

Environmental quality benchmarks—the good, the bad, and the ugly

Peter M. Chapman¹

Received: 4 July 2016 / Accepted: 16 October 2016 / Published online: 24 October 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Environmental quality benchmarks (EQBs) such as water or sediment quality guidelines comprise one line of evidence for assessing the potential harm from chemicals and other stressors (physical, biological). They are useful but not perfect tools, should not always be used, and should never be used alone for final decision-making. The “good” can be designed to be situation-specific and can provide understandable scientific input to decision-makers. The “bad” includes perception that they are absolutes (i.e., definitive binary decision points), no or limited adaptability based on good science or common sense, and protection of individual organisms not populations of organisms. The “ugly” includes benchmarks based on simplistic indices (information loss, misleading results), misuse of biomarkers, and misapplication of EQBs. Other factors to be considered include the following: appropriately deriving EQBs, uncertainty, the laboratory is not the field, contaminant uptake and cause-effect, and specifics regarding sediment quality benchmarks (i.e., their specific “good,” “bad,” and “ugly” components). EQBs are not always needed or useful.

Keywords Sediment quality · Water quality · Benchmarks · Toxicity · Guidelines · Criteria · Standards

Introduction

Environmental quality benchmarks (EQBs) are tools, one line of evidence (LoE) for assessing potential harm from chemicals and other stress, both physical and biological. EQBs can be developed for individual stressors, for mixtures of stressors (e.g., considering contaminant competition and interaction at biological uptake sites), and in different matrices (e.g., air, effluent, water, sediment, soil, and tissue). They are not perfect; there are no perfect tools (i.e., there are no individual tools or LoE that provide all necessary biological and chemical information for assessment and decision-making). However, they can be useful tools. But, like all tools, they can be misused and should not always be used. They and their usage can be “good,” “bad,” or “ugly” as discussed below.

The “good”

Appropriately derived EQBs provide two useful decision points: negligible concern if the EQB is not exceeded and possible concern if the EQB is exceeded. Appropriately derived EQBs can assist in determining stressor(s) of potential concern, chemical contaminants, and physical and biological stressors.

EQBs should not be used alone for final decision-making, but should incorporate uncertainty (ranges in recognition of uncertainty, not single numbers in denial of uncertainty), modifying factors, and consider all stressors. Chemical contaminants can include inorganic and organic substances, dissolved oxygen, pH, etc. Physical stressors can include habitat change or loss, temperature changes, etc. Biological stressors can include invasive species, eutrophication, harmful algal blooms, etc. Because of variability between different stressor effects, considering the combined effects of different stressors ideally

Responsible editor: Kenneth Mei Yee Leung

✉ Peter M. Chapman
peter@chapmanenviro.com

¹ Chapema Environmental Strategies Ltd., 1324 West 22nd Street, North Vancouver, BC V7P 2G4, Canada

requires narrative, not solely numeric EQBs for single substances. It must be recognized that we are protecting populations and communities, not individual species unless they are unique, threatened, and/or endangered. Ideally, we are using EQBs to protect ecosystem services (the benefits humans obtain from ecosystems).

EQBs can be designed to be situation-specific, for example, specific to a particular geographic location, combination of ecosystem services, or combination of stressors. They are as much “art” as “science.” They can provide understandable scientific input to decision makers. Specifically, they allow clarity regarding uncertainty, urgency, priorities, and significance regarding further actions, provided they are not too prescriptive, misrepresented, or misunderstood (Chapman 2000; Johnson and Sumpter 2016). Allowance by regulatory agencies for the development of site-specific EQB derivation is definitely “good” (e.g., CCME 2003, 2007).

A good example of how EQBs have developed based on scientific developments is provided by metals in water as described by Chapman (2008). Initially, only total metal concentrations in water were measured, then dissolved metal concentrations, modifying factors such as hardness were applied, and, finally, the biotic ligand is now used to set benchmarks for some metals in water. Similar developments apply to metals in sediment, where acid volatile sulfides have been measured as a component of divalent metal sediment quality benchmarks (Campbell et al. 2006; Chapman 2008).

The “bad”

There are four “bad” perceptions or beliefs related to EQBs from some users and other stakeholders: the perception that they are absolutes (i.e., definitive binary decision points); the belief that laboratory toxicity tests, the basis of many EQBs, provide real-world results; the belief that correlation equals causation (i.e., exceedance of an EQB explains an adverse biological effect); and the belief that EQBs are based on protecting all individual organisms. These perceptions and beliefs are incorrect.

EQBs are not absolutes For example, sediment quality values derived in different jurisdictions and by different researchers can vary by orders of magnitude for the same substance (Chapman and Mann 1999; Chapman et al. 1999). Water quality values similarly can vary by orders of magnitude for the same substance (Hahn et al. 2014).

Laboratory toxicity tests do not provide real-world results

The laboratory is not the field. For example, Proulx and Hare (2008) and Martin et al. (2008) found that uptake of metals in larvae of two *Chironomus* species in the laboratory was different than observed in field-collected larvae. They

determined that this was due to the different feeding strategies of the two species. One species feeds mainly within surface toxic sediment; the other feeds within deeper, anoxic sediment. Laboratory sediment toxicity tests are conducted on homogenized sediment from both surface and deeper sediment; the laboratory sediment to which the larvae were exposed bore no relationship to the field sediments.

Another example is provided by Colombo et al. (2016), who investigated the effects of sediment bioturbation by the aquatic oligochaete *Lumbriculus variegatus* on zinc chemistry and toxicity to the epibenthic chironomid larvae, *Chironomus tepperi*. They found that the presence of *L. variegatus* significantly decreased zinc toxicity to *C. tepperi*. This was due to a combination of geochemical and ecological processes. Bioturbation reduced pore water zinc and bioavailability in overlying water. It also modified microbial abundance and structure, resulting in more food for the chironomid larvae but also more zinc adsorption by the microbial community, reducing bioavailability via water uptake to the larvae. Single species laboratory toxicity tests do not consider these (or other) processes.

Laboratory tests provide a range of numbers, due to the innate variability of such testing; for instance, replicate tests of the same substance by the same laboratory will not provide exactly the same result. They are typically conservative (i.e., overprotective as they provide the worst case information). Test organisms are naïve (i.e., not previously exposed to the stressors they are tested against); there is no allowance for compensation and regulation that can result in tolerance (acclimation and/or adaptation). Exposure conditions are conservative (e.g., food and behavior are restricted, exposure is maximized). Exposure and toxicity modifying factors are typically absent (e.g., aging of contaminants in sediments or soils as opposed to laboratory spiking; reduced bioavailability in water (e.g., hardness, pH, etc.)), and laboratory cultures can be unduly sensitive (e.g., due to in-breeding, nutritional deficiencies). Burton (2016) noted that “laboratory-based guidelines are overly conservative, ignoring spatial-temporal exposure and chemical bioavailability dynamics, the influence of refugia, ecosystem-context, and the artifacts associated with sediment homogenization and sediment spiking.”

Correlation is not causation. Correlation of effects with measured sediment contaminants does not necessarily indicate causation, which could be due to unmeasured contaminants, combinations of measured or unmeasured contaminants, or other stressors. Correlation is only indicative; causation can only be definitively determined by subsequent investigative studies (e.g., Environment Canada and Ontario Ministry of the Environment 2008; Chapman 2016).

EQBs are not based on protecting individual organisms that are neither threatened nor endangered. They are based on protecting populations and communities (i.e., protecting community function, not necessarily structure) except for threatened or endangered species, which merit individual protection.

The “ugly”

Uncertainty is endemic in all scientific endeavors, including the use of EQBs. Uncertainty has four different forms, only one of which can be addressed directly for EQBs. Human error can be reduced by implementing appropriate quality assurance/quality control during data development and assessment.

Two of the other three different forms of uncertainty can only be reduced by considering other lines of evidence (LoE), i.e., not relying solely on EQBs. Uncertainty related to imperfect knowledge can be reduced by obtaining necessary additional knowledge from other LoE. Uncertainty related to simplification of the real world can be reduced by increased realism based on other LoE (i.e., going beyond numbers based on laboratory studies). The last of the four different forms of uncertainty, stochasticity (natural variability; “noise”), cannot be reduced. However, its boundaries (temporal and spatial) can be estimated/described with other LoE.

It is “ugly” when there is limited opportunity for changes to EQBs based on good science or common sense; in other words, there is no adaptability to EQBs once they are developed despite, for instance, scientific advances that not only merit but also require changes. For example, development of selenium benchmarks based on the state-of-the-science for both water and tissue concentrations took almost two decades in the USA (USEPA 2016); Canada has not revised its selenium water quality guideline, developed in 1981 (<http://www.env.gov.bc.ca/wat/wq/BCguidelines/selenium/selenium.html>). Similarly, it is “ugly” when there is no adaptability despite common sense that would dictate otherwise; for instance, evidence that the EQBs are ineffective or inappropriate. Unfortunately, regulatory EQBs are not rapidly or sometimes ever revised.

Conducting testing at environmentally unrealistic concentrations is not only irrelevant but also misleading. Phuong et al. (2016) noted that, in the case of microplastic (MP), most laboratory experiments are “performed with MP concentrations of a higher order of magnitude than those in the field... [they do] not mimic the natural environment.” This is an excellent example of unsupported speculation that can be incorrectly used to support EQBs based on the suggestion, not the proof, of adverse effects.

It is equally “ugly” to use toxicological and other data to develop EQBs that are basically simplistic indices that scale data to provide definitive binary decision points. Indices are the result of information loss and typically provide misleading results (Chapman 2011; Green and Chapman 2011). Environmental complexity must be recognized and respected, neither ignored nor simplified.

EQBs serve for screening, not for definitive decision-making without consideration of other LoE. They should definitely not be misused alone to, for instance, identify the need

for chemical source control measures, trigger management and/or regulatory actions, or set remediation objectives.

Example: sediment and water quality benchmarks

Sediment quality benchmarks (SQBs) provide specific examples of the “good,” “bad,” and “ugly”; similar examples exist for water quality benchmarks (WQBs). There are five major “good” aspects to correctly derived and used SQBs and WQBs: using cause-effect relationships between contaminant concentrations and biological effects; associating a chemical exceeding a field-derived SQB or WQB with a potential biological effect, such that an exceedance is not necessarily indicative of an actual biological effect (Johnson and Sumpter 2016); adequate data entries for both effects and no-effects data (e.g., >20 of each); not combining freshwater and saltwater chemical and biological data without technically defensible justification; and not evaluating contaminants separately, without accounting for the potential presence of elevated concentrations of other contaminants.

There are three major “bad” aspects to SQBs and WQBs: relying on correlative relationships between contaminant concentrations and biological effects, inadequate data entries for both effects and no-effects data, and inappropriately combining freshwater and saltwater chemical and biological data. Correlative relationships do not consider unquantified chemicals or other stressors that may have been responsible for observed biological effects. For example, some SQBs derived using field data are based on the presence or absence of a benthic species—this is not how benthic communities are assessed by ecologists; they are assessed based on consideration of all species present not the presence or absence of individual species. Finally, correlative relationships ignore the presence of exposure and toxicity modifying factors (ETMFs) and other confounding factors, for instance, physical stressors (e.g., habitat, scour), chemical stressors (e.g., ammonia, sulfide and in particular acid volatile sulfides for divalent metals, total organic carbon, black carbon, water quality), and biological stressors (e.g., competition, predation).

There are five major “ugly” aspects to SQBs and WQBs. First, databases may be inadequate; they may contain elevated concentrations of multiple co-occurring contaminants or a preponderance of relatively low contaminant concentrations. Second, screening criteria for comparing mean chemical concentrations in toxic samples compared to mean chemical concentrations in non-toxic samples can be variable (e.g., factor-of-one or factor-of-two screening). Third, SQBs and WQBs from lightly contaminated sites can inappropriately be applied to more heavily contaminated sites. Fourth, SQBs and WQBs can be set such that there is limited opportunity for changes based on changing scientific evidence, resulting in reliance on dated SQBs (i.e., not based on the current state-of-the-

science). Fifth, SQBs and WQBs can be developed with no consideration of bioavailability or modifying factors—this is typically how they are developed, based solely on total chemical concentrations despite the fact that frameworks exist allowing developers to take bioavailability into account with, for instance, a range of SQBs depending on such factors as particle size and organic carbon content (Chapman 2008).

Conclusions

EQBs must be developed, presented, and recognized as adaptable (i.e., subject to change as the state-of-knowledge advances). They should be used to filter out sites, cases, and situations of negligible concern from those of possible concern; prioritize remaining sites, cases, and situations for further investigation; determine stressors of potential concern for further investigation; and provide, when combined with other LoE, necessary information for management decision-making. They should not be used when they are wrong (e.g., definitive binary decision points with no consideration of uncertainty), inappropriate (e.g., not based on the state-of-the-science), or unnecessary (e.g., at either end of the stressor-affected spectrum of adverse biological effects). EQBs should never be developed for their own sake (e.g., to add to publication lists).

EQBs are a powerful, but imperfect LoE for assessing potential harm from chemicals and other stressors. They can be a useful tool, but can also be misused. They need to be developed and used appropriately, incorporating both good science and common sense. This is critically important for environmental decision-making and environmental protection, which should be focused on solutions to real and pressing issues rather than wasting time and resources on relatively insignificant issues.

Acknowledgments I thank Professor Kenny Leung for inviting me to give a Keynote Lecture at EQSPA-2016 (International Conference on Deriving Environmental Quality Standards for the Protection of Aquatic Ecosystems, 18–20 June 2016, the University of Hong Kong). I also thank two anonymous referees for their useful comments.

References

Burton GA (2016) Breaking from tradition: establishing more realistic sediment quality guidelines. Abstract Book, EQSPA 2016, International Conference on Deriving Environmental Quality Standards for the Protection of Aquatic Ecosystems, June 18–20 2016, The University of Hong Kong, Hong Kong, p 41

- Campbell PGC, Chapman PM, Hale B (2006) Risk assessment of metals in the environment. In: Hester RE, Harrison RM (eds) Chemicals in the environment: assessing and managing risk. Issues in environmental science and technology volume 22. Royal Society of Chemistry, Cambridge, pp. 102–131
- CCME (Canadian Council of Ministers of the Environment) (2003) Canadian water quality guidelines for the protection of aquatic life: guidance on the site-specific application of water quality guidelines in Canada: procedures for deriving numerical water quality objectives. Winnipeg, MB. <http://ceqg-rcqe.ccme.ca/>
- CCME (2007) A protocol for the derivation of water quality guidelines for the protection of aquatic life 2007. Winnipeg, MB. <http://ceqg-rcqe.ccme.ca/>
- Chapman PM (2000) Why are we still emphasizing screening level numbers? *Mar Pollut Bull* 40(465):466
- Chapman PM (2008) Environmental risks of inorganic metals and metalloids: a continuing, evolving scientific odyssey. *Human Ecol Risk Assess* 14:5–40
- Chapman PM (2011) Indices—attractive delusions. *Integr Environ Assess Manage* 7:313
- Chapman PM (2016) Chapter 9: ecological risk and weight of evidence assessments. In: Blasco J, Chapman PM, Campana O, Hampel M (eds) *Marine ecotoxicology: current knowledge and future issues*. Elsevier, London
- Chapman PM, Mann GS (1999) Sediment quality values (SQVs) and ecological risk assessment (ERA). *Mar Pollut Bull* 38:339–344
- Chapman PM, Wang F, Adams W, Green A (1999) Appropriate uses of sediment quality values for metals and metalloids. *Environ Sci Technol* 33:3937–3941
- Colombo V, Pettigrove VJ, Hoffmann AA, Golding LA (2016) Effects of *Lumbricus variegatus* (Annelida, Oligochaeta) bioturbation on zinc chemistry and toxicity to the epi-benthic invertebrate *Chironomus tepperi* (Diptera: Chironomidae). *Environ Pollut* 216: 198–207
- Environment Canada and Ontario Ministry of the Environment (2008) Canada-Ontario decision-making framework for assessment of Great Lakes contaminated sediment. Ottawa, ON
- Green R, Chapman PM (2011) The problem with indices. *Mar Pollut Bull* 62:1377–1380
- Hahn T, Diamond J, Dobson S, Howe P, Kielhorn J, Koennecker G, Lee-Steere C, Mangelsdorf I, Schneider U, Sugaya Y, Taylor K, Van Dam R, Stauber JL (2014) Predicted no effect concentration derivation as a significant source of variability in environmental hazard assessments of chemicals in aquatic systems: an international analysis. *Integr Environ Assess Manage* 10:30–36
- Johnson AC, Sumpter JP (2016) Are we going about chemical risk assessment for the aquatic environment the wrong way? *Environ Toxicol Chem* 35(7):1609–1616
- Martin et al (2008) Explaining metal concentrations in sympatric *Chironomus* species. *Limnol Oceanogr* 53:411–419
- Phuong NN, Zalouk-Vergnoux A, Poirier L, Kamari A, Châtel A, Mouneyrac C, Lagarde F (2016) Is there consistency between the microplastics found in the field and those used in laboratory experiments? *Environ Pollut* 211:111–123
- Proulx I, Hare L (2008) Why bother to identify animals used for contaminant monitoring? *Integr Environ Assess Manage* 4:125–126
- USEPA (U.S. Environmental Protection Agency) (2016) Aquatic life ambient water quality criteria for selenium freshwater 2016. EPA 822-R-16-006. https://www.epa.gov/sites/production/files/2016-07/documents/aquatic_life_awqc_for_selenium_-_freshwater_2016.pdf