml/yr)	60% (2 cfu/100ml/yr)
After 20 years	30% (0.04 cfu/100ml/yr)	40% (0.6 cfu/100ml/yr)



Even though the performance criteria are not being met, the minimum detectable trend over 5 years is extremely low in absolute terms for the bay stations (0.4 cfu/100ml/yr). The enterococci concentrations at these stations are low (3 cfu/100ml) relative to the water quality standard (35 cfu/100ml, geomean). Therefore, monthly monitoring at bay stations is sufficient because trends of deteriorating water quality would be detected well before the standard was violated. A more reasonable performance criteria for bay stations should be the ability to detect trends of 1 cfu/100ml/yr over five years.

In the tributaries, the monthly monitoring frequency is not adequate. The baseline enterococci concentrations in the Squamscott River are approximately equal to the geometric mean water quality standard (30 vs. 35 cfu/100ml, respectively). The minimum detectable trend from monthly sampling in absolute terms is 8 cfu/100ml/yr, which means that over 5 years only changes greater than +/-40 cfu/100ml would be detectable. However, smaller concentration shifts (such as +/-10 cfu/100ml) will determine whether the water quality meets or violates the standard. Therefore, more frequent monitoring is needed at these stations in order to be able to detect subtle trends of deteriorating or improving water quality. The performance criteria for monitoring at the tributary sites should be the same as for the bay sites (ability to detect 1 cfu/100ml/yr over five years).

Weekly sampling is the highest frequency of monitoring that could reasonably be achieved given field staff availability. Monte Carlo modeling is needed to determine whether weekly sampling at tributary sites would meet the new performance criteria for this indicator. In order to estimate the likely cost of increased bacteria monitoring at the tributary sites, it has been assumed that weekly monitoring will be sufficient. This assumption will be checked using Monte Carlo modeling by the end of 2002.

BAC3: Trends in Wet-Weather Bacteria Indicator Concentrations

- Are the correct parameters being measured?

If discharges to estuarine waters are tested, fecal coliforms and enterococci should be monitored.

- Are the measurements being made in the correct locations?

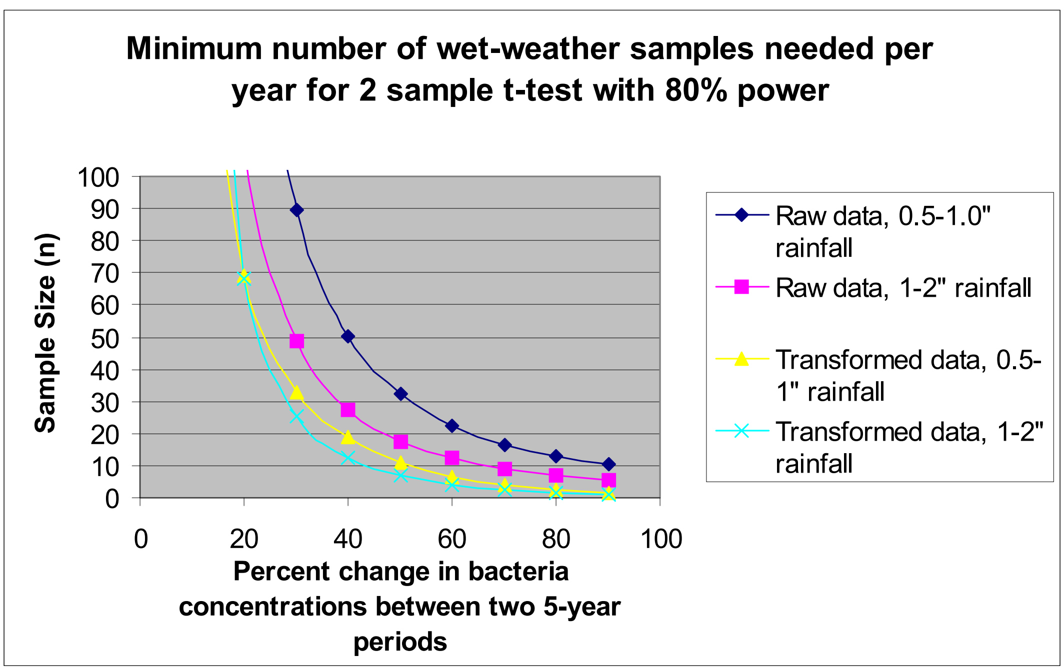
No, the interim plan for a wet-weather monitoring program was to only monitor for wet-weather effects at the HOT tributary sites around Great Bay. Monitoring in Hampton-Seabrook Harbor, Little Harbor, and Portsmouth is needed because these shellfish growing areas are also affected by stormwater discharges of bacteria. However, the hydrography and stormwater discharges are different at HOT tributary sites than at open harbors so the monitoring approach may have to be different in different areas.

- Are the measurements being made frequently enough to provide good statistical power?

The existing monitoring programs collect bacteria samples monthly and are not targeted at wet-weather conditions. As a result, only 1-2 samples per year are collected during wet-weather, which is inadequate for any trend detection. If monthly *wet-weather* samples were collected, approximately 10 years would be required to be able to detect 50% to 100% (8-15 cfu/100ml/yr) changes in wet-weather concentrations depending on the type of statistical test. Over a five year period, only major changes (175% change or 53 cfu/100ml/yr) would be detectable. This trend detection capacity is much worse than the performance criterion for this indicator from the Monitoring Plan (0.8 power for detecting 10% change from 2000 concentrations over 5 years).

Minimum detectable trend with power of 0.8

Duration	Trend analysis by SKT with monthly sampling	Two-sample t-test on log-transformed data, 5 year groups, 10 samples/yr
5 years	175% (53 cfu/100ml/yr)	NA
10 year	100% (15 cfu/100ml/yr)	50% (8 cfu/100ml/yr)
20 years	60% (5 cfu/100ml/yr)	NA



For the dry-weather bacteria indicator (BAC2), it was possible to increase the frequency of monitoring to increase the trend detection power. This is not possible with wet-weather bacteria trends. The frequency of wet-weather samples per year cannot be more than approximately monthly because that is the frequency of occurrence for adequate storms (>0.5 inch or rainfall during daylight, weekday hours).

The baseline concentrations of *E. coli* in wet-weather samples from station 7-Cch is 150 cfu/100ml, which is higher than the geomean water quality standard of 126 cfu/100ml. The original performance criteria of detecting 10% change over 5 years corresponds to the ability to detect changes of 3 cfu/100ml/yr. Given the high variability in wet-weather samples and the baseline concentrations relative to the standard, a more reasonable performance criteria would be the ability to detect trends of 10 cfu/100ml/yr over five years. This would provide the opportunity to document changes from baseline of 25-30% over 5 years. However, even this more reasonable performance criterion is not feasible with monthly wet-weather sampling, which can only detect changes of approximately 50 cfu/100ml/yr.

Detecting the effects of NHEP actions to eliminate wet-weather pollution should be a high priority for the NHEP Monitoring Plan. However, due to the extreme variability in the wet-weather concentrations and the frequency of storm events, detecting significant trends or differences over time scales of management interest is impossible. A new approach is needed to monitor the effects of NHEP-funded improvements in wet-weather discharges.

BAC4: Tidal Bathing Beach Postings

No evaluation needed.

BAC5: Trends in Bacteria Concentrations at Tidal Bathing Beaches

- Are the correct parameters being measured?

Yes.

- Are the measurements being made in the correct locations?

Yes. All public tidal bathing beaches are tested.

- Are the measurements being made frequently enough to provide good statistical power?

No. Beaches are tested weekly during the summer months which results in 5-10 samples per beach per year. The performance criteria for this indicator in the Monitoring Plan is "0.8 power for detecting 10% changes from 2002 concentrations over 5 years". With the existing monitoring frequency, the minimum detectable trend after 5 years is 150%. Since baseline enterococci concentrations at tidal beaches are 4 cfu/100ml, the minimum detectable trend in absolute terms is approximately 1 cfu/100ml/yr.

Minimum detectable trends at power=0.8

	5 samples per year	10 samples per year
After 5 years	300% (2.4 cfu/100ml/yr)	150% (1.2 cfu/100ml/yr)
After 10 years	175% (0.7 cfu/100ml/yr)	100% (0.4 cfu/100ml/yr)

* assuming baseline concentrations of 4 #/100ml

Achieving the original performance criteria is infeasible, because this would require approximately 2,000 samples per year (6 per day). However, given the low concentrations of enterococci, the existing monitoring program is sufficient to detect any deterioration of the water quality before the standard is exceeded. A more reasonable performance criterion is ability to detect trends of 1 cfu/100ml/yr, which is achieved with a sampling frequency of 10 times per year. This performance criterion is the same as was recommended for the dry-weather bacteria concentrations at bay stations.

BAC6: Violations of Enterococci Standard in Tidal Waters

- Are the correct parameters being measured?

Yes

- Are the measurements being made in the correct locations?

Yes. The Assessment Units (AU's) for the 305b report will be the growing areas of the DES Shellfish Program. The stations for monitoring enterococci were checked against the AU map to determine how many AUs are covered by the current monitoring program.

Area	Number of Assessment Units for 305b	Number of Enterococci Stations	Comments
Open ocean	12	0	
Tidal Beaches	9	9	See BAC4
GB tidal tributaries	6	6	Sfr, Cch, Lmp, Sqm, Oys, Blm
Great Bay/Little Bay	11	3	
Piscataqua River	2	2	
Portsmouth Harbor/Back Channel	5	3	
Little Harbor	3	2	
Hampton/Seabrook Harbor	14	3	
TOTAL	62	28 (45%)	

Although less than half of the AU's contain an enterococci station, the spatial coverage of the stations is adequate. Many of the AU's are small slivers of water around particular sources that would be impractical to monitor. Importantly, all the major areas of the estuary have at least one enterococci monitoring station. There are no offshore monitoring stations but the tidal beaches are monitored regularly and these should represent worst-case conditions for the offshore zone.

- Are the measurements being made frequently enough to provide good statistical power?

Yes. The current monthly monitoring frequency more than satisfies the data requirements of the 305b assessment methodology for 2002. The power for trend detection will be same as for the BAC2 indicator.

BAC7: Freshwater Bathing Beach Postings

No analysis necessary.

BAC8: Bacteria Load from Wastewater Treatment Plants

- Are the correct parameters being measured?

Yes.

- Are the measurements being made in the correct locations?

Yes. All WWTFs are required to monitor bacteria loading.

- Are the measurements being made frequently enough to provide good statistical power?

Data on bacteria loading are reported monthly by each WWTF. The performance criterion is to be able to detect 10% changes from 2000 levels over 5 years with power of 0.8. With monthly reporting, the minimum detectable trends are much higher than this (70-100%).

Minimum detectable trend with power of 0.8

	By SKT	By Linear Regression
After 5 years	70% (4 cfu/100ml/yr)	>100%
After 10 years	40% (1 cfu/100ml/yr)	70% (4 cfu/100ml/yr)
After 20 years	30% (0.4 cfu/100ml/yr)	50% (1 cfu/100ml/yr)

* Based on baseline concentrations and variance in Hampton WWTF effluent 1990-2000.
* Based on variance in monthly average concentrations as reported on DMRs. Raw data from individual samples could not be obtained.

Using the Hampton WWTF as an example, trends of 4 cfu/100ml/yr from the baseline total coliform concentration of 26 cfu/100ml would be detectable after 5 years. The permit limit for the WWTF is 70 cfu/100ml. Therefore, although the performance criteria from the Monitoring Plan are not being met, using the monthly values reported by the WWTFs appears to provide enough trend detection power relative to the baseline concentrations and the permit limits.

The performance criteria for this indicator from the Monitoring Plan should be reconsidered. Uniform criteria for all the WWTFs cannot be developed because the WWTFs on the seacoast use a variety of different bacteria indicator species for their permits and test at different frequencies.

BAC9: Microbial Source Tracking

- Are the correct parameters being measured?

Yes.

- Are the measurements being made in the correct locations?

This depends on the objective. The past studies have been focused, short-term studies on portions of the estuary to identify and characterize specific sources. No long-term MST monitoring has been done. Establishment of one or two long-term trends sites for MST monitoring (e.g., Hampton Harbor and Great Bay) would provide information on how the sources of bacteria to the estuaries as whole are changing over time.

- Are the measurements being made frequently enough to provide good statistical power?

Uncertain. Data to evaluate the statistical power of MST monitoring programs are not yet available.

<u>Toxic Contaminants Indicators</u>

TOX1: Shellfish tissue concentrations relative to FDA Standards

- Are the correct parameters being measured?

Yes.

- Are the measurements being made in the correct locations?

Yes. The Gulfwatch stations that have been tested in the past provide good coverage of the entire estuary. Collection and analysis of oysters in 2001-2002 expanded the geographic coverage into Great Bay. Oysters and clams will not be tested again as part of the Gulfwatch Program without NHEP-support.

Since oyster and clam tissue provide critical information regarding public health risks from consuming popular species, oyster and clam assessments should be repeated every 2 years.

- Are the measurements being made frequently enough to provide good statistical power?

Yes. The performance criterion for the indicator states that it should be able to detect a 1.0 ug/g difference between the lead concentrations and the FDA standard with 0.8 power. Lead was chosen because lead concentrations have historically been closest to the FDA standards. A one sample t-test power analysis with the median standard deviation for lead at station MECC shows that 4 subsamples per year have 0.83 power of detecting differences of 1.0 ug/g between the readings and the FDA standard of 11.5.

TOX2: Public Health Risks from Toxic Contaminants in Shellfish Tissue

This indicator uses data from TOX1. There is no independent sampling for this indicator, so the review questions about parameters, stations, and frequency are not relevant.

The only performance criterion for this indicator is that the data should be useable in a public health risk assessment. The following table shows that the Gulfwatch Program detection limits are much lower than FDA standards and lower than EPA risk-based criteria (when available).

Compound	Gulfwatch MDL (dry weight)	FDA Standard (dry weight)	EPA Risk Based Standard (dry-weight)
Ag	0.1 ug/g	---	---
Al	3.0 ug/g	---	---
Cd	0.2 ug/g	25 ug/g	4.7 ug/g
Cr	0.3 ug/g	87 ug/g	---
Cu	0.6 ug/g	---	---
Fe	6.0 ug/g	---	---
Hg	0.1 ug/g	6.7 ug/g	1.5 ug/g (Me-Hg)
Ni	1.2 ug/g	533 ug/g	---
Pb	0.6 ug/g	11.5 ug/g	---
Zn	1.5 ug/g	---	---
PAHs	3.6-12.6 ng/g	---	21 ng/g (BaP)
PCBs	0.7-2.8 ng/g	13,000 ng/g	80 ng/g
Pesticides	0.9-2.0 ng/g	700 ng/g (mirex)*	10 ng/g (dieldrin)*

* Lowest of available standards for: DDT, Chlordane, Dieldrin, Aldrin, Heptachlor, Heptachlor epoxide, and Mirex.
Source for MDLs is Chase et al. (2001)
Sources for FDA standards are listed in the NHEP Monitoring Plan
Source for EPA risk-based values is EPA (2000) with the following assumptions: 85% moisture content to convert wet-weight standards to dry weight, 4 meals per month exposure, 2 significant figures rounding. If criteria were available for both cancer and non-cancer risks, the lower (more stringent) of the two criteria was used.

TOX3: Trends in Shellfish Tissue Concentrations

- Are the correct parameters being measured?

Yes.

- Are the measurements being made in the correct locations?

Yes, but there is only one trend site in the estuary: MECC in Portsmouth Harbor. More trend sites would be helpful, perhaps in Great Bay and Hampton Harbor. Since the methods for analysis and interpretation are better established for mussel tissue, the trend sites should be mussel beds, not clam or oyster beds.

- Are the measurements being made frequently enough to provide good statistical power?

The performance criterion for this indicator is to be able to detect trends of 10% change over 5 years with 0.8 power. The existing monitoring program collects 4 replicate samples per year at a trend site. The minimum detectable trend from this monitoring program is 35% after 5 years. If the number of replicates is increased to 8 per year, the minimum detectable trend would drop to 25%. In absolute terms, the existing monitoring program with 4 replicates/year has the power to detect trends in mercury of 0.05 ug/g/yr, in PAHs of 0.6 ng/g/yr, and in PCBs of 3 ng/g/yr. For all three of these priority contaminants, baseline concentrations are no more than 50% of relevant public health criteria. It would take 10-20 years of increasing concentrations at the minimum detectable rate for the public health criteria to be exceeded. Therefore, even though the original performance criteria is not being met, the existing monitoring program is sufficient. It is not necessary to increase the number of replicates per year at Gulfwatch stations. New performance criteria for this indicator should be considered, such as the ability to detect trends in Hg of 0.05 ug/g/yr, in PAH of 1 ng/g/yr, and in PCBs of 5 ng/g/yr.

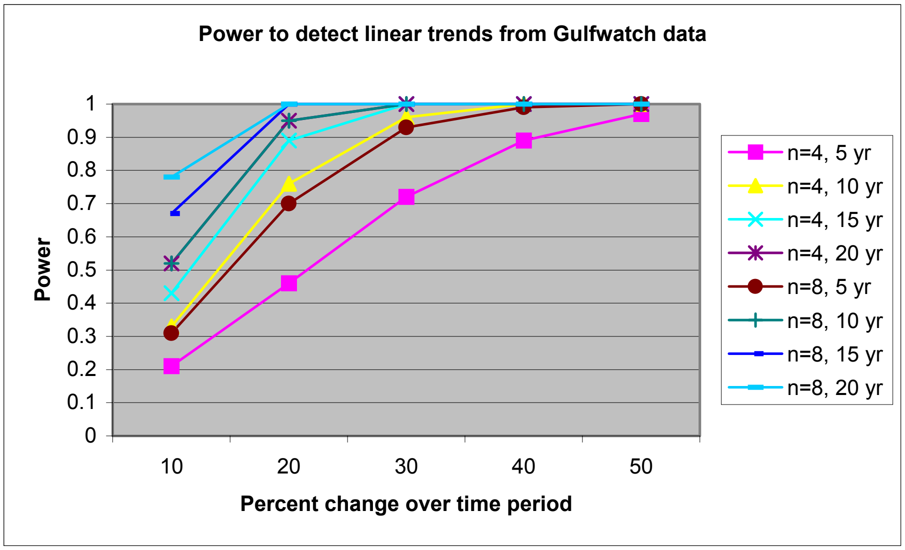
Minimum detectable trends at a station with power=0.8

4 replicates/station/yr

Duration	Percent Change	Absolute Change Hg (ug/g/yr)	Absolute Change PAH (ng/g/yr)	Absolute Change PCB (ng/g/yr)
5 years	35%	0.05	0.6	3
10 years	20%	0.01	0.2	0.8
20 years	15%	0.005	0.06	0.3
Baseline	---	0.65 ug/g	8 ng/g	42 ng/g
Health criteria	---	1.5 ug/g (EPA)	21 ng/g (EPA)	80 ng/g (EPA)

8 replicates/station/yr

Duration	Percent Change	Absolute Change Hg (ug/g/yr)	Absolute Change PAH (ng/g/yr)	Absolute Change PCB (ng/g/yr)
5 years	25%	0.03	0.4	2
10 years	15%	0.1	0.1	0.6
20 years	10%	0.003	0.04	0.2
Baseline	---	0.65 ug/g	8 ng/g	42 ng/g
Health criteria	---	1.5 ug/g (EPA)	21 ng/g (EPA)	80 ng/g (EPA)



TOX4: Trends in Finfish Tissue Contaminant Concentrations
TOX5: Sediment Contaminant Concentrations Relative to NOAA Guidelines
TOX6: Trends in Sediment Contaminant Concentrations

These indicators cannot be evaluated yet because not all of the data from the National Coastal Assessment from 2000-2001 is available. TOX4 and TOX6 have similar performance criteria as the other trend indicators in the Monitoring Plan. Since many of the other trend indicators have not been able to meet these criteria, the criteria for TOX4 and TOX6 should be reviewed for feasibility as well.

Nutrients and Eutrophication Indicators

NUT1: Annual load of Nitrogen to Great Bay from WWTF and Watershed Tributaries

- Are the correct parameters being measured?

Yes, but more sensitive analytical methods are needed for the HOT tributary sites. Most of the ammonia and half of the nitrate samples from the HOT tributary sites in 2001 were below the detection limit of the method. Having non-detected values introduces additional error into loading estimates because the non-detects must be arbitrarily assigned a value between zero and the method detection level (assignments of one-half the method detection level are common).

- Are the measurements being made in the correct locations?

Yes for tributary stations, but monitoring for WWTFs throughout the watershed is needed. Without monitoring data on all the WWTFs, the total nitrogen concentration in WWTF effluent must be assumed. Making this assumption could introduce as much as 80% bias into the total load calculation for WWTFs. Because this would be a bias and not a random error, the error would have to be added linearly, not by the root-mean-square (RMS) method. A current NHEP-funded study by JEL will analyze effluent from all 8 NH and 3 ME (if possible) WWTFs that discharge directly into estuarine waters for NH_4^+, NO_3^- and dissolved organic nitrogen (DON). A similar study of the remaining 6 WWTFs in the coastal watershed is also needed.

- Are the measurements being made frequently enough to provide good statistical power?

Yes for tributary loading, uncertain for WWTF loading. The performance criteria for this indicator is to be able to calculate the total nitrogen load to the estuary from tributaries and WWTF with an accuracy of

+/- 10%. Uncertainty in the WWTF load due to lack of data was discussed above. For fluvial loads, an error of 40% for monthly load estimates for individual rivers was estimated due to temporal heterogeneity and analytical errors. However, because all the errors are random and independent, the individual errors can be summed by the root-mean-square (RMS) method such that the error in the total load estimate is lower. With uncertainty of 40% for each monthly load estimate, the total annual load estimate from all the rivers would be +/-10%.

NUT2: Trends in Estuarine Nutrient Concentrations
NUT3: Trends in Estuarine Particulate Concentrations

- Are the correct parameters being measured?

Yes, but two other parameters are needed in order to comply with EPA's initiative to develop nutrient criteria for estuaries: total nitrogen (TN) and total light extinction. There are two ways to increase the nutrient monitoring to capture TN. Total Kjeldahl Nitrogen (TKN) could be added (TN=NO3+NO2 + TKN, assuming particulate inorganic nitrogen to be negligible). Alternatively, measurements of dissolved organic nitrogen (DON) and particulate organic nitrogen (PON) could be added to provide the total nitrogen estimate (TN=NO3+NO2+DON+PON, assuming particulate inorganic nitrogen to be negligible). Light extinction over 1 meter depth for photosynthetically active radiation (PAR) is also needed to provide data on "response variables" for the nutrient criteria.

- Are the measurements being made in the correct locations?

Yes. After a few years of data are collected, it may be possible to reduce the number of stations being tested for long-term trends if the stations are exhibiting the same response.

- Are the measurements being made frequently enough to provide good statistical power?

The existing monitoring programs collect samples monthly to capture the seasonal changes in the estuary. The performance criteria for this indicator from the Monitoring Plan is to have 0.8 power to detect 10% changes from 2000 concentrations over 5 years. Using the variance observed in the 1988-2001 dataset, the monthly monitoring frequency can only detect trends of 50% or greater after 5 years. For dissolved inorganic nitrogen (DIN) and total suspended solids (TSS), the minimum detectable trend over five years in absolute terms would be 1-2 uM/yr and 1-3 mg/l/yr, respectively. Although the original performance criteria are not being met, the absolute trends that could be detected are probably as low as possible given uncertainty introduced by the laboratory methods (assumed to be +/- 10% of baseline, see table below). Therefore, the existing monthly monitoring program is sufficient. The performance criteria should be changed to reflect what is achievable with monthly monitoring.

Minimum detectable trend at power=0.8 using the Seasonal Kendall Test with monthly samples

	Percent Change	Absolute Change DIN (uM/yr)	Absolute Change TSS (mg/l/yr)
After 5 years	50-60%	1 (bay), 2 (tribs)	1 (bay), 3 (tribs)
After 10 years	30-40%	0.4 (bay), 0.6 (tribs)	0.4 (bay), 1 (tribs)
After 20 years	20-30%	0.2 (bay), 0.2 (tribs)	0.1 (bay), 0.3 (tribs)
Baseline	----	10 (bay), 16 (tribs) uM	10 (bay), 21 (tribs) mg/l

NUT4: Eelgrass Distribution within Tidal Tributaries of Great Bay

- Are the correct parameters being measured?

Yes. Coverage of eelgrass is mapped using aerial photography and groundtruthing from a boat. However, additional groundtruthing parameters (e.g., canopy height, biomass, and shoot density) could be added to improve accuracy and provide an estimate of eelgrass biomass, not just areal coverage.

- Are the measurements being made in the correct locations?

Yes. The entire estuary is mapped. Groundtruthing stations for additional parameters would be needed to conduct biomass assessments. The number of groundtruthing stations needed to obtain adequate statistical power is not clear at this time.

- Are the measurements being made frequently enough to provide good statistical power?

Since 1987, JEL has mapped the areal distribution of eelgrass in Great Bay annually. These eelgrass maps represent the longest, continuous record of estuarine habitat for the Great Bay. Assessments of biomass have also been completed but the dataset for biomass is not as long or as consistent as the dataset for areal distribution. It is not possible at this time to assess whether the methods employed for areal mapping have an accuracy of +/- 10% as called for in the performance criteria for this indicator.

In addition to the NUT4 indicator, eelgrass biomass is one of the NHEP's indicators for critical species and habitats. Eelgrass biomass is a measure of the total mass of eelgrass in the system, while the NUT4 indicator of eelgrass distribution is a map of the locations of eelgrass beds in the system. These two indicators are related, since the eelgrass distribution map is a necessary first step for assessing eelgrass biomass. Therefore, it would make sense to combine the NUT4 indicator with the eelgrass biomass research indicator to have one indicator for eelgrass in the Critical Species and Habitats section of the NHEP Monitoring Plan.

NUT5: Violations of Instantaneous Dissolved Oxygen Standard
NUT6: Violations of the Daily Average Dissolved Oxygen Standard

- Are the correct parameters being measured?

Yes.

- Are the measurements being made in the correct locations?

Yes, but, pending results of pilot study this summer, it may be necessary to install a datasonde in the Salmon Falls River permanently.

- Are the measurements being made frequently enough to provide good statistical power?

Yes. For NUT5, the performance criteria is that the sensor have an accuracy of +/- 0.2 mg/l for instantaneous dissolved oxygen readings. This criteria is being met. For NUT6, the performance criteria is to have 80% power for detecting differences of 5 units (dosat%) between daily average concentrations and the state standard of 75% using 0.05 as the level of the test. Using data from the sondes from 1995-2000, this power was calculated and determined to be 99-100%.

NUT7: Trends in Biological Oxygen Demand Loading to Great Bay

- Are the correct parameters being measured?

Yes, but the analytical method for BOD5 at the HOT tributary sites is not sensitive enough. 78% of samples in 2001 were below the detection limit.

- Are the measurements being made in the correct locations?

Yes, all tributaries to the Great Bay are being monitored.

- Are the measurements being made frequently enough to provide good statistical power?

Yes, if the criteria is changed to match the nitrogen loading indicator (NUT1). The Monitoring Plan lists the performance criteria for this indicator to be "0.8 power for detecting a 10% change from 2000 concentrations over 5 year period with 0.10 as the level of the test". The other loading indicator in the Monitoring Plan is for nitrogen loading which has a performance criterion of an accuracy of +/-10% in the total load estimate. These two indicators should have the same type of performance criteria for consistency. The "accuracy-based" criteria seems more appropriate than the "trend-based" criterion. The monitoring frequency for BOD loading from rivers is identical to the monitoring frequency for total nitrogen. Assuming that BOD has the same or lower coefficient of variation as total nitrogen, then the error analysis for nitrogen loading should produce the same result (10% total error).

Shellfish Resource Indicators

SHL1: Area of Oyster Beds in Great Bay

- Are the correct parameters being measured?

Yes.

- Are the measurements being made in the correct locations?

Yes, but two of the six oyster beds being tracked by the NHEP were not mapped during the 2001-2002 assessment (Piscataqua River-12.8 acres, and Squamscott River-1.7 acres).

- Are the measurements being made frequently enough to provide good statistical power?

Uncertain. The performance criterion for this indicator is an accuracy in the area estimates of +/-0.5 acres. The accuracy of the most recent methods for mapping cannot be assessed until the 2002 data are reported. The mapping methods will be reviewed before next round of mapping. The oyster bed mapping project should be repeated in 2005-2006.

SHL2: Density of Harvestable Oysters in Great Bay Beds

- Are the correct parameters being measured?

Yes.

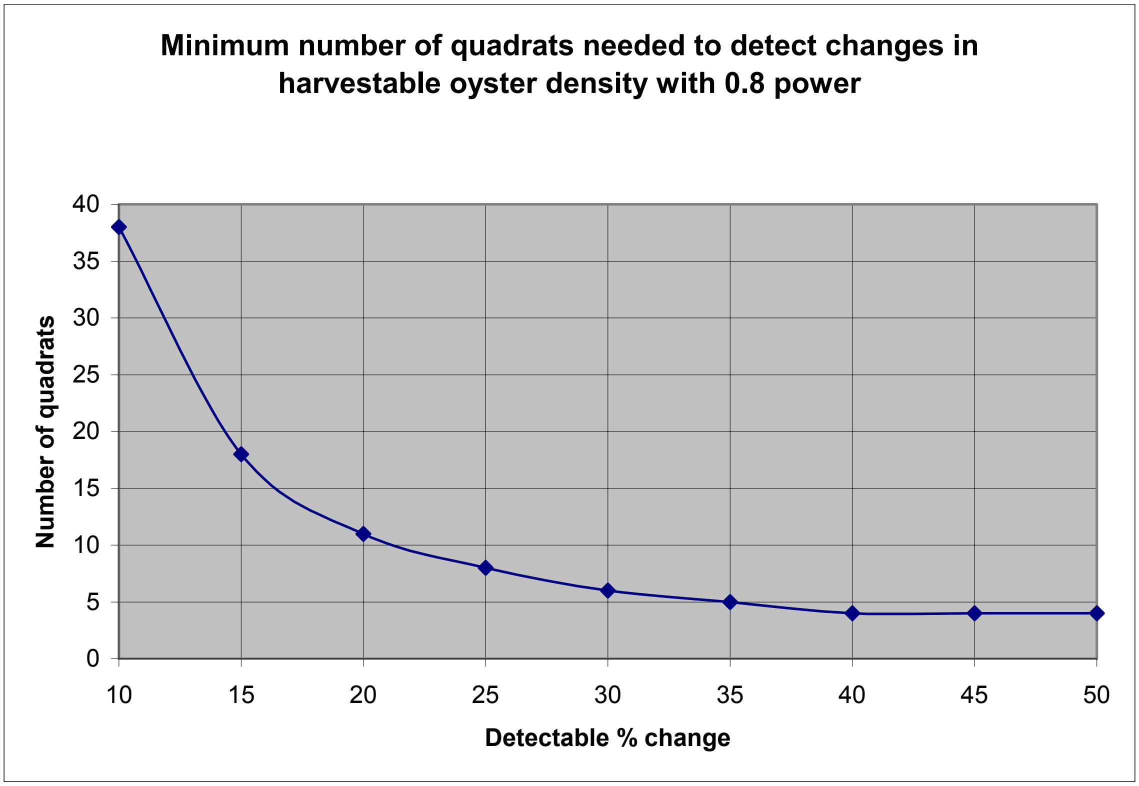
- Are the measurements being made in the correct locations?

Yes. All six of the beds being tracked by the NHEP were assessed within the last 3 years.

- Are enough measurements being made each year to provide good statistical power?

The performance criterion for this indicator is to have 0.8 power to detect 5 #/m2 differences between mean harvestable oyster densities in a bed and the goal (1997 levels). The difference of 5 #/m2 is equal to a 10% change from 1997 levels at Nannie Island. The most recent method for assessing oyster densities is to collect approximately 5 quadrats from randomly selected sampling locations in each bed. Based on the variance between quadrats that has been observed historically and in 2001 (see figures and table below), the method has power of 0.8 to detect changes from 1997 levels of 30% or greater. If the number of quadrats per bed were increased to 10, the program would be able to detect changes of 20%

from 1997 levels (differences of 10 #/m2). 40 quadrats per bed would be needed to meet the original performance criteria (detect 10% changes from 1997 levels), but this is not feasible. Therefore, it is recommended that the number of quadrats be increased to 10 and the performance criteria be changed to ability to detect differences of 10 #/m2. As illustrated below, the variance in oyster density is uniform across all the major recreational harvest beds (i.e., Adams Point, Nannie Island, Woodman Point, Oyster River) so 10 quadrats would be needed from each of these beds. Fewer quadrats could be taken from the Squamscott River and Piscataqua River beds because the variance within these beds appears to be lower.

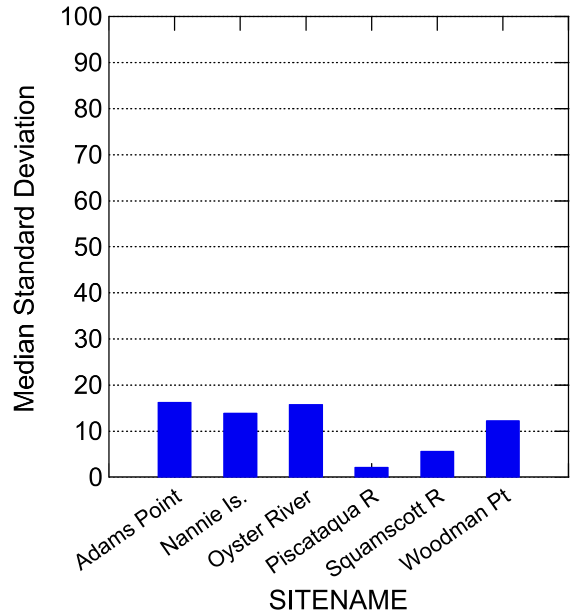


Standard deviation of harvestable oyster density from yearly NHF&G surveys

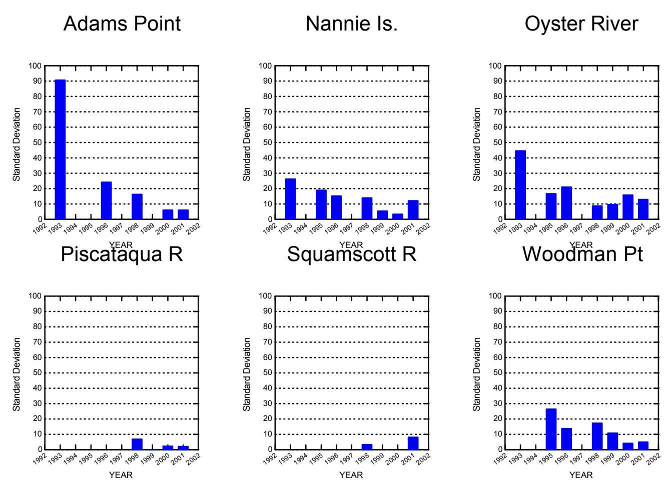
Oyster Bed	Median Standard Deviation	Comments
Adams Point	16.2 (n=5)	
Nannie Island (1993-2001)	13.9 (n=7)	
Nannie Island (2001)	12.0	Value used in power analysis
Woodman Point	12.2 (n=6)	
Oyster River	15.8 (n=7)	
Piscataqua River	2.2 (n=4)	
Squamscott River	5.6 (n=2)	

The following figures illustrate the median values for the standard deviation at each bed and the standard deviation values for each year at each bed.

Standard Deviation in Harvestable Oyster Density



Standard Deviation of Harvestable Oyster Density



SHL3: Density of Harvestable Clams at Hampton Harbor Flats

- Are the correct parameters being measured?

Yes.

- Are the measurements being made in the correct locations?

Yes.

- Are the measurements being made frequently enough to provide good statistical power?

Assumed to be yes. Raw data from clam flat surveys was not available for this assessment. NAI uses power calculations to determine whether enough quadrats have been collected each year. Therefore, it is likely that the performance criteria of having 0.8 power to detect 5 #/m2 differences between the mean density and the goal will be met.

SHL4: Area of Clam Flats in Hampton Harbor

- Are the correct parameters being measured?

Yes.

- Are the measurements being made in the correct locations?

Yes. All three of the clam flats in Hampton Harbor being tracked by the NHEP are mapped by NAI for the Seabrook Station Environmental Monitoring Program. The flats were last mapped in 1995 but will be mapped again in 2002. It may be necessary for the NHEP to map these beds in 2004 or 2005 unless NAI increases the frequency of their mapping exercises.

- Are the measurements being made frequently enough to provide good statistical power?

Uncertain. The performance criteria for this indicator is that the flat area estimates should be accurate to +/- 5 acres. This criteria cannot be checked until the 2002 maps are produced. The Monitoring Plan calls for repeat mapping of the clam flats in Hampton Harbor every 2 years in order to document rapid changes in flat shape and area.

SHL5: Standing Stock of Harvestable Oysters in Great Bay
SHL6: Standing Stock of Harvestable Clams in Hampton Harbor

These indicators are calculated from the results of other indicators. No evaluation is needed.

SHL7: Abundance of Shellfish Predators

- Are the correct parameters being measured?

No. Green crabs are the parameter being measured and these crabs are the dominant predator for soft-shell clams in NH's estuaries. However, there is no systematic monitoring for the dominant predators for oysters (oyster drill, mud crabs). Monitoring for the abundance of these predators is needed.

- Are the measurements being made in the correct locations?

Yes, for green crabs. Between several programs there is good coverage of green crab abundance monitoring stations throughout the estuaries.

- Are the measurements being made frequently enough to provide good statistical power?

The performance criteria for this indicator is and accuracy of +/-10% in the abundance estimates at each station. It is unclear how this performance criteria can be checked since measures of abundance differ based on the type of gear and method used to collect the sample. The value of this indicator is in illustrating trends in the relative abundance of shellfish predators, which the current monitoring programs provide. Annual monitoring by Seabrook Station has been able to detect 29-365% increases (statistically significant) in green crab abundance over the past 20 years. Therefore, the existing monitoring program for green crabs in Hampton Harbor is sufficient. The performance criteria for this indicator in the Monitoring Plan should be changed.

SHL8: Clam and Oyster Spatfall

- Are the correct parameters being measured?

Yes.

- Are the measurements being made in the correct locations?

Yes. Spatfall is monitored at all the clam flats and oyster beds being tracked by the NHEP.

- Are the measurements being made frequently enough to provide good statistical power?

The performance criteria for this indicator is and accuracy of +/-10% in the spatfall estimates at each bed. If the number of quadrats per bed for assessing adult oyster density is increased (see SHL2), the accuracy of oyster spatfall estimates will be improved as well. Because the densities of harvestable oyster and clams are the most important indicators for the NHEP, the performance criteria for oyster and clam monitoring should be based on harvestable oyster and clam density (see SHL-2 and SHL-3). There should not be a performance criteria for this spatfall indicator.

SHL9: Recreational Harvest of Oysters

- Are the correct parameters being measured?

Yes. It is necessary to estimate the actual harvest of oysters in order to assess the impact of recreational harvesting on oyster populations.

- Are the measurements being made in the correct locations?

Yes.

- Are the measurements being made frequently enough to provide good statistical power?

Recreational harvest of oysters is assessed through a mail-in survey of oyster harvest licensees. The last survey was completed in 1997. NHF&G plans to conduct another survey in 2002 or 2003. The Monitoring Plan calls for a repeat of this survey every 3 years. Because the survey is not statistically based, the uncertainty in the final results cannot be calculated to compare to the performance criteria of an accuracy of +/- 25%. Recent data shows that recreational oyster harvest is a small percentage (5%) of the total standing stock of harvestable oysters. Therefore, it is not imperative that the oyster harvest survey be very accurate. Given the non-statistical nature of a mail-in survey, this indicator should not have a quantitative performance criteria.

SHL10: Recreational Harvest of Clams

- Are the correct parameters being measured?

Yes. The observations of digger trips on the Hampton Harbor flats can be used to estimate the total catch by assuming that each digger catches the limit.

- Are the measurements being made in the correct locations?

Yes. All the flats being tracked by the NHEP are monitored.

- Are the measurements being made frequently enough to provide good statistical power?

For each Friday-Saturday period for which the flats are open for harvesting, Seabrook Station makes observations of digger trips on Friday and then infers the number of digger trips on Saturday. The total harvest is calculated by assuming that each digger catches the limit (10 quarts of clams). This approach involves two assumptions and one source of random error. The first assumption is that the digger trips on Saturday can be inferred from the observed digger trips on the preceding Friday. The second assumption is that each clammer catches exactly 10 quarter of clams. The observations of digger trips on Friday are also subject to random error. It is not possible to assess the accuracy of this program to determine if the performance criteria of an accuracy of +/- 25% is being met. However, data from the past 20 years indicate that the harvest estimates by this method can have high error when the standing stocks are falling or are low. For instance, in 1986 the standing stock of harvestable clams in Hampton Harbor was 2,793 bushels, but the total estimated harvest for the year was estimated to be 3,384 bushels. This discrepancy could be due to errors in the standing stock or harvest estimates, or both.

Data from the past 20 years indicate that clam harvests in Hampton Harbor have taken a high percentage of the clam standing stock during periods of peak clam density. Therefore, recreational harvest is an important indicator. To improve the accuracy of this indicator, observers could be dispatched to the harbor on both Friday and Saturday during the harvesting season. This modification of the monitoring program would eliminate the need to assume that the digger trips on Saturday are related to the digger trips on Friday. The assumption that all harvesters catch the limit would still add error to the estimate. However, it is not feasible to poll each harvester regarding his catch. A random survey of harvesters might be possible to provide some quantitative information on catch amounts with less effort. These and other options for improving the clam harvest estimates should be discussed with NHF&G and Seabrook Station.

SHL11: Prevalence of Oyster Disease

- Are the correct parameters being measured?

Yes. MSX and DERMO are both measured.

- Are the measurements being made in the correct locations?

Yes. All the major oyster beds being tracked by the NHEP are monitored. The beds at Adams Point, Nannie Island, Woodman Point, and Oyster River are assessed yearly. The Piscataqua River and Squamscott River beds are not assessed on a set frequency – but should be monitored at least once every 3 years.

- Are the measurements being made frequently enough to provide good statistical power?

Yes. The Quality Assurance Project Plan for the NHF&G Oyster Disease Testing Program determined that the sampling and analytical methods used meet the data quality objectives for this indicator.

SHL12: Prevalence of Clam Disease

- Are the correct parameters being measured?

Yes. Neoplasia, the dominant clam disease in NH's estuarine waters, has been periodically monitored in Hampton Harbor. However, the relative trends in neoplasia prevalence are difficult to discern because the results have been reported differently over the years (e.g., sometimes as percent infected, sometimes as percent systemically infected). A common definition of neoplasia prevalence is needed to facilitate trend detection.

- Are the measurements being made in the correct locations?

Yes. The major clam resource in Hampton Harbor is being monitored.

- Are the measurements being made frequently enough to provide good statistical power?

The accuracy of the methods cannot be assessed relative to the performance criteria of 10% accuracy. Neoplasia has been monitored infrequently over the past decade. NAI plans to test the flats in 2002. It may be necessary for the NHEP to fund neoplasia assessments in Hampton Harbor if the Seabrook Station monitoring continues to be infrequent. The Monitoring Plan calls for yearly assessments of neoplasia in clams from Hampton Harbor.

Land Use and Development Indicators

LUD1: Impervious Surfaces in Coastal Subwatersheds
LUD2: Rate of Sprawl – High Impact Development
LUD3: Rate of Sprawl – Low-Density, Residential Development
LUD4: Rate of Sprawl - Fragmentation

- Are the correct parameters being measured?

Yes. These indicators will be assessed using land use GIS data and other records of development. The imagery needed for this indicator may also be needed by other coastal programs. Therefore, the NHEP should coordinate with other groups to eliminate redundant work.

- Are the measurements being made in the correct locations?

Yes. These assessments will cover the entire coastal watershed using land use GIS data.

- Are the measurements being made frequently enough to provide good statistical power?

All of these indicators have a performance criteria of +/- 10% accuracy in the final product. It is not possible to assess the accuracy of the methods for these indicators at this time. These indicators will be calculated in 2003 for the periods 1990 and 2000. GIS data will have to be acquired in 2006-2007 to assess conditions in 2005.

Critical Species and Habitats Indicators

All the indicators in this chapter are considered Research Indicators until they can be more fully defined, which the exception of the eelgrass indicator discussed in NUT4. Therefore, these indicators could not be assessed relative to performance criteria. All of these Research Indicators as well as Research Indicators from other chapters are assessed in the next section to identify high-priority indicator development projects for 2003.

Research Indicators

The NHEP Monitoring Plan contains a number of Research Indicators. Research Indicators are ideas for either indicators or supporting variables to be added to the plan but which need to be developed first. In some cases, development of a Research Indicator would require creation of an entirely new monitoring program to collect the necessary data. In other cases, data are available to compute the indicator but the methods and interpretation have not been developed.

All the Research Indicators from all the chapters of the Monitoring Plan were listed in Table 1. Each indicator was then categorized based on the following questions:

1. Will this Research Indicator report on a Management Objective that currently does not have any indicator associated with it?
2. Does this Research Indicator require a new monitoring program and NHEP funding?

Research indicators for which the answer to question 1 was "yes" should have highest priority for development since these are needed in order to be able to report on environmental status and trends for all aspects of the NHEP Management Plan. The second question separated the Research Indicators for which monitoring is needed because only these need to be considered for allocating the 2003 NHEP monitoring funds. Five of the 26 Research Indicators met both criteria, making them the highest priority for indicator development funding in 2003. They are:

*Finfish and Lobster Edible Tissue Concentrations Relative to FDA Guidelines
Toxic Contaminants in Stormwater
*Open Shellfish Beds in Estuarine Waters
*Salt Marsh Extent
Freshwater Wetland Functions

Of these five, the three with an asterisk involve discrete projects that could be completed with the extra funding available for 2003 to establish baseline conditions. These three should be considered for the extra funds available for 2003. The other two would involve some ongoing monitoring which might not be available after 2003.

Details for Potential Research Indicator Development Projects for 2003

The objective of the finfish and lobster monitoring program would be to provide up-to-date information on toxic contaminants in the *edible tissues* of finfish and lobster from NH's estuary. The existing public health advisory against consumption of lobster tomalley is based on limited data of PCBs in lobsters from Little Bay in 1991 and information from the State of Maine on dioxin in lobster tomalley. There are currently advisories on all marine finfish for mercury (based on a nation-wide FDA advisory) and advisories against consumption of bluefish and striped bass for PCBs (based on regional monitoring data).

To determine the correct target species for this monitoring program, the NHEP surveyed environmental and public health departments of east coast states about their marine finfish monitoring programs. This survey identified 4 possible target species: winter flounder, lobster, bluefish, and striped bass. Bluefish and striped bass are highly migratory, and are only in NH waters for weeks at time. It was not recommended that bluefish and striped bass should be part of a NH monitoring program because the same fish are already being caught and tested by Massachusetts and Maine officials. Therefore, the two species to consider for a NHEP monitoring program are flounder and lobster. At a minimum, these species would need to be tested for PCBs and mercury. The total cost of this project will depend on the monitoring objectives and sampling design. A subgroup of the TAC will be convened during the fall of 2002 to establish the work scope and cost of this project.

The Natural Resources Assessment Group from the University of Massachusetts was solicited to submit a proposal for mapping salt marsh extent using aerial imagery. This group conducted the aerial surveys and mapping of salt marshes and benthic habitats for the Narragansett Bay Estuary Project. The total cost of the proposal ranged from $21,000 to $32,000 depending on the type of imagery used and the overhead rate. The proposal is to provide coastal wetland and deepwater habitats inventory and photoanalysis of degraded coastal wetlands, including invasive species and freshwater wetlands which are potentially former salt marsh, for the New Hampshire estuary project area (approximately 6,200 acres). The total cost is for a final deliverable as a rectified and digitized data base on a CD. All estimates include field work to support photointerpretation (3 days for 2 staff).

For the “Open shellfish beds in estuarine waters” indicator, the critical datagap that could be filled in 2003 is a uniform and comprehensive map of clam habitat in the Great Bay Estuary. In the NHEP Management Plan, Shellfish Goal #1 states that the percentage of shellfish beds open for harvesting should be increased to 75% of all beds. Objective SHL1-2, set a specific goal of 2,502 acres of open clam flats based on an estimate of the total acres of clam flats (3,369 acres). The TAC has concluded that a more accurate inventory of the total acres of shellfish resource areas (clam and oyster) in the estuary is needed before this goal can be adopted. Data on the oyster beds in Great Bay and clam flats in Hampton Harbor are readily available from other indicators (“Area of Oyster Beds in Great Bay” and “Area of Clam Flats in Hampton Harbor”, respectively). However, a uniform and comprehensive assessment of clam *habitat* in Great Bay must be completed. This project is still undefined so an accurate cost estimate is not possible. The previous work by Banner and Hayes (1996) should be reviewed as a starting point for scoping out the workplan for this project.

During the fall of 2002, the TAC will develop the scope of work for these projects and estimate their costs in order to develop a workplan for Management Committee approval in December 2002.

SUMMARY AND RECOMMENDATIONS

1. In the previous sections, all the indicators from the NHEP Monitoring Plan were reviewed to assess whether changes were needed in the existing monitoring programs (both NHEP-funded and funded by other agencies) in order to meet performance criteria. In addition, research indicators from the NHEP Monitoring Plan were reviewed for potential indicator development projects in 2003. All of the possible improvements to the NHEP Monitoring Plan as identified by these assessments have been summarized in Table 2 along with the estimated cost to implement them above current funding levels. *Therefore, Table 2 should be considered a relatively complete list of options for how to allocate NHEP monitoring funds for 2003 and beyond These recommended monitoring actions should be prioritized and used to develop the workplans for NHEP monitoring funds for 2003.*

2. If all the recommended monitoring programs were funded, approximately $75,000 to $85,000 would be needed for the programs that are repeated annually. In addition, $20,000 to $30,000 would be needed each year to keep up with assessments that are repeated on longer intervals than annually. Two important examples are the impervious surface mapping and oyster bed mapping projects, which are both repeated every 5 years.

3. Many of the indicators for assessing trends in the estuary did not meet their performance criteria. In most cases, the criteria were not feasible and should be changed. The TAC should review the performance criteria for the following indicators and develop more reasonable performance goals:

Indicator	Performance Criteria in Monitoring Plan	Suggested New Criteria
BAC2: Trends in Dry-Weather Bacteria Indicators Concentrations	Ability to detect 10% change over 5 years with power of 0.8 and using 0.1 as the level of the test	Ability to detect linear trends over 5 years of 1 cfu/100ml/yr with power of 0.8 and using 0.1 as the level of the test
BAC5: Trends in Bacteria Concentrations at Tidal Bathing Beaches	Ability to detect 10% change over 5 years with power of 0.8 and using 0.1 as the level of the test	Ability to detect linear trends over 5 years of 1 cfu/100ml/yr with power of 0.8 and using 0.1 as the level of the test
BAC8: Bacteria Load from Wastewater Treatment Plants	Ability to detect 10% change over 5 years with power of 0.8 and using 0.1 as the level of the test	TBD
TOX3: Trends in Shellfish Tissue Concentrations	Ability to detect 10% change over 5 years with power of 0.8 and using 0.1 as the level of the test	Ability to detect linear trends over 5 years of 0.05 ug/g/yr for Mg, 1 ng/g/yr for PAHs, and 5 ng/g/yr for PCBs with power of 0.8 and using 0.1 as the level of the test.
TOX4/6: Trends in Finfish Tissue/Sediment Contaminant Concentrations	Ability to detect 10% change over 5 years with power of 0.8 and using 0.1 as the level of the test	TBD
NUT2/3: Trends in Estuarine Nutrient/Particulate Concentrations	Ability to detect 10% change over 5 years with power of 0.8 and using 0.1 as the level of the test	TBD

SHL2: Density of Harvestable Oysters in Great Bay Beds	Ability to detect changes greater than 5 #/m2 with power of 0.8 and using 0.05 as the level of the test.	Ability to detect changes greater than 10 #/m2 with power of 0.8 and using 0.05 as the level of the test.

4. For some indicators, it was impossible to judge whether the performance criteria were being met because the data were not reported in a format that could be compared to the criteria. For instance, the criteria for data on shellfish predators was an accuracy of +/-10% but the absolute abundance values depend on the type of gear used. The indicators with this problem were mostly either mapping projects or supporting variables related to shellfish biology. The TAC should determine whether these indicators need performance criteria and, if they do, identify criteria that are appropriate to the types of data being collected:

Mapping Indicators	• NUT4: Eelgrass Distribution within Tidal Tributaries of Great Bay • SHL1: Area of Oyster Beds in Great Bay • SHL4: Area of Clam Flats in Hampton Harbor • LUD1-4: Impervious Surfaces/Fragmentation in Coastal Subwatersheds
Shellfish Supporting Variables	• SHL3: Density of Harvestable Clams at Hampton Harbor Flats • SHL7: Abundance of Shellfish Predators • SHL8: Clam and Oyster Spatfall • SHL9: Recreational Harvest of Oysters • SHL10: Recreational Harvest of Clams • SHL11: Prevalence of Oyster Disease • SHlL12: Prevalence of Clam Disease
Other	• BAC9: Microbial Source Tracking • NUT7: Trends in Biological Oxygen Demand Loading to Great Bay

5. The power analysis showed that it is not feasible to detect trends in wet-weather bacteria indicator concentrations (BAC3) given the high variability and limited number of storms that occur each year. Improving wet-weather bacteria loads to the estuary is a major focus of the NHEP. Therefore, the TAC should determine another way to monitor/document improvements in wet-weather water quality due to NHEP-funded stormwater pollution prevention projects.

6. The NHEP should coordinate all its aerial imagery needs in order to eliminate redundancy. The NHEP should also work with other coastal partners to seek out cost efficiencies in imagery acquisition.

7. Monte Carlo simulations should be used to determine the power to detect trends at tidal tributary sites for indicator BAC2 (Trends in Dry-Weather Bacteria Indicator Concentrations). The modeling should test the effectiveness of a weekly monitoring frequency at reaching the performance criteria for this indicator.

8. The scope of work and cost for each of the research indicator projects for 2003 should be developed by the TAC during the fall of 2002.

REFERENCES

Banner A and Hayes G (1996) Important Habitats of Coastal New Hampshire. U.S. Fish and Wildlife Service, Gulf of Maine Project, November 1996.

Chase et al. (2001) Evaluation of Gulfwatch 1998: Eighth year of the Gulf of Maine Environmental Monitoring Plan. University of New Hampshire, Durham. December 2001.

EPA (200) Guidance for assessing chemical contamination data for use in fish advisories, Volume 2, Third edition. EPA 823-B-00-008. Office of Water, Washington, DC. November 2000. Available at http://www.epa.gov/ost/fishadvice/volume2/index.html.

Gerrodette T (1987) A power analysis for detecting trends. *Ecology*, **68**(5): 1364-1372.

Gerrodette T (1991) Models for power of detecting trends – A reply to Link and Hatfield. *Ecology*, **72**(5): 1889-1892.

Hirsch RM, Slack JR, and Smith RA (1982) Techniques of trend analysis for monthly water quality data. *Water Resources Research* **18**(1): 107-121.

Hirsch RM and Slack JR (1984) A non-parametric trend test for seasonal data with serial dependence. *Water Resources Research* **20**(6): 727-732.

Kline, SJ (1985) The purposes of uncertainty analysis. *J. Fluids Engineering* **107**: 153-160.

Link WA and Hatfield JS (1990) Power calculations and model selection for trend analysis: A comment. *Ecology* **71**(3): 1217-1220.

NHEP (2002) NHEP Monitoring Plan. New Hampshire Estuaries Project, Portsmouth NH. May 2002.

Peters, DG, Hayes, JM, and Hieftje, GM (1974) Chemical Separations and Measurements: Theory and Practice of Analytical Chemistry. Saunders Golden Sunburst Series. 1974.

APPENDIX A

Power Analysis Methods

WHAT IS POWER ANALYSIS?

All statistical tests are based on a null hypothesis which is rejected if the probability of it being true becomes small enough. The probability below which the null hypothesis is rejected is called the level of the test and it is typically set at 0.05. So, if the null hypothesis is rejected, there is still a small probability (<0.05) that the null hypothesis was actually true and it was falsely rejected. This is why the level of the test is also called the Type I error or the "false rejection error".

The other possible error that can be made in statistical test is failing to reject the null hypothesis when, in fact, it should be rejected. For example, if there is a small trend in bacteria concentrations but too few samples were collected to detect it, the null hypothesis of "no trend" would be retained and the information about the real trend would be lost. This type of error where the null hypothesis is falsely accepted (or "not rejected" in statistics lingo) is called Type II error or "false acceptance error". A classic way to present these two types of errors is in the following form.

	Ho is not rejected by test	Ho is rejected by test
Ho is true in reality	Correct decision - No error	Type I Error
Ho is false in reality	Type II Error	Correct decision - No error

* Ho = Null Hypothesis

** Since Type I and II Errors are probabilities, they lie on the range of 0 to 1.

In statistical tests, the Type I error is assigned, but the Type II error is a function of the Type I error, the sample size, the variability in the population being monitored, and the smallest change that needs to be detected. Calculation of the Type II error is lengthy for all but the simplest statistical tests. Therefore, power analysis uses computer programs to compute the Type II error given inputs for Type I error, sample size, variance, and minimum difference.

The power of a test is 1-Type II Error. A power of 0.8 or greater is considered acceptable. Therefore, if a test has a 0.1 probability for a Type II error, the power would be said to be 0.9 and the interpretation would be that the test had sufficient power to detect the minimum difference needed for the study.

TYPES OF POWER ANALYSES

For the assessment of the NHEP Monitoring Plan, three types of analyses were conducted: power to detect trends, power to detect differences, and error analysis for summations.

Power to Detect Trends

To evaluate a monitoring program, power curves were generated that plotted the power of the test on the y-axis vs. the magnitude of the trend on the x-axis. Different curves were generated for different durations (e.g., 5 years, 10 years, and 20 years). At the point where each curve intercepted power=0.8, the magnitude of the trend was read off the x-axis. This value was considered the "minimum detectable trend at power=0.8" and was compared to the performance criterion for all trend indicators (ability to detect trends of 10% change over 5 years with 0.8 power).

The method used to assess the trend detection power of a monitoring program depended on the type to statistical test that would be used to test for trends. For linear regression, the computer programs

MONITOR and TRENDS were used. For the non-parametric Seasonal Kendall Test (SKT), power was estimated from power curves published in Hirsch et al. (1982). For step change functions, a two-sample t-test power analysis was conducted using SYSTAT v.10 software.

As a final check on the power analysis results, linear regression and the SKT were run on the available data from the existing monitoring programs. None of the tests were able to detect trends below the minimum detectable trend predicted by the power analysis. This result was what would be suspected and supported the conclusions from the power analysis.

Linear Regression

If the data satisfy the assumptions inherent with linear regression (normality, homoscedascity, no serial correlation), it is possible to investigate trends using linear regression. Two software programs were used to estimate the power of monitoring programs to detect trends using linear regression: MONITOR by James Gibbs of SUNY-Syracuse and TRENDS by Tim Gerrodette at NOAA. Both programs require inputs for: number of samples collected per year, number of years of data, coefficient of variation in dataset (standard deviation divided by the mean), magnitude of trend to detect, and Type I error of the test. Based on these inputs, the power of the test was estimated.

The major difference between the two programs was that TRENDS estimated the power for a linear regression through the annual means, while MONITOR estimated the power for a linear regression using all the data. Despite these different approaches, the power estimates produced by the two programs were approximately the same given the same inputs.

MONITOR v.7 can be downloaded from:
http://www.mp1-pwrc.usgs.gov/powcase/monitor.html

TRENDS v.3 can be downloaded from:
http://swfsc.nmfs.noaa.gov/prd/software/Trends.html

Non-Parametric Seasonal Kendall Test

Many of the trend indicators do not satisfy the assumptions for linear regression, so non-parametric (distribution assumption free) methods are needed. The NHEP Monitoring Plan calls for the Seasonal Kendall Test for trend analysis in these cases. The lack of assumptions that makes the SKT useful for handling environmental data unfortunately makes power calculations very difficult. The only way to independently verify the power of this test is to simulate thousands of synthetic datasets with a known magnitude of random variability and known magnitude of trend, run the SKT on each dataset, and record the percent of datasets where the test correctly rejected the null hypothesis of no-trend. This work has been done for graduate student theses or for publications in the scientific literature, but it was beyond the scope of what could be done for this monitoring plan review.

Fortunately, one paper (Hirsch et al., 1982) that has been published on this subject provided power curves for monthly monitoring programs lasting 5, 10, and 20 year durations. Different curves were provided for different types of datasets, such as "lognormal independent" data or "normal independent with seasonal cycle" data to illustrate the effectiveness of the test for different types of data. Most importantly, the magnitude of the trends for the analyses were reported in units of "standard deviations per year". Since most of the NHEP monitoring programs occur at approximately monthly frequency, it was possible to estimate the power of the existing programs to detect trends using the standard deviation from the data that has already been collected. The major downside to this approach is that it is not possible to estimate how the power would change if the sampling frequency were changed from monthly. To determine the power at different monitoring frequencies more Monte Carlo simulations would be needed. This type of more sophisticated modeling can be done if the TAC deems this necessary.

Step-Function Trend

Another option for assessing change over time is to compare the populations of data from two periods of time to determine if there is a statistically significant difference. This approach is applicable if some change has occurred between the two periods which should result in a change in concentrations. Another way that this method is helpful is because it simplifies translation between log-space and normal-space if log transformations are needed. For instance, bacteria concentrations exhibit large variability in concentrations and extreme non-normality. A log transformation is a simple way to normalize the data. However, once data are log transformed, it is difficult to interpret the results of trend analyses because the estimated trends are no longer linear but exponential. Using a step function trend test does not have this problem because it is easy to determine the log of the target value that represents a fixed percentage change from the original values.

The power of this approach was estimated using SYSTAT v.10 software with the following inputs: pooled standard deviation (assumed to be equal to the standard deviation of the dataset since both populations were assumed to have the same standard deviation), sample size in each group, minimum detectable difference, Type I error, and the null hypothesis. The minimum detectable trend was reported as the percent of change from the baseline level divided by the duration of the two time periods.

Power to Detect Differences

For some of the indicators in the Monitoring Plan, the performance criteria were stated in terms of having "0.8 power to detect a 10% difference from 1997 levels". For datasets that are normally distributed, a one-sample t-test was used to test for changes. The power of the one-sample t-test was determined using SYSTAT v.10 software in the same manner as discussed above for the two-sample t-test.

Error Analysis for Summations

Some of the indicators in the NHEP monitoring plan had performance criteria that stated that the error bars on the overall estimate should be less than a fixed percentage of the value. The indicator for which this criterion was most important was the total annual nitrogen load to Great Bay. This indicator calculates the total load of nitrogen to the Great Bay from WWTFs and tributaries in the coastal watershed. The total load must be calculated from monthly measurements of nitrogen and flow for each of the WWTFs and each of the tributaries. All of these measurements individually have uncertainties due to analytical and sampling error that must be propagated through the loading equations to determine the error in the total load estimate.

As an example of the error propagation process, the method that was used to determine the total error in the nitrogen loading from tributaries was:

- Determine the average error associated with different sampling frequencies. A study by Loder and Penfold measured nitrogen species at three Great Bay tributaries at a daily frequency for three years. The data from this study was used to determine the average percent error in the monthly load estimate associated with monthly and weekly sampling frequencies. The Loder-Penfold dataset contained daily data on both flow and concentration, and, therefore, nitrogen load. Therefore, the average sampling error was determined in terms of nitrogen load, not in terms of the components of flow and concentration.
- Determine the average percent error associated with analytical methods. Information on laboratory analytical uncertainty was obtained from the DES Laboratory. Error in the area transposition process

for estimating flow at the tidal dams was assumed to have the same uncertainty as the laboratory methods. Since flow and concentration are multiplied to estimate the load, the percent error in a monthly flow estimate due to the combination of the analytical errors can be determined by root-mean-square summation of the percent error from laboratory methods and flow methods.

- Combine the sampling and analytical percent errors for monthly measurements using RMS summation to obtain a total percent error in each monthly loading estimate.
- Multiply the total error in monthly measurements by average values for monthly loads to obtain the actual errors on a month-by-month basis over a year. Sum the actual errors over the year using RMS summation and compare to the sum of the monthly loads to estimate the percent error in the yearly load estimate from one source.
- Multiply the percent error in the yearly load estimate from one source to average values for all sources to obtain the actual errors in the yearly load from all sources. Sum the actual errors over all the sources using RMS summation and compare to the sum of the yearly loads from all the sources to estimate the percent error in the total load estimate.

Technically, assessing the accuracy of the loading values is not the same as assessing power of a statistical test. However, the error bars for the loading value can be used to determine whether loading values from different years could be considered different within their level of uncertainty. Therefore, this assessment accomplishes many of the same objectives as the power analyses

Table 1: Research Indicators from the NHEP Monitoring Plan

Section	Indicator	Needed to fill Mgmt Objective Gap?	Monitoring needed for indicator development?
Toxic Contaminants	Sediment toxicity and benthic community index of Biological Integrity (IBI)	N	N
Critical Species and Habitats	Juvenile Fish	N	N
Critical Species and Habitats	Anadromous Fish returns	N	N
Critical Species and Habitats	Benthic community index of Biological Integrity (IBI)	N	N
Critical Species and Habitats	Abundance of Lobsters	N	N
Critical Species and Habitats	Abundance of Wintering Waterfowl	N	N
Critical Species and Habitats	Phragmites in Salt Marshes	N	N
Bacteria	Microbial Pathogens and Harmful Algae	N	Y
Nutrients and Eutrophication	Frequency and duration of phytoplankton blooms in Great Bay	N	Y
Nutrients and Eutrophication	Nuisance Macroalgae	N	Y
Nutrients and Eutrophication	Eelgrass Nutrient Pollution Index	N	Y
Critical Species and Habitats	Adult Finfish	N	Y
Critical Species and Habitats	Eelgrass Biomass	N	Y
Toxic Contaminants	Public Health risks from toxic contaminants in finfish and lobster tissue	Y	N
Critical Species and Habitats	Shoreland Development	Y	N
Critical Species and Habitats	Protected, undeveloped Tidal shorelands	Y	N
Critical Species and Habitats	Protected Forest Blocks	Y	N
Critical Species and Habitats	Protected Wetlands with High Water Habitats	Y	N
Critical Species and Habitats	Protected, undeveloped freshwater shorelands	Y	N
Critical Species and Habitats	Protected, rare and exemplary Natural Communities	Y	N
Critical Species and Habitats	Conservation Lands	Y	N
Critical Species and Habitats	Forest Blocks	Y	N
Toxic Contaminants	Finfish and Lobster Edible Tissue Concentrations Relative to FDA Guidelines	Y	Y
Toxic Contaminants	Toxic contaminants in Stormwater	Y	Y
Shellfish	Open shellfish beds in Estuarine Waters	Y	Y
Critical Species and Habitats	Salt Marsh Extent	Y	Y
Critical Species and Habitats	Freshwater Wetland Functions	Y	Y

Table 2: List of suggested improvements to monitoring programs for NHEP indicators and incremental costs above 2002 baseline funding ($25,000).

Indicator	Indicator Type*	Suggested Improvements	Extension of existing programs?	Frequency	Estimated Incremental Cost	Comments
NUT5/6	EI	Deploy new sonde in Salmon Falls River starting in 2003	Y	Yearly	$5,000	Estimated cost, based on cost of 2002 pilot study
SHL2	EI	Increase the number of quadrats collected from each bed to 10 to improve accuracy in estimates of the mean density of harvestable oysters.	Y	Yearly	$0	Augment existing NHF&G program. No extra funds needed.
NUT1	EI	Switch to more accurate analytical methods for nitrogen species at HOT sites	Y	Yearly	$300	Cost difference between methods assumed to be 25% of current cost $30/sample
NUT7	EI	Switch to more accurate analytical methods for BOD5 at HOT stations.	Y	Yearly	$300	Cost difference between methods assumed to be 25% of current cost $30/sample
NUT1	EI	Monitor total nitrogen concentrations in the effluent from the 6 WWTFs in the coastal watershed not being tested by the 2002 JEL study.	Y	One-time in 2003	$10,000	Estimated cost, based on 2002 study of other WWTFs
SHL1	EI	Map Squamscott River and Piscataqua River beds to complete 2001-2002 set of oyster bed maps.	Y	One-time in 2003	$10,000	Estimated cost, based on cost of 2001-2002 oyster bed mapping project
SHL1	EI	Repeat oyster bed mapping every 5 years starting in 2006.	Y	Every 5 years starting in 2006	$20,000	$14,000 to map 4 beds in 2001 plus an estimated $5,000 to map 2 beds in 2003.
LUD1-4	EI	Repeat impervious surface maps of the coastal watershed and acquire land use GIS data every 5 years starting in 2007.	Y	Every 5 years starting 2007	$25,000	Based on $21,485 cost in 2002.
BAC3	EI	Develop a new approach for assessing changes in wet-weather bacteria concentrations. It is not possible to detect trends or differences through monitoring due to high variability.	N	Yearly	$12,000	High priority. Investigate other options for sampling and statistical models.
BAC2	EI	Increase frequency of bacteria sampling at HOT and tributary estuarine stations to at least weekly in order to detect more subtle trends.	N	Yearly	$7,500	3 add'l samples of FC, EC, Ent per mo. @ 8 estuarine stns, 3 add'l samples per mo. of EC at 9 HOT stns, $10/sample. Does not include staff time.
BAC9	SV	Establish long-term trend stations for microbial source tracking.	Y	Yearly	$12,000	$75/isolate, 20 isolates/sample, quarterly samples, 2 stations (GB, HH)
NUT4	SV	Add biomass ground-truthing stations to eelgrass assessments to for eelgrass biomass estimates (one of the NHEP's research indicators for habitat).	Y	Yearly	$5,800	Quote from JEL for sample 12 sites for canopy height, biomass and shoot density for a basic comparison to cover distribution to estimate biomass distribution in the estuary
NUT4	SV	Support aerial imagery and boat groundtruthing for eelgrass mapping to ensure that the long-term record of estuarine habitat is continued.	Y	Yearly	$5,200	Quote from JEL to acquire aerial photography of the estuary, map eelgrass distribution based on cover estimates, and conduct eelgrass ground truth observations from a boat.
NUT2/3	SV	Add total nitrogen and light extinction to suite of parameters at estuarine stations monitored by NERR and NCA.	Y	Yearly	$6,200	19 stns, 12 mo/yr., $25/sample (for TKN), plus $500/yr for maintenance and calibration of PAR sensors.
TOX3	SV	Direct baseline NHEP funding to maintaining benchmark sites in Portsmouth Harbor and Great Bay. Increase funding to add one more trend site in Hampton Harbor.	Y	Yearly	$2,120	$450 per rep, 4 reps at MECC, NHDP, and NHHS each, minus 2002 funding of $3280
SHL7	SV	Add a monitoring program for oyster drill and mud crabs to provide information on oyster predators.	Y	Yearly	$0	Discuss options for monitoring programs with NHF&G.
SHL10	SV	Add observations of digger trips on Saturdays during the harvest season to improve the accuracy of recreational clam harvests.	Y	Yearly	$4,320	Intern to observe Hampton Harbor flats on Saturdays, 8 hr/day, for 9 months, @ $15/hr. This would augment the existing observations by Seabrook Station on Fridays.
SHL12	SV	If not covered by Seabrook Station, initiate annual surveys for neoplasia in clams from Hampton Harbor.	N	Yearly	$1,000	It is uncertain whether this will be necessary. Based on cost estimates from Hampton Harbor Juvenile Clam Mortality Study.
NUT2/3	SV	Purchase PAR sensor to allow for light extinction measurements at estuarine stations.	Y	One-time in 2003	$3,000	Field test so no yearly cost except maintenance. Would need to purchase equipment in 2003. Estimated cost based on LICOR quote of $2900.
TOX1	SV	Repeat oyster and clam tissue monitoring for toxics by the Gulfwatch Program every 2 years starting in 2004	Y	Every 2 years starting in 2004	$3,600	$450 per replicate, 4 replicates/site, 2 sites (Nannie Island, Middle Ground)
SHL9	SV	Conduct oyster harvest survey in 2003 and then repeat every 3 years.	N	Every 3 years starting in 2003	$0	NHF&G will conduct this survey in 2002 or 2003.
SHL4	SV	If not covered by Seabrook Station, repeat clam flat maps every 2 years starting in 2004.	N	Every 2 years starting in 2004	$5,000	It is uncertain whether this will be necessary. Cost of aerial imagery was $2000 in 1995. Additional costs for processing and analysis.
TOX	RI	Monitoring program for toxic contaminants in edible tissue of finfish and lobsters	N	One-time in 2003	TBD	The scope of this project will be defined by the TAC during the fall of 2002.
SHL	RI	Comprehensive assessment of clam habitat in Great Bay	N	One-time in 2003	TBD	The scope of this project will be defined by the TAC during the fall of 2002.
HAB	RI	Mapping salt marsh extent using aerial imagery	N	One-time in 2003	TBD	The scope of this project will be defined by the TAC during the fall of 2002.

* EI = Environmental Indicator, SV=Supporting Variable, RI=Research Indicator