A Review of the Gulfwatch Program 1993 – 2004

Submitted by

The Regional Association for Research on the Gulf of Maine

Currently at:

School of Marine Sciences, University of Maine
Gulf of Maine Research Institute
350 Commercial Street
Portland, ME 04101

To

The Gulf of Maine Council on the Marine Environment

Care of:

Cindy Krum, Executive Director
Association of US Delegates to the Gulf of Maine Council on the Marine Environment
PO Box 2246
So. Portland, ME 04116

Review Panel Members

John Farrington
Walt Galloway
Jocelyne Hellou
Tom O'Connor
William Robinson
James Shine
Bruce Tripp
Gordon Wallace (Chair)

Introduction

The Gulf of Maine Council's Gulfwatch Program is, to our knowledge, **the only ongoing long-term monitoring program measuring contaminants of environmental concern in both US and Canadian coastal waters in the Gulf**. The Program monitors organic compounds and inorganic elements in the mussel *Mytilus edulis*, a commonly used sentinel for estimating contaminant availability in coastal seawater. As such it offers the only such data for the coastal regions of the Gulf of Maine and provides managers and policy makers in the adjacent states and provinces guidance on the current state of the Gulf and trends in the uptake of chemicals of environmental concern. The Council is to be commended for its commitment to this unique effort in assessing water quality in the Gulf.

This review is performed under the guidance of the Regional Association for Research on the Gulf of Maine (RARGOM), a consortium of academic and government institutions in the Gulf of Maine region with a long-term goal of enhancing Gulf of Maine regional research activities, resultant scientific knowledge and its use to those with interests in the Gulf of Maine. Members of the review committee reflect a wide range of experience and expertise in the biogeochemistry and toxicology of contaminants of environmental concern in the marine environment. The goal of the review is to provide independent constructive criticism and recommendations to strengthen the Gulfwatch Program and enhance its value to the Gulf of Maine Community.

We assume for the purposes of this review that the Gulf of Maine Council's criteria against which the accomplishments of the Gulfwatch Program are being measured are those identified on the Council's Web page:

- Protect and restore habitats
- Foster environmental and human health
- Support vibrant communities

Stated goals of the monitoring program are to

- Evaluate the status, trends, and risks of contaminants to the Gulf of Maine ecosystem.
- Evaluate the human health risks from contaminants in the Gulf of Maine ecosystem.
- Provide monitoring information to resource managers that will allow both efficient and effective management action and evaluation of such action.

Specific null hypotheses addressed by the monitoring program are as follows:

- Concentrations of chemical contaminants in mussel tissues are the same at all sites in the Gulf of Maine
- No changes in chemical contaminant concentrations occur in indigenous mussel tissue over time

To test these hypotheses, metal and organic contaminant concentration data for Mussel tissues have been obtained at a number of different sites throughout the Gulf of Maine since 1991 although data used to test the hypotheses are derived from data obtained between 1993 and 2004. The findings have been presented in two reports, one entitled

The Gulfwatch Program: 1993 – 2001, January 2007, 148 pp plus Appendices

and a second data report with minimal interpretation entitled

Gulfwatch 2002 – 2004 Data Report: Twelfth-Fourteenth Years of the Gulf of Maine Environmental Monitoring Plan, September 2006, 491 pp.

Data for the earlier part of the program (1993 – 2001) are also present on the Gulfwatch website:

http://www.gulfofmaine.org/gulfwatch/

The results of the first nine years of monitoring have also been published in the peer-reviewed literature (Chase et al, 2001; Jones et al, 2001).

The Review Process

The Regional Association for Research on the Gulf of Maine (RARGOM) was asked to organize and conduct a peer review of the Gulfwatch Program. Initially it was to focus on only the 1993 – 2001 data report. However RARGOM suggested that, given the fact that sampling had continued through 2005, all of the data be made available for review. The review was conducted within a framework that focused on a series of questions about the Program that the GOMC felt appropriate. Outside reviewers were asked to address these questions as part of a team that included expertise in inorganic and organic geochemistry, toxicology, and former agency personnel managing marine monitoring programs in the past. This review is organized around this framework. Individual reviewer comments are provided to facilitate use of the review and enable those seeking further information to contact the specific reviewers identified. Reviewers and their affiliation are listed in Appendix I.

Overall Recommendations of the Committee

The Committee makes the following general recommendations pending a more complete discussion with the Gulf of Maine Council's Environmental Quality Monitoring Committee and interested parties. These recommendations were synthesized from the specific and general comments that follow that are in response to the specific questions addressed in this review.

I. The Program should continue but be improved by taking the following steps:

1. **The goals of the Environmental Quality program need to be clarified and clearly identified and consistent with stated Council goals. This review could have benefited from such guidance in the context of a mutually agreed upon mission statement for the program before initiation of the review.**
2. **There does not appear to be a mechanism to identify, on a regular basis, shifts in management needs. This is a critical issue and should be addressed jointly by the program scientists, managers and others using the data and the funding authorities.**
3. **A focused workshop is recommended to accomplish the above and would benefit both the Council and its Environmental Quality Monitoring Committee and should be held as soon as feasible.**
4. **Engagement of stakeholders, NGOs, academic and government, are needed to facilitate rapid identification of emerging issues**

II. Annual review of the program should occur with emphasis on identifying new indicators/methods that might be employed

III. There needs to be timely data analysis and reporting along with frequent updating of the website

IV. Resources appear to be a problem. To this end staffing requirements should be clearly identified and the use of volunteers to participate in the program be encouraged where appropriate.

V. Sampling design and indicators related to defined goals need to include linkage between site specific concerns and statistical validity in providing a GOM region –wide assessment

VI. Justification of the current sampling design was difficult to identify or at least to understand despite the length of time the Environmental Quality Monitoring Committee apparently spent on this effort. The current design needs to be reevaluated perhaps in the context of the workshop suggested above.

VII. The committee recommends a review of the need for depuration before analysis by an appropriate working group to establish when and if such depuration is needed and the methodology required to achieve it without alteration of tissue levels for any analyte measured.

Part I – Program Review

Documents used to support the review have been identified above. Only recent financial data was made available to the committee and consequently our comments regarding adequate support are predicated on a very limited data set. The context of this review assumes a priori that sufficient resources were made available to conduct the program as designed. If not, many of the comments below may reflect the result of inadequate or marginal funding. With that caveat, each of the following questions were addressed by one or more reviewers. Initials in front of the comments identify the respective reviewer.

Ia. Are the reported findings justified?

Organics

JF: I believe that majority of the reported findings are justified by the data and interpretation. I have noted the exceptions in comments below. Specific comments are noted here for various sections of the report that contain some questionable statements or might be misleading.

WR: The conclusions and reported findings appear to be justified, since the interpretation of the data is extremely conservative. By this we mean that the Environmental Quality Monitoring Committee (EQMC) has documented the spatial extent of contaminants in the Gulf of Maine, and have demonstrated several instances where temporal declines of particular contaminants were discerned in the data. This is a very obvious use of the monitoring data. Employing the non-parametric Mann-Kendall correlation analysis for identifying spatial and temporal patterns insures that only the most robust pattern will be disclosed, whereas more subtle changes will be missed. Our biggest quarrel is not with the actual conclusions that were drawn (which are very limited in scope), but rather with the extent of the analyses, interpretation and conclusions. They do not go far enough, and do not adequately address the three monitoring goals established by EQMC (see section below on the three goals).

TO: I disagree with WR because I don't think the data allow any more interpretation. Gulfwatch can and does say something about status and trends of contaminants. But despite the stated goals, Gulfwatch can say nothing about risks (other than comparing with human health guidelines) or about habitats or the health of marine communities**.**

JF: Specific Comments:

Draft for Peer Review January 2007.

i) P.25 section 2.5.3 Comparison with NOAA Mussel Watch Results. "mussels collected even a few days apart can differ in contamination burden (O'Leary and Breen, 1998). ----- Nevertheless, qualitative comparisons are made in this

report. ---". I suggest that the extent of the magnitude of the differences reported by O'Leary and Breen be stated and placed within the context of what is found in the Gulfwatch data. The O'Leary and Breen data, if there are large differences (I do not have the reference handy), not only apply to the comparisons between NOAA Mussel watch and Gulfwatch data, but also apply to the year to year Gulfwatch data. Does this mean that year to year variability will be such that temporal trends and certain of the geospatial trends are not significant? I do not believe so, but the situation needs to be clarified. More than "qualitative" comparisons are made for temporal and spatial trends within the Gulfwatch data. This also applies to the 2002-2004 Data Report.

ii) P102. Section 4.2.11. Overview re DDT and Metabolites. The statement about the amount of DDT used in 1950s to 1960s for Spruce budworm spraying needs a reference. Note that the Dimond and Owen, 1996 reference cited at the bottom of page 103 is missing in the references. Although I cannot cite a specific published reference at this time, I believe that DDT was used extensively in spraying coastal wetlands in Massachusetts during the 1950s and perhaps the early 1960s. While a young lad at Crescent Beach in Mattapoisett, MA on Buzzards Bay in summers during that time, I distinctly remember aerial over flight spraying of the coastal areas near marshes (including nearby houses, cars – people were warned to stay indoors during the early AM hours when this happened). It may be the marsh spraying residuals in sediments that are the main sources of present contamination in some areas.

iii) P. 105 Goldberg, 1975 is missing in reference list and I believe that Goldberg was influenced by Harvey and Steinhauer (1974) Atmospheric Transport of polychlorinated biphenyls in the North Atlantic. *Atmos. Environ. 8,* 387-388. However, without knowing the exact Goldberg reference I cannot be certain.

iv) P.110. PCBs Background section 4.3.2 second paragraph. The statement that "PCBs are synthetic chemicals used in the 1930s as heat and pressure-resistant lubricants in electrical capacitors ---." Primarily PCBs were used in transformers and capacitors as insulation (not lubricants) because PCBs have excellent insulation qualities (high electrical resistivity, favorable dielectric constants) and have low flammability. (Brinkman and de Kok, 1980 Chapter 1. in Renate D. Kimbrough (editor), Halogenated biphenyls, terphenyls, naphthalenes, and dibenzodioxins and related products. Topics in environmental health Volume 4. 1980 Elsevier). I believe the definition or description the authors used is the same as from a web site which has it wrong and continues to misinform the web site readers.

v) Bottom page 11 bottom paragraph. "Certain PCB congeners are considered toxic---." State which ones and how many of these are among the analytes measured in the Gulfwatch program and why.

vi) Page 112 middle paragraph just before the section 4.3.3. It would be worthwhile including here the time trend data for DDT and PCB in cod liver oil from the southern Baltic 1971-1989 (Kannan et al, 1992). This is a powerful data set that makes the point intended in the paragraph very nicely.

vii) Page 113 top full paragraph. Inclusion of a synopsis in a figure of Pala et al 2003 data would be a plus and be helpful to users of the review and report.

viii) PCBs had various uses at one time in ship activities such as bottom paints and in electrical components in ships. Is there any proximity to shipyards or marinas of the elevated concentration sites in some Gulf areas? PCBs were also used in paperless copy paper and became contaminants in the environment when paper was recycled in some places, e.g. the Superfund Site in Fox River near Green Bay Wisconsin. Are there any sites associated with reprocessing recycled paper associated with paper mills in the Gulf watershed?

ix) Section 4.3.4 page 114 bottom to top of page 115. The FDA 2ug/g wet weight guideline is for total PCBs as measured by an estimate of Aroclors and summed in that way, unless something has changed. The sum of PCB congeners measured in the Gulfwatch program measures only a subset of what is estimated by the FDA 2 ppm wet weight guideline. This should be clarified and with a specific reference to the FDA guideline and how the PCBs are measured for that guideline compared to Gulfwatch sum of congeners in order to justify the statement about mussel burdens of PCBS being below FDA guideline concentrations. The authors are correct and have a done job after this in making the point about the specific congeners.

Also the issue of mussel burdens and how these may relate to higher or lower burdens in other harvested and human consumed sea food from the same areas should be addressed. For example, do the mussel data indicate that lobsters in the area may have higher concentrations of CB congeners of concern? (See Pruell et al, 2000. Organic contaminant distribution in sediments, polychaetes (*Nereis virens*) and American Lobster (*Homarus americanus*) from a laboratory food chain experiment (Marine Environ. Res. 49:19-36).

This would be a case of alerting authorities about needs for more extensive sampling in some areas, or that the case had been considered and was not warranted. I believe that this is one way in which the Gulfwatch data should be used.

Comments on the POLYCYCLIC AROMATIC HYDROCARBONS sections.

x) p. 116 section 4.4.2. It would be better to provide references in addition to

Environment Canada, 1998. I accessed the Environment Canada Site on line and after fifteen minutes of searching was unable to find the reference using advanced search and keywords. Since I do not have the reference, I am at a disadvantage in suggesting that the exact source for perylene via diagenesis is unknown. However, Page et al (1995) mention only bacterial processes – non-specific, and my recollection of the literature leads me to the conclusion that thus far the exact origin of perylene is unknown. However, perylene is present in combustion product PAH [Lima et al (2005), Combustion-Derived Polycyclic Aromatic Hydrocarbons in the Environment- A Review." Environmental Forensics 6: 109-131] and probably in some crude oils as well as being in sediments due to diagenesis.

xi) Page 118 NRC Oil in the Sea II, 1985 in the middle of the page. This reference is missing from the list of references. Note that there is a recent update of this report – NRC 2003 Oil in the Sea III available from the US National Academy Press online and in print that should be cited and has data pertinent to this current draft report.

xii) Page 119 top lines 5-6. The bioavailability of PAH to mussels includes PAH in water – both dissolved and colloidal. These would be added to the list.

xiii) Page 119 last paragraph before 4.4.3.2. There is something confusing here in the comparisons i.e**. similar in range to moderate levels** with 700-6600ng/g and 100-3800 ng/g cited and **somewhat lower than** PAH levels reported for the southern Med 25-390 ng/g. These statements are contradictory. Needs rewording and a clearer explanation.

xiv) Page 121 top paragraph re Boothbay Harbor. The NRC Oil in the Sea III report (2003) noted above has a section in it on the significant inputs of petroleum and fossil fuel hydrocarbons from small craft (pleasure craft and fishing boats). Trying some transplants of mussels closer to marina sources or in a series of transects away from a large marina would be a worthwhile exercise and provide good policy and management information.

xv) Page 121 New Brunswick. The St. Johns creosote site. Again this would be a good site for mussel cage transplants in a series of transect away from the site to assess bioavailability and if the site is the major source of the slightly elevated concentrations. The 120 and 180 ng/g do not seem very elevated concentrations to me. Same with Nova Scotia Broad Cove p. 121 bottom to page 122 top.

xvi) Page122 4.4.3.3 Temporal Trends. There are indications in some areas of the United States from sediment core records that PAH flux to sediments in a few areas has increased over the past few years. (Lima et al, (2003), High-resolution record of pyrogenic polycyclic aromatic hydrocarbon deposition during the 20th century, Environmental Science and Technology 37: 53-71 and

references therein). This is consistent with increases in fossil fuel use by diesel trucks and a few other sources.

It would be worthwhile as a special project to provide an assessment of fossil fuel use and use type in the Gulf of Maine watershed states and provinces, and if possible, specific coastal counties or specific coastal areas, to discern what is happening with use patterns. This should include fuel use and type of fuel by small craft and recreational craft and fishing boats.

xvii) Page 123 bottom paragraph. If this is the case, do we have some record of how the site-specific risk assessments were accomplished? This is critical information for interfacing the Gulfwatch data with policy and management /risk assessment issues. Also note that a recent paper by Booth et al (2007), Unresolved Complex Mixtures of Aromatic Hydrocarbons: Thousands of Overlooked Persistent, Bioaccumulative, and Toxic Contaminants in Mussels. **Environmental Science and Technology 2007, 41,** 457-464, raises the issue that the PAH being measured are only capturing a fraction of the real or potential biological effects associated with PAH in the samples. Much more needs to be said about the relatively small number of PAH being analyzed compared to the large number present and why that is. In my view, it started with the fact that there were few standards of PAH available and so only compounds for which there were standards and also PAH relatively easily resolved by capillary GC (e.g. parent PAH and a few of the alkylated naphthalenes and phenanthrenes) were quantitatively measured. We are stuck in this early 1980s analysis mode while there have been significant advances in analytical methods.

xviii) Summary. P. 124. Temporal Trends. The statement that "all significant temporal trends indicate declining concentrations for metals and organic chemicals" assumes that the only "significant" trend is an increase or a decrease. This is not the case for PAH. Various actions have been taken to reduce inputs. However, perhaps because of increasing use of fossil fuel and more cars and trucks, and the legacy of PAH contamination in coastal sediments, the PAH data suggests no statistically significant changes increasing or decreasing. Is my interpretation correct? This certainly seems to be the case when viewing the data for the entire Gulfwatch set from the beginning through the 2002-2004 data report (see figures therein).

A Few Specific Comments on the Gulfwatch 2002-2004 Data Report.

a. Page 57 of 490. Section 4.2.3.1 last statement. Formal trend analyses of the Gulfwatch data is an important next step ---? When and how will this be done? It should have been part of the 2002-2004 data effort. This segmentation of data collection and then waiting for "formal" assessment is resulting in interpretations that are not timely.

b. Page 202 of 490. Top of page. How are the lipid percentages determined? This is crucial since different methods of lipid analyses yield different percentages. Then the increasingly common practice of normalizing organic contaminants to lipids and comparing data from different programs often results in incorrect comparisons because the lipid methods were different. The greater lipid content of the mussels collected in northern Maine and New Brunswick sites may be related to the statement elsewhere in the reports that some of the mussels in this geographic grouping still had ripe gonads. Is there anyway of cross correlating this?

c. Page 210 of 490. next to last paragraph. Middle of the paragraph the statement: "The value of long-term environmental monitoring is evident from the analysis of the present data set. " This needs amplification. Value to whom? Scientists? Managers? And exactly how is this valuable?

d. Page 211of 490. Similar to the previous comment. (3). The statement "Gulfwatch provides a unique and invaluable source of information for management decisions on issues related to toxic contamination in the near coastal waters of the Gulf of Maine." What management decisions? Who made them? When? Why? I believe that the Gulfwatch program has value, but I cannot let such blanket statements go unchallenged. With the last statement of the page – what are the emerging contaminant concerns of coastal resource managers? – again we should also ask the coastal resource managers as well as the research scientists.

Metals

GW: Interpretation of the data in the first report covering 9 years suggest decreasing trends for Ag, Hg, Pb and Cr at one or more sites. However, high detection limits for Ag, Hg and Pb and inclusion of the 2002-2004 data raise significant uncertainty with respect to these observations. In 2003, Gulfwatch contracted with the Battelle lab in Sequim WA to perform the analyses. The limit of detections for data obtained using Battelle's methods were ~ an order of magnitude lower than previous method detection limits (MDL) (Appendix A, Tables A-2 and A-3). While Ag and Hg data suggest a downward trend, the conclusion that there are downward trends for Pb and Cr are in question when the 2002 – 2004 data are considered. In fact, the 2003 and 2004 concentrations of Cr and Pb at MASN (Sandwich, MA) and at NSDI (Nova Scotia, Digby Island), two of the benchmark stations visited annually, appear to have increased.

Unfortunately analytical problems (high MDLs) for these metals in the early years of the program pose significant difficulties in interpreting the data. Of particular concern was the lack of overlap between the two laboratories to assure the transition between labs was not a factor in the interpretation of the results. Both the digestion and analytical methodologies changed in that transition and thus, to assure continuity in data quality, an intercomparison study should have been conducted. Problems with the quality of the Hg

and Ni data were noted in the first report (Section 3.1.1). For example the Hg values in 2002 analyzed by the first laboratory are distinctly higher than those observed in the 2003 and 2004 samples (Figure 11 in 2002 – 2004 report) analyzed by the second laboratory. Analysis of the 2002 and 2003 samples by both laboratories could alleviate some of the uncertainty in interpreting year to year variations over this time period. Such a intercalibration exercise was mentioned in the first report (Section 3.1.1, p32) but the results of this effort, if exercised, were not included in either report. Given the importance of mercury from both an ecosystem and human health perspective, and coastal management efforts to reduce Hg release into the Gulf, it is important that analytical doubt be relegated to a negligible role in the interpretation of the data. That unfortunately is now not the case and can only be resolved by revisiting archived samples, if available, using methodology with lower detection limits routinely achievable in laboratories using trace-metal-clean techniques.

Discussions of Al and Fe concentrations and their interpretation are generally not warranted given the lack of a total digestion of the samples and resulting low recovery of these metals. They could be of use in interpreting ingested sediment contributions to concentrations of metals as non-depurated mussels analyzed, especially in areas with high levels of sediment contamination and also in mussels from cleaner areas but containing high amounts of sediment (Robinson et al. 1993, Gut contents: a significant contaminant of Mytilus edulis whole body concentrations. **Arch. Environ. Contam. Toxicol. 25**: 415-421).

The differences in Cu, Pb and Cr between Gulfwatch and NOAA Mussel watch samples are significant (as much as a factor of 2) and raise concerns about the intercomparability of the Gulfwatch and NOAA Musselwatch data. Was there an attempt to investigate this further to explore differences in time of collection and/or location? Was there any comparison with the MWRA caged Mussel data?

GW: Specific Comments on January 2007 Draft Report

Page 19, Sec 2.3.1:
- EQMC 1991 reference is missing
- MOAA should be NOAA

Page 19, Sec 2.4
- Tables and Figures are mislabeled – e.g. Table 3.3.9 should be 2.3.9 etc. This is a problem in this chapter and throughout the report. All table and figure references need to be checked and corrected.

Page 20, Sec 2.4.1
- Analytical procedures for Hg and Al are not given.

Page 21, Sec 2.5.1
- It is not clear how standardization (or censoring) was accomplished. Needs clarification.

Page 22, Sec 2.5.1.3
- Reference is made to Section 3.4.2.1 which as far as I was able to determine does not exist.

Page 24, 2nd ¶, line 3

Special should be spatial.

Page 32, Sec 3.1.1

Was the "soon-to-be-completed analysis….) for Hg completed and if so what were the results?

Page 33, Sec 3.1.2.4

Reference to a $P<0.1$ as marginally significant is not normally used. A $P<0.05$ is the usual divider between significant and non-significant differences. The abrupt change in Ag concentration (Fig 3.1.2, p 44 and in subsequent years (Fig 58 in Gulfwatch 2002-2004 data report) may reflect changes in analytical methodology rather than a real decrease. See comments later in report re the need to overlap analyses when changing methodologies.

Page 33, Sec 3.1.3.1

Reference to the Al concentrations as tissue concentrations is inappropriate and should be referred to as tissue plus ingested sediment Al concentrations.

Page 34, Sec 3.1.4.1

Should be "Cadmium (detection limit ± 0.2 µg/g DW)…

Page 97, Sec 4.1.3.1, 2nd ¶

The Trowbridge 2006 and Capuzzo and Anderson 1973 references are missing.

TO: I have not tried to peruse the actual data because I don't have it on a computer. GW is probably correct in pointing out differences that could be due to analysis rather than reality. However, it is not at all unexpected for concentrations of any chemical to show a trend from 1993-2001 but to no longer follow that trend for 2002-2004. Even within the 1993-2001 time series there are years that don't fit the overall trend. Moreover, if a trend is a decrease, it cannot continue *ad infinitum*. Metals will reach natural levels and synthetic organics will become undetectable.

Summary of temporal trends

	MASN	NHDP	NHCC	MEKN	NBHI	NSDI	NSFI
n	9	5/6	9	9	9	8	6/5
Ag	D	nd	nd	nd	nd	nd	nd
Al	-	-	-	-	-	-	-
Cd	-	-	-	-	-	-	-
Cr	D	-	-	-	-	D	-
Cu	-	-	-	-	-	-	-
Fe	D	-	-	-	-	I	D
Pb	D	-	-	-	D	D	D
Zn	-	-	-	-	D	-	-
tDDT	-	-	D	-	-	-	-
CHCs	nd	nd	nd	nd	nd	nd	nd
tPCB	-	-	-	-	-	-	-
tPAH	-	-	-	-	-	-	-
LMW	-	-	-	-	-	-	-
HMW	-	-	-	-	-	-	-

nd = too often nd to test for trend

Ib. Are the Gulfwatch data collection, analytical and statistical methods used (and proposed changes) sound and appropriate?

TO: Statistical analysis

Use of non-parametric Mann-Kendall statistics and medians is a good idea for all the reasons given by Gulfwatch. I would prefer, though, that they had used the Spearman test of ranks. The square of the Spearman correlation coefficient (r) is a measure of how much the independent variable (concentration) depends on the dependent variable (year, geographic sequence, latitudinal sequence, watershed, or jurisdiction). I cannot extract the same information from the Kendall-tau coefficients (they are not predictions) but I am fairly certain that some of the statistically significant correlations based on large n's are still fairly weak correlations. Of course the geographic correlations are only significant because Massachusetts Bay lies at the southern end of the Gulf (or Nova Scotia at the northern end in the cases of Al and Fe). If high concentrations happened to have been in the middle of the Gulf there would have been no linear trends.

Is replication worth doing?

The Gulfwatch protocol requires 4 replicates per site per year. Since about 20 sites per year are sampled that equates to about 80 analyses. I suggest that they sample all 57 sites per year but only analyze a single replicate

Appendix K includes a calculation that concludes that three samples must be per station every second year for 20 years (n=10) in order to have 80% statistical power of finding a temporal trend whereby concentrations change by 25% over those 20 years. I have a small quarrel with that calculation (see below) but do not think Gulfwatch needs that power. What would be a benefit is the ability to test for trends at more sites.

With a single analysis Gulfwatch would lose the ability to say anything with any confidence about a single measurement at a single site in a single year. Gulfwatch would gain the ability to find patterns over time because the "n" in any statistical test is the number of years with data. The increase in "n" would not affect the spatial trend analyses because the way the data are lumped means that the n's are fixed. For the geographic and latitudinal trends n = 57 and for the watershed trends n = 35.

I think it is the pattern that Gulfwatch is after. Nowhere in the January 2007 report has Gulfwatch made any statistical comparisons between any two years at one site or between two or more sites in any year. Gulfwatch is looking for trends rather than statistically significant differences.

As it is Gulfwatch is making little use of its four replicates. For each site and year there is one median concentration for each chemical. Those medians are used for the five sites at which trends are sought. For all the geographic trends the site medians are the medians of all the yearly medians. The lumping becomes thicker for the watershed medians and even more so for the provincial medians because these are based on medians of all site medians (already lumped over years) within the watershed or province. Adding more data might change small changes to those site or watershed or even province medians but I doubt that the spatial trends would change. The sites, watersheds, and provinces with the high concentrations now will continue to exhibit the highs. There might be an odd "high" at a low site and vive-versa but the overall picture is not going to change.

One interesting test of the Gulfwatch data would be to randomly select one of the replicates at each site/year combination and to then proceed with the data analysis. Would different spatial or temporal trends be found?

The temporal trends would be more meaningful if they could be found to occur at more than a few sites. Decreasing Ag, Cr, and Pb at MASN is interesting but is that the only site in Massachusetts with a trend? If more MA sites had been sampled more often we would know if the MASN trends were more than just of local interest.

<u>Should we expect temporal trends to follow first order kinetics?</u>

My small quarrel with the 25% trend is in the use of percent changes. The basic model for such a change is

$dC/C = kdt$ where k is the rate of change.

Why should the amount of change (dC) be at all dependent on C? There is no reason to assume that concentrations will obey first order kinetics.

The calculation of power in appendix J used the variability established with the existing set of replicates (4 per site/year) and was a Monte Carlo simulation of an underlying 25% trend over 20 years. I have no alternative to using percentages in this example because it would be impossible to run simulations for a whole series of different starting concentrations. Nonetheless the percent change idea is equivalent to

$\ln C = \ln Co + kt$

which is the model being tested whenever the Cs are log transformed prior to calculating a regression. So for the same reason I don't think one should assume first order kinetics, I don't think that log-transformed data should be used. The transformation is harmless if one is only testing for the existence of a trend but if the actual rate of change is to be used there needs to be some justification for assuming the model. Gulfwatch has not assumed such a model but I cannot tell whether or not the Monte Carlo simulations did so.

TO: I must amend my suggestion of one sample per site every other year, if Gulfwatch changes the sampling scheme (as per Appendix K) to annual sampling at 3 sites (all in NH), biennial sampling at 11 sites, 3-year sampling at 2 sites, and 6-year sampling at 32 sites. This scheme retains trend sampling at 14 sites and single samples would serve in those cases. It would be better to take single samples biennially at more sites because finding trends at neighboring sites would indicate more than just local increases or decreases. However the introduction of 34 3-and 6-year sites tells me that Gulfwatch is no longer sampling the Gulf of Maine but rather sampling each site for its own local sake. With that change replication becomes necessary. Now Gulfwatch will want to say something about each of those 34 sites in the year it is sampled. They will be using the data at these 34 sites to compare among sites and to compare with earlier data at the same site. Regardless of whether they use parametric or non-parametric statistics to test the significance of any differences they will need replication. As it stands now Gulfwatch does not need replication because they are not testing for differences among individual sites or among years at any one site

JS: The statistical analysis that was done was perfectly fine, but there is more that can be done. First of all, I didn't see any basic descriptive stats on the mean and variance of contaminants. A goal of an assessment program is to assess the distribution of contaminants in the environment. What is the average level of contaminants in the Gulf of Maine? What is the variance (and what fraction of locations exceed some kind of threshold)? Has the average level of contaminants changed over time? Has the fraction of samples exceeding a threshold changed over time? If we believe in the sample design, each sampling event is truly an estimate the distribution of contaminants out there in the Gulf of Maine. This simple information is easily interpretable and can be communicated to the lay public. You can make statements such as the average level of contaminants in the Gulf of Maine (as measured by uptake into mussels) has gone down over the last 10 years. Or, the incidence of contaminant hotspots exceeding a level of concern have been

reduced by a factor of X over the last 10 years. These metrics can be done for the Gulf of Maine as a whole, or can be done within administrative boundaries (Massachusetts, New Hampshire, Maine, etc.).

The second data analysis issue concerns the non-parametric smoothing of the data. The smoothing techniques were perfectly justified, but could use a bit of tweaking. First of all, the stations were lined up categorically from South to North, with no regard for the distance between stations. This is OK, but it would be much nicer if each station could be assigned a distance metric (i.e., station 1 is at 0 km, the next sample at 5 km, the next at 20 km, etc.). I understand that this can be difficult given the tortuous contour of the coastline. If we can do this, however, we can do some more sophisticated non-parametric smoothing that weights adjacent points based on their distances from each other. Right now, adjacent points are weighted equally regardless of whether they are 1 km or 100 km apart.

The second statistical idea concerns how spatial trends are assessed. The way the analysis was set up, you only the monotonic south-north trend in the data was examined. This doesn't make sense, as there are likely many local maxima and minima in concentrations (such as high hits in Boston, Portsmouth, or Portland) that make looking at monotonic trends nonsensical. Granted, because the smoothing is non-parametric, one can't simply differentiate a parametric fit curve to see where interesting maxima and minima occur. However, there is a rich body of literature in an area called "bump hunting", techniques to identify local maxima (or minima) in non-parametric curves. These techniques can be applied to this data. If desired one could even try multivariate non-parametric smoothing to control for differences in condition index on spatial variation in contaminant levels. Bump-hunting can be used to seek maxima in non-parametric hypercubes from multi-dimensional data. There is a lot of good information yet to be mined from this data!

Organics

JF: The current methods look fine to me with the exception of the question about how are the lipid analyses done (see my earlier comment above). I am also concerned about the need to broaden the suite of PAH analytes as noted above with the reference to Booth et al (2007). This has been partially addressed with the proposed addition of alkylated PAHs as in Appendix K3.

JH: Treatment of nds for aggregated organic groups such as tPCB was to let tPCB=0 if none of the 22 congeners was detected. It is not clear what was done in cases where at least one congener was detected. Were ½ MDLs used for the non detected congeners? If so the tPCBs in those cases are probably more above 0 than it should be.

TO: Gulfwatch should use zeros for all undetected compounds when creating sums of those compounds. If they are using ½ MDLs sometimes, and not at other times, they are confusing the issue. It also confuses the issue to use ½ MDLs all the time because it

raises the lowest sum to where it will be barely distinguishable from cases where the sum includes only one or a few actually detected concentrations.

Metals

WR: The authors of the January 2007 report (the "Nine year report") state that analyte medians and median absolute deviations are "superior" to the use of arithmetic means and standard deviations (p 21). However, they did not reference any peer-reviewed articles or grey literature to support this contention. Adopting medians seems to be an abrupt change to their procedures (which is also reflected in the Gulfwatch Interactive Mapping Tool on their web page), but the rational for this change has not been clearly explained. Even in their September 2006 Gulfwatch 2002-2004 Data Report they relied on means and standard deviations, indicating that this change was made rather recently. The Monte Carlo simulations recently run to determine whether four replicates of each site should be maintained assumed a normal distribution of contaminant concentrations based on means and standard deviations (of each site? of each year? of each year at each site?).

The procedures used for analyzing the metals and organic contaminants appear to be adequate. The EQMC noted the problems they had with the analysis of Ni, Hg and Al (Section 3.1.1, page 32), and are trying to correct some of the problematic data. This may prove not to be possible. This, however, points out why stringent QC/QA procedures are absolutely mandatory in monitoring programs, and why a timely assessment of the QC/QA results is needed before the data are statistically analyzed, and certainly before new field samples are taken. It appears that the realization that there was a problem with the Ni, Hg and Al data occurred only after many years of sampling and needless analysis of these analytes. Funds were wasted in the endless analysis of these problem analytes.

As noted in the January 2007 report, an analysis of whether four replicate samples (of 20 mussels per metal and organic replicates) need to be analyzed has recently been conducted. This analysis should have been done long ago. I am not convinced, however, that the criteria used - the ability to detect a 25 % change in a contaminant concentration over time - is realistic. For some contaminants (many metals, for example), a 25 % decrease may never be achieved, although a smaller decrease might be an important indication of an improvement in contaminant loads. For some of the highly persistent contaminants (PCBs, chlorinated pesticides), a 25 % change may take multiple decades. The 25 % change criterion seems to be too high a bar to be realistically met in the relatively short term. For some contaminants, a 10 % change may be highly significant. However, whether the 25 % change criteria is appropriate or not (and the ultimate decision as to whether four replicates need to be analyzed at each site) really depends on what the actual goals of the program are. As noted in section f below, the existing program does not meet the monitoring goals that have been stated. Changing the number of replicates will not address this issue. Until the actual goals are clarified, the choice of sample size cannot be made.

The criteria for site selection in the 9-year Gulfwatch program seem very broad, and the choice of stations is left to each of the five jurisdictions to decide for their own area. Although the January 2007 report stated that "...well-defined characteristics must be present at each chosen sample site" (page 13-14), details of these characteristics were not included in the materials sent to the reviewers (these characteristics are apparently contained in Crawford & Sowles 1992 and Sowles and Crawford 1993). Even in the newly proposed Program, the criteria used for choosing sites seem quite broad (almost any station can be easily justified), and the actual choice of sites are once again left to the individual jurisdictions.

GW: In addition to my comments in section 1a, there are some apparent inconsistencies between data reported in the first report and shown in the second report. For example Cr data in Figure 3.1.5 in the first report appear to be different than data shown in Figure 60 in the second report. A similar observation can be made with respect to Pb in Figure 3.1.13 and Figure 64 respectively. Clearly if changes in the database have occurred there should be clear documentation of the justification for doing so.

Proposed changes in the monitoring program are presented on a jurisdictional basis for the next 12 years. All of the benchmark sites would be sampled on two-year cycle rather than annually with the exception of the site in New Hampshire where annual sampling would occur, and in New Brunswick where the former benchmark station would be sampled only on a 6 year rotating basis. Various additions of sites are proposed to address site specific management concerns, e.g. the condition of St. John's Harbor and in the Bay of Fundy. A number of sites would be discontinued. For example the MAPR site in Massachusetts, reflecting conditions in an urban impacted environment would be dropped. It is not clear why such a station should be dropped and not monitored for trends in view of the site criteria established in Appendix K. Two of the criteria stated for sites used in trend analysis are 1) significant contamination and 2) high population/industrial activity, both of which apply to this site. The argument made in the narrative seems to support keeping the site rather than dropping it!

In general it is not clear whether there has been an in-depth examination of proposed site changes with respect to either the spatial or temporal framework of the program. The power analysis conducted was based on the ability to detect a 25% increase over 20 years with biennial sampling given uncertainties about the mean for the replicates over the 1993 - 2001 samples for two metals and 3 classes of organic contaminants. With these assumptions, the number of reps required for metals could be three rather than the current four. The analysis should probably be repeated with a broader perspective. Analytical detection limits and precision have changed using the newer methodology and thus might produce different results. More importantly the analysis was designed to evaluate increases and not decreases in concentrations, the latter of which is needed to confirm management success in controlling sources of these contaminants to the Gulf. On a more fundamental level the question as to whether a 20 year time period is acceptable basis for making management decisions should be addressed.

In addition to reevaluating the above analysis using the improved analytical uncertainties, there should also be a more integrated approach to evaluating site selection that includes source strength analysis, physical circulation patterns, and the more abundant information on sediment concentrations. Clearly this is a nearshore monitoring program and while high resolution physical models have not yet been established for shallow coastal embayments in the Gulf, more information should be integrated into the process with input from managers on source strengths and high growth areas, aquatic and sediment chemists, physicists and modelers, and biologists to identify critical habitats most sensitive to point and non-point inputs of contaminants. Perhaps this exercise has already been conducted but if so is not reflected in the reports reviewed.

TO: Chemical analysis

These are small things:

Total extraction of Ag requires using a large excess amount of HCl (yes HCl, Daskalakis, et al. 1997, Evaluation of digestion procedures for determining silver in mussels and oysters, Environ. Sci. Technol. 31, 2303-2306.). They know this at Battelle at Sequim because Eric Crecilius is one of the et als.

As Gulfwatch notes most of the Al in mussels is within ingested particles. Total extraction of Al requires HF and is not usually used in the digestion of mussels (however see Robinson et al. 1993 Arch. Environ. Contam. Toxicol. 25: 45-421). The Al concentrations are compromised by incomplete extraction but Al in mussels is not affected by human activity (i.e. it is not a contaminant)

Cr is also a structural element of particles whose complete extraction requires HF. Nonetheless I do accept that the high Cr concentrations found in areas under the influence of tanneries is Cr in tissues. The contribution from particles will only appear when the tissue level is low.

Ic. Are there alternative data analysis methods that should be included?

WR: Gulfwatch has collected data on mussel condition indices (CIs) and shell growth over the years, but have apparently done nothing with this data. It is unclear as to how shell growth was determined. It may have only been measured in the early transplant studies, which were not reported on here. It was stated on page 10 (paragraph 2) that growth rate is a "... fundamental measure of physiological fitness/performance....and, therefore, a direct integrative measure of impairment to physiology." If this is so important, where are the data and where are the graphs depicting temporal and spatial trends? Where is the analysis of this growth data? The authors of the January 2007 report acknowledge that the condition index data could be useful in interpreting contaminant body burdens (page 14, paragraph 4), but have not used this data in their spatial or temporal trend analysis. As stated on page 14-15, "CI data are not included in this review because no large differences were seen." However, in the Gulfwatch 2002-2004 Data Report, it was stated (p204) that "ANOVA performed on CI means was

significant (p,0.05)." This statement seems at odds with the previous statement. (I'm also not sure what 'performing ANOVA on means' really means.) Where is the analysis? What is a "large" difference (10%? 20%? 50%?). Why are the data not available on the Gulfwatch web site? in the Appendices to the January 2007 report? A summary table (Table 10, page 205-6 was included in the 2002-2004 Data Report, but the individual data that the means and standard deviations were based on was not reported? Would a multi-variable non-linear regression technique be useful for including these data in the analysis of contaminant trends? This approach could also use the wet and dry weight data. If the data is not going to be used, why collect it?

Organics

JF: I believe that it would be worthwhile to conduct an assessment of spatial and temporal trends of major individual PAH instead of lumping them together in sum PAH or LMW PAH and HMW PAH. An assessment of ratios would be worth considering as well. The LMW PAH and HMW PAH groupings may be missing some important trends and just because this is how NS and T started doing it does not make it the only way or necessarily the correct way to do interpretation of PAH data.

Metals

WR: Since mussels were not allowed to depurate gut contents prior to analysis, "...the common crustal elements, aluminum and iron, were measured to help assess whether elevated trace metal concentrations were associated with direct tissue contamination or with the ingested sediment particles." However, no analysis was presented to estimate the degree to which gut contents effected metal concentrations. This issue was raised at the start of the Program, and again prior to the fifth year review. A paper by Robinson, Ryan and Wallace (1993. Arch. Environ. Contam. Toxicol. 25: 45-421) was given to the committee at the time as an example of how metal concentration could be corrected for gut sediment. It was noted on page 97 that two depuration studies, apparently by Gulfwatch participants, showed that only Fe, Al and possibly Cr body burdens were affected by depuration, yet there was apparently no analysis of how these local results could be applied to entire Gulfwatch Program. Neither of these studies (Chase et al. 2002, nor Jones et al., 2005) were included in the list of References Cited. If the conclusion is that no correction is needed, this analysis needs to be presented. A small number of additional samples (depurated vs non-depurated mussels; analysis of water column particulate matter, etc.) could be done, either infrequently or on a one-time basis, to address this question. This question, nevertheless, needs to be addressed. Also, the possible contribution of organic contaminant binding to particulate organic matter should be assessed.

Problems with the analysis of Ni, Hg and Al were noted in Section 3.1.1. on page 32. Will these problems be resolved in time for this year's sampling?

GW: Of prime importance is the completion of the analysis of all of the data to reevaluate trends over the entire time period sampled. Trends in metal concentrations

need to be reevaluated before changing the design of the program. In addition there should be some attempt to evaluate the relative sensitivity of the sites to ambient changes in concentrations. Sites which are actively flushed with ambient GOM water will be sensitive only to very large excursions in contaminant loading whereas those located in more poorly flushed areas will be much more sensitive to potential changes in loading. Other than a few well chosen reference sites it would be to the program's benefit to choose sites based on their sensitivity to potential changes in loading as well as the criteria listed in Ib. above.

Id. Was information valuable to coastal and ocean managers discerned from the results disseminated?

Organics

JF: The report of Appendix I. and the Appendix J report of the Environmental Quality Monitoring Committee (EQMC) of the Gulf of Maine Council on the Marine Environment provide evidence that the data and interpretations have been useful in informing an impressive spectrum of policy and management actions. At some point soon it might be worthwhile to conduct a social science/political science type survey of managers and policy folks in a blind or double blind manner that would be a more objective way to assess the utility of the Gulfwatch program in the policy and management arena. The EQMC has made an excellent suggestion (Appendix K4) of having some new policy and management people involved in the EQMC to avoid an inbred aspect of the program. This suggestion applies as well to peer reviewers and scientific advisors to the program. A balance of continuity with an added influx of new ideas is the optimal path to choose.

There is a specific point that needs addressing. On page 1 of Appendix I in I.4.1 second paragraph, last sentence. "It is not known if any shellfish beds have been opened or closed as a direct result of Gulfwatch data." This seems to me to be a relatively straightforward survey problem. Two decades ago at a scientific meeting I heard the statement that most shellfish beds where chemical contaminants were high enough to warrant closure would already be closed due to simultaneous contamination by human pathogen vectors or indicators. Does this hypothesis hold for the Gulfwatch Data?

Metals

WR: See Ig below.

GW: Extensive use of the Gulfwatch data is documented in Appendix I of the first report. Specific uses of the Gulfwatch data has been given and it is apparent that the Gulfwatch program has impacted decision making throughout the region in both the US and Canada. Such visibility makes it imperative that the data and its analysis be of highest quality. This has not yet been achieved for several important metals in the dataset

(Hg, Pb of importance to human health concerns and Ag as an important tracer of sewage).

Ie. Is the program efficient and cost-effective when considered in light of the Council's criteria?

WR: Complete financial data was not included, so we cannot determine whether or not the program is cost-effective. Based on the limited information that can be gleaned from the materials sent to us, it appears that the program is resource-starved.

GW: To our knowledge this program is the only intensive long-term regional monitoring program for metal and organic contaminants in the Gulf of Maine. NOAA's National Status and Trends Program has measured contaminants in the sediments at a number of sites in the Gulf of Maine (n = 15) that are sampled every other year. Of the total 9 are located in Massachusetts coastal waters, one in New Hampshire and 5 in Maine. In contrast Gulfwatch has sampled 5 sites every year, 27 sites on a three year rotating basis, 6 sites every 4-6 years and an additional 18 sites on a less frequent basis. The results of the first seven years of monitoring have been published (Chase et al 2001) and a Gulfwatch web site provides access to data collected from 1993 to 2001.

Costs of the Gulfwatch Program were only made available for the last 4 years and are summarized in the table below. Assessment of the efficiency and cost-effectiveness of

Year	Cost
07/01/03 – 07/30/04	$61,165
07/01/04 – 06/30/05	$108,094
07/01/05 – 06/30/06	$140,667
07/01/06 – 06/30/07	$110,440
	$420,366

the program is difficult to judge in absolute terms. Total expenditures were variable over the last four years for reasons unknown to the reviewers and a detailed breakdown of specific expenditures over the entire conduct of the program being reviewed was not provided. Environment Canada laboratories were used for the analysis of organic contaminants consistently over the period of monitoring covered in these reports (1993 – 2004). In contrast the metals analysis were provided by two different laboratories over the same period, the State of Maine Health and Environmental Testing Laboratory (1993 – 2002) and Battelle Marine Science Laboratory in Sequim, WA, the latter having ~ an order of magnitude lower detection limits for most metals and the ability to accurately determine Ag and Hg concentrations that were frequently below the detection limit of the State of Maine laboratory. The change in detection limits is addressed in other sections of this report.

The data do provide some insight into the status and trends of contaminants over the lifetime of the program and are complimented by data from other programs that have

been used by Gulfwatch for comparison to their own data. Review of the trends in metal and organic contaminants are addressed below.

JF: I cannot answer this without access to the funding levels and when funding is made available as compared to when data and interpretations are expected and then delivered

If. Are the three original monitoring goals being met through the current program and if not, where are the deficiencies? (Refer to monitoring goals in the Gulf of Maine Environmental Monitoring Plan [Hayden, 1991])[1]

JF: Thus far, for the first two goals I would say mostly yes. The exception is to add emerging contaminants such as the polybrominated biphenylethers and other similar emerging organic contaminants to the list of analytes for at least a quick survey in one year to discern levels and then perhaps follow up with more intensive analyses. The other question is the extent to which other programs (other than Gulfwatch) provide harmful algal bloom monitoring data and also human health pathogen data for beaches and shellfish areas. A more complete matrix of how human health concerns as assessed by various states, provinces, and agencies within the Gulf of Maine region is needed to fully answer the question posed. A similar recommendation has been made by the EQMC in Appendix K4.

GW: Gulfwatch data has the partially realized potential to assist in the **evaluation of the status, trends, and risks of contaminants to the Gulf of Maine ecosystem** as the only long-term intensive monitoring program on a region-wide basis. Further improvements in the analysis for metals combined with a better assessment of the representativeness of the sites chosen (see Sec.Ic.) and inclusion of the 2002-2004 data will be important. **Evaluation of the human health risks from contaminants in the Gulf of Maine ecosystem** will require follow up in areas where Gulfwatch has demonstrated elevated levels of contaminants. Gulfwatch provides direction to coastal areas impacted by contaminants requiring further investigation, but by itself cannot be used to assess human health risks except where concentrations of the contaminant in edible mussel tissue analyzed exceed criteria. As such Gulfwatch does **provide monitoring information to resource managers that should allow both efficient and effective management action and evaluation of such action** although more timely reduction and analyses of the data are needed.

WR: The Environmental Quality Monitoring Committee (EQMC) evidently sees the Program "objectives" as much narrower than the three Gulfwatch monitoring "goals" that were stated in section 1.1. There is a disjunction between the stated goals of Gulfwatch and the objectives that Gulfwatch has set for itself. The EQMC apparently sees the program objectives as "...assess[ing] the spatial extent and temporal trends of chemical contamination in mussel tissue at sites along the coast of the entire Gulf of Maine." This is reflected in their two null hypotheses (page 10). They view success as meeting these objectives, not necessarily in meeting the programs three goals.

Unfortunately, none of the three goals have been adequately addressed. The EQMC apparently realizes that the Program (the first 9 years of it) is a "first step" and has a limited scope (page 10, paragraph 2; Appendix J Section 2.2, page 25). Nevertheless, it is unclear why the committee has conducted a program for over 9 years that does not match the stated goals of the program. This appears to be predominantly a resource issue rather than a criticism of the Environmental Quality Monitoring Committee who have conducted a strong monitoring program for over 15 years with limited financial support. Nevertheless, as reviewers we are being specifically asked to determine whether or not the Program has met its three goals. The answer is 'no.'

Goal #1: ***To provide information on the status, trends, and sources of risk to the marine environment in the Gulf of Maine*** **(Section 1.1, page 9)**

The Gulfwatch findings do address the status, trends and sources of the usual suite of contaminants in the Gulf of Maine. This is important and useful information, and the long-term monitoring of these contaminants is the major strength of the Gulfwatch program. However, how the status and trends of contaminants relate to the Gulf of Maine environment, and therefore the ecosystem, has not been addressed. The risks of contaminants to the Gulf of Maine ecosystem are also not addressed [FOLLOWING REF ADDED](e.g. comparison to critical body residue data such as in Jarvinen & Ankley, 1999. Linkage of Effects to Tissue Residues: Development of a Comprehensive Database for Aquatic Organisms Exposed to Inorganic and Organic Chemicals. SETAC Press; or ecological risk assessment). In addition, if Goal #1, as stated, is really the mandated goal of the Program, then risks and trends in addition to chemical contaminants should also be addressed (e.g. habitat loss/degradation, overfishing, invasive species, human population growth, etc.). These risks to the marine environment were not addressed in the materials sent to us.

Goal #2: ***To provide information on the status, trends and sources of marine based human health risks in the Gulf of Maine*** **(Section 1.1, page 10)**

There was only a cursory examination of human health risks. Three metals (Cd, Cr, and Pb) were compared to the 1993 FDA guidance levels for Cd, Cr, Pb and Ni. (p 100-101). Pb levels exceeded this level at three sites (Boston Harbor, Portland Harbor and Boothbay Harbor). For DDTs and other chlorinated pesticides, no "toxic threshold" values were given or referenced, so a health risk was not presented. For PCBs, mussel concentrations were compared to the Health Canada and FDA guideline of 2 μg/g wet wt (or approximately 10 μg/g dry wt, assuming 80% water content). All Gulfwatch sample were below this threshold. It was noted in Appendix I, page 3 that Environment Canada has guidelines for contaminant levels in shellfish taken from areas of known chemical contamination that prevent these shellfish from being used as human food, yet these guidelines were not applied to the Gulfwatch results. It should also be pointed out that there are several other organic contaminant guidelines that FDA has published that are applicable to Gulfwatch data, but were not mentioned (e.g.dieldrin, DDTs, heptachlor, heptachlor epoxide; see http://www.cfsan.fda.gov/~comm/haccp4:/html).

It should be noted that FDA guidelines and guidance levels tend to be much higher than other notable threshold values. Threshold values from other countries can be found in Nauen (1983, FAO Fisheries Circular No 764). While dated, Nauen's manuscript is still used by the international community and provides comparative human health thresholds for many of the inorganic and organic contaminants measured by Gulfwatch. In this compilation, for example, FDA's total mercury limit (1.0 mg/kg wet tissue wt) is at the high end of the spectrum which ranges from 0.1 to 1.0 mg/kg wet tissue wt). Environment Canada has compiled comparative threshold values for PCB levels in seafood (Summary of Environmental Criteria for Polychlorinated Biphenyls (PCBs), 1987. Technical Report # En 49-4/4-1, Ministry of Supply and Services, Canada. Bearuegard Press, Ltd.). In this compilation, total PCB limits in various countries ranged from 0.1 to 5 mg/kg wet tissue wt in fish and shellfish, with FDA's level (2 mg/kg wet tissue wt) at the higher end of the range. It would be better to use some of these more conservative guidelines, and/or follow EPA's "Guidance for Assessing Chemical Contamination Data for Use in Fish Advisories", Volumes I through IV 1994. EPA 823-B-94-004 June 1994 and subsequent volumes for a more realistic assessment of human health impacts.

Another means of addressing human health risks is readily available for the Gulfwatch PCB data since individual congeners were routinely measured and several coplanar congeners were analyzed in separate studies. Toxic Equivalents (TEQs) can be calculated using the congener-specific concentration data and Toxic Equivalent Factors (TEFs) compiled for human health risk. This would allow the toxicity of the individual PCB congeners to be summed and compared to a human health benchmark.

Goal #3: ***To provide appropriate and timely information to environmental and resource managers that will allow both efficient and effective management action and evaluation of such action*** **(Section 1.1., page 10).**

Gulfwatch data and their interpretation do not appear to be provided sufficiently quickly enough to be of use to environmental managers (Since this section is an issue that is also important for the redesign/redirection of the monitoring program, see section g below for a discussion of this point)

JH: Goal 1: Status, trends and risks of contaminants to the Gulf of Maine Ecosystem

Mussels are exposed to water soluble and particle bound chemicals and will reflect the state of the environment they inhabit. When animals are sampled in the field, they contain two kinds of materials that are extracted and measured as bioaccumulation in this and many other MW programs. However, some of the quantified chemicals are present in a temporary state, as particles in the visceral mass and can be readily eliminated as faeces. A second part of the detected chemicals have undergone desorption from particles or were bioaccumulated directly from the water soluble phase and are bound within the soft tissues. Both types of chemicals are presently measured in the GOM MW. An unknown

proportion of the determined body burden is actually not representative of the health status of the animals because it represents a transient load that will not affect the animals.

There is a discussion of how to deal with the bioaccumulation of metals by normalising the results; however in my opinion, it would be much easier to make sense of the data by implementing a modification in the mussels' collection. It has been shown that in the case of organics, a 24 hrs depuration is sufficient to eliminate organically enriched particles. In the case of metals, a 48 hrs depuration is needed to eliminate more inorganically enriched particles. An analytical challenge that would result from implementing a depuration period would be a lower bioaccumulation of chemicals. In the case of organics, this would lead to lower detected levels of bioaccumulation. It is not stated how many of the mussel samples display non detectable concentrations, only that half the detection limit is used to calculate the means. This might also contribute to larger standard deviations and misrepresent the true state of the environment, the level of variability.

The amount of particles present in mussels' tissues and within the measured bioaccumulation will depend on the water circulation in the sampled area. The currents, turbidity and sediment transport is affecting the various results in an unknown manner. If there was no major change on the shore line surrounding the sampling site, over the years of sampling, then it is safe to compare the chemical data obtained at one site, over time. If there was shoreline construction changing the physical oceanography in proximity of the sampling, then the scale of the effect on the data is also unknown.

It is very nice to see that all animals were collected over the same two weeks period of early June. I am sure this restricted sampling time represented a major commitment from the volunteers involved in the mussels' collection. It is mentioned and justifiably expected that the condition index (CI) of the animals will be representative of the health status of the animals. The question is which proportion of the CI solely reflects the gonad development, as opposed to the feeding conditions? Sorting out what level of variability in the CI results is indicative of a difference in the gonad development or of the feeding conditions faced by the sampled mussels represents another unknown. In my experience, differences in water temperature are blamed for differences in gonad development; however this is not necessarily the case for all samples.

Goal 2: Human health risks from contaminants in the Gulf of Maine ecosystem

This item is closely linked to the previous one, with the difference that regardless of the lack of depuration, there are low concentrations of contaminants in most of the bivalves, except for a few sites (e.g. Table 8).

It would seem logical to monitor the sites displaying a higher chemical body burden as a priority. It does not seem necessary to check on the human health risks where levels are <DL.

The guidelines for human health were not outlined in any parts of the reports. To the best of my knowledge, PAH are not on such a list in the US or Canada, but there are guidelines in some European countries, such as Germany.

Goal 3: Monitoring information to resource managers that will allow both efficient and effective management action and evaluation of such action.

There is discussion of communicating the results of the MW. The plan of action relative to higher managerial levels must depend on the involved jurisdictions.
Managerial decisions are not similar across jurisdictions and unpredictable because of changes in politicians/governing political parties, within the US and Canada.

TO: We have had an email discussion of depuration. I recommend against it for the practical reason that it's a logistic nuisance that won't change the measured concentrations except for three elements that are much more highly concentrated in particles than in mussels; Al, Fe, and Cr. (Al and Fe are not contaminants in the sense that their concentrations in mussels can be affected by human activity.) In addition, there is also a theoretical reason for allowing the mussels to retain their fecal material. Yes, some of that consists of particles inadvertently ingested as the mussels filter their food from suspended matter. And, yes, the chemicals on those particles should not count as part of a mussel's body burden. However, fecal material also contains undigested food (i.e. plankton) and the chemicals associated with it are part of the environment that we are asking mussels to monitor for us. At any moment the measured concentration of a chemical in a mussel is the resultant of ingestion, assimilation, growth, and defecation. Ideally this is all at steady-state at the time of collection but probably not (this is why it's a good idea to always sample at the same time of year as Gulfwatch does). Once a mussel is sampled, ingestion ends and so should defecation.

JS: Human health outcomes are clearly of importance, but to frame human risks via direct consumption of mussels via comparison to FDA or WHO thresholds may not tell the whole story. As others have pointed out, mussels generally take up contaminants from the water (through both dissolved and particulate routes of exposure). Since contaminants generally reside in the sediments and not the overlying water, adverse seafood concentrations will generally be seen either in benthic species or higher trophic level pelagic species through combined benthic-pelagic food chain coupling. The concentrations of contaminants in mussels shouldn't be seen as a manifest (direct) measurement of human health risks, but rather as a latent (indirect) indicator of the presence of contaminants in the local environment and therefore potential human health risks. The link between mussel concentration and human health risks depends on the water/mussel partitioning of contaminants (fairly well constrained), water/sediment partitioning of contaminants (not so well known), sediment/benthic organism partitioning (reasonably understood), benthic/pelagic coupling (not so easily modeled), and trophic dynamics (potentially can be modeled, but may be site specific and thus quite variable). All of these uncertainties lead to an unclear relationship between contaminants measured in mussels and human health risks. Again, contaminant levels in mussels should be

looked upon as latent indicators of human health risk, not necessarily as direct indicators of risk.

Ig. Has there been a consistent and timely review of the data to permit the full utility of the data to be used in redesign/redirection of the monitoring program? (See Id. Above.)

Organics

JF: If the delay in getting this review going and the debate about including data from years 2002-2004 in a 2007 review is any indication, timeliness is not a major plus for this program in terms of data review and redesign. How this is related to question (d) above is unknown to me for reasons stated in the answer to question (d.).

Metals

GW: No. Whether a function of delays in receiving the analytical data or a lack of dedicated resources supporting data analysis is the problem, it needs to be corrected. The program could be more effective if there was a mechanism for annual systematic updates and interpretation of the data designed into the program. Much of the framework of the reporting function of the program is in place to permit rapid updates and interpretation if properly funded. The website and mapping tools are of use but only if updated with current data.

WR: No. This is a major shortfall of the program and is a problem that was acknowledged by the EQMC itself in the January 2007 report. In order to be useful to both managers and to the investigators of the program, the data needs to be generated quicker and analyzed faster. There is no indication in the materials provided that the data or interpretation of the data have been externally reviewed (or even internally reviewed by the Environmental Quality Monitoring Committee).

It appears that the Gulfwatch Program itself has not been externally reviewed since a 5-year review was conducted (in 1998). As I remember it, this 5-year review report (which included the 2-year pilot project, 1991-1992, plus the first three years of Gulfwatch, 1993-1995) was hastily drafted and contained an incomplete analysis of the data. I am not aware of other program reviews.

As indicated in Appendix K, the Gulfwatch program was slated for a program review after the initial 9 years of operation (1993-2001) so that the program design could be assessed and modifications made as needed. While waiting for this review, the initial program design was followed for three additional years (2002, 2003 and 2004). Finally, in the summer of 2005, the Environmental Quality Monitoring Committee met and developed a new design for the program which was initiated in the Fall 2005 sampling season. It appears that the program was modified in 2005 even though a formal review of the program was not conducted (until now, in 2007). It is not clear what effect this current review will have on the currently used monitoring scheme. So, it is clear from the

materials received that there has not been a "consistent and timely review of the data to permit the full utility of the data to be used in redesign/redirection of the monitoring program." This raises several critical questions. Was it because resources were not made available to the EQMC by the Gulf of Maine Council to conduct the review? Or, was it an issue of efficient allocation of resources (money and time) by the EQMC itself? Why was the program modified prior to the benefit of an extensive external review to help assess the program's existing strengths and weaknesses?

We have been asked (in April 2007) to review the first 12 years of the program, and were sent an in-depth report on the first 9 years (1993-2001), plus a data report for the three year period 2001-2004. The reviewers did not even get a complete set of hardcopy texts until May for an early June deadline. In addition, the copy that was received (as well as the Gulfwatch 2002-2004 Data Report, September 2006) are rift with typos, missing axis labels on numerous figures, poor figure and table captions, and erroneous references in the text to figures and tables. While the analytical results are probably still being assembled for the Fall 2006 sampling period, it is surprising that the 2005 data could not be incorporated into the report, and even more surprising that the 2001-2004 data was not included in the in-depth analysis of the program. If it takes well over a year for analysis, quality control checks, data compilation and cursory analysis to be completed in order to include data into a "Data Report", and over three years for the data to be incorporated into an in-depth analysis, how can the Program claim that the data has been used to modify the Program in a timely manner or that it is useful to environmental managers?

In the past, it has taken far too long draft, finalize and disseminate Gulfwatch reports:

Report	Sampling	Report Date	Approx. time
Evaluation of 1991 Pilot Project	Fall 1991	Oct 1992	~1 yr
Evaluation of Gulfwatch 1992	Fall 1992	June 1994	~1.5 yr
Evaluation of Gulfwatch 1993	Fall 1993	Nov 1996	~3 yr
Evaluation of Gulfwatch 1994	Fall 1994	Feb 1997	~2.25 yr
Evaluation of Gulfwatch 1995	Fall 1995	Feb 1997	~1.25 yr
First Five Years of Gulfwatch, 1991-1995	Fall 1991-95	1998	~3 yr
Gulfwatch 2002-2004 Data Report	Fall 2002-04	Sept 2006	~2 yr
The Gulfwatch Program: 1993-2001	Fall 1993-01	June 2006?; Jan 2007?	~4.5-5 yr

It was not stated in the report as to how quickly the data are made available on the Gulfwatch web site, or upon request by interested researchers and environmental managers. However, upon checking the Gulfwatch web site (http://www.gulfofmaine.org/gulfwatch/results.asp) it was apparent that only data for 1991 through 2000 are available in tabular form for each replicate (Metals, PAH, PCB, pesticides), and only data from 1993 through 2001 are available through the Gulfwatch Interactive Mapping Tool (and these mapped data omit PAH, PCB and pesticides, and omit 18 sites that were infrequently sampled). If the most recent data available is from

2001, how can the Program claim that it is providing data to managers in a timely manner?

The 2002, 03 and 04 data do not appear to have been analyzed sufficiently. Although a "data report" (dated September 2006) containing these data was included in the materials sent to us, these data were not included in the overall 9-year program review document (January 2007). It is not clear why the review document limits the review to the first 9 years rather than to 12 years. In addition, no reference was made to the status of the data for 2005 and 2006.

The January 2007 document cites various Gulfwatch reports, fact-sheets, and peer-reviewed articles, but does not include a table of all these materials. Much of this information was footnoted in Appendix J. We know of a number of Gulfwatch documents that were not cited (e.g. yearly "evaluations" for 1991, 1992, 1993, 1994, 1995, etc.) and so suspect that there may be other reports that were omitted as well. It would be very important for the Program to document its productivity and outreach by including a table of all its reports and publications.

JH: There were modifications done between the 1993-2001 and the 2001-2005 phases. These involved less sampling and a gigantic move to analyse replicates. Progress can only be accomplished by examining the data and adapting the approach.
There is a proposal for a different sampling frequency outlined at the end of the 9 years report, with two different cycles, a spatial (every six years) and temporal one (every two years). A justification is provided for each of the states and provinces. In my opinion, sampling has to address needs and it is not possible for a group to make sense, direct or contradict the outline. The GoM MW should address the issues of concern to the various jurisdictions. It should adapt to needs expressed because of a concern.

A major addition should be considered: collecting mussels that will be depurated for 24 hrs prior to organic analyses and 48 hrs prior to metal analyses; compare these results to pools analysed readily as described for the past sampling of the GoM MW.

GW: Clearly, there has been no consensus regarding the need for depuration before analysis. The committee recommends a review of this subject by an appropriate working group to establish when and if such depuration is needed and the methodology required to achieve it without alteration of tissue levels for any analyte measured. See recommendations.

Ih. What efforts have been made to integrate the results of the program with other ongoing monitoring efforts?

Organics

JF: There have been clear efforts, as noted and referenced in the reports, to compare Gulfwatch data with US National Standards and Trends Mussel watch data and also

Massachusetts Water Resources Authority (MWRA) Data for Massachusetts Bay. Also, the Gulfwatch Data are compared in a few instances with data from a few other coastal regions of the world for illustrative purposes. In some cases a bit more explanation and detailed comparisons of actual data in the reports would be helpful as illustrations (e.g. for MWRA Data) rather than just a reference that requires readers to sort through all the various data reports or websites to see the actual data comparisons. See also answer to f) above.

Metals

GW: An effort to couple the Gulfwatch data with the abundant sediment data available for the Gulf (http://pubs.usgs.gov/of/2002/of02-403/) might be worthwhile as both are used as indicators of contaminants. Normalization of the metal and organic data to sedimentary organic carbon might reveal consistency between the two indicators. In addition to the efforts to compare results with the NOAA status and trends program, efforts to examine the results of the recent EPA coastal assessment program and identify other potential sources of species-specific contaminant data for the GOM. Is there a potential to expand the monitoring program to other key species of interest in the Gulf?

WR: Based on the material presented in Appendix I ("Uses of Gulfwatch"), the Program appears to have made considerable efforts to provide Gulfwatch data to a variety of organizations who are involved in environmental monitoring in the region. Substantial evidence that Gulfwatch data has been incorporated in a variety of reports and analyses made by these environmental organizations was provided. While the January 2007 report mentions that private aquaculture businesses have used Gulfwatch data in their citing applications, specific names or numbers of companies were not listed.

In addition, Gulfwatch has integrated the results of NOAA's NS&T Program into its data analysis for the January 2007 report.

Part II - Future Program Design and Direction

IIa. Are the sampling frequency, number of samples and spatial array justified? If not why not?

JH: In my opinion, the examination of environmental variability demonstrated that this is not an issue. One sample per site seems sufficient for the future. In view of the goals of the program and the results, it does not seem necessary or useful to monitor contaminants every year.

TO: My comments on this point have been given in Section Ib. above.

Organics

JF: I believe that a less frequent analysis of contaminants for benchmark and other station samples can be justified given the current temporal trends and geographic sets of data. I recommend consideration of more emphasis on sampling and archiving the samples. In future years, if some indication of a temporal trend seems to be happening, then back fill with analyses if needed. This would allow for expansion of analytes for emerging contaminants for archived samples and a few benchmark stations, as well as expanding geographic coverage with in situ samples or transplant mussels at some of the problematic or questionable locations identified in the data interpretation in the reports. I endorse the very careful analysis of the power of four versus three replicates to detect temporal trends as presented in Appendix K3.

It is not possible to go beyond this sort of general statement and recommendation without knowing the budget levels available for funding.

Metals

GW: See comments in section Ib.

WR: This question cannot really be answered since the choice of sampling frequency, number of samples, the actual measurement endpoints chosen, etc. all depend on the specific goals of the program. The 9-year program did not address the three stated goals of the program (as discussed in section f above). If the goals of the future program remain are the same, then the proposed changes will not address these goals (for the same reasons as discussed in section f). It is not clear who sets these goals – whether it is the Gulf of Maine Council or the Monitoring Committee. In either case, the program needs to be designed to meet the stated goals or else the goals need to be evaluated in light of available resources and a new set of goals developed. The Program should then be designed to meet these new goals.

The proposed program describes a sampling scheme where each jurisdiction will sample 2 stations (each on alternating years) for temporally-intensive (GOM-T) sampling and 6 stations (one per year on a 6 year cycle) for spatial coverage (GOM-S). On any particular year, 2 stations per jurisdiction (one GOM-T and one GOM-S from each jurisdiction; therefore 10 stations throughout the Gulf of Maine) will be sampled. This design is reasonably blocked, and should make statistical analysis of the data uncomplicated. However, since the choice of stations is left to the discretion of each jurisdiction, almost none of the jurisdictions are following this sample regime (taken from Appendix K):

Jurisdiction	GOM-T	GOM-S
MA	2 alternating years	4 staggered
NH	1 Gulfwatch + 2 paid for by NHEP	7 staggered
ME	3 staggered	10 staggered
NB	2 alternating years	8 staggered
NS	4 (1 for metals only; 1 for organics only	6 staggered

Little justification for this outcome was given. For example, it is unclear why Massachusetts will only have 4 GOM-S stations instead of 6. In comparison, and considering the extent of shoreline, why does New Hampshire have 7? Why set up a sampling regime and then immediately abandon it? Why can't the committee come to an agreement on what stations should be sampled throughout the Gulf of Maine, rather than allowing each jurisdiction to choose its own stations?

The choice of sampling a small number of stations every two years in order to address temporal changes in contaminant concentrations was not justified in the report. Since the analysis that was done to determine the proper number of replicates per station used a 25 % change over a time period, it appears that the EQMC has some expectation that a 25 % change (hopefully a decline) is an achievable target. It is unlikely, however, that many of the metals will exhibit this extent of change, and that many of the persistent organic contaminants will show this degree of change over 2, 4 or even 6 year periods. It seems to me that, if the purpose is to discern long term declines in contaminant body burdens, that the sampling interval for a set number of stations should be 3 - 5 years. For example, Goldberg et al. (1983. Estuar. Coast. Shelf Sci 16: 69-93), when assessing the original US Mussel Watch Program for metals and radionuclides, suggested that U.S. Mussel Watch stations should not be sampled every year, but at 3-year intervals in order to monitor for temporal changes.

IIb. Is the funding level consistent with program goals and expectations?

WR: We cannot say, since adequate information on the cost of the program was not provided. It appears that funding is used to provide analytical services (two external labs), and one staff person that is involved in report writing, outreach activities, data analysis, etc.

We suspect that the analytical service budget constricts the number of samples that can be analyzed. The Program needs to explicitly present a rationale for the minimal number of samples that must be analyzed each year based on the goals of the Program, and an argument for what the benefits would be if the number of analyses is increased to higher and higher levels (cost-benefit analysis).

Based on the inability of the Program to analyze data and complete reports in a timely manner, it is clear that additional resources are needed to staff the Program. Basing these important tasks on the volunteer work provided by the Environmental Quality Monitoring Committee is inefficient in terms of both time and productivity.

Organics

JF: I cannot answer this question because I have no information about the funding levels.

Metals

GW: Budget information was insufficient to assess the efficiency of the program**.** Based on the total expenditures, which were variable from year to year, and estimates of the analytical contract costs there appears to be little left to consistently support the program. Given that there were eleven authors of the 2002 – 2004 report, it appears that there are significant contributions of uncompensated time devoted to the project. More consistent support with appropriate outreach (webmaster, communication with managers and regulators in the five jurisdictions) could enhance the value and use of the program.

IIc. Is the rationale for new analytes or indicators, if any, adequately documented and appropriate?

WR: In Appendix K, several new analytes were discussed and recommended for addition to the list of currently analyzed contaminants. These additional analytes included several "emerging contaminants" such as PBDEs, PFOS, surfactants, nitro-musk compounds and pharmaceuticals (which are included in what is currently called "PPCPs," pharmaceuticals and personal care products). The list also included several traditionally recognized contaminants, included coplanar PCBs, PCDDs and PCDFs, organotins, and toxaphene (although it is not clear why toxaphene was listed, since it has only consistently been quantified in California, Georgia and Texas coastal areas that have been monitored; National Academy of Science. 1991. Seafood Safety). All of these contaminants were entered into a Hazard/Risk Evaluation Matrix (Table 6.1, appendix K) and each attribute was apparently ranked as either High (H), Medium (M) or Low (L). However, there was no indication as to how each of these attributes were prioritized or weighted, nor was there any ranking of the order of which new analytes should be added to the current suite. While the discussion that followed mentioned alkylated PAHs as a possible new analyte (although no justification for this selection was made), this class of compounds was not included in the Hazard/Risk Evaluation Matrix. It is unclear which analytes are being proposed as additions, and what the criteria are that lead to this decision.

It was noted in Appendix J Section 3.2 that some samples have already been analyzed for coplanar PCBs, PCDDs and PCDFs (3 in 1993; 17 in 1996; 13 in 1997; 10 in 1998). These results were not discussed in the materials send to the reviewers. It is unclear

whether the results that have already been obtained would justify the inclusion of these analytes in the new program.

Although the Program has at least considered the addition of new analytes, no consideration was presented for the addition of new indicators, biomarkers, or ecotoxicological endpoints to the Gulfwatch Program. Since one of the stated goals of the Program is to provide information on the risk to the Gulf of Maine environment (and therefore, its ecosystem), some measurement endpoint at either the population, community or ecosystem level of biological organization will need to be added in order to address this goal. If not, the goal should be abandoned. Since bioaccumulation in itself is not indicative of toxicity, mussel body burden data is not useful for addressing environmental (or ecosystem) impacts.

Organics

JF: Other than the 2002-2004 addition of Condition Indices and Lipid percentage levels, there have been no new analytes to my knowledge. The CI addition was explained adequately, although it is not yet clear how these will be interpreted in view of the influence of reproductive status on CI in some of the samples. The measurement of lipids was not adequately explained in terms of why it was done.

I endorse the suggestions in Appendix K3 for new analytes and the approach to be used as starting point for identifying and providing a priority for analyses of new analytes.

TO: New chemicals

Appendix K lists 8 possible new chemicals.

Co-planar PCBs, dioxins, and furans
PBDE
Toxaphene
Organotins
Polyfluorohydrocarbons
APE surfactants
Nitro-musk compounds
pharmaceuticals

The appendix grades each chemical group in terms of its analytical cost, environmental concern, and hazard.

Only the coplanars seem to be under serious consideration for being added to the program. My bet is that the Gulfwatch may add these chemicals ONCE. They will find that with the exception of Boston Harbor they will only have nds to show for a large expense.

My comments on replication and on new analytes are above. With regard to emerging contaminants Gulfwatch should note that organic compounds with pKows < about 3.5 don't bioaccumulate and are not susceptible to monitoring via mussels

I mostly agree with the Gulfwatch assessment of the advisability of not adding the other chemical groups but PBDEs might prove a worthwhile addition. It would be good to have an organic analyte whose use has not been banned for at least 10 years.

Metals

GW: Organotins were chosen as a possible new analyte based on a set of criteria that considered environmental concern, human health risk factors and other criteria (Appendix K). Arsenic and methylmercury were also considered but not given high priority. The rationale could be strengthened by greater specific examples of concern by managers in the region as well as documenting the criteria (rather than the H,M and L) used to justify addition to the list. For example Gulfwatch might be a means to document whether organotins are declining with time in response to regulatory action partially banning their use. As the saying goes it is better to do a few things well rather than a lot poorly. Decisions to add analytes should first consider needs of the existing program and its spatial and temporal resolution capabilities.

Archiving of samples for revisiting and determination of new analytes should be given serious consideration for this reason. Storage of the relatively small number of processed samples should not pose a great problem given the investment already made in their acquisition and their potential for providing information on new or emerging contaminants of environmental concern.

IId. Have critical management issues changed and what mechanism has been put in place to make these issues known?

WR: In the summer of 2005, the Environmental Quality Monitoring Committee met to discuss and incorporate changes into the Gulfwatch program. During this meeting, they developed a new working draft of the Program's Mission Statement:

> *"Using mussel tissue monitoring as a starting point, provide high quality and relevant data to allow for characterization of the condition of ecosystems in the GOM for enhancing marine resource management and protecting public health"* (Appendix K)

This proposed new Mission Statement is a significant departure from the mission statement that was given to them by the Gulf of Maine Council on the Marine Environment:

> *"It is the mission of the Gulf of Maine Environmental Quality Monitoring Program to provide environmental resource managers with information to*

support sustainable use of the Gulf and allow assessment and management [of] *risk to public and environmental health from current and potential threats"* (Section 1.2, page 9)

Note that the proposed Mission Statement now addresses the "...condition of ecosystems...," thereby explicitly pointing out the need for population, community or ecosystem measurement endpoints (as referred to c above).

While the materials provided to the reviewers did not mention any changes in critical management issues, it appears that the Environmental Quality Monitoring Committee perceives that changes are needed in both the scope and purpose of the Program. It is not clear how changes may be made, or who has the ultimate authority to enact changes.

Organics

JF: There is no section of any of the reports that were given to us for review that addresses these issues. These are important ands critical issues that should be addressed by those charged with funding and oversight of the program. I suspect that there is a certain "chicken and egg" aspect to these questions being asked in this review. **These are the sorts of issues that the program scientists and managers who use the data, along with the funding authority, should work out in a first draft prior to asking for the type of review that is currently being pursued.**

Metals

GW: Because the GOM Council has significant participation from the management, regulatory and policy community in the region and their inclusion as members of the Council's Environmental Contaminants, Gulfwatch and Ecosystem Indicators Partnership Committees the opportunity to communicate new concerns should be readily achievable. If there is not a current mechanism to facilitate such interaction one should be put in place.

IIe. What process in development of the 2007-2018 design will be used to provide results that will address the updated critical management and scientific questions that reflect the hypotheses of the program?

WR: While the Environmental Quality Monitoring Committee has outlined proposed changes to the program and has suggested a modified Mission Statement, it is unclear as to what process will be followed to institute these changes, or even who has the responsibility for deciding if changes are necessary.

Organics

JF: See comment in IId. above.

Metals

GW: A primary mechanism to address these concerns could be achieved by suitable archiving of samples taken by Gulfwatch that would allow a retrospective examination of the contaminants identified as a critical need. As noted in their plan for the future (Appendix K) Gulfwatch has indicated that samples may be collected and archived for analyses at a later time as warranted. Section K.4 in that Appendix also identifies some necessary actions to enhance the program.

IIf. What improvements would the Panel suggest to address critical management and scientific needs if not being addressed by the existing plan?

WR: The Program is necessary and should be continued. However, a number of changes should be implemented:

1. Institute a strict Chain-of-Custody procedure to ensure that samples are not inadvertently lost (as was evident in Table 2.3.7, page 28).

2. Step up the analytical analysis so that the data is available within 4 months of sampling. This problem was acknowledged in the January 2007 report as a major challenge that needs to be addressed (page 126).

3. Provide sufficient funds for appropriate data analysis and report preparation to (a) support and improve the program itself, and (b) rapidly get the information to environmental regulators. Analysis should include a complete analysis of QC/QA (blanks, reference materials, re-analysis of selected samples, etc.), trend analysis, and spatial analysis (jurisdiction, latitudinal, etc.).

4. Narrow the temporal window in which samples are collected. As stated in section 2.3.1 (page 18) field sampling is conducted sometime between mid-September and early November, a span of almost two months. This seems excessive, particularly at this time in the year. Alternatively, sampling could be keyed to a percentage decline in water temperature, so that mussels from all sites would more likely be in a similar physiological state.

TO: I agree that samples should not be lost but a formal chain-of-custody is a very arduous procedure required for legal proceedings. Results from sample analyses can be thrown out of court if there is not a strict written record of who handled the sample at each stage from collection to analysis and then who handled the data. Gulfwatch does not need to get into all of that. If Gulfwatch data ever do become part of a legal proceeding they might be discarded because there was no legal chain-of-custody but that's true of every monitoring program. Sampling after oil spills is a different matter, those folks do chain-of-custody.

Organics

JF: See comment in IId. above.

Metals

GW: Facilitate better interaction between the Council Committees and the scientific community to bring the best expertise in management and science possible in assessing emerging issues. Both sectors need to better understand the framework in which the Council must work in to succeed in achieving its stated goals. Current evolution of coastal ocean observing infrastructure for the Gulf of Maine should recognize management needs for the region and a frank open dialog should be initiated immediately if the Council's needs are to be considered.

II g. Are improvements compatible with existing funding levels? If inadequate identify tier 1, 2 and 3 levels of activity with tier 1 the most important etc.

WR: Probably not! Let's prioritize after everyone has submitted suggested improvements.

Organics

JF: See comment in IId. above.

Metals

GW: There needs to be a deliberate discussion of this supported by information not available to the reviewers. This should be discussed at the Committee meeting

GENERAL COMMENTS

JH: General Comments: 1993-2001 Report

-Data quality assessment is nicely presented.

-bottom para on p.16 is repeated from an earlier page, but is indicative of the reality. Extreme values are few and from sites where work could be needed if management has the ability and funds to take action. Not a minor issue.

-Analytical interferences: due to the separation of lipids? the load on the GPC, where clean up was not sufficient? Or related to the state of the MS? Overload and carry over? Preventive action?

-Appendix H seems useful, but even with the short descriptions it is difficult to judge the reasoning. I agree that sampling was too frequent at some sites, but the 4 replicate analyses seem like a statistically justified, but analytical overkill because of the inherent variability associated with analyses, best lab practices and state-of-the-art QA/QC.

General Comments: 2002 - 2004 Data Report

-p. 3, item 5: this justification is not strictly correct. Mussels are only representative of the exposure that other sessile suspension feeders would be exposed to. They are not representative of animals burrowing in sediments or of filter feeders or of fish living in the water column. They are good for an inter-comparison of mussels, but there would be added value if the role of particles within the measured concentrations was eliminated prior to performing the chemical analyses.

-p. 4, bottom para: not sure what line 2 refers to? Was other data gathered than CI? Which shell growth?

-Table 4: Analytical variability is increased by choosing to use half the DL in calculating means. It remains that the level of variability is low for many of the metals such as Cd, Cu and Ni, broadly ranging from 10 to 30%, which is within analytical variability for concentrations that are well above (such as 5-10 times) the detection limits (not Ag or Hg) and not affected by the background or amount of particles, such as would be the case for Al and Fe for example. This also means that replicate analyses answered the question of environmental variability, it is very low and similar to analytical variability, and in my opinion, should not be continued since it is not cost effective. The QA/QC is very good, summarized in tables B, especially Table B2.1, p. 226 describes the precision in the analyses.

Instead of replicates, more samples should be collected if mussels are near a point source of contamination, such as Boston Harbor or the St John River. In those cases, the approach could be aiming to examine the geographical extent of contamination or the gradient observed over space. Temporal monitoring would then be justified, especially if action will be taken to address the issue of reducing the foot print of environmental contamination.

-Table 8, pesticides 21 vs 15, include the DDT family. These carry most of the body burden detected in samples. Perhaps since the results are non detectable levels, these analyses can be performed once every 12yrs, not 2 or 6. It is good to demonstrate that contamination is low, but does not justify a time trend, unless there is a reason to expect change. PCB and DDT are representative of regulated persistent organic pollutants and their levels would only increase if there is a disposal of material or unearthing of buried drums. This issue should be discussed in terms of what is accomplished by accumulating more ND? How far is the level requiring management action that would justify continuing to analyze mussels?

-To reduce the amount of analytical effort and increase the turn around time for data, the lab could analyze for a smaller number of organic contaminants. This choice can rely on the detection frequency summarized at the end of the 1993-2001 report. For example, for the PCB, IUPAC congeners 153, 136, 118, 105, … can be chosen in this decreasing order of quantifying targets. If 153 is not present, then forget about having to examine the other ones. In terms of pesticides, quantify p,p'-DDE followed by p,p'-DDT and then set a priority for what else should be followed up. For the PAH, fluoranthene, pyrene and phenanthrene are the three predominant compounds detected in many matrices. If these levels are <DL, there is no need for additional data. Reducing the number of targets would provide added value, would waste less time.

JF: General Comment:

Overall, I find the effort to be very good and the data to be of a quality that is very good to excellent with only minor exceptions. What puzzles me, and I believe is the root cause behind what concerns various reviewers and potential users of the data may have is the statement on page 10, lines 5-7 of the Draft for Peer Review January 2007.

> "In support of the mission and as a first step towards meeting the desired goals and address a significant knowledge gap, the Gulfwatch Program was established to measure chemical contamination Gulfwide (Barchard, 1991, Barchard and Johnson-Hayden, 1990)."

The words are "as a first step". It seems to me that no other steps have been taken over the intervening time to meet the three goals and objectives. If there have been such steps, they are not identified in any of the reports or information provided to the reviewers.

Furthermore, in Appendix K of the report Gulf Watch Program: 1993-2001, the summer, 2005 meeting of the Environmental Quality Monitoring Committee prepared a draft of a new Mission Statement.

> "Using mussel tissue monitoring as a starting point, provide high quality and relevant data to allow for characterization of the condition of ecosystems in the GOM for enhancing marine resource management and protecting public health."

We are now 15 years into this Gulfwatch program and we have moved from **"first step"** to **"a starting point".** This is indicative of a fundamental problem – most likely lack of required financial resources to get beyond step 1. What ever the cause for the delay in expanding the program to what is really needed to address the initial stated mission and expectations of 1993, these causes should be explicitly recognized and addressed. It is not acceptable to keep talking about a starting point or first step.

WG: General Commments:

I agree with much of WR's comments. I don't agree with JH on depuration, because it's just too risky and complicated in my view. But then the Al/Fe numbers should be used to account for sediment.

I agree with others that the multiple sample reps are unnecessary.

As far as number of stations and locations, my comment is it's totally dependent on the questions actually being addressed. To supplement WR's comments on this, I would add that if the goal is GOM-wide assessment, then the design needs to be probabilistic and random (are you surprised?). Having said that, it's clear that the data collected to date have been very useful to at least some GOM managers and groups and I can see that they would be (I'm an original mussel watch guy, after all.) **What that means is I think the goals (or questions being answered) need to be revisited and clarified.**

I suggest that a full accounting of actual expenditures, including the clearly substantial in-kind contribution of volunteers, is necessary to answer the questions related to budget.

I also think the report should be a bit more forthright about whatever the analytical issues were that caused the change in organic analytical labs, since that would indicate some compromised data, it seems to me.

GW: General Comments:

Overall the report provides valuable information on the levels of contaminants of environmental concern in mussels from a variety of sites and with some capability of discerning trends, but the latter where only large excursions might occur. To my knowledge, it is also the only regional-scale international contaminant monitoring program being conducted in the Gulf of Maine. The resource intensive QA/QC reporting task may have overwhelmed the MW staff in their attempts to produce the reports in a timely fashion. I also suspect resources are mostly consumed at the analytical level at the expense of the data interpretation level, but in the absence of additional financial data, this conclusion is necessarily speculative.

Whatever the reason(s) there has been an inability for the interpreters to keep up with the analytical data although we note there were substantial delays in delivering the analytical data in a timely fashion as well. While I think the program should continue there needs to be careful consideration of the following points:

- Information on sources of contaminants, analysis of sensitivity of different coastal areas to perturbation and use of other available data on contaminants, particularly sediment concentrations, in the Gulf of Maine nearshore coastal zone should be updated before any redesign of the program occurs.
- A protocol should be established such that any change in analytical methodology should be thoroughly validated to be consistent with previous data. Usually this

involves overlapping analyses of samples using both the old and new methodologies.

- In this reviewer's opinion there is a benefit to expanding the spatial coverage (using info gathered as described in the first bullet above) while keeping the 5 benchmark stations but with sampling at a less frequent level. This also would require greater communication with managers and policy makers.
- Careful assessment of the cost and benefits of adding additional analytes should be undertaken.

Concluding Remarks

While the Review Committee finds that there are a number of areas that could be improved in Gulfwatch we reiterate its value and utility to the community it serves. As noted earlier it is the only such program for monitoring the Gulf of Maine on a transboundary region-wide basis. It appropriately focuses on the nearshore coastal waters of the Gulf where the impacts of the introduction of contaminants of environmental concern should be most obvious and is of greatest consequence to those who depend on this region and those charged with its management and stewardship. However, the goals of the Environmental Quality program need to be clarified and clearly identified and consistent with stated Council goals. The review committee notes that the lack of identifiable trends in the data are equally as valuable as demonstrating either declining or increasing contaminant concentrations provided high quality data is generated by the monitoring program.

Monitoring by its very nature seems to be lackluster and increasingly difficult to maintain, although meeting the original purposes of the monitoring program. It is critical that such a program maintains contact with the community it serves, be adaptable in the introduction of new methods and indicators, integrates its results with other monitoring and research efforts in the region, and recognizes emerging issues at an early stage as it matures. To this end we urge the Council to encourage the conduct of a workshop to revisit proposed changes in the design of the program. We are also recommending a session in the upcoming Gulf of Maine Symposium being organized by RARGOM and others that will specifically focus on the potential of new technologies, analytical techniques and sensors that might enhance such efforts. It is also critical that there be a concerted effort to make sure that the nearshore interests of the GOM stakeholders be reflected in the development of the GOM regional ocean observing system. Such collaboration should lead to improved use of limited resources and contribute to the maintenance of a high quality and long-term commitment to monitoring the environmental quality of the Gulf of Maine waters by the Council and other private, academic and government agencies. The committee hopes this review is a step in this process and will be of use in the future development of the program.

Appendix I

Contact List for Gulfwatch Reviewers

John Farrington
Marine Chemistry and Geochemistry Department
MS#8
Woods Hole Oceanographic Institution
360 Woods Hole Road
Woods Hole, MA 02543

Email: jfarrington@whoi.edu
Phone: 508-289-3911

Walt Galloway
National Health and Environmental Effects Research Laboratory
27 Tarzwell Drive, Narragansett, RI 02882

Email: Galloway.Walt@epamail.epa.gov
Phone: 401-782-3096

Jocelyne Hellou
Ecosystem Research Division
Bedford Institute of Oceanography
Fisheries and Oceans
Dartmouth, NS, Canada B2Y 4A2

Email: HellouJ@mar.dfo-mpo.gc.ca
Phone: 902-426-7451

Tom O'Connor
5123 Pheasant Ridge Rd.
Fairfax, VA 22030

Phone: 703-383-9878

Jim Shine
Exposure, Epidemiology, and Risk Program
Department of Environmental Health
Harvard School of Public Health
665 Huntington Avenue
Boston, Massachusetts 02115

Phone: 617-384-8806

Bruce Tripp
MS #2
Woods Hole Oceanographic Institution
Woods Hole, MA02543

Phone: 508-289-2900
Fax: 508-457-2172

William Robinson
Department of Environmental, Earth and Ocean Sciences
University of Massachusetts at Boston
100 Morrissey Blvd.
Boston, MA 02125

Email: William.Robinson@umb.edu
Phone: 617-287-7456

Gordon Wallace
Department of Environmental, Earth and Ocean Sciences
University of Massachusetts at Boston
100 Morrissey Blvd.
Boston, MA 02125

Email: Gordon.Wallace@umb.edu
Phone: 617-287-7447