Technical Memorandum

Review of EPA Science Advisory Board responses to proposed conductivity benchmark and comparison to GEI review

Introduction

The following is a comparison and analysis of EPA's Science Advisory Board (SAB; draft report dated September 28, 2010) and GEI Consultants' (GEI) responses to the charge questions posed regarding the draft report, *A Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams* (EPA 2010).

It is important to understand that the underlying bases for the SAB and GEI reviews were fundamentally different and, as such, responses to the charge questions vary. In particular, the SAB focused on the technical details and clarity of the report itself and did not necessarily question the overall premise of the benchmark (i.e., that conductivity was the best predictor of benthic macroinvertebrate impairment). In contrast, GEI questioned the validity of this underlying assumption and focused its review on testing this hypothesis and evaluating the extent to which the water quality criteria-based methodology, which is designed to produce criteria protective of all aquatic life and their uses, was implemented appropriately.

In a general sense, though there is little overlap between GEI's and the SAB's primary concerns, both groups agreed that there were flaws with both the conceptual approach as well as implementation of the benchmark derivation. The SAB's criticisms focused primarily on the transparency by which an argument presented, and provided less emphasis on the technical sufficiency of the arguments. Transparency issues raised by the SAB included making sure that EPA included all the details on sampling, providing better details for their validation study, the boot-strapping techniques used, how EPA derived their 95% confidence interval around the benchmark value, etc. SAB ultimately provided numerous suggestions for improving the benchmark document, but stopped far short of challenging the overall rigor and validity of the benchmark. Suggested improvements from the SAB included:

- EPA's failure to use taxa other than invertebrates (particularly mayflies)
- Use of a more conservative response variable such as taxonomic "depletion concentration" rather than the more extreme endpoint of "extirpation" used by EPA[1]
- An assertion that "conductivity itself is not a pollutant, but is a surrogate measure for the constituent ions in the mixture"

[1] It should be noted that EPA did not clearly define how their "extirpation coefficient" is actually just a mathematical representation of capture probability. Therefore, use of the term "extirpation" by EPA is technically not correct regardless of whether or not SAB's alternative suggestion is more or less conservative.

GEI Consultants, Inc./Ecological Division
4601 DTC Boulevard, Suite 900, Denver, CO 80237
303.662.0100 fax: 303.662.8757
www.geiconsultants.com

- Lack of rigorous, parametric statistical tests to better and more comprehensively identify all possible causal variables
- The transparency and completeness with which EPA addressed additional potential confounding factors

In contrast, GEI had fewer problems with the lack of transparency (although the issues raised by the SAB were valid and, if resolved, would improve the document substantially), but instead focused on our disagreements with the overall foundation and methods used to derive the benchmark given the patterns observed in the data.

Details supporting these general observations are provided below. For each charge question, a table comparing the SAB's and GEI's key conclusions was prepared. An introduction to and summary of the significant similarities and differences between the two groups' findings are presented before each table. Conclusions are presented at the end of this memo that specifically identify our departures from SAB's review as well as concerns that both GEI and the SAB share.

Charge Question 1

Charge Question 1 posed the following: *The datasets used to derive a conductivity benchmark (described in Section 2 of this report) were developed primarily by two central Appalachian states (WV and KY). Please comment on the adequacy of these data and their use in developing a conductivity benchmark.*

GEI and the SAB generally agree that the dataset used is acceptable in quality; however, GEI does not believe that it is appropriate for derivation of the conductivity benchmark, given that significant abiotic and biotic interactions are present that potentially confound the results. SAB concludes that the data are sufficient for the use, although they add that there needs to be additional discussion of many sampling-related details as well as acknowledgment of several concerns associated with data compilation and benchmark derivation.

The SAB also was concerned about the difference in background conductivity levels between Ecoregions 69 and 70, anticipating that there might be a difference in the tolerance of the invertebrate communities due to those background levels, and noting that the suite of taxa in reference sites in Ohio and Pennsylvania may vary naturally even within an Ecoregion.

GEI's contentions that the conflicting stressor-response profiles do not represent an internally consistent dataset for constructing the Species Sensitivity Distributions (SSDs) to quantify a benchmark value were neither acknowledged or addressed by the SAB. The SAB also did not acknowledge GEI's independent statistical analysis of the dataset[2] which indicated that substrate composition, ionic composition, and channel features were better predictors of benthic macroinvertebrate community structure than use of conductivity alone.

[2] It should be noted that the draft version of the GEI report that was submitted in advance of the SAB meeting in July 2010 did not contain a complete version of this statistical analysis. The complete version of this analysis is presented in GEI's final report from September 2010.

Table 1: Comparison of GEI and SAB responses to Charge Question 1.

GEI	SAB
Dataset is generally acceptable in terms of quality. However, its use in deriving a benchmark is not appropriate for the following reasons:	Data thought to be adequate to establish a quantitative relationship between conductivity and benthic community responses with the caveats discussed in the comments presented below.
1. Multiple conflicting stressor-response profiles are exhibited by the genera used to derive the benchmark and, thus, do not represent an internally consistent dataset.	No corresponding comment.
2. Field data limits the reliability of the approach to derive a regulatory threshold for a composite water quality measurement such as conductivity.	Use of pre-existing field data is reasonable, timely, and cost-effective, given: 1. QA/QC at time of collection were adequate, 2. enough data remained available after removing confounded data, and 3. sufficient number of reference sites studied. These assumptions appear to be largely met, though more information on QA/QC would be helpful.
3. EPA largely takes it as a given that conductivity is the best predictor of biological responses. Our independent analysis indicates other hypotheses are supportable.	No corresponding comment.
	The following comments relate to data-quality: 1. Information on macroinvertebrate sampling methods should to be included in the report. 2. Information related to the timing of sampling relative to flow should be included in the report to ensure ephemeral streams were adequately sampled. 3. Information on month and/or season of sampling should be included in the report to ensure impacts on insect emergence were not missed.
EPA did not sufficiently demonstrate that the rare taxa were rare due to conductivity or any other water quality effect, and not from general rarity itself.	The following comments relate to data compilation for use in deriving the benchmark: 1. Restrict analysis to major watersheds that contain coal/where coal mining occurs, rather than use data from entire state. 2. Justify the exclusion of genera at <30 sites. In doing so, the analysis essentially excluded rare taxa which may be rare due to increased sensitivity to environmental stressors.

GEI	SAB
GEI also recognized the limitations of deriving an aquatic life benchmark only using benthic macroinvertebrates, rather than additional aquatic taxa to ensure a broader level of aquatic life protection, much as is done with more traditional numeric aquatic life criteria. GEI did not comment on the indirect ecological effects identified by the SAB given that this is a constraint inherent to all numeric aquatic life criteria, and so is not an issue that is unique to conductivity or major ion toxicity. It should be noted that EPA did not clearly define how their "extirpation coefficient" is actually just a mathematical representation of capture probability. Therefore, use of the term "extirpation" by EPA is technically not correct regardless of whether or not SAB's alternative suggestion is more or less conservative.	The following comments relate to the nature of the benchmark itself: 1. Only macroinvertebrates were used to develop the benchmark. Recommend the following: a) Compare differential responses among most sensitive species within organism groups (e.g., fish, amphibians, mollusks) based on data from nearby states. b) Discuss potential for indirect effects on aquatic-dependent mammals, birds, and bats. c) Discuss interactions between pH and algal production and associated potential for increased toxicity due to metals. d) Discuss amphibian population effects in headwater streams. 2. Extirpation represents an extreme ecological effect and not a chronic response. A "depletion concentration", defined as the level of a stressor that results in a specified reduction in abundance, may be a more appropriate endpoint. 3. Data Quality Objectives (DQOs) and intended uses for the benchmark should be defined. Important because the intended uses influence the degree of uncertainty tolerable to decision-makers.
These are valid points raised by the SAB that we agree should be considered by EPA.	The following comments relate to background conductivity levels: 1. Explain the marked difference in background conductivity between ecoregions 69 and 70 and the expected differences in macroinvertebrate responses and associated benchmarks. 2. Because the suite of taxa at reference sites may vary even within an ecoregion (see OH and PA data), genera comprising the SSD might need to be adjusted for subregions with high natural background (i.e., where species associated with low conductivity and hardness would not be expected). 3. USDA Forest Service and EPA water quality data from both mined and unmined first-order streams in the Appalachian coal fields could be used to expand the data on background conductivity levels.

Charge Question 2

Charge Question 2 posed the following: *The derivation of a benchmark value for conductivity was adapted from EPA's methods for deriving water quality criteria. The water quality criteria methodology relies on a lab-based procedure, whereas this report uses a field-based approach. Has the report adapted the water quality criteria methodology to derive a water quality advisory for conductivity using field data in a way that is clear, transparent and reasonable?*

GEI believes that the report is reasonably transparent but does not agree that implementation of the water quality criteria methodology is appropriate in this context. In particular, some of the most significant shortcomings with the methods used by EPA that were identified in GEI's report included the conflicting stressor-response profiles evident in the database as well as the limited attempts by EPA to confirm that conductivity is the best predictor of mountain top mining and valley fill (MTM-VF)-related effects on the benthic macroinvertebrate communities in this region. However, the SAB does not acknowledge the conflicting stressor-response profile issue and it suggests that the goal of the report was not to determine the best indicator of effects, but rather to determine a methodology for deriving a protective threshold specifically for conductivity. The SAB, therefore, does not rigorously question the central assumption underlying EPA's analysis that conductivity is the best predictor of biological response and so is valid for use in deriving a regulatory benchmark.

The SAB does, however, question some aspects of the methodology (e.g., use of the HC_{05}) and also feels that there is a significant lack of transparency regarding many of the decisions made and assumptions incorporated into the benchmark derivation. Many of these assumptions were also questioned by GEI, including the use of conductivity rather than its constituent ions and the singular focus of the evaluation on mayflies. The SAB further feels that the analysis of the KY data used to help validate the conductivity benchmark should be presented in its entirety—a comment we agree with fully.

Table 2: Comparison of GEI and SAB responses to Charge Question 2.

GEI	SAB
Document appears to be generally clear and transparent, but we do not believe the approach is reasonable and does not appropriately follow water quality criteria methodology for the following reasons:	Field-based approach was justified but not perfect. The benchmark is an improvement over one that would have been derived from laboratory data but will require more justification/explanation to this effect as well as consideration of key deviations from the traditional water quality criteria approach as described below.
1. Conflicting stressor-response profiles are not appropriate for use in an SSD.	No corresponding comment.
2. Factors other than conductivity were not fully or adequately evaluated and so EPA concludes that conductivity is the best/most appropriate predictor of impairment.	Goal of report was to develop a benchmark to protect benthic communities from the adverse effects of elevated conductivity, thus the dataset was highly caveated by 10 different

GEI	SAB
	minimizing criteria. Since the goal was not to describe the causes of variability in benthic communities across the study area, this data reduction process was justified and the result is a benchmark that is relevant to effects of conductivity, *mostly* in the absence of other stressors" (emphasis in original). Causal analysis/confounding factors analysis for KY data not been done/presented.
While the GEI report did not address these issues so much from the point of transparency, technical issues were raised on many, including use of ion ratios, focus on mayflies for benchmark derivation and causal analysis, etc.	The following comments relate to transparency: 1. Compare/contrast these methods with 1985 Guidelines. 2. Explain limitations in extrapolation from lab to field methods. Explain why field methods more relevant (two examples provided). 3. Justify the use of conductivity rather than ion concentrations or ion ratios. 4. Relate conductivity to endpoints other than just mayflies. 5. Be clear about the extent to which data come from perennial streams only. Applying the benchmark to other types of streams (e.g., ephemeral, intermittent) is acceptable and should be acknowledged. 6. Discuss effect of each assumption used in culling the dataset on the benchmark e.g., effect on benchmark if rare species requirements are relaxed, effect of including genera not appearing at reference sites, effect of seasonality, effect of including fish data, effect if ion concentrations or ratios used instead of conductivity, effect if abundance-weighted analyses used instead of P/A, effect if nutrient numerical limit methods used. 7. Details on analysis of KY data are lacking, including a similar causal analysis (i.e., goes straight from methods to conclusion that method is "robust" without presenting results and/or discussion).
GEI did not address the adequacy of the HC_{05} value itself because it is a standard threshold used in derivation of regulatory aquatic life criteria by EPA and numerous other agencies worldwide. EPA aquatic life criteria, by design, are not intended to protect all species at all times, but rather the overall integrity of the aquatic community. We do not agree that	The following comments relate to benchmark derivation methodology: 1. Use of the HC_{05}, as set forth in the water quality criteria methodology, may not be appropriate in this context as it accepts the loss of 5% of genera. This could lead to the elimination of entire groups of related species that may be key to certain

GEI	SAB
additional levels of conservatism (e.g., using a lower threshold percentile as SAB appears to be suggesting) is necessary or appropriate. GEI did not address this second issue, but recognizes that a weight of evidence approach using additional lines of evidence could indeed help account for confounding factors or better validate causality.	ecological functions. Should deviate from the traditional approach by using subject knowledge to modify the benchmark if necessary to conserve food-web-important taxa. 2. Should compare benchmarks derived using other approaches, like the nutrient criteria approaches document that used quantile regression, logistic regression, conditional probability analysis, Weight of Evidence (WOE) approach, etc. to lend support to field-derived SSDs.

Charge Question 3

Charge Question 3 posed the following: *Appendix A of the report describes the process used to establish a causal relationship between the extirpation of invertebrate genera and levels of conductivity. Has the report effectively made the case for a causal relationship between species extirpation and high levels of conductivity due to surface coal mining activities?*

GEI and the SAB do not agree on this issue, primarily because GEI does not necessarily accept the causal relationship as valid (as supported by the independent analysis presented), and the SAB did not go beyond evaluating whether the case for this causal relationship was sufficiently convincing. The SAB concluded that EPA's arguments were compelling, but further acknowledged that factors other than conductivity (e.g., metals or habitat) or factors actually comprising conductivity (e.g., specific ions and their ratios) may be influencing the reported effect. SAB does not acknowledge that alternative WOE scoring is supportable or that additional lab (or even mesocosm) studies are needed to validate the relationship.

Table 3: Comparison of GEI and SAB responses to Charge Question 3.

GEI	SAB
No, the report has not adequately made this case for the following reasons:	Two requirements for building a strong case for causality were met as described below.
1. The report did not adequately test the hypothesis that conductivity is the best predictor of impairment.	No corresponding comment.
2. Causal analysis conclusions were subjective, alternative scoring is supportable.	No corresponding comment.
3. Lab studies with appropriate test duration, species, ionic composition, hardness are needed to validate the species-specific responses to given levels of conductivity or its related ion composition.	No corresponding comment.

GEI	SAB
4. Abiotic and biotic parameters as important predictors of community structure are poorly addressed. GEI agrees that conductivity is not a pollutant, and that the concentrations of individual ions most directly related to toxicity (e.g., sulfate or chloride) should instead be used for deriving regulatory criteria, as has been done in other U.S. states.	Conductivity is not a pollutant. Should account for the possibility that other factors (while not primary) may be confounding the conductivity-taxa loss relationship. Should instead relate specific ionic concentrations/ratios that are actually causing stress. (Repeated below.) Metals can provide a similar effect (loss of genera), so "alteration" is not as specific as EPA indicates. (Repeated below.)
In general, GEI agrees that these *correlations* exist, but that the analyses presented in Appendix A were based on relatively few data, and so are not entirely convincing.	**1. EPA linked stream conductivity and the amount of valley fill in the upstream catchment.** EPA adequately demonstrated a correlation between valley fill extent (size, % of area) and conductivity, particularly in waters with high SO_4 and high HCO_3^- and low Cl^-.
GEI disagrees that EPA rigorously supported the existence or consistency of these relationships. Statistical relationships between conductivity and biological response were highly variable, and based mostly on a single group of aquatic insects. General physiological mechanisms were plausible, but taxonomic specificity of the responses are not well established, and effect concentrations in controlled laboratory studies are variable and largely not consistent with the proposed benchmark level.	**2. EPA linked elevated stream conductivity and loss of benthic macroinvertebrate genera.** 1. Analyses focus on Ephemeroptera (E), but supporting evidence from other groups included and negative correlations between # E genera and total # of genera. 2. Plausible physiological mechanisms presented. 3. Demonstrated consistency of patterns of loss of specific taxa within this study and in another published study. 4. Taxa were exposed to sufficient concentrations to elicit an effect. However, the following comments/suggestions were offered: 1. Conductivity is not a pollutant. Should account for the possibility that other factors (while not primary) may be confounding the conductivity-taxa loss relationship. Should relate specific ionic concentrations/ratios that are actually causing stress—use EMAP mixture percent contribution equations to better explain ion role and contribution. 2. Literature cited to support conductivity effects in ionoregulation should be relevant (i.e., not K-based, since K is not important in WV). 3. Metals can provide a similar effect (loss of genera), so "alteration" is not as specific as EPA indicates.

Charge Question 4

Charge Question 4 posed the following: *In using field data, other variables and factors have to be accounted for in determining causal relationships. Appendix B of the report describes the techniques for dealing with confounding factors. Does the report effectively consider other factors that may confound the relationship between conductivity and extirpation of invertebrates? If not, how can the analysis be improved?*

Similar to charge question 3, GEI does not agree that the causal relationship between conductivity and invertebrate "extirpation" (again, an incorrect use of this term) has been rigorously established, and so GEI does not agree that the confounding factors evaluation presented in Appendix B was implemented appropriately. While the SAB concluded in general that EPA "did a credible job" to identify confounding factors, they present several suggestions wherein important and significant modifications to the evaluation are recommended (see below).

Table 4: Comparison of GEI and SAB responses to Charge Question 4.

GEI	SAB
EPA does not effectively consider other factors related to impairment, but based on their analysis in Appendix A, takes it "as a given" that conductivity is the primary stressor (see Charge Question 3 response). GEI report discusses several causal factors that we suggest merit additional study, including habitat, influence of species rarity, and the singular focus on mayflies in EPA's analysis. GEI agrees with many of SAB's suggested improvements, but we suggest that instead of being used as just refinements to EPA's analysis, these steps should instead be used to more critically evaluate the influence of these potential confounding factors. In particular, GEI agrees that additional statistical tests are needed to provide a more objective framework. The GEI report presents an independent statistical analysis that suggests the influence of factors other than conductivity provide a more comprehensive and ecologically realistic view of factors most strongly associated with changes in the benthic community.	EPA did a credible job to carefully consider factors that could confound the relationship—many potential factors were removed during removal of individual sites, and a WOE approach was undertaken to evaluate correlations between confounding factors, conductivity, and mayflies. SAB agreed that conductivity was primary variable explaining mayfly richness as presented in Appendix A, and so could not be falsified due to weight of evidence. However, the following comments/suggestions were offered: 1. EPA needed to clarify the relationship between conductivity and the matrix ions that generate conductivity. They suggested that EPA partition the ionic concentrations/ratios instead of just conductivity to decide which is the most appropriate causal factor. Also suggested that EPA present SSDs for individual ions thought to play an important role in extirpation. 2. Address additional potential confounding factors such as selenium, trace metals, dissolved organic carbon, or stream flow. SAB also recommended more detailed analysis of substrate composition and vegetation, which greatly affect macroinvertebrate communities. 3. Include some rigorous, parametric, statistical tests, which could provide a more objective framework for data analysis.

Charge Question 5

Charge Question 5 posed the following: *Uncertainty values were analyzed using a boot-strapped statistical approach. Does the SAB agree with the approach used to evaluate uncertainty in the benchmark value? If not, how can the uncertainty analysis be improved?*

GEI did not explicitly evaluate this approach but concluded that the high uncertainty should have been discussed more in the section of the report describing the final benchmark. The SAB "commends the Agency for providing a characterization of the uncertainty in the benchmark" but also asked for a much more in-depth description of the techniques used and the associated uncertainties.

Table 5: Comparison of GEI and SAB responses to Charge Question 5.

GEI	SAB
GEI did not evaluate this specifically, but this generally seemed reasonable, although the effect of the uncertainty—which was often quite large for any particular genus in the SSD—was not addressed in the final benchmark. The more critical uncertainty of conflicting stressor-response profiles was not addressed by SAB in their response.	SAB commended EPA for providing a characterization of the uncertainty in the benchmark but offered the following comments/suggestions: 1. EPA should briefly discuss other examples of using bootstrapping in relevant water resources applications 2. EPA did not provide an adequate description of how the 95% confidence interval of 95-305 µS/cm around 300µS/cm was derived. 3. EPA should address the boot-strapping technique more completely. 4. SAB acknowledged uncertainties in the benchmark not addressed using bootstrapping (e.g., assignment of cause and effect between conductivity and macroinvertebrate extirpation).

Charge Question 6

Charge Question 6 posed the following: *The field-based method results in a benchmark value that the report authors believe is comparable to a chronic endpoint. Does the Panel agree that the benchmark derived using this method provides for a degree of protection comparable to the chronic endpoint of conventional ambient water quality criteria?*

GEI concluded that the benchmark value is not necessarily comparable to a chronic endpoint, primarily because the data used to derive the benchmark are field-based, which encompass endpoints on multiple levels with many abiotic and biotic interactions. The SAB feels the benchmark is comparable to or more sensitive than a chronic endpoint, but feels the lack of attention to rare species, focus on mayflies, and use of extirpation as the endpoint are all problematic.

Table 6: Comparison of GEI and SAB responses to Charge Question 6.

GEI	SAB
The field-based benchmark is not necessarily comparable to the chronic endpoints used in conventional water quality criteria, for the following reasons: 1. Field populations are influenced by and integrate chronic survival, growth, and reproduction endpoints as well as acute endpoints. 2. Chronic endpoints are generally organismal, whereas field data integrate population, community, and ecosystem-level interactions of all possible abiotic and biotic combinations. GEI agrees that the primary focus on Ephemeroptera throughout the EPA benchmark report limits persuasiveness and renders the benchmark much less defensible.	General approach is sound and provides a degree of protection comparable to or better than a conventional ambient water quality criterion derived from traditional chronic toxicity testing. However, the report should more explicitly highlight the strengths of field data over lab-based chronic tests. However, rare species and fish and freshwater mussels were not included in the SSD, and these are important (as Threatened and Endangered [T&E] species, etc.). Furthermore, focusing on one sensitive group of invertebrates (Ephemeroptera) may limit persuasiveness and render the benchmark less defensible. Finally, extirpation as an endpoint results in a loss of sensitivity, compared to a 50% decline in abundance, for example. Should include a safety factor, subject knowledge, or other protocol for added protection.
	Use of field data results in a benchmark that is more sensitive to changes in conductivity than lab tests and is more realistic in terms of protecting communities.

Charge Question 7

Charge Question 7 posed the following: *As described, the conductivity benchmark is derived using central Appalachian field data and has been validated within Ecoregions 68, 69, and 70. Under what conditions does the SAB believe this method would be transferable to developing a conductivity benchmark for other regions of the United States whose streams have a different ionic signature?*

Because GEI does not agree that the currently-proposed benchmark has been validated adequately, it is not appropriate for application elsewhere (or even in these regions). The SAB feels that the method (but not the benchmark) can be applied elsewhere <u>assuming a lengthy list of assumptions is met</u>. Meeting these assumptions will be difficult and both GEI and the SAB's comments on related questions suggest that some of these assumptions have actually not yet been met with either the WV and/or KY databases.

Table 7: Comparison of GEI and SAB responses to Charge Question 7.

GEI	SAB
Unless or until method validated adequately (GEI believes it has not been), it is not appropriate to transfer to other regions or even apply to these regions. In particular, until the synergistic effects of different ionic signatures, hardness, and other likely stressors such as habitat, temperature, and flow are better understood for a wide range of species, it is very premature to extrapolate this method of using field correlations for conductivity from one region to another region.	The method (but not the benchmark) are transferable, assuming minimum data requirements are met, as follows:
Agree with this requirement.	1. High quality reference sites should be available (including if necessary, minimally-impacted sites as surrogates for reference sites if none are available).
Agree with this requirement.	2. Fauna of reference sites should reflect a common regional pool.
Agree with this requirement.	3. Existence of good prior knowledge of environmental requirements of the regional pool.
Agree with this requirement, as long as ionic ratios are consistent; this assumption has not been met in the WV database (see below)	4. Background conductivity should be similar across reference sites.
Agree with this requirement. However, the GEI report emphasizes that despite EPA's assertions to the contrary, the composition of ions related to elevated conductivity is not consistent throughout the WV dataset. In particular, the ratio of sulfate:alkalinity varies over two orders of magnitude, and does not remain constant with elevated conductivity.	5. Relative ionic composition of elevated conductivity should be consistent across region.
Agree with this requirement. However, the GEI report concludes that the confounding factors analysis was not adequate primarily because the basic assumption that conductivity is the primary or best predictor of impairment is not adequately tested.	6. Potential confounding factors should be understood and addressed. The process EPA used in WV and KY provides a good framework for other regions.
Agree with this requirement.	7. Large dataset needs to be available.
Agree with this requirement.	8. Needs to be validated (using independent dataset, jackknifed data, or subset of data not used in derivation from a large dataset).
Agree with this requirement. In particular, GEI agrees that ion composition even within the WV dataset used by EPA is too variable to use conductivity as a surrogate for the major ions that may be more directly related to biological response.	9. Don't extrapolate the benchmark too far geographically where insufficient data are available for validation. Even within an ecoregion may be too far.

Charge Question 8

Charge Question 8 posed the following: *The amount and quality of field data available from the states and the federal government have substantially increased throughout the years. In addition, the computing power available to analysts continues to increase. Given these enhancements in data availability and quality and computing power, does the Panel feel it feasible and advisable to apply this field-based method to other pollutants? What issues should be considered when applying the method to other pollutants?*

GEI concludes that, to date, the use of field-based data for benchmark derivation has been unsuccessful for a variety of other pollutants (including the conductivity example in EPA's report). The SAB instead concludes that such an approach can be applied to other pollutants given several caveats, some of which have not been met by the conductivity benchmark.

Table 8: Comparison of GEI and SAB responses to Charge Question 8.

GEI	SAB
This is not so much a question of large datasets or computing power but rather whether we can determine if individual stressors are powerful and reliable enough to be good predictors of biological response (separate from all other components of the environment) and that the responses are causal. This requires taking hypotheses generated from large datasets, "data mining", and objectively testing these hypotheses to ensure the patterns are reliable and causally related. To date, we have not yet seen this done adequately for any stressor including iron, nutrients, or conductivity. So until sufficient studies are conducted to validate any given predictor in a rigorous way that is causally related, it is premature to apply the method to other pollutants, just like it should not be applied to conductivity.	Yes, it can be, given the following caveats:
Agree with this caveat. Using a combination of laboratory, mesocosm, and field approaches would greatly help validate not only the predictive patterns, but also help confirm mechanisms and causality.	1. Recognize that it should include applicable lab tests to demonstrate sensitivity of certain species, results from mesocosm or other novel controlled approaches, and robust field data; recognize that field-based data offers advantages over lab-based data when toxicant is difficult to test in lab.
Agree with this caveat.	2. Follow natural classifications, since some stressors may even be reach-specific.

GEI	SAB
Agree with this caveat. As emphasized in the GEI report, physiological and laboratory toxicity data do not strongly support the conductivity benchmark value, and abiotic and biotic confounding factors were not adequately addressed by EPA.	3. Examine mode of effect. Where there is a direct physiological effect, this approach may be good, but where there isn't it might not work. Also, in most of those cases, there can be significant confounding factors in abiotic and biotic interactions.
Agree with this caveat.	4. A large, robust dataset, including stressed and reference sites, is absolutely necessary. Availability of a large validation set is also important. Size of dataset would need to be larger as number of confounding factors increases.
Agree with this caveat. GEI further emphasizes that because the EPA benchmark is only related to benthic macroinvertebrate communities (and, primarily, mayflies) that the differential sensitivity of other aquatic life and their uses has not been addressed.	5. Might need to account for tiered aquatic life use classifications (e.g., removal from SSD of taxa that aren't appropriate to a given tier).

Conclusions

We are pleased that the draft SAB review points out several portions of the draft EPA conductivity benchmark report that require additional work or explanation, including many areas of technical deficiency that were also pointed out in GEIs technical review. However, given the importance of many issues raised by the SAB, we are disappointed that their review stops well short of challenging the overall reliability of the approach used and the ultimate benchmark value that was derived. There were several main areas of technical concern noted by the SAB with which we generally agree, but that we suggest are of sufficient concern to question the overall reliability of the conductivity benchmark:

- *Accepting the overall premise that conductivity is the best predictor of benthic macroinvertebrate community impairment* – As noted in GEI's review, evidence supporting the primary underlying premise of EPA's analysis is a substantial shortcoming that we suggest was not adequately questioned in SAB's review. GEI feels that these analyses were insufficient due to the fact that EPA failed to adequately test the primary hypothesis that conductivity was the best predictor of impairment; however, the SAB concluded that, for the most part, these analyses yielded convincing arguments for causality as well as a lack of confounders. It is worth noting that the SAB did request that additional confounding factors be evaluated using more robust statistical techniques and that there are other parameters (i.e., the constituent conductivity ions and/or metals) that may be contributing to the observed response.

- *"Conductivity is not a pollutant"* – This appeared to be one of the SAB's largest concerns. The SAB repeatedly emphasized the need to "distinguish between

conductivity and the matrix ions that generate conductivity" and "partition the ionic concentrations/ratios instead of just conductivity to decide which is the most appropriate causal factor." GEI agrees that a benchmark based on a composite variable is a flawed approach given the potential for many ionic combinations to yield the same conductivity level. The SAB further emphasized this in their recommendations for implementation of the method in other regions. One of the assumptions they require be met is that relative ionic composition of elevated conductivity should be consistent across the region; however, we believe in practice this "consistency" may be both difficult to define and achieve. Indeed, we suggest that this assumption has not even been met within the WV dataset used to derive the conductivity benchmark. Ultimately, use of single ion criteria for key ions such as sulfate and chloride is a superior approach that has been used successfully in other states, and also provides a broader level of aquatic life protection for species other than benthic macroinvertebrates.

- *Lack of attention to technical details regarding water quality criteria derivation methodologies* – While the SAB appears to have considered certain aspects of the approach traditionally used to derive water quality criteria (i.e., they repeatedly raise concern over the fact that the benchmark is solely based on mayflies), they do not consider other factors that are equally, if not more, important. In particular, the presence of conflicting stressor-response profiles in the dataset used to derive the SSD and associated HC_{05} was not addressed by the SAB, in contrast to the strong emphasis on this problem in GEI's comments. In selecting the 5th percentile of an SSD compiled with XC_{95} values for species that respond optimally (i.e., probability of capture was low on either end of a specified conductivity range), favorably (i.e., probability of capture increased with increasing conductivity), and negatively (i.e., probability of capture decreased with increasing conductivity), the benchmark is not protective of all species beyond the 5th percentile as would be the case in a water quality criteria derived using standard toxicity data. In particular, species at the upper end of the distribution (those who respond to conductivity favorably) will be negatively impacted by a benchmark selected at the low end of the conductivity distribution since they prefer environments with a high conductivity.

- *Failure to identify that the "extirpation" endpoint is a misnomer* – SAB expressed a concern that extirpation is too extreme an endpoint to use in a regulatory context because, in doing so, one accepts the potential loss of key taxa from the ecosystem. However, though the EPA report uses this language, the benchmark is actually based on capture probabilities which, without a priori knowledge of the pre-disturbance population dynamics, cannot predict "extirpation." That is, extirpation assumes the species were historically present and we do not have data to support this assumption. It is not enough to base extirpation on comparisons of capture probabilities to a reference site since, as the SAB points out, the suite of taxa found at any one site may vary widely even within an ecoregion.

- *Provisional acceptance of the use of this methodology for other regions and pollutants* – We conclude that many of the caveats that the SAB suggests must be met prior to application in a different setting have not yet even been met for the currently-proposed conductivity benchmark. Therefore, GEI concludes that it is premature to apply this methodology to other pollutants or regions at this time.

- *Transparency of the document* – Many of the SAB's comments were focused on the need for EPA to better explain the process they followed and the decisions they made. Though GEI's comments did not focus on this as much, we agree with the suggestions the SAB made that more completely justifying certain aspects of the method is needed to ensure that any resulting benchmark is scientifically defensible.

- *Need for defining DQOs for the benchmark* – The SAB requested that EPA go through the formal DQO process for determining how this benchmark will be implemented. This comment came up in the context of uncertainty, since the degree of uncertainty a regulator is willing to accept will vary depending on the application of the regulation. We feel that the uncertainty in the XC_{95} calculations was quite large for the majority of taxa and lack of acknowledgement of this uncertainty in the final discussion of the benchmark derivation was a significant shortcoming.

After reviewing SAB's draft comments on EPA's conductivity benchmark, we reiterate the overall conclusion from GEI's original technical review that the relationship between conductivity and changes in benthic macroinvertebrate community structure is neither strong nor reliable enough to warrant derivation of a regulatory benchmark at this time. Combining the technical issues raised by the SAB, GEI, and other respondents during the public comment period makes a powerful case that much work is yet needed before a regulatory benchmark based on conductivity can be considered scientifically defensible, if indeed this is even possible for a composite variable such as conductivity. We hope that the SAB and EPA will carefully consider the many technical concerns raised by all parties before proceeding any further with development or application of this benchmark.

Reference

U.S. Environmental Protection Agency (EPA). 2010. *A Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams.* Draft EPA/600/R-10/023A, U.S. Environmental Protection Agency, Washington, DC.

Steven P. Canton
Certified Senior Ecologist

Robert W. Gensemer, Ph.D.
Senior Ecotoxicologist