

Using Relative Risk to Compare the Effects of Aquatic Stressors at a Regional Scale

John Van Sickle · John L. Stoddard ·
Steven G. Paulsen · Anthony R. Olsen

Received: 4 August 2005 / Accepted: 19 February 2006
© Springer Science+Business Media, Inc. 2006

Abstract The regional-scale importance of an aquatic stressor depends both on its regional extent (i.e., how widespread it is) and on the severity of its effects in ecosystems where it is found. Sample surveys, such as those developed by the U.S. Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP), are designed to estimate and compare the extents, throughout a large region, of elevated conditions for various aquatic stressors. In this article, we propose relative risk as a complementary measure of the severity of each stressor's effect on a response variable that characterizes aquatic ecological condition. Specifically, relative risk measures the strength of association between stressor and response variables that can be classified as either "good" (i.e., reference) or "poor" (i.e., different from reference). We present formulae for estimating relative risk and its confidence interval, adapted for the unequal sample inclusion probabilities employed in EMAP surveys. For a recent EMAP survey of streams in five Mid-Atlantic states, we estimated the relative extents of eight stressors as well as their relative risks to aquatic macroinvertebrate assemblages, with assemblage condition measured by an index of biotic integrity (IBI). For example, a measure of excess sedimentation had a relative risk of 1.60 for macroinvertebrate IBI, with the meaning that poor IBI conditions were 1.6 times more likely to be found in streams having poor conditions of sedimentation than in streams having good sedimentation conditions. We show how stressor extent and

relative risk estimates, viewed together, offer a compact and comprehensive assessment of the relative importances of multiple stressors.

Keywords Relative survey · Environmental stressor · EMAP · Stream monitoring · Sample survey

Introduction

The activities of human beings have altered the structure and function of aquatic ecosystems in a variety of ways. Assessing the ecological condition of streams and lakes, therefore, involves an evaluation both of the biota and of the environmental factors that have direct and indirect effects on biota. Although it is possible to assess the condition of the biota by itself (e.g., Karr 1981, Karr and Chu 1997), or to conduct assessments of individual stressors such as acidic deposition (e.g., Stoddard and others 2003) or nutrients (e.g., U.S. Geological Survey 1999), it is becoming increasingly common (and necessary) to assess biota and stressors simultaneously (e.g., U.S. Environmental Protection Agency 1994, 2000). A straightforward strategy for such assessments is to use biological data to access the ecological condition of the aquatic ecosystems, and also to use chemical, physical, and biological data to evaluate the relative importance of various stressors.

At a regional scale, the relative importance of an aquatic stressor depends both on its relative extent throughout the region (that is, how widespread or common it is) and on its relative severity of effect (that is, its consequence for the biota at stressed sites within the region). Sample surveys conducted across large regions, with randomized site selection and uniform sampling protocols, can directly estimate the relative extent of elevated stressor conditions with known reliability (U.S. Environmental Protection

J. Van Sickle (✉) · J. L. Stoddard · S. G. Paulsen · A. R. Olsen
National Health and Environmental Effects Research
Laboratory, Western Ecology Division, U.S. Environmental
Protection Agency, 200 SW 35th Street, Corvallis
Oregon 97333 USA
E-mail: VanSickle.John@epa.gov

Agency 1994, 2000; Boward and others 1999; Herlihy and others 1990, 2000). In this article, we focus on the problem of estimating the severity of the effect of stressors.

A stressor’s severity of effect can be measured by the strength of its empirical association with significant biological change, as observed in regional survey data. Thus, the effect of a stressor is deemed to be relatively severe if differences across sites in stressor condition are strongly associated with substantial differences in biotic condition. Bivariate associations between stressors and biological responses have frequently been modeled using correlation or regression analysis (Van Sickle 2003). In addition, multiple regression models have been used to relate a single response variable to multiple stressors (e.g., Dyer and others 2000; Gordon and Majumder 2000; Yuan and Norton 2004), and multivariate methods can relate multiple responses to multiple stressors (e.g., Comeleo and others 1996; Tong 2001). The correlation coefficients or standardized partial regression coefficients of such models are used to express the relative strengths of association between stressors and responses.

Regression and correlation coefficients are best suited to describing associations among continuous measures of ecological condition. However, we believe that few people, apart from stream ecologists, can clearly and directly interpret numerical values of a continuous variable such as total phosphorus concentration or riparian vegetative cover, in terms of stream ecological condition. The public relies instead on ecologists to identify the ranges of phosphorus or riparian cover that delimit generally interpretable classes of condition with class labels such as “good,” “fair,” or “poor.” Thus, we believe that environmental survey results and resulting assessments can be conveyed more effectively to some audiences by using discrete condition classes, rather than continuous variables, for all stressors and responses.

In this article, we suggest simple and compatible statistics for stressor extent and severity of stressor effect that are designed for dichotomous stressor and response variables. To measure a stressor’s extent, we estimate the proportion of all streams within a region that are in “good” or “poor” condition for that stressor. To measure the severity of a stressor’s effect, we estimate “relative risk,” a statistic that is widely used in human health assessments and is familiar to a large part of the intended audience for ecological assessments.

Here, we illustrate the use of relative risk to assess the severity of effect for eight potential stressors on the macroinvertebrate communities in small streams of the Mid-Atlantic region of the United States. We describe how relative risk and corresponding relative extent estimates offer complementary evaluations of relative importance for the eight stressors. Our estimates of severity and extent, and confidence bounds for those estimates, are derived

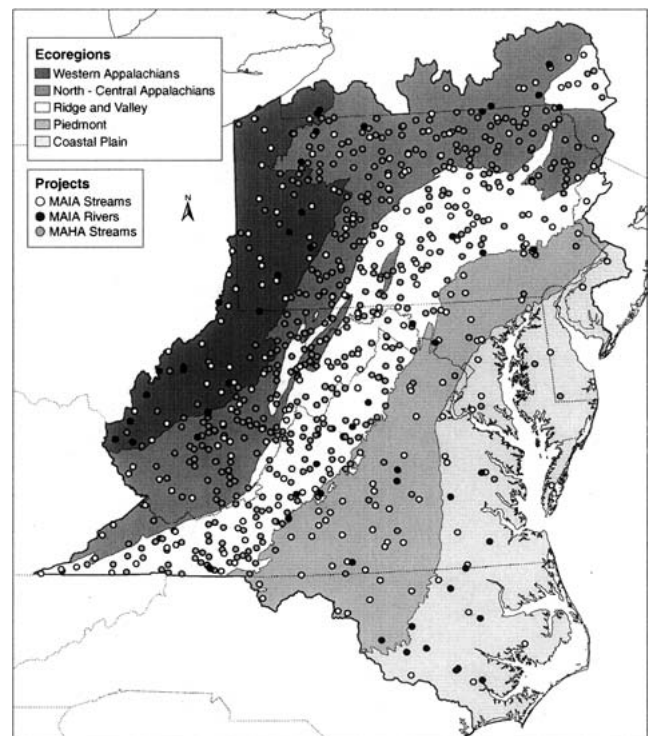


Fig. 1 The Mid-Atlantic region with sampling sites on small upland streams sampled by the Mid-Atlantic Highlands Assessment (MAHA) project, small regional streams sampled by the Mid-Atlantic Integrated Assessment (MAIA) project, and large rivers sampled by MAIA

from the design of our Mid-Atlantic sample survey (Herlihy and others 2000; Stoddard and others 2006a), which included unequal-probability selection of the survey sites.

Methods

We used data collected as part of the Mid-Atlantic Integrated Assessment (MAIA; Stoddard and others 2006a). The MAIA study collected biological (fish, macroinvertebrate, and periphyton assemblages), physical habitat, and chemical data from streams and rivers throughout the Mid-Atlantic Region between 1993 and 1998 (Figure 1). The region encompasses approximately 310,000 km² and extends from the Atlantic Ocean in the east to the Ohio River in the west, and from the headwaters of the Delaware and Susquehanna drainages in New York in the north to the Roanoke/Chowon drainage in North Carolina in the south (Figure 1). It includes all of the states of Delaware, Maryland, Pennsylvania, Virginia and West Virginia, and parts of New Jersey, New York, and North Carolina.

Design of the MAHA/MAIA Stream Surveys

During 1993 and 1994, EPA researchers used sample survey techniques to select small streams (1st through 3rd

Strahler order streams on 1:100,000 USGS maps) throughout the upland portions of the Mid-Atlantic region (Herlihy and others 2000). The biological, chemical, and physical habitat sampling of those streams resulted in the Mid-Atlantic Highlands Streams Assessment (U.S. Environmental Protection Agency 2000), the first comprehensive assessment of the ecological condition of streams in any region using both statistical site selection and biological indicators. Sampling for this survey (referred to as MAHA) continued in 1995 and 1996, and those additional data are used in this article.

The MAHA survey sampling design defined all 1st through 3rd Strahler order stream traces shown on 1:100,000 USGS maps, comprising 230,400 km of stream length, as the statistical target population (Herlihy and others 2000). A sampling grid was laid over the entire region, to achieve regional-scale spatial balance. Sites on all stream traces within 40-km² areas surrounding the grid points were used as candidates for sampling. However, 1st-order streams alone comprise 59% of the total length of all stream traces (Figure 2). Thus, a completely random sample of sites would be dominated by those on first-order streams. At the same time, sites on larger streams would be much less likely to be selected in a completely random sample, possibly resulting in poor characterization of large-stream attributes.

To avoid this problem, the MAHA design employed unequal probabilities for site selection, based on the stream orders of candidate sites. This approach provided a more even distribution of samples among 1st to 3rd order streams, and it also increased the sample size of larger streams and rivers relative to that expected from a completely random sample (Figure 2). In addition, the geographic density of sample sites was increased within the North-Central Appalachian and Ridge and Valley ecoregions (Figure 1) to better characterize acidic deposition effects, requiring a further adjustment of site selection probabilities (Herlihy and others 2000). Once sites were selected, they were assigned sampling weights that were proportional to the inverse of their selection probabilities and were normalized to sum to the total target stream length. A site's sampling weight could then be interpreted as the total length of stream (in kilometers) within the target population that is represented by that site. Sampling weights were incorporated into estimates of regional averages and totals from sample data, thus providing unbiased estimates of ecological condition indicators for the statistical population of flowing waters in the Mid-Atlantic region.

In 1997 and 1998, data collection was expanded to include all nontidal streams and rivers (all Strahler orders) of the Mid-Atlantic region (Figure 1). This larger-scale project was known as MAIA. Site selection for the 1997–1998

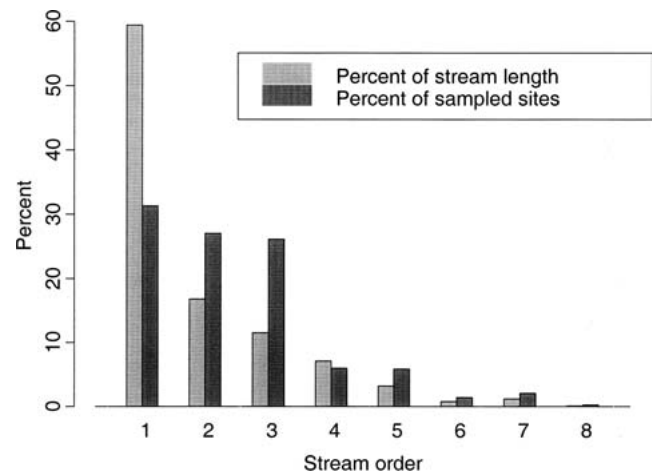


Fig. 2 Distributions of stream length for MAHA/MAIA target population and for sampled sites, by Strahler stream order

survey used the same strategy as employed in MAHA. A major emphasis in MAIA was to extend the sampling methods developed for small streams in MAHA to the large rivers included in MAIA. The result is a set of sampling protocols, all based on identical principles and producing directly comparable information, for all sizes of streams and rivers (Lazorchak and others 1998, 2000). These sampling protocols were used for all of the data collection reported here.

We adjusted the site sampling weights used in MAIA and MAHA to give an overall assessment of Mid-Atlantic streams and rivers based on the two surveys combined. Hereafter, we refer to the overall assessment and its combined data as “MAIA” (Stoddard and others 2006a). The combined data consists of 773 sites sampled for macroinvertebrates and corresponding stressor variables between 1993 and 1998 (Figure 1). Our sampling grid approach geographically dispersed the selected sites, resulting in 99.9% of the 319,600 straight-line intersite distances in Figure 1 being greater than 9 km. This degree of site separation gave us confidence that the potential for spatial correlation effects was minimized and that sites could be considered mutually independent for statistical analyses.

Macroinvertebrate Sampling and Index of Biotic Integrity

Of the 773 sampled sites, 699 were on wadeable (generally 1st through 4th order) streams as described in Lazorchak and others (1998). Each wadeable stream site's sampled reach, equal in length to 40 times the stream's wetted width (or a minimum of 150 m), was divided into 11 equally spaced transects, with one kicknet sample collected from each of the (9) interior transects (the upstream and downstream extremes were not sampled). Separate pool and

Table 1. Thresholds of condition classes for macroinvertebrate IBI and eight stressor indicators

Variable	Poor ^a	Good	Geographic restrictions ^b	Basis for thresholds ^c	Measurement (units)
Macroinvertebrate IBI	<41	≥62	None	RD (122)	Unitless index, range = [0,100]
Excess sedimentation	<-2.0	≥-1.5	P, CP ecoregions	BPJ, in P and CP ecoregions	Log 10 (Relative bed stability). Unitless index
	<-0.9	≥-0.3	All other ecoregions	RD, in all other ecoregions (42)	
Lack of large wood	0	≥2.2	None	RD (50)	Percent of wetted stream area covered by large wood
Riparian habitat	<0.5	≥0.61	None	RD	Unitless index, range = [0, 1]
Total nitrogen	>1500	≤750	CP ecoregion	In CP, use criteria from USEPA (2000).	Total N (μg/L)
	>1200	≤425	All other ecoregions	RD, in all other ecoregions (123)	
Total phosphorus	>100	≤50	CP ecoregion	For CP, use criteria from USEPA (2000).	Total P (μg/L)
	>63	≤14	All other ecoregions	RD, for all other ecoregions (123)	
Mine drainage	>1000	≤400	NCA, WA ecoregions	BPJ	SO ₄ (μeq/L)
	>5000	≤1000	All other ecoregions		
Acid deposition	ANC < 0, and Mine drainage = "Good" ^d	Mine drainage ≠ "Good", or ANC ≥ 50	None	BPJ	ANC = Acid neutralizing capacity (μeq/L)
Acid mine drainage	ANC < 0, and Mine drainage ≠ "Good" ^d	Mine drainage = "Good," or ANC ≥ 50	None	BPJ	ANC = Acid neutralizing capacity (μeq/L)

^aValues lying between the "Good" and "Poor" thresholds define the "Marginal" condition

^bEcoregion codes: P = Piedmont, CP = Coastal Plain, NCA = North and Central Appalachians, WA = Western Appalachians

^cBPJ = Best professional judgment, RD = Reference distribution, with count of reference sites given in parentheses

^dMarginal condition defined as 0 ≤ ANC < 50 μeq/L, with same dependence on mine drainage as for Poor condition

riffle composite samples were created from each stream reach. All resulting macroinvertebrates were preserved, and a fixed count of 300 organisms from each composite sample was identified to the lowest practical taxonomic resolution, usually genus. Macroinvertebrate samples were also collected from 74 nonwadeable sites using similar sampling protocols as applied to the wadeable near-shore ends of their transects (Lazorchak and others 2000).

We used a benthic Index of Biotic Integrity (IBI) to measure the condition of each sampled macroinvertebrate assemblage (Klemm and others 2003; Stoddard and others 2006a). The benthic IBI calculates a score between 0 and 100 for each site, with 100 denoting the best attainable condition.

Sampling of Stressors

A comprehensive set of chemical and physical stressors was measured at each site at the same time as biological

indicators were sampled. We used total phosphorus and total nitrogen concentrations to indicate nutrient stress (Table 1). Acid neutralizing capacity and sulfate concentrations were used to indicate stresses of acid deposition, mine drainage, and acid mine drainage (Herlihy and others 1990). We also used excess fine sediments, riparian condition, and the lack of large woody debris as physical habitat stressors (Lazorchak and others 1998; Kaufmann and others 1999).

Condition Classes for Macroinvertebrate IBI and Stressors

We defined "poor," "marginal," and "good" classes of stressor or response condition as meaning "different from," "possibly different from," or "not different from," respectively, the stressor or response values expected at least-disturbed reference sites (Reynoldson and others 1997; Bailey and others 2004; Stoddard and others

2006b). Here, we summarize the procedures for setting class thresholds (Table 1). Full details are given by Stoddard and others (2006a).

Whenever possible, thresholds for the condition classes were based on the distribution of indicator values obtained in samples from a set of least-disturbed reference sites (Stoddard and others 2006b). We first identified a separate set of reference sites for each indicator based on stream and watershed attributes that suggest minimal human disturbance for that indicator. Candidates for reference sites included all sites sampled by the MAIA survey, as well as 58 hand-picked sites. For example, reference sites identified for developing the macroinvertebrate IBI were nonacidic, with low sulfate, chloride, and nutrient levels, and high overall habitat quality (Waite and others 2000; Stoddard and others 2006a). However, to avoid circularity, we did not use macroinvertebrate IBI scores themselves to help identify reference sites for the macroinvertebrate IBI. A similar strategy was employed for most stressors (Stoddard and others 2006a). For example, we chose reference sites for the large wood stressor to be those having (a) less than 10% of their watershed area in some combination of urban, agriculture, or mining land uses, and also (b) riparian habitat indices greater than 0.8, on a 0–1 scale.

Given a set of reference sites, we defined condition classes based on quantiles of the distribution of indicator values at those sites (Stoddard and others 2006a, 2006b). If higher values of an indicator denoted improved condition (e.g., amount of large wood), then scores lower than the 1st percentile of the reference site distribution were classified as “poor.” Scores between the 1st and 25th percentiles for reference sites were classified as “marginal”, and those higher than the 25th percentile of reference sites were classified as “good.” On the other hand, if increased indicator scores denoted worse condition (e.g., phosphorus concentration), then the “good”–“marginal” and “marginal”–“poor” thresholds were set-by the 75th and 99th percentiles, respectively, of the reference site distribution.

For the acid deposition, mine drainage, and acid mine drainage stressors, we used acid neutralizing capacity (ANC) and sulfate (SO_4) criteria based on prior research, rather than reference site distributions, to set thresholds of condition classes (Herlihy and others 1990, 1991). We first assigned condition classes for mine drainage, based on sampled SO_4 concentrations in stream water (Table 1). We then assigned acidification status based on measured values of ANC. For sites showing evidence of mine drainage (either marginal or poor condition for this stressor), we assumed that a sampled value of low ANC (high acidity) was the result of acid mine drainage, and either poor or marginal conditions of the acid mine drainage stressor were assigned depending on ANC (Table 1). Where there was little or no evidence of mine drainage (mine drainage

condition = “good”), a low ANC level was assumed to be evidence of acid deposition, and the acid deposition stressor was assigned to be poor or marginal depending on ANC.

Thresholds for the nitrogen, phosphorus, and sedimentation stressors were determined separately by ecoregions, to capture strong ecoregional differences in least-disturbed conditions (Table 1, Figure 1; Omemik 1987; Woods and others 1996). Moving from left to right in Figure 1, the Western Appalachians ecoregion has low rounded hills, low-gradient streams, and extensive wetland areas. In contrast, the North-Central Appalachian ecoregion is higher elevation, more rugged, and more forested. The Ridge and Valley ecoregion consists of a sequence of limestone or shale valleys running northeast to southwest, separated by forested ridges. Moving east, the Piedmont ecoregion is a transition between the higher Appalachians and the Coastal Plain, and is currently reverting to woodland and urban/suburban land use after being widely cultivated. Finally, the flat Coastal Plain ecoregion has low-gradient, sandy-bottom streams in landscapes ranging from marshes to urban areas to woodlands. More detailed ecoregion descriptions are given by Stoddard and others (2006a). For some stressors and some ecoregions, such as nitrogen and phosphorus in the Coastal Plain ecoregion, suitable reference sites were either rare or nonexistent, and we instead relied on best professional judgment to set thresholds (Stoddard and others 2006a, 2006b).

Relative Extent of Poor Condition

We define the *relative extent* of poor condition as the total stream length that was found in poor condition, expressed as a proportion of the total stream length in all condition classes (good, marginal, or poor). We report the confidence interval for each extent estimate that would apply if that single estimate had been made in isolation. Details of confidence interval estimation are given in the Appendix.

Relative Risk of Poor Macroinvertebrate Condition

We used relative risk estimates to describe the associations between macroinvertebrate IBI condition and the condition of each of the eight stressors. In this section, we define relative risk and illustrate its estimation by using excess sediment as an example stressor. A general formulation, along with methods for estimating confidence intervals, is given in the Appendix.

The relative risk of poor macroinvertebrate IBI condition, given poor sediment condition, is defined as a ratio of conditional probabilities (Lachin 2000; Woolson and Clarke 2002):

Table 2. Estimated lengths of stream (km) and sampled site counts (in parentheses), for combinations of good and poor macroinvertebrate IBI and sediment condition

Macroinvertebrate IBI condition	Sedimentation condition	
	Good	Poor
Good	22,700 (78)	3,930 (8)
Poor	27,450 (50)	27,680 (55)

$$R = \frac{\text{Pr}(\text{poor IBI, given poor sediment condition})}{\text{Pr}(\text{poor IBI, given good sediment condition})} \quad (1)$$

The relative risk ratio can be estimated from a contingency table containing the estimated lengths of streams in various combinations of IBI and sediment condition (Table 2). Because sample sites are unequally weighted (see Appendix), the occurrence probabilities of Equation 1 are most accurately estimated from the length estimates of Table 2, rather than from the Table’s raw counts of sampled sites (Rao and Thomas 1988; Lohr 1999).

In estimating relative risk to macroinvertebrates for a particular stressor (Equation 1 and Table 2), we excluded all sites that were in marginal condition either for macroinvertebrate IBI or for that stressor. We did this because our classes of “good,” “marginal,” and “poor” are discretizations of inherently continuous gradients of stressor and biotic condition, so that sites on either side of, and close to, a condition class boundary may have quite small differences in actual condition. By contrasting only “poor” and “good” condition classes, we ensure that there is little or no overlap in actual condition between sites assigned to the two classes. In addition, our relative risk estimates are based on the two extremes of class conditions, good and poor, and should thus represent the largest observed severities of stressor effects.

We now illustrate the calculation of relative risk for the data in Table 2. The probability, or “risk,” of finding poor rather than good macroinvertebrate IBI condition, given that sites have poor sediment conditions, is estimated by the proportion of those streams with poor sediment that also had poor IBI. From Table 2, that proportion is $27,680 / (3,930 + 27,680) = 0.876$. Likewise, the risk of finding poor rather than good IBI in streams, given that they had good sediment conditions, is estimated by $27,450 / (22,700 + 21,450) = 0.547$. Thus, poor IBI conditions had a greater risk of occurring when sediment conditions were poor (risk = 0.876) than when sediment conditions were good (risk = 0.547). Relative risk expresses this relationship as the ratio $R = 0.876 / 0.547 = 1.60$ (Equation 1). In summary, a poor (rather than good) macroinvertebrate IBI

score was estimated as 1.60 times more likely to occur in streams having poor sediment condition than in streams having good sediment condition.

A relative risk of 1.0 denotes independence between stressor and response classes. That is, if $R = 1.0$, then poor IBI condition is just as likely to occur under poor sediment conditions as it is under good sediment conditions. A 90% confidence interval for the sediment–macroinvertebrate relative risk is (1.11, 2.30) (see Appendix for methods). This confidence interval does not include 1.0, giving us statistical evidence that the risk of poor IBI in the stream population is indeed elevated in poor-sediment streams, as compared with good-sediment streams.

Correlated Stressors

If two or more stressors are strongly correlated, then their effects are confounded and cannot be clearly assessed by a bivariate association measure such as relative risk. To explore this potential problem, we estimated a correlation matrix for the categorical stressor variables in Table 1. We calculated the product-moment correlation, r , between each pair of stressor variables after receding their “poor” and “good” classes to 1’s and 0’s (Bishop and others 1975). This correlation between binary variables, also known as Cramer’s Φ coefficient (Zar 1999), has the same interpretation as a conventional correlation coefficient. We incorporated sampling weights into our estimates of r (Sarndal and others 1992).

Results

The estimated product-moment correlation between nitrogen and phosphorus stressor classes was 0.58, and the estimate for large wood versus riparian habitat was 0.48. These correlations are high enough to indicate some confounding between stressors. Thus, we interpret nitrogen and phosphorus stressor results together as a generalized nutrient loading effect, and we also interpret large wood and riparian habitat results together. All other pairs of stressors had correlation magnitudes less than 0.24, suggesting little confounding, so that their relative risks could be directly compared.

Only 26% of Mid-Atlantic stream length was estimated to be in good condition for macroinvertebrate IBI, with another 37% that was marginal (Figure 3). For stressors representing physical habitat (sedimentation, lack of large wood, and riparian habitat) and nutrient loading (phosphorus and nitrogen), between 40% and 57% of stream length was estimated to be in good condition. With between 90% and 99% of stream length estimated in either good or marginal stressor conditions, serious mining and

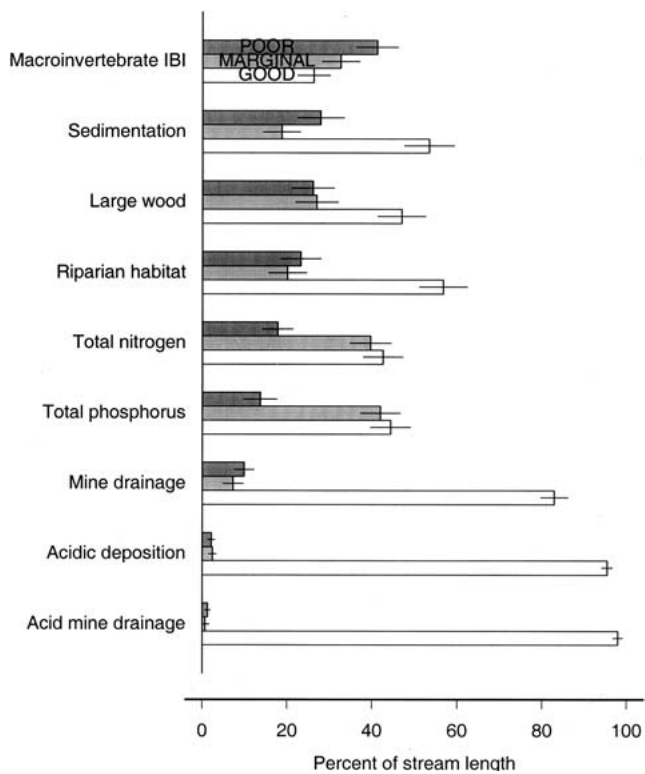


Fig. 3 Estimated extents of good, marginal, and poor-condition streams, for macroinvertebrate IBI and eight stressors. Error bars denote 90% confidence intervals for a single estimate of extent

acidification impacts were relatively rare across the entire population of Mid-Atlantic streams (Figure 3).

The relative importance of stressors can be more fully assessed by estimating their relative risks for macroinvertebrates and comparing these risks with each other and with stressor extent estimates (Figure 4). Poor conditions of mine drainage, acid deposition, and acid mine drainage occurred in only a small proportion of Mid-Atlantic streams. However, where these stressors were in poor condition, they all showed significant relative risks for macroinvertebrates (Figure 4). Poor conditions of large woody debris, riparian habitat, nitrogen, and phosphorus occurred in 14–26% of streams. Although these four stressors had estimated relative risks ≥ 1 , their confidence intervals for relative risk included 1.0, suggesting that their associations with poor macroinvertebrate condition were weak, at best. Overall, excess sedimentation appears to be the stressor of greatest concern for macroinvertebrates, with the highest estimated relative extent of poor condition (28% of streams) and an estimated relative risk of 1.60 (Figure 4).

Discussion

We believe that relative risk is a simple and interpretable measure of severity for a stressor's effect. Figure 4 illus-

trates the complementary roles of relative extent and relative risk in assessing stream ecological conditions at a regional scale. Based only on their relative extents of poor condition, one would conclude that physical features of stream habitat, such as sedimentation, riparian habitat, and large wood, are the most prevalent, and hence most important, stressors in Mid-Atlantic streams. Likewise, acidification and mining effects might be viewed as negligible because poor conditions of these stressors occur in so few streams. However, these conclusions may be altered if one considers both stressor severity and extent, where severity is measured by relative risk to macroinvertebrates. Among the three physical habitat stressors, only sedimentation had a relative risk that clearly exceeded 1.0. Although poor mining and acidification conditions are rare, when they do occur their relative risk to macroinvertebrates is significant, exceeding all other stressors except sedimentation. Based on both its severity of effect and its extent, sedimentation appears to be the stressor of overall greatest concern for macroinvertebrates.

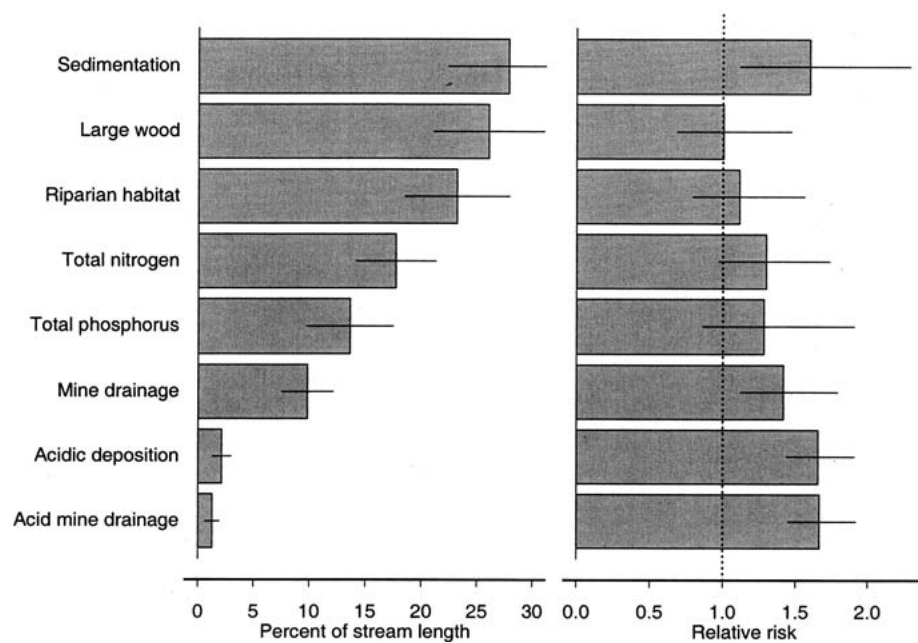
We emphasize that the relative risks estimates of Figure 4 pertain only to an IBI for aquatic macroinvertebrate assemblages. Stoddard and others (2006a) also developed IBI's for fish and periphyton assemblages sampled during the MAHA/MAIA surveys, at the same sites and times as macroinvertebrates. The reference-condition approach was again used to classify periphyton and fish IBI scores as either "good," "marginal," or "poor." Unlike the risks to macroinvertebrates, relative risks to fish IBI were highest for nutrients, lack of large wood, and acidic stressors (acid deposition and acid mine drainage), whereas the sedimentation risk was very near 1.0 (Stoddard and others 2006a). The periphyton IBI also had high relative risks from nutrients (>2.5), and its sedimentation risk was 1.5, whereas all other stressors showed nonsignificant relative risks. These results support the expectation that different components of stream ecosystems (fish, macroinvertebrates, and periphyton) would respond differently to various aquatic stressors.

Alternative Risk-Based Measures of Stressor Effect

Alternatives to relative risk are available for measuring an association between dichotomous stressors and a response. Because Environmental Monitoring and Assessment Program surveys are cross-sectional sampling designs, as opposed to retrospective or prospective designs, the conditional and unconditional probabilities underlying these alternatives can be estimated without bias from the basic contingency table in Table 2 (Ramsey and Schafer 1997; Lachin 2000; Woolson and Clarke 2002).

For example, an odds ratio is similar to a relative risk ratio, except that the numerator and denominator are expressed as the odds, rather than the risk, of poor IBI,

Fig. 4 Estimated relative extent of poor stressor condition (left panel), and the estimated relative risk of poor macroinvertebrate IBI condition under poor stressor condition (right panel). Error bars denote single-estimate 90% confidence intervals for extent and Bonferroni-corrected 90% confidence intervals for relative risk. Vertical line denotes the “No association” value (1.0) for relative risk



conditional on poor (numerator) or good (denominator) stressor condition. The odds of poor IBI is defined as the probability (i.e., risk) of poor IBI condition, divided by the probability of good IBI condition (Agresti 1990; Woolson and Clarke 2002). For our Table 2 example, the estimated odds of poor IBI condition in poor-sedimentation streams is calculated as $(27,680/(3930 + 27,680))/(3930/(3930 + 27,680)) = 27,680/3930 = 7.04$. That is, poor IBI is 7.04 times more likely than good IBI to be found in poor-sediment streams. Similarly, the odds of finding poor, rather than good, IBI in good-sediment streams is given by $27,450/22,700 = 1.21$. The odds ratio, given by $7.04/1.21 = 5.81$, then expresses how much greater is the odds of finding poor IBI in streams with poor, rather than good, sediment conditions. We prefer using the relative risk ratio because we believe that, for most people, the risk, or probability, of an event occurring is easier to understand than the odds of an event occurring.

As a second example, the absolute conditional risks that appear in the numerator and denominator of the relative risk ratio may be individually informative. If all sampled sites having poor sediment condition were to be included in its estimate, then the numerator risk of Equation 1 would be the regional extent of stream length with poor IBI, conditional on streams also having poor sediment conditions. The relative risk denominator has a similar interpretation of conditional extent. For our particular application, however, these useful interpretations are not valid because we excluded sites with marginal IBI and/or marginal stressor conditions when estimating absolute and relative risks.

A single absolute risk appears easier to interpret than a ratio of two risks. Thus, one may be tempted to interpret

the relative risk numerator, by itself, as a measure of stressor severity. However, a single absolute risk may give a misleading picture of severity, because, by definition, an “effect” can only be produced by a change in stressor intensity, e.g., from good to poor condition. The only reliable way to quantify that effect is to measure the change in the response as a function of the change in the stressor. For this reason, the starting point in the design of human epidemiological studies is the comparison of two groups of people: those experiencing a “risk factor” (i.e., stressor), and those who are free of it. Thus, the severity of effect for a risk factor can be accurately assessed only by comparing two risks, using either a difference or a ratio (Dunn and Everitt 1995; Lachin 2000; Woolson and Clarke 2002).

Other variations on relative risk may be useful. For example, by reversing the roles of stressor and response variables in Table 2, one could estimate the risk of observing poor, rather than good, sediment conditions in streams with poor IBI conditions, relative to that risk in streams with good IBI conditions. It is also possible to carry out chi-squared tests of independence for contingency tables, employing specialized methods to account for unequal sampling probabilities (Rao and Thomas 1988; Lohr 1999). We recommend that contingency tables be made available whenever reporting relative risks, to enable estimation of alternative risk-based measures.

Interpreting Relative Risk

Our proposed application of relative risks should be interpreted carefully, for at least three reasons. First, comparisons of relative risks must recognize that “poor”

and “good” conditions do not have identical meanings for all stressors or responses. In our MAIA example, we employed a unique set of thresholds to define the condition classes for each stressor and response. In addition, those thresholds were determined using best professional judgment for some stressors, and using quantiles of reference-site distributions for others. Regardless of how we determined the class thresholds, our objective was to define stressor or response levels near or beyond the extremes of the distributions at reference sites as denoting “poor” conditions. It is in this restricted and approximate sense that our “poor” conditions can be interpreted as a common currency across stressors.

Secondly, if stressors are correlated, then their individual effects cannot entirely be teased apart, and relative risk estimates may give a misleading picture of their relative effect severities. We addressed this problem by calculating correlations between all pairs of stressor variables and then interpreting relative risks for moderately-to-highly-correlated stressors as representing confounded effects. Alternatively, one could model the joint effects of several categorical stressors on a categorical response, using multiple logistic regression or multiway contingency tables (Bishop and others 1975, Agresti 1990). In principle, the standardized partial effect of a stressor within such a model can measure its relative importance. However, in the presence of strong multicollinearity, estimated partial regression effects can, like relative risk, give a misleading assessment of stressor importance (Neter and others 1990, Montgomery and others 2001).

Thirdly, for many people the language of “relative risk” implicitly asserts a causal link between stressor and response. We recognize that monitoring efforts such as the MAIA survey provide associational data, which, by itself, does not give conclusive evidence of causality (U.S. EPA 1998, Shipley 2000, Adams 2003). Thus, in applying the relative risk approach, we have been careful to say that poor stressor and poor biota conditions tend to be “found” or to “occur” in the streams, emphasizing that our relative risk estimates are strictly associational. However, we also believe that analyses of associations based on monitoring data, combined with conceptual models and other evidence for or against competing causal mechanisms, can offer strong inferences regarding likely causes. We encourage potential users of the relative risk approach to be aware of the causality issue, and to recognize that a well-designed monitoring program can provide strong, albeit not conclusive, evidence about causes of ecosystem impairment.

Finally, whether one focuses on relative extent, relative risk, or a combination of the two depends, in part, on the management or policy question being addressed. If the management question is on decisions about a specific site, then relative risk would be an appropriate focus. If one is

more interested in which stressors should be the focus of policy for a region or the nation with the intent of producing the greatest improvement in length of streams in good biological condition, then using both relative extent and relative risk is more appropriate.

Acknowledgments We thank Susan Norton, Bob Ozretich, and three anonymous referees for valuable comments on the manuscript. Alan Herlihy, Bob Hughes, Phil Kaufmann, and Phil Larsen also contributed their insights. This research was funded by the U.S. Environmental Protection Agency. The manuscript has been subjected to review by the Western Ecology Division of EPA’s National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Appendix: Standard Errors and Confidence Intervals for Relative Risk and Relative Extent

Estimating Relative Risk and Its Standard Error

From quantities in Table A1, relative risk is estimated as:

$$R = \frac{T_2/T_4}{T_1/T_3} \quad (A1)$$

We used large-sample approximations to estimate standard errors and construct confidence intervals for R . Because the domain of $\log(R)$ is symmetric about the no-association value of $\log(1.0)$, and confidence bounds on $\log(R)$ cannot escape that domain, we estimated standard errors and resulting confidence intervals on the log-transformed scale (Lachin 2000). From Equation A1, $\log(R)$ is given by:

$$\log(R) = \log(T_2) - \log(T_4) - \log(T_1) + \log(T_3) \quad (A2)$$

The variance of $\log(R)$ can be estimated by applying first-order Taylor linearization to each term on the right-hand side of Equation A2 to give (Sarndal and others 1992):

$$Var[\log(R)] = \sum_{i=1}^4 a_i^2 Var(T_i) + 2 \sum_{i=1}^4 \sum_{j<i} a_i a_j Cov(T_i, T_j) \quad (A3)$$

where $a_i = \pm \frac{\partial[\log(T)]}{\partial T_i} = \pm \frac{1}{T_i}$. The exact sign on $1/T_i$ is equal to the sign on the i th term in Equation A2, and partial derivatives are evaluated at point estimates of T_i .

If a simple random sampling design is employed, then the probability that a site is sampled is equal for all sites, and the inclusion of any site in the sample is independent of which other sites are sampled. In this case, the quantities in

Table A1. The contingency table of stressor condition versus response condition can be written as:

Response condition	Stressor condition	
	Good	Poor
Good	A	B
Poor	T ₁	T ₂
Sum	T ₃ = A+T ₁	T ₄ = B+T ₂

Table A1 and Equation A1 are raw site counts or their sums. The estimated variance of $\log(R)$ reduces to (Lachin 2000, Woolson and Clarke 2002):

$$Var[\log(R)] = \frac{A}{T_1 T_3} + \frac{B}{T_2 T_4} \tag{A4}$$

Alternatively, if sample inclusion probabilities for sites are not equal, as in the MAIA and other EMAP surveys, then cell entries in Table A1 are stream lengths, estimated by summing the sampling weights for all sampled sites in each cell (Stehman and Overton 1994, Lohr 1999). For this case, we used Horvitz-Thompson estimators of the variances and covariances in Equation A3, because of their applicability to any unequal-probability sampling design (Sarndal and others 1992, Stevens and Olsen 2003). In applying the Horvitz-Thompson estimators, we assumed that sites were independently and randomly sampled, so that pairwise joint inclusion probabilities are directly proportional to the product of the individual inclusion probabilities (Stevens and Olsen 2003). For spatially balanced survey designs, it is possible to reduce the magnitudes of variance and covariance terms in Equation A4 by using local neighborhood information in their estimation rather than the Horvitz-Thompson method (Stevens and Olsen 2003).

Confidence Intervals

Given an estimate of the standard error, $SE_{\log R} = \sqrt{Var[\log(R)]}$, Normal distribution theory was used to construct a large-sample confidence interval (CI) for $\log(R)$ (Lachin 2000). Specifically, a two-sided CI at the $100*(1-\alpha)$ percent confidence level is given by $\log(R) \pm Z_{(1-\alpha/2)}(SE_{\log R})$, where Z denotes a percentile of the standard normal distribution. We applied a Bonferroni correction when computing multiple, nonindependent CIs on the relative risks between macroinvertebrate IBI and several stressors. For a family of K confidence intervals, a familywise confidence level of $100*(1-\alpha)$ percent is obtained by using the $[1-\alpha/(2K)]$ percentile of Z (Ramsey and Schafer 1997). Endpoints of a CI on $\log(R)$ were back-transformed to give a CI for R itself. Because the back-transformation is nonlinear, a two-sided CI on the

relative risk scale is not symmetric about the point estimate of R .

Stressor Extent

The estimated relative extent of poor stressor condition is a ratio of two totals—the total estimated stream length in poor stressor condition, and the overall total estimated length of streams. Each total was calculated by summing the weights of sampled streams. We approximated the standard error for stressor extent using Taylor linearization and Horvitz-Thompson estimation, by following steps similar to those given above for relative risk but omitting the logarithmic transformation (Sarndal and others 1992).

Free, R-language software (Ihaka and Gentleman 1996) for calculating relative risk and extent estimates, and their confidence intervals, from unequal-probability survey data is included in the `psurvey.analysis` package (available at <http://www.epa.gov/nheerl/arm/>).

References

Adams S. M. 2003. Establishing causality between environmental stressors and effects on aquatic ecosystems. *Human and Ecological Risk Assessment* 9:17–35

Agresti A. 1990. *Categorical data analysis*. John Wiley and Sons, New York

Bailey R. C., R. H. Norris, T. B. Reynoldson 2004. *Bioassessment of freshwater ecosystems: Using the reference condition approach*. Kluwer Academic Publishers, New York

Bishop Y. M. M., S. E. Feinberg, P. W. Holland 1975. *Discrete multivariate analysis: theory and practice*. MIT Press, Cambridge, Massachusetts

Boward D. M., P. F. Kazyak, S. A. Stranko, M. K. Hurd, and T. P. Prochaska. 1999. From the mountains to the stream: the state of Maryland’s freshwater streams. EPA/903/R/99/023, Maryland Department of Natural Resources, Monitoring and Non-tidal Assessment Division, Annapolis, Maryland

Comeleo R. L., J. F. Paul, P. V. August, J. Copeland, C. Baker, S. S. Hale, R. W. Latimer. 1996. Relationships between watershed stressors and sediment contamination in Chesapeake Bay estuaries. *Landscape Ecology* 11:307–319

Dunn G., B. Everitt 1995. *Clinical biostatistics: an introduction to evidence-based medicine*. Edward Arnold, London

Dyer S. D., C. White-Hull, G. J. Carr, E. P. Smith, X. Wang. 2000. Bottom-up and top-down approaches to assess multiple stressors over large geographic areas. *Environmental Toxicology and Chemistry* 19:1066–1075

Gordon S. I., S Majumder. 2000. Empirical stressor-response relationships for prospective risk analysis. *Environmental Toxicology and Chemistry* 19:1106–1112

Herlihy A. T., P. R. Kaufmann, M. E. Mitch, D. D. Brown. 1990. Regional estimates of acid mine drainage impact on streams in the mid-Atlantic and southeastern United States. *Water Air and Soil Pollution* 50:91–107

Herlihy A. T., P. R. Kaufmann, M. E. Mitch. 1991. Chemical characteristics of streams in the Eastern United States, 2. Sources of acidity in acidic and low ANC streams. *Water Resources Research* 27:629–642

- Herlihy A. T., D. P. Larsen, S. G. Paulsen, N. S. Urquhart, B. J. Rosenbaum. 2000. Designing a spatially balanced, randomised site selection process for regional stream surveys: The EMAP Mid-Atlantic Pilot Study. *Environmental Monitoring and Assessment* 63:95–113
- Ihaka R., R. Gentleman. 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5:239–314
- Karr J. R. 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6:21–27
- Karr, J. R., and E. W. Chu. 1997. Biological monitoring and assessment: using multimetric indexes effectively. EPA/235/R97/001, University of Washington, Seattle, Washington
- Kaufmann, P. R., P. Levine, E. G. Robison, C. Seeliger, and D. Peck. 1999. Quantifying physical habitat in Wadeable streams. EPA/620/R-99/003, U.S. Environmental Protection Agency, Washington, D.C
- Klemm D. J., K. A. Blocksom, F. A. Fulk, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, W. T. Thoeny. 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. *Environmental Management* 31:656–669
- Lachin J. M. 2000. *Biostatistical methods: the assessment of relative risk*. John Wiley and Sons, New York
- Lazorchak, J. M., D. J. Klemm, and D. V. Peck (eds.) 1998. *Environmental Monitoring and Assessment Program—Surface Waters: field operations and methods for measuring the ecological conditions of Wadeable streams*. EPA/620/R-94/004F, U.S. Environmental Protection Agency, Cincinnati, Ohio
- Lazorchak, J. M., B. H. Hill, D. K. Averill, D. V. Peck, and D. J. Klemm (eds.) 2000. *Environmental Monitoring and Assessment Program—Surface Waters: field operations and methods for measuring the ecological condition of non-wadeable rivers and streams*. EPA/620/R-00/007, U.S. Environmental Protection Agency, Washington, D.C
- Lohr S.L. 1999. *Sampling: design and analysis*. Brooks/Cole, Pacific Grove, California
- Montgomery D. C., E. A. Peck C. G. Vining. 2001. *Introduction to linear regression analysis* (3rd ed.). John Wiley and Sons, New York, New York
- Neter J., W. Wasserman, M. H. Hunter 1990. *Applied linear statistical models* (3rd ed.). Irwin, Homewood, Illinois
- Omernik J. M. 1987. Ecoregions of the coterminous United States. *Annals of the Association of American Geographers* 77:118–125
- Ramsey F. L., D. W. Schafer 1997. *The statistical sleuth: A course in methods of data analysis*. Duxbury Press, Belmont, California
- Rao J. N. K., D. R. Thomas. 1988. The analysis of cross-classified categorical data from complex sample surveys. *Sociological Methodology* 18:213–269
- Reynoldson T. B., R. H. Norris, V. H. Resh, K. E. Day, D. M. Rosenberg. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society* 16:833–852
- Sarndal C.-E., B. Swensson, J. Wretman 1992. *Model-assisted survey sampling*. Springer-Verlag, New York
- Shiple B. 2000. *Cause and correlation in biology*. Cambridge University Press, Cambridge, UK
- Stehman S. V., W. S. Overton 1994. Environmental sampling and monitoring. In: G. P. Patil, C. R. Rao (eds). *Handbook of statistics*, volume 12. Elsevier, New York. pp 263–306
- Stevens D. L. Jr., A. R. Olsen. 2003. Variance estimation for spatially-balanced samples of environmental resources. *Environmetrics* 14:593–610
- Stoddard, J. L., J. S. Kahl, F. A. Deviney, D. R. DeWalle, C. T. Driscoll, A. T. Herlihy, J. H. Kellogg, P. S. Murdoch, J. R. Webb, and K. E. Webster. 2003. Response of surface water chemistry to the Clean Air Act Amendments of 1990. EPA/620/R-03/001, U.S. Environmental Protection Agency, Corvallis, Oregon
- Stoddard, J. L., A. T. Herlihy, B. H. Hill, R. M. Hughes, P. R. Kaufmann, D. J. Klemm, J. M. Lazorchak, F. H. McCormick, D. V. Peck, S. G. Paulsen, A. R. Olsen, D. P. Larsen, J. Van Sickle, and T. R. Whittier. (2006a). *Mid-Atlantic Integrated Assessment (MAIA)—State of the Flowing Waters Report*. EPA/620/R-06/001, U.S. Environmental Protection Agency, Washington, DC
- Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson and R. H. Norris (2006b). Setting expectations for the ecological condition of streams: The concept of reference condition. *Ecological Applications* 16:1267–1276
- Tong S. T. Y. 2001. An integrated exploratory approach to examining the relationships of environmental stressors and fish responses. *Journal of Aquatic Ecosystem Stress and Recovery* 9:1–19
- USEPA (U.S. Environmental Protection Agency). 1994. *National water quality inventory: 1992 report to Congress*. EPA/841/R-94/001, Washington, DC
- USEPA (U.S. Environmental Protection Agency). 1998. *Guidelines for ecological risk assessment*. EPA-630-R-95-002F. Washington, DC
- USEPA (U.S. Environmental Protection Agency). 2000. *Mid-Atlantic Highlands streams assessment*. EPA/903/R-00/015, U.S. Environmental Protection Agency, Region 3, Philadelphia, Pennsylvania
- USGS (U.S. Geological Survey) 1999. *The quality of our nation's waters—nutrients and pesticides*. Circular 1225, Reston, Virginia
- Van Sickle J. 2003. Analyzing correlations between stream and watershed attributes. *Journal of the American Water Resources Association* 39:717–726
- Waite I. R., A. Herlihy, D. P. Larsen, D. P. Klemm. 2000. Comparing strengths of geographic and nongeographic classifications of stream benthic macroinvertebrates in the Mid-Atlantic Highlands, USA. *Journal of the North American Benthological Society* 19:429–441
- Woods, A. J., J. M. Omernik, D. D. Brown, and C. W. Kiilsegaard. 1996. Level III and IV ecoregions of Pennsylvania and the Blue Ridge Mountains, the Ridge and Valley, and the Central Appalachians of Virginia, West Virginia, and Maryland. EPA/600R-96/077, U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Corvallis, Oregon
- Woolson R. F., W. R. Clarke 2002. *Statistical methods for the analysis of biomedical data* (2nd ed.). John Wiley & Sons, New York
- Yuan L. L., S. B. Norton. 2004. Assessing the relative severity of stressors at a watershed scale. *Environmental Monitoring and Assessment* 98:323–349
- Zar J. H. 1999. *Biostatistical analysis* (4th ed.). Prentice-Hall, Upper Saddle River, New Jersey