A broadly applicable, objective freshwater mussel assemblage health index (MAHI)

Wendell R. Haag[1,5], Angela K. Burrow[2,3,6], Traci P. DuBose[2,7], and Steven J. Price[4,8]

[1]United States Forest Service, Southern Research Station, Frankfort, Kentucky, USA
[2]Oak Ridge Institute for Science and Education (ORISE) Postdoctoral Research Fellow, United States Forest Service, Frankfort, Kentucky, USA
[3]Present affiliation: Michigan State University, Department of Fisheries and Wildlife, East Lansing, Michigan, USA
[4]University of Kentucky, Department of Forestry and Natural Resources, Lexington, Kentucky, USA

Abstract: Freshwater mussel declines range from dramatic to subtle, but there are few methods for quantifying the extent of decline and overall assemblage health. Available assessment tools are specific to particular streams or regions and depend on untested assumptions or subjective assessments of species' tolerance to various factors. We developed a broadly applicable, objective mussel assemblage health index (MAHI). We compiled a dataset of mussel survey data from 50 streams across eastern North America and evaluated the ability of 19 candidate metrics to discriminate healthy and degraded assemblages. MAHI includes 4 metrics representing fundamental aspects of assemblage or population health (species loss, recruitment, abundance, and population growth). Metrics are scaled from 0 to 10 (representing increasing health), and the composite score is the unweighted mean of the 4 metrics. MAHI effectively discriminated healthy (median score = 8.3) and degraded (3.2) assemblages. The composite score is robust to missing values—omitting any single metric resulted in a median score change of ≤0.7. Composite scores were similar to independent scores provided by experts, but MAHI eliminates variation due to differences in individual perception or experience. MAHI can be applied to streams across eastern North America using commonly collected mussel survey data, and it may be adapted to other regions. MAHI is free of untested assumptions or subjective assessments, and it accounts for expected differences in mussel assemblages due to stream size, biogeographic region, and other intrinsic factors. MAHI will be a valuable tool for early detection of mussel declines, better understanding the causes of declines, and evaluating the effectiveness of conservation strategies and management actions. We developed a Microsoft Excel application that calculates MAHI scores, and we provide a link to that resource.
Key words: bioassessment, environmental tolerance, monitoring, Unionoida, impairment, population dynamics, recruitment

Assessment of the integrity or health of ecological assemblages is widely used to evaluate the extent to which ecosystems are affected by anthropogenic stressors. For example, bioassessment of aquatic insect and fish assemblages is used worldwide as an indicator of water and habitat quality in streams (Barbour et al. 1999). Many bioassessment methods are multimetric indices that provide a single, composite score derived from metrics that each represent a different aspect of biological integrity (Lydy et al. 2000, Stoddard et al. 2008). Index scores typically are interpreted by comparison with scores from high quality reference sites, and indices often must be calibrated to account for regional faunal differences (Stoddard et al. 2008). Some bioassessment methods are based on the demonstrated or proposed tolerance of different organisms to various stressors under the assumption that low quality sites will be dominated by tolerant species (Lenat 1993, Hill et al. 2000, Wang and Lyons 2003). In addition to making inferences about water and habitat quality, bioassessment is used to monitor ecological assemblages over time and evaluate the effectiveness of management or restoration efforts (Vowell 2001).

Freshwater mussels (order Unionoida) are highly imperiled animals and are the focus of conservation efforts worldwide (Aldridge et al. 2022). Surveys and monitoring of mussel assemblages are key components of most mussel conservation programs and are used extensively to assess the relative health of assemblages, document faunal declines, evaluate potential causes of declines, and monitor the success of

E-mail addresses: [5]wendell.haag@usda.gov; [6]burrowa1@msu.edu; [7]tracipdubose@gmail.com; [8]steven.price@uky.edu

Received 5 April 2023; Accepted 15 November 2023; Published online 16 April 2024. Associate Editor, Carla L. Atkinson.

  https://doi.org/10.1086/730378

management actions (Haag and Williams 2014). However, interpretation of survey or monitoring data is subjective and not guided by consistent methods, and it can vary according to the perceptions and experience of individual biologists.

Multimetric indices for mussel bioassessment

Several multimeric indices have been developed in the United States to provide formalized bioassessment methods for mussels (see Herman et al. 2021), but all were developed for specific regions. The mussel community assessment tool (MCAT) was developed for application in the upper Mississippi River and consists of 10 metrics representing the conservation status and tolerance of constituent species, taxonomic composition, mortality, recruitment strength, abundance, and diversity of assemblages (Dunn et al. 2020). The mussel community index was developed for 4^{th}- and 5^{th}-order streams in Illinois and consists of 4 metrics representing species richness, species tolerance, abundance, and recruitment strength (Szafoni 2002). The mussel index of biotic integrity was developed for a wadable stream in Ohio and consists of 10 metrics representing species conservation status and tolerance, species richness, recruitment strength, abundance, and diversity (Hoggarth and Grumney 2013). The stream species quality index provides an overall assessment of biotic integrity in Minnesota streams, but it also provides a standalone assessment of mussel assemblages consisting of 4 metrics representing species tolerance, species richness, recruitment strength, and abundance (Minnesota Department of Natural Resources; https://www.dnr.state.mn.us/whaf/about/scores/biology/streamspc.html). Other indices incorporate more limited aspects of mussel assemblages (species tolerance or richness) in broader assessments of overall stream community health within specific geographic regions (e.g., benthic index of biotic integrity, Kerans and Karr 1994; tiered aquatic life unit, Daniel et al. 2014).

Beyond the United States, we are aware of bioassessment indices for mussels only in Germany. However, German bioassessment indices provide rankings for individual mussel populations, not assemblages (Stoeckl et al. 2020). These indices rank populations based on population attributes (population size and age structure), habitat quality, and disturbance in the watershed. However, these indices may perform poorly for some species (Stoeckl et al. 2020).

Existing mussel assemblage indices share many attributes, and their metrics can be classified into 5 broad groups, each of which represents a fundamental aspect of assemblage composition or population structure (Dunn et al. 2020; Table 1).

Tolerance and conservation status The use of species tolerance is based on the well-established idea in bioassessment that higher quality habitats should support a higher percentage of intolerant species than degraded habitats, which should be dominated by tolerant species (Lenat 1993, Wang and Lyons 2003). The use of conservation status as a metric is similarly motivated by the assumption that imperiled species (e.g., species listed as threatened or endangered under the United States Endangered Species Act or by individual states) are less tolerant to anthropogenic stressors than nonimperiled species, and, therefore, higher quality habitats should support a higher percentage of imperiled species than degraded habitats (e.g., Dunn et al. 2020). Both assumptions are intuitively reasonable, and metrics in this group are included in most mussel indices (Table 1).

Table 1. Comparison of existing bioassessment indices for freshwater mussels. See text for index sources. B-IBI = benthic index of biotic integrity, CPUE = catch per unit effort, MCAT = mussel community assessment tool, MCI = mussel community index, m-IBI = mussel index of biotic integrity, SSQI = stream species quality index, TALU = tiered aquatic life unit.

Broad metric grouping	Example metrics	Index
Tolerance and conservation status	% federally listed species	MCAT, m-IBI
	% tolerant species	MCAT, MCI, SSQI, B-IBI
Taxonomic or life-history composition	% Lampsilini	MCAT
	% burrowing species	m-IBI
	% invasive species	m-IBI
Population attributes	Recruitment strength	MCAT, MCI, m-IBI, SSQI
	Recent mortality	MCAT
Abundance	Mussel density (no./m^2)	MCAT
	Mussel abundance (CPUE)	MCI, SSQI
	Raw number of mussels	m-IBI
Richness and diversity	Pielou's evenness (J')	MCAT
	Species richness	MCAT, MCI, m-IBI, TALU, SSQI
	Shannon–Weiner index (H')	m-IBI

Taxonomic or life-history composition The motivation for metrics related to taxonomic or life-history composition is similar to that for tolerance and conservation status. It is assumed that some broad taxonomic groups (e.g., tribe or genus) or life-history strategies have higher or lower tolerance to anthropogenic stressors and therefore reflect habitat quality. For example, the MCAT uses the percentage of species in the tribe Lampsilini as a metric, assuming that either high or low percentages are reflective of poor-quality habitat (Dunn et al. 2020). Similarly, species in the tribe Anodontini and those with an opportunistic life-history strategy are generally considered more tolerant than other tribes or life-history strategies, and higher abundance of these groups is considered indicative of degraded habitats (Morris and Corkum 1996, Metcalfe-Smith et al. 1998, Hornbach et al. 2019).

Population attributes Population attributes of species comprising an assemblage are fundamental measures that should be strongly affected by, and indicative of, habitat quality. For example, low recruitment and high mortality across many species should indicate declining populations likely reflective of habitat impairment. Recruitment or mortality typically are not incorporated in other bioassessment methods because of the relatively short life spans and high annual mortality of most aquatic insects and fishes. However, it is essential to incorporate recruitment in mussel bioassessment because of the long lifespan of many species (often >30 y). Many streams support large mussel populations composed exclusively of old individuals with no evidence of recent recruitment, which portends a steep decline in the future (Haag 2012). For this reason, nearly all mussel indices include a metric that represents recruitment strength in some way (Table 1). A measure of mortality is used only by the MCAT.

Abundance Mussel abundance is perhaps the most fundamental measure of assemblage health and habitat quality: high quality habitats are expected to support more mussels than degraded habitats. Obtaining some measure of mussel abundance is a primary goal of most mussel surveys, and most mussel indices incorporate some type of abundance metric (Table 1).

Richness and diversity Species richness and diversity are intuitively appealing metrics because higher quality habitats are expected to support richer and more diverse assemblages. Indeed, species richness or diversity are fundamental metrics in many bioassessment methods for aquatic insect and fish assemblages (Barbour et al. 1999). Most existing mussel indices incorporate some measure of species richness, and 2 incorporate measures of diversity (Table 1).

Weaknesses of existing indices

Existing mussel assemblage indices have several potential weaknesses. First, because all existing indices were created specifically for a particular region or stream, they may not perform well in other areas with different faunas. Second, some indices require data that are unavailable or not commonly collected by mussel surveys. For example, assessments of tolerance are not available for most North American mussel species, and the MCAT requires an estimate of mortality, which is not often measured during routine mussel surveys (Dunn et al. 2020; see Population attributes and Methods: Candidate metrics). Third, most available assessments of species tolerance are based on untested assumptions about species' responses to various factors (see Discussion), and existing assessments for the same species often differ among sources (Haag 2012). Finally, only the MCAT has been evaluated or published in the peer-reviewed literature.

We describe a broadly applicable, objective mussel assemblage health index (MAHI). Our goals were to develop an index that is 1) applicable in streams across eastern North America, 2) calculated based on commonly collected mussel survey data, and 3) objective and largely free of untested assumptions. We discuss the extent to which MAHI discriminated healthy and degraded assemblages, how well MAHI represented professional opinion, the strengths and weaknesses of MAHI, and how it may be adapted to other regions beyond eastern North America.

METHODS

We developed MAHI by compiling a list of 19 candidate metrics in all 5 broad groups. We compiled 2 survey datasets to develop and test MAHI. We evaluated the ability of each candidate metric to discriminate between healthy and degraded mussel assemblages. We retained discriminatory metrics, scaled them from 0 to 10 (representing increasing health), and computed a composite score (the mean of scores for each metric) for all assemblages. We evaluated the extent to which MAHI scores reflected professional opinion by asking experts familiar with each stream to score them from 0 to 10 based on their field experience.

Dataset

We obtained mussel survey data for 50 stream reaches in eastern North America from published literature, unpublished reports, and colleagues (Tables 2, S1). We limited our dataset to 1 reach/stream unless widely separated reaches of a stream clearly differed in physical conditions or other attributes. For each reach, we attempted to obtain density estimates for all species in the assemblage, as obtained from quadrat sampling. If density estimates were unavailable, and if mussel density appeared to be low, we estimated density based on catch per unit effort (CPUE; see Methods: Candidate metrics: Abundance). We also obtained the shell length of each mussel encountered in the samples. If data from multiple sites were available within a relatively homogeneous stream reach, we used mean

Table 2. Stream reaches in Georgia (GA), Kentucky, (KY), Minnesota (MN), North Carolina (NC), New York (NY), Tennessee (TN), Virginia (VA), Alabama (AL), Indiana (IN), Pennsylvania (PA), and West Virginia (WV), USA, and Ontario (ON), Canada, used to develop and test the mussel assemblage health index (MAHI) and scaled scores for each stream. * indicates that mussel density was estimated from catch per unit effort. – indicates that the data were not available to calculate the metric. Median values across stream reaches for each assemblage health status category are reported, whereas mean coefficient of variation (CVs) across metrics are reported. See Table S1 for data sources.

		Individual metric scores					
Assemblage condition	Stream reach, state/province	Proportion historical richness	Assemblage recruitment strength	Density relative to stream size	Population growth	Composite MAHI score	CV
Degraded	Conasauga River, GA*	4.8	0.2	6.1	3.1	3.6	0.72
	Horse Lick Creek, KY	2.2	0.0	2.6	0.0	1.2	1.16
	Little River, KY*	2.9	5.3	1.7	–	3.3	0.56
	Little South Fork Cumberland River, KY*	1.8	0.0	3.2	0.0	1.2	1.22
	Nolin River, KY*	1.0	0.0	0.2	–	0.4	1.33
	Rockcastle River, KY	5.3	0.3	6.1	1.6	3.3	0.85
	Minnesota River, MN	5.8	10.0	3.4	–	6.4	0.52
	Little Tennessee River, NC[a]	8.6	0.0	5.1	0.0	3.4	1.23
	Hudson River, NY[b]	1.3	0.0	7.8	3.2	3.1	1.11
	Buffalo River, TN*	1.1	0.0	2.4	–	1.1	1.04
	East Fork Stones River, TN	1.0	7.3	6.7	2.3	4.3	0.72
	Tellico River, TN	0.8	0.0	3.2	–	1.3	1.26
	Clinch River, VA[c]	2.2	1.3	5.5	1.7	2.7	0.72
	Copper Creek, VA	2.9	2.5	5.6	2.8	3.5	0.41
	Middle Fork Holston River, VA	5.5	0.0	5.1	0.4	2.8	1.07
	North Fork Holston River, VA[d]	1.3	4.7	4.8	–	3.6	0.55
	Powell River, VA/TN	6.7	3.0	5.2	3.0	4.5	0.41
	Median	2.2	0.0	5.1	1.7	3.2	0.88[i]
Healthy	Sipsey River, AL	10.0	–	9.5	9.3	9.6	0.04
	Tippecanoe River, IN	7.4	10.0	8.3	–	8.6	0.15
	Green River, KY	8.7	5.0	8.1	7.6	7.3	0.22
	Licking River, KY	7.9	6.3	8.4	9.2	7.9	0.15
	Chippewa River, MN	7.1	10.0	7.1	7.7	8.0	0.17
	St Croix River, MN	10.0	10.0	6.8	6.8	8.4	0.22
	Little Tennessee River, NC[e]	10.0	10.0	8.9	–	9.6	0.06
	Delaware River, NY/PA	10.0	–	9.6	–	9.8	0.03
	Hudson River, NY[f]	7.5	–	10.0	–	8.8	0.20
	Neversink River, NY	10.0	10.0	7.2	10.0	9.3	0.15
	Grand River, ON	8.1	6.8	7.7	10.0	8.1	0.16
	Sydenham River, ON	9.4	7.8	9.1	6.9	8.3	0.14
	Thames River, ON	6.7	9.3	9.2	10.0	8.8	0.17
	Clinch River, TN[g]	8.8	2.4	9.7	9.4	7.5	0.46
	Duck River, TN	8.7	7.3	8.7	7.4	8.0	0.10
	Elk River, WV	8.2	1.9	5.8	5.7	5.4	0.48
	Little Kanawha River, WV	6.6	4.5	7.8	9.1	7.0	0.28
	Median	8.7	7.4	8.4	9.1	8.3	0.19[i]
Undetermined	Beech Fork Salt River, KY	6.1	10.0	8.9	–	8.3	0.24
	Big South Fork Cumberland River, KY	9.1	5.9	6.4	2.8	6.0	0.43
	Rolling Fork Salt River, KY	9.2	4.0	9.6	7.4	7.6	0.36
	Cannon River, MN	6.9	10.0	5.7	4.5	6.8	0.35

Table 2. (*Continued*)

Assemblage condition	Stream reach, state/province	Individual metric scores				Composite MAHI score	CV
		Proportion historical richness	Assemblage recruitment strength	Density relative to stream size	Population growth		
	Pomme de Terre River, MN	6.7	10.0	5.6	10.0	8.1	0.28
	Snake River, MN	9.6	2.7	6.6	5.8	6.2	0.46
	Straight River, MN	7.3	5.4	5.7	7.1	6.4	0.15
	Tuckasegee River, NC[h]	6.7	0.0	0.0	4.5	2.8	1.20
	Shawangunk Kill, NY	8.0	3.7	7.5	–	6.4	0.37
	Webatuck Creek, NY	8.8	3.5	7.1	–	6.4	0.42
	Saugeen River, ON	10.0	0.8	7.2	10.0	7.0	0.62
	South Thames River, ON	9.4	10.0	6.0	7.0	8.1	0.23
	Collins River, TN	6.3	7.5	6.4	–	6.7	0.10
	Meathouse Fork, WV	9.0	1.7	5.3	6.2	5.6	0.54
	Middle Island Creek, WV	9.3	0.1	6.8	0.3	4.1	1.14
	West Fork, WV	4.2	2.7	7.2	–	4.7	0.50
	Median	8.4	3.7	6.5	6.6	6.4	0.46[i]

[a] After mussel declines coincident with appearance of the invasive bivalve, *Corbicula fluminea* (Jarvis 2011).
[b] After mussel declines coincident with appearance of the invasive bivalve, *Dreissena polymorpha* (Strayer and Malcom 2007a).
[c] After mussel declines prior to 2014 (Jones et al. 2018).
[d] Downstream of industrial pollution sources at Saltville, VA (Henley et al. 2016).
[e] Prior to mussel declines coincident with appearance of the invasive bivalve, *Corbicula fluminea* (Jarvis 2011).
[f] Prior to mussel declines coincident with appearance of the invasive bivalve, *Dreissena polymorpha* (Strayer and Malcom 2007a).
[g] Prior to mussel die-off beginning in 2016 (Richard et al. 2020).
[h] Population growth calculated using catch per unit effort data.
[i] Mean value reported.

density estimates and aggregated shell lengths across sites. For many sources, it was necessary to contact the authors to obtain shell length data, and we asked the authors of all surveys to provide additional information about the mussel assemblages in the reach (see Population attributes; Richness and diversity).

We used a subset of reaches to create a development dataset, which consisted of 17 reaches that were regarded by the survey authors to support exceptionally healthy mussel assemblages and 17 reaches that were regarded as having exceptionally degraded assemblages. The full dataset included all reaches from the development dataset as well as 16 other reaches for which the assemblage could not easily be classified as healthy or degraded.

Candidate metrics

We compiled a list of 19 candidate metrics from all 5 broad groups identified by Dunn et al. (2020; Table 3), including at least 2 metrics in each group. We attempted to evaluate as many metrics as possible that are used or were evaluated by existing indices. However, we simplified or otherwise adapted several metrics, and we eliminated others for various reasons discussed in subsequent sections. All the metrics we evaluated except 1 can be calculated from data from a single mussel survey event. Population growth is the only metric that requires repeated surveys.

Conservation status and tolerance We evaluated 2 metrics in this group (Table 3). Following the MCAT and the mussel index of biotic integrity, we evaluated a metric representing the proportion of individuals belonging to species listed as threatened or endangered under the United States Endangered Species Act. To provide a broader assessment of imperilment, we also evaluated a metric representing the proportion of United States threatened and endangered species and species currently proposed for listing. We did not evaluate a metric representing species listed as imperiled by United States state agencies or as Species at Risk in Canada because regional lists often do not reflect broad-scale conservation status, and inclusion of multiple state or provincial listings would complicate metric calculation across a broad area. We did not evaluate a metric representing environmental tolerance because assessments of tolerance were not available for many species in our dataset.

Table 3. Candidate metrics evaluated during development of the mussel assemblage health index and results of Mann–Whitney tests (U, p-value) of metric scores between degraded and healthy mussel assemblages. Median scores are raw, unscaled scores. Broad metric groupings were adapted from Dunn et al. (2020). * indicates metrics retained in the final version of mussel assemblage health index. n_1 = degraded assemblages, n_2 = healthy assemblages, prop. = proportion.

					Median scores	
Broad metric grouping	Metric	n_1, n_2	U	p	Degraded	Healthy
Tolerance and conservation status	Prop. threatened and endangered	17, 17	197	<0.10	0.000	0.035
	Prop. threatened and endangered plus under review for listing	17, 17	159	>0.20	0.055	0.074
Taxonomic or life-history composition	Prop. Lampsilini	17, 17	201	<0.10	0.746	0.431
	Prop. Anodontini	17, 17	231	<0.005	0.000	0.024
	Prop. Quadrulini	17, 17	368	<0.02	0.000	0.104
	Prop. host generalists	17, 17	216	<0.02	0.045	0.376
	Prop. opportunistic life-history strategists	17, 17	210	<0.05	0.000	0.008
	Prop. periodic life-history strategists	17, 17	180	>0.20	0.393	0.312
Population attributes	Mean species recruitment strength	16, 14	176	<0.01	0.009	0.120
	Prop. of species with recruitment	17, 17	202	<0.001	0.300	1.000
	Assemblage recruitment strength*	16, 14	184	<0.005	0.002	0.116
Abundance	Density relative to stream size*	17, 17	281	<0.001	−2.544	0.172
	Population growth*	11, 13	143	<0.001	−0.103	0.050
Richness and diversity	Current species richness	17, 17	251	<0.001	5	23
	Shannon–Weiner index (H')	17, 17	224	<0.01	1.250	1.852
	Pielou's evenness (J')	15, 17	154	>0.20	0.636	0.617
	Simpson's D	17, 17	205	<0.05	0.683	0.828
	Effective number of species (Hill $N-1$)	13, 16	107	>0.20	1.262	1.183
	Prop. of historical richness present*	17, 17	279	<0.001	0.222	0.865

Taxonomic or life-history composition We evaluated 6 metrics in this group (Table 3). Following the MCAT, we evaluated 3 metrics representing the proportion of individuals belonging to tribes Lampsilini, Anodontini, and Quadrulini. We did not evaluate genus-level metrics because, except for *Lampsilis*, none of the genera evaluated by Dunn et al. (2020) are native to all streams in our dataset, and none were discriminatory in MCAT development. We evaluated 3 metrics related to life-history strategy that are not included in existing indices (proportion of host generalists, proportion of individuals belonging to opportunistic and periodic life-history strategies). We assigned species to host use (generalist or specialist) or life-history groups following Haag (2012). We considered host generalists to be species whose glochidia can complete metamorphosis on fishes from multiple families.

Population attributes We evaluated 3 metrics in this group, all of which were related to mussel recruitment (Table 3): mean species recruitment strength, proportion of species with recruitment, and assemblage recruitment strength. We did not evaluate a metric representing mortality, as included in the MCAT, for 2 reasons. First, the MCAT mortality metric represents the percentage of recently dead shells vs the number of live mussels in a sample, but most mussel surveys do not count dead shells. Second, determining the time of death for shells is based on the extent of shell decay, but decay rates differ widely among streams and species according to water chemistry and other factors (Strayer and Malcom 2007b, Illari et al. 2019), which limits the comparability of mortality estimates made in this way. Other methods for estimating mortality (e.g., mark-recapture) are beyond the scope of routine mussel surveys.

Our 3 metrics each depict a different facet of recruitment, and all 3 can be calculated from a single sampling event. The 1st metric was mean species recruitment strength across species based on the shell length distributions of all individuals of each species encountered in samples. We used a definition of <10 y old to identify recruits. The MCAT uses a definition of ≤5 y, and more restrictive definitions often are used to identify recruits in mussel sampling or monitoring (e.g., <1–2 y old, 5th percentile for length; Haag and Warren 2010, Henley et al. 2016, Hornbach et al.

2018). Our less restrictive definition has 2 advantages. First, recruits based on more restrictive definitions may not be consistently detectable by common mussel survey methods, which are strongly biased against small individuals (Vaughn et al. 1997). Second, stable mussel populations can be sustained by low annual recruitment or periodic high recruitment (Payne and Miller 2000, Villella et al. 2004, Haag 2012). Therefore, the chances of encountering younger recruits (even if detectable) can be low unless sample sizes are large or sampling takes place in a year of strong recruitment. Our less restrictive definition improves the ability to obtain a useful measure of recruitment integrated across the previous 10 y.

In addition to defining a recruit, 2 practical issues affect development of a recruitment metric. The 1st is how to estimate the age of individuals. The ideal way is to directly estimate age from external annuli, as used by the MCAT (Dunn et al. 2020). However, in many species or habitats, external annuli are indistinct or obscured by shell erosion, and interpretation of annuli in the field adds considerable time to surveys. Therefore, shell length is a more widely applicable and easily measured proxy for age. The 2nd issue is how to estimate age from shell length. The ideal way is to use length-at-age relationships for each species in the assemblage to identify recruits (Haag and Warren 2010). However, this method is impractical for widespread application because length-at-age relationships do not exist for all species, and they must be generated for each river to account for spatial variation in size and growth.

We determined a shell length criterion indicative of individuals <10 y old with von Bertalanffy growth parameters for 92 mussel species and populations, representing all North American mussel tribes (Haag and Rypel 2011). For each species or population, we calculated the predicted age of an individual that was 0.4× the maximum observed size (L_{max}), 0.5 × L_{max}, and 0.6 × L_{max}. We calculated predicted age using an inversion of the von Bertlanffy equation to solve for age based on length—this inversion performs poorly for large, older individuals, but it provides accurate age estimates for small to medium-sized mussels (Haag 2009). If L_{max} was not reported, we used asymptotic maximum length (L_{∞}). We examined frequency distributions of predicted age across all 92 species and populations for each of the 3 multipliers of L_{max} to determine the most appropriate criterion (Fig. S1). The criterion of 0.6 × L_{max} yielded age estimates <10 y old for 85% of the species. The criteria of 0.4 × L_{max} and 0.5 × L_{max} yielded age estimates <10 y old for 99% and 95% of the species, respectively. Although 0.4 × L_{max} identified individuals <10 y old slightly more often, the distribution of predicted ages was strongly right skewed, and age estimates for 79% of species were ≤4. The distribution of predicted ages for 0.5 × L_{max} was more uniform and better represented a range of ages between 1 and 9 (57% of species ≤4 y old). We chose 0.5 × L_{max} as our criterion for this reason and because its larger multiplier provides a greater chance of encountering recruits for the greatest number of species.

We calculated recruitment strength for each species as the proportion of all individuals in the sample that were ≤0.5 × L_{max} observed for the sample. For L_{max} we used the 95th percentile of observed lengths to reduce the influence of very large individuals, and we calculated recruitment strength only for species represented by ≥5 individuals. Our metric was mean species recruitment strength, calculated as the mean of recruitment strength estimates across all species at the reach. We calculated this metric only for reaches at which quadrats were sampled visually, followed by light substrate disturbance to reveal buried mussels. We did not calculate the metric for reaches at which quadrats were sampled by whole-substrate excavation and subsequent sieving of the material. The latter method yields higher numbers of small mussels than visual inspection, but it is laborious and not practical for routine application in surveys or monitoring (Miller and Payne 1988). Inclusion of data collected by sieving likely would result in higher estimates of recruitment that are not comparable with estimates made using data collected by less intensive methods.

The 2nd recruitment metric was the proportion of species in the assemblage that exhibited evidence of recent recruitment. We could make inferences about this metric for species represented by ≥5 individuals (see previous) but not for species with low sample sizes. We asked the survey authors to provide for each species their best professional opinion about whether recent recruitment had occurred at the reach (yes or no). Experienced biologists can make useful assessments of whether recruitment is occurring because they may have observed recruits during sampling for goals other than monitoring or at nearby sites, and our results supported the ability of biologists to make accurate assessments (see Results). We developed this metric to overcome our inability to assess recruitment strength of rare species and to lessen the chances of resulting erroneous conclusions. For example, consider 2 assemblages, each composed of 4 species. In 1 assemblage, 40% of individuals of 1 species are recruits, but no recruitment has occurred for the other 3 species. In the other assemblage, 10% of all 4 populations are composed of recruits. These 2 assemblages would receive identical metric scores for mean species recruitment strength (0.1), yet variation in recruitment among species indicates that the assemblages differ in overall health. However, on its own, the proportion of species with recruitment can also lead to erroneous conclusions. Consider 2 assemblages, 1 in which all species are recruiting, but recruits make up an average of 1% of individuals. In the other assemblage, all species are recruiting, but recruits make up 20% of individuals. These 2 assemblages would receive identical metric scores for the proportion of species with recruitment, yet variation in recruitment strength indicates differences in assemblage health.

The 3rd metric, assemblage recruitment strength, was the product of the first 2 recruitment metrics. This metric represents the expected proportion of individuals in the entire assemblage (including rare species) that indicate recruitment in the last 10 y. Integrating the first 2 metrics avoids potential erroneous conclusions inherent in species recruitment strength and proportion of species with recruitment when used on their own. However, because assemblage recruitment strength is the product of mean species recruitment strength and proportion of species with recruitment, values for the metric will be 0 if no recruits are observed for any species, even if biologists determine that some species are experiencing recruitment. Observed estimates of recruitment strength were 0 for 47 populations of species in our dataset that were determined by biologists as having recruitment (see Results). The median sample size for these populations was $n = 11$ (range = 5–104), which suggests that recruits may have been present but were not detected because of small sample sizes. Nondetection of recruits may result in inaccurately low recruitment strength estimates and biased metric scores. Because biologists accurately assessed the occurrence of recruitment (see Results), we developed a method to correct for probable nondetection of recruits. For all observations of recruitment strength for species determined as experiencing recruitment and when $n \geq 5$ and ≤ 28 (the median sample size for our dataset), we replaced zero-recruitment-strength estimates with the value corresponding to the 5th percentile of observed nonzero-recruitment-strength estimates in our dataset (0.017; see Appendix S1 and Fig. S2). This threshold resulted in replacement of 40 out of 47 observed zero-recruitment values representing 14% of total populations (40/285) in our dataset. After correcting values, we then calculated mean recruitment strength across species as described previously. In the special case in which no species at the site had $n \geq 5$, we set mean species recruitment strength for the assemblage to 0.017.

Abundance We evaluated 2 metrics in this group (Table 3). The 1st metric was overall mussel density (no./m^2, all species combined) at the reach relative to stream size (relative density). Existing indices use various measure of abundance, but we used mean mussel density because it is straightforward and provides more robust abundance estimates than CPUE (Vaughn et al. 1997, Strayer and Smith 2003). We used CPUE (no. mussels/h) only for 5 reaches where mussel abundance was extremely low such that quadrat sampling was inefficient (<10/h; Table 2). We estimated mean mussel density from CPUE with the relationship log (density) = 1.104(log [CPUE]) – 1.245; $p < 0.001$, $R^2 = 0.73$), derived from estimates of density and CPUE at 49 sites on 20 streams (Table S2, Fig. S3). CPUE typically is an imprecise predictor of mussel density (Strayer et al. 1997). However, for the purposes of this metric, the relationship we derived provided useful estimates of density at sites with low CPUE, but imprecise estimates when CPUE was higher (Fig. S3).

Mussel abundance typically is highest in large streams that support dense mussel beds and much lower in smaller streams that support few and scattered mussel aggregations (Haag 2012), and this intrinsic variation precludes comparison of metric values among streams of different sizes. We accounted for the influence of stream size on mussel density by compiling an independent dataset of mussel density estimates from 51 stream sites across eastern North America (Table S3). We excluded from this dataset streams where mussel density was thought to be greatly reduced by anthropogenic stressors. We estimated watershed area (km^2) upstream of each sampling reach as a proxy for stream size and computed a linear regression of mussel density on watershed area (ln mussel density = 0.338[ln watershed area] – 0.180; $p < 0.001$, $R^2 = 0.57$; Fig. S4). Our metric was the residual between observed ln mussel density at a reach and ln mussel density predicted by the regression based on watershed area.

Our 2nd metric was population growth, representing the change in mussel density (no./m^2, all species combined) over time (see Strayer and Smith 2003). Temporal trends in abundance are perhaps the single most informative measure of assemblage health and habitat quality. Declining abundance is expected to indicate poor assemblage health in response to habitat impairment, whereas increasing abundance should indicate favorable conditions. However, no existing indices incorporate direct measures of temporal trends in abundance (Table 1).

Survey data for at least 2 time periods are required to calculate population growth, and 35 reaches in our dataset had such data. Among these sites, the median number of sampling events was 3 (range = 2–15), and the median interval between events was 5.5 y (range = 1–29). The number and frequency of sampling events varies greatly among monitoring programs because of differences in goals, time since establishment, and available resources. Consequently, we standardized this metric by time to account for this variation. We used the slope of the linear regression of ln density on time to provide an estimate of the instantaneous rate of change (r) in density/y. Unlike the slope of raw density on time, the ln-normal relationship provides the same value for 2 populations changing at the same rate regardless of the magnitude of density.

We recognized that biologists often have monitoring data over time as CPUE, but they may not have density data. Consequently, we evaluated the relationship between slope estimates based on density regressed on slope estimates based on CPUE. The relationship was linear and passed approximately through the origin, and the slope was 1.104 (see the relationship between CPUE and density; previous text and Fig. S3), showing that estimates of population growth based on CPUE were consistently underestimated by 10.4% compared with estimates based on density (Appendix S2, Fig. S5). Accordingly, if monitoring data were available as

CPUE but not as density, we estimated population growth as described previously but multiplied resulting values by 1.104. We estimated population growth in this way for only 1 reach (Table 2). Although estimating density from CPUE when CPUE is high provides imprecise estimates (Fig. S3), estimating population growth from CPUE data should be unaffected by this error, assuming that timed search methods are consistent across years.

Richness and diversity We evaluated 6 metrics in this group (Table 3). Five of these metrics are used by or were evaluated by existing indices: current species richness, Shannon–Weiner index of diversity (H'), Pielou's evenness, Simpson's D, and the effective number of species (Hill $N - 1$) (see Dunn et al. 2020). We evaluated Pielou's evenness only for the species level and not for the tribe level (Dunn et al. 2020). We did not evaluate the expected number of species based on sample size (via rarefaction; Dunn et al. 2020) because the necessary data usually were not available to us. We minimized the effects of sample size on species richness estimates by using professional opinion in addition to quadrat data. We asked the people who conducted each survey to provide us their best professional opinion about which species currently occur at the reach or at nearby sites (see subsequent). Quadrat sampling is necessary to provide robust abundance estimates, but it is less efficient for estimating species richness (Strayer et al. 1997, Vaughn et al. 1997). However, biologists often encounter evidence of species occurrence during sampling for goals other than surveys or monitoring.

We developed an additional metric not used by existing indices, the proportion of historical species richness present, to represent the extent of species loss. This metric is a measure of taxonomic completeness based on the ratio of observed/expected species (Hawkins 2006). In addition to data about the current assemblage, this metric requires data sources such as historical literature, museum records, or the presence of relic shells to allow reconstruction of the probable historical species richness at the reach (expected richness). These data often do not exist for a specific study reach, but for the purposes of this metric, reasonable inference can be made using data from comparable, nearby sites. For riverine mussels, beta diversity (the difference in species richness or occurrence among sites) typically is low among sites with similar habitats (e.g., similar stream size) and within a biogeographic region (Haag 2012, Cai et al. 2019). We asked the people who conducted each survey to provide us their best professional opinion about which species occurred at the reach historically based on literature, museum records, or relic shells at the reach or nearby sites with similar habitats.

Evaluation of metrics

We calculated scores for all 19 metrics for all reaches in our development dataset. We used visual inspection of data scatter and 2-tailed Mann–Whitney tests (Excel, Microsoft Office 365, Microsoft Corporation, Redmond, Washington) to evaluate the ability of each metric to discriminate healthy and degraded assemblages (see Stoddard et al. 2008, Ma et al. 2020). In determining which metrics to retain, we considered metrics with stringent evidence of a difference between healthy and degraded assemblages (i.e., with Mann–Whitney p-values < 0.01), but some metrics that met this criterion had scores that overlapped widely between the 2 groups and provided poor visual discrimination. We therefore also considered other factors about the range and distribution of metric values (see Results), and we discarded from further consideration metrics that we considered nondiscriminatory.

Final index development

We retained 4 metrics as nonredundant and discriminatory (see Results) and calculated scores for those metrics for all reaches in our full dataset. We then scaled each metric from 0 to 10 based on the observed range of variation in our full dataset and other factors, as described subsequently. For any scaled score <0 we set the value to 0, and for any scaled score >10 we set the value to 10. Proportion of historical richness originally was on a scale of 0 to 1. We scaled this metric for a maximum value of 10 (i.e., scaled score = raw score × 10), and we left the minimum value as 0 to allow for streams that had lost all their historical richness, even though such streams were not present in our dataset. Assemblage recruitment strength had several extreme values (maximum = 0.4) that caused most other streams to have low to intermediate scores. Consequently, we scaled this metric from 0 to the 80th percentile of observed values (scaled score = [raw score / 0.157] × 10). This approach resulted in a score of 5 for an assemblage in which all species were recruiting and the mean species recruitment strength was ~0.08. Relative density had no extreme values, and we scaled this metric based on the observed range of variation (scaled score = [(raw score + 6.782) / 8.322] × 10).

Scaling population growth based on the observed range or a reduced symmetrical range (e.g., percentiles) resulted in a low score (3.8) for a stable assemblage with a growth rate (r) of 0. We considered this undesirable because stable populations should be considered indicative of a relatively healthy assemblage. Furthermore, we recognized that monitoring estimates often have high error, particularly when density is low (Strayer and Smith 2003), which can result in erroneous conclusions about population growth. Consequently, we scaled this metric between values of r representing a 4× decrease ($r = -0.139$) and a 2× increase ($r = 0.069$) over 10 y (scaled score = [(raw score + 0.1386) / 0.2079] × 10). Our rationale for this approach was that an estimated decrease of >4× probably represents a severe decline even

considering sampling error, and an estimated increase of >2× probably represents a rapidly increasing population. This scaling approach yielded a score of 6.7 for a stable population ($r = 0.000$) and a score of 7.7 for the median value of population growth across healthy assemblages in our dataset ($r = 0.021$).

After scaling all metrics, we calculated a composite MAHI score for each reach as the unweighted mean of all metric scores. We calculated the CV across metric scores within a reach to examine variability around the composite score.

Evaluation of MAHI performance

We conducted a sensitivity analysis to evaluate the extent to which variation in each metric affected composite MAHI scores. We used the Sobol' method, a variance-based sensitivity analysis, with the *sensitivity* package (version 1.28.0; Iooss et al. 2022) in R (version 4.0.2; R Project for Statistical Computing, Vienna, Austria) to build 100,000 random samples encompassing the observed range of raw metric scores and calculate Sobol' indices (S_i) with 100 bootstrap replicates (Saisana et al. 2005). We report only the 1st-order indices because MAHI metric scores are entirely additive and have no interactions.

We conducted an uncertainty analysis to evaluate the extent to which composite MAHI scores were affected by missing metric values (Saisana et al. 2005, Carboni et al. 2007). We used all reaches in our full dataset that had values for all 4 metrics ($n = 33$). We omitted 1 metric at a time and recalculated MAHI scores for each reach. We compared the absolute difference in MAHI scores with each metric omitted and the MAHI score with all 4 metrics. We tested for differences among metrics with a Kruskal–Wallis rank sum test and post hoc pairwise Dunn's test with Bonferroni adjustment (*dunn.test* package, version 1.3.5; Dinno 2017) in R.

We did not use an independent dataset to evaluate the ability of MAHI to discriminate healthy and degraded assemblages. Because our goal was to facilitate meaningful application of MAHI across a large area and in many contexts, we thought it was more important to evaluate the observed range of variation in each metric at as many sites as possible and to scale metric values accordingly, rather than to divide sites of known condition into a smaller development dataset and an evaluation dataset. Instead, we evaluated how well composite MAHI scores in our full dataset represented expert opinion by polling 24 biologists active in mussel conservation. We provided the following guidance to each biologist:

> Please provide a score for the native mussel assemblage in each stream reach that reflects the condition of the assemblage on a scale of 0–10. Zero represents a stream that has lost its entire mussel assemblage. Ten represents a fully intact healthy assemblage with no known species extirpations, mussel density that is equal to or higher than expected in that habitat type, evidence of vigorous recent recruitment for all species, and stable or increasing populations over time. You may use any method you choose to develop your scores, from initial impressions to examination of recent survey data. Please provide scores only for those streams with which you have experience and feel qualified to provide a meaningful score. You may share this poll with other experts qualified to score any of the streams, but please do not discuss your scores or scoring method with other people, and do not attempt to arrive at a consensus score with others.

We compiled scores from each respondent. We evaluated the relationship between expert scores and MAHI scores by first evaluating the effect of nonindependence of scores from each expert. We tested the following mixed-effects linear regression models and compared them with Akaike information criterion (AIC): Model 1, expert score = MAHI score; Model 2, expert score = MAHI score + expert identity; Model 3, expert score = MAHI score + expert identity + MAHI score × expert identity. We considered expert identity and the interaction random terms (*nlme* package version 3.1-152; Pinheiro et al. 2021). Model 1 had the lowest AIC score (402.7; Model 2, AIC = 403.3; Model 3, AIC = 407.1), showing that neither expert identity nor the interaction term improved model fit. Consequently, for simplicity, we used linear regression (SAS version 9.4; SAS Institute, Cary, North Carolina) to evaluate the relationship between expert scores and MAHI scores (residuals were normally distributed, and there were no influential observations).

RESULTS

Evaluation of individual metrics

Nine of 19 metrics in 4 of the 5 broad groups had values that differed between healthy and degraded assemblages (Table 3, Figs 1, 2). We retained 4 of these 9 metrics (in 3 broad groups) in the final version of MAHI.

Conservation status and tolerance and taxonomic or life-history composition Neither metric in the group conservation status and tolerance were discriminatory, and values overlapped widely between healthy and degraded assemblages. In the group taxonomic or life-history composition, the proportion of Anodontini was the only metric for which metric values differed between healthy and degraded assemblages, and values were higher for healthy assemblages. However, we did not retain this metric in the final index because the range of values for both groups overlapped completely and median values for both groups and the difference in median values (0.024) were low (see Stoddard et al. 2008), as well as for conceptual reasons (see Discussion). Values for all other metrics in the group taxonomic or life-history composition overlapped widely between healthy and degraded assemblages.

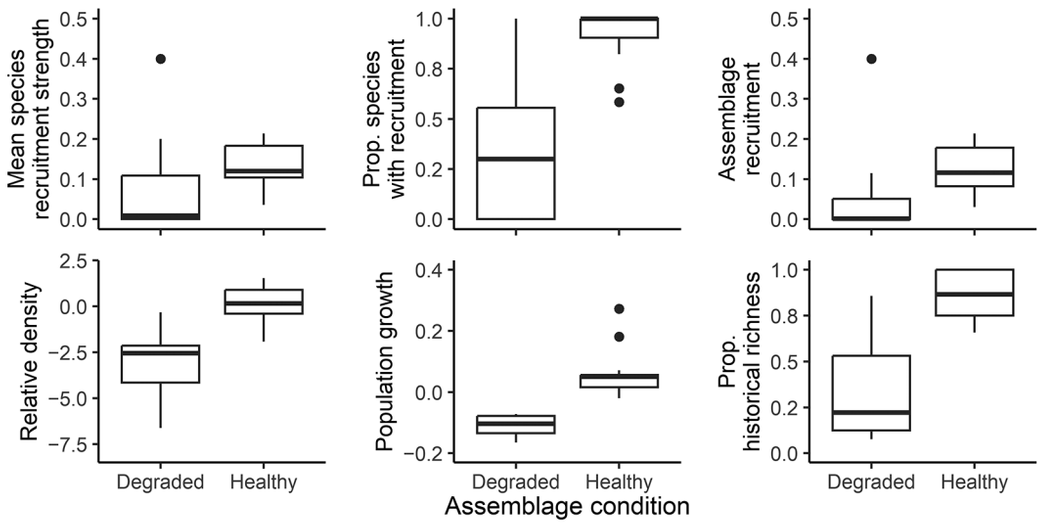

Figure 1. Boxplots showing distribution of raw values for candidate metrics that were considered discriminatory and included in the mussel assemblage health index. Horizontal lines are medians, boxes are the 25th and 75th percentiles, whiskers are 1.5× the interquartile range, and black dots are outliers. Mean species recruitment strength and proportion (prop.) of species with recruitment are not included directly in the mussel assemblage health index, but their product is assemblage recruitment strength. See Table 3 for statistical test results and sample sizes for each metric.

Population attributes Values for all 3 metrics related to population attributes differed between healthy and degraded assemblages. Values for mean species recruitment strength overlapped widely between the groups, but there was better separation for the proportion of species with recruitment and assemblage recruitment strength. For mean species recruitment strength and assemblage recruitment strength, the Minnesota River was a strong outlier among degraded assemblages. To avoid redundant metrics in the index, we retained only assemblage recruitment strength.

Estimates of recruitment strength for individual populations supported the ability of biologists to make accurate assessments of the proportion of species with recruitment. For samples with $n \geq 5$ individuals, only 4% (1/23) of populations of species classified as having no recruitment showed evidence of recent recruitment based on observed lengths, and the single exception had low estimated recruitment strength (recruits made up 1.1% of individuals). In contrast, 82% of populations of species classified as having recruitment (215/262 populations) showed evidence of recent recruitment based on observed lengths.

Correcting mean species recruitment strength for nondetection due to small sample size (see Appendix S1) resulted in a mean increase in the scaled assemblage recruitment strength metric score of only 0.2 points (range = 0–0.4). However, this correction could be more important in some situations when biologists determine that most species in an assemblage are recruiting but samples sizes are low for all species.

Abundance Both abundance metrics differed between healthy and degraded assemblages. There was minimal overlap between the assemblage groups in values of relative density, and the median residual value for healthy assemblages (0.172) was close to 0, at which predicted density = observed density. Values for population growth were nonoverlapping between healthy and degraded streams, and all but 1 healthy assemblage (Elk River) showed positive population growth. We retained both metrics in this group because they represent different aspects of assemblage condition.

Richness and diversity Values for 3 of the 6 metrics in the group richness and diversity differed between healthy and degraded assemblages. Values of H' were higher in healthy assemblages, but we did not retain this metric in the final index because values overlapped widely across a broad range and for conceptual reasons (see Discussion). Current richness and proportion of historical richness both were higher in healthy assemblages. To avoid having redundant metrics in the final index, we retained only proportion of historical richness because it provided better visual separation between the groups than current richness and for conceptual reasons (see Discussion). The other 3 measures of richness or diversity did not differ between healthy and degraded assemblages, and values for all metrics overlapped widely between the 2 groups.

Evaluation of MAHI performance

Composite index scores almost perfectly discriminated healthy and degraded assemblages (Table 2, Fig. 3). Scores

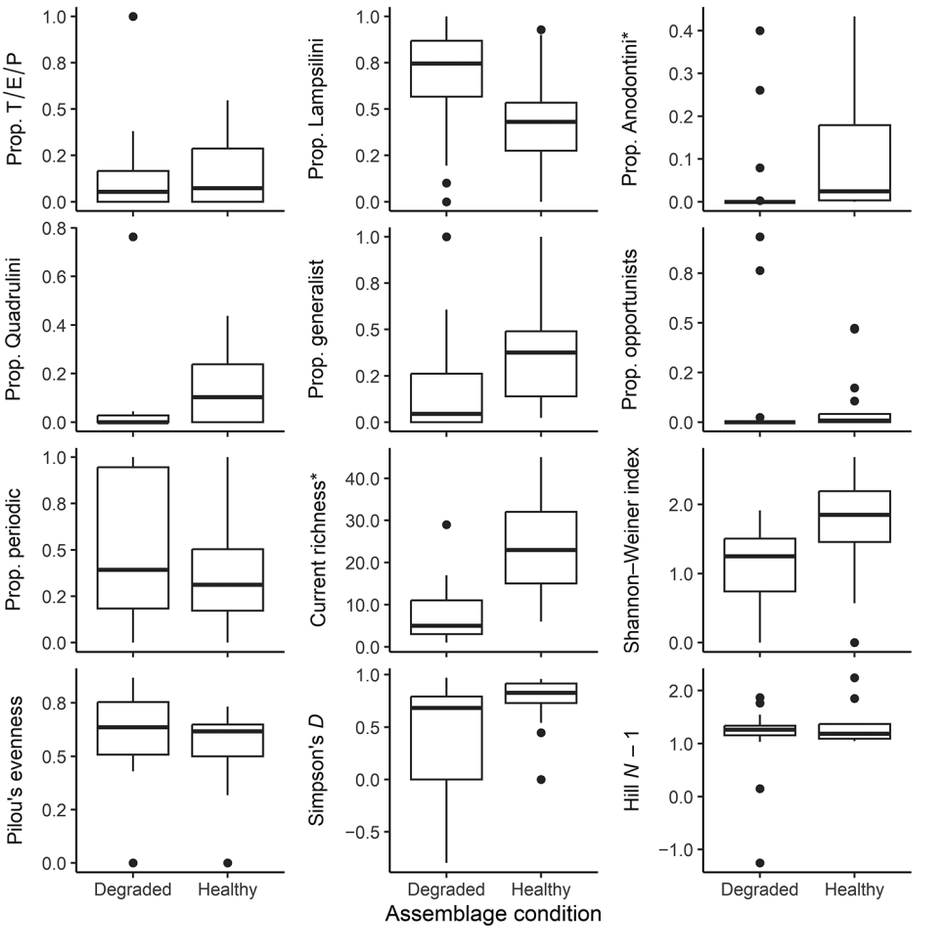

Figure 2. Boxplots showing distribution of raw values for candidate metrics considered nondiscriminatory or otherwise not included in the mussel assemblage health index. Horizontal lines are medians, boxes are the 25th and 75th percentiles, whiskers are 1.5× the interquartile range, and black dots are outliers. Metrics for which values differed between degraded and healthy assemblages ($p < 0.01$) are denoted by *. T/E/P represents species listed as threatened (T) and endangered (E) and species proposed (P) for listing under the United States Endangered Species Act. Results for the proportion of threatened and endangered species were similar to T/E/P and are not shown. Hill $N - 1$ represents the effective number of species. See Table 3 for statistical test results and sample sizes for each metric. Prop. = proportion.

were different between the 2 groups (median, healthy = 8.3, degraded = 3.2; Mann–Whitney test $U = 224$, $n_1 = n_2 = 17$, $p < 0.001$). The only assemblages that overlapped with the other group were Minnesota River (degraded, 6.4) and Elk River (healthy, 5.4). Individual metric scores were more variable for degraded assemblages (mean CV = 0.88) than healthy assemblages (mean CV = 0.19). For example, the Little Tennessee River (after the appearance of the invasive bivalve, *Corbicula fluminea*) retained most of its historical richness (score = 8.6), but recruitment and population growth metrics were both 0, and density had an intermediate score (5.1). In contrast, metric scores were relatively uniform for most healthy assemblages. MAHI scores for undetermined assemblages ranged from 2.2 to 8.3, and most scores were intermediate between healthy and degraded assemblages (median = 6.4).

Sensitivity analysis showed that the metric population growth had the greatest contribution to composite MAHI scores ($S_i = 0.34$, 95% CI = 0.33–0.34), followed by assemblage recruitment (0.235, 0.230–0.244), proportion of historical richness (0.215, 0.208–0.221), and relative density (0.211, 0.204–0.217). The following weights result in approximately equal values of S_i for each metric: population growth × 0.220, assemblage recruitment × 0.250, proportion of historical richness × 0.265, and relative density × 0.265. Weighting composite indices is a complex process without a completely objective method (Greco et al. 2019), and the weights indicated by our analysis were ~0.250 for all metrics and had little effect on composite MAHI scores (mean change in score ~0.01 points). Consequently, we retained our method of calculating the composite MAHI score as a simple unweighted mean of all metric scores.

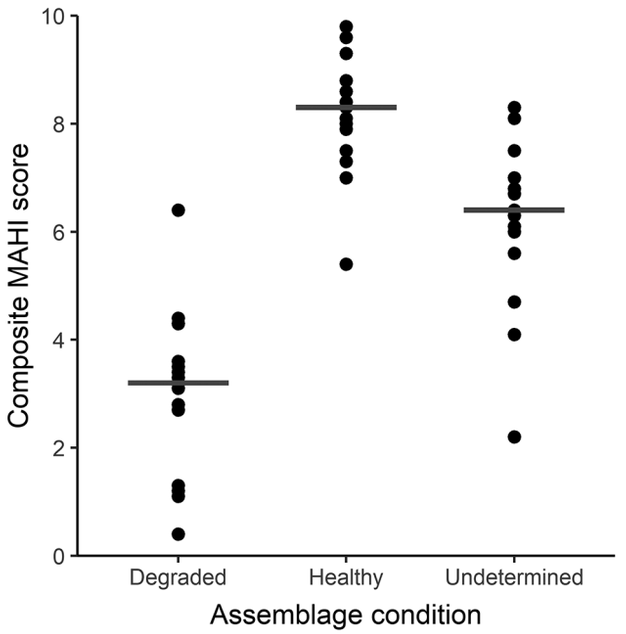

Figure 3. Composite mussel assemblage health index (MAHI) scores (black dots) for 17 degraded and 17 healthy assemblages and 16 assemblages of undetermined condition. Horizonal lines are median values.

Composite MAHI scores were robust to missing metric values (Fig. 4). The median absolute difference in index scores between indices missing 1 metric and indices with all 4 metrics ranged from 0.5 to 0.7, and no values were >2.0. The absolute difference for population growth was different than for assemblage recruitment strength, but no other metrics differed from each other (population growth vs assemblage recruitment strength: Kruskal–Wallis H_3 = 10.3, $p < 0.02$; Dunn's test $z > 2.96$, $p < 0.009$).

We received 109 scores from 22 experts, and we received at least 1 score for every reach in our dataset (mean number of scores/reach = 2.2, range = 1–5). We received scores from >1 expert for 39 reaches. Expert scores were positively related to MAHI scores, and MAHI explained 61% of the variation in expert scores (n = 109, $p <$ 0.0001, R^2 = 0.61; Fig. 5). The slope of the relationship was different than 1 (0.715, $p < 0.0001$, 95% CI = 0.606–0.825), and the intercept was different than 0 (1.413, $p <$ 0.0001, 95% CI = 0.727–2.098). Some experts scored assemblages consistently lower than MAHI, and others scored assemblages consistently higher (Fig. 6). The mean absolute difference between expert scores and MAHI scores was 1.3 (range = 0.0–4.8). The mean maximum absolute difference among expert scores for the same reach was 1.5 (range = 0.0–6.0).

Appendix S3 provides step-by-step instructions for calculating and scaling the 4 metrics that make up MAHI and calculating a composite MAHI score, and Appendix S4 provides R code for making these calculations. We also developed a Microsoft Excel application that allows users to input required data and that returns individual metric scores, a composite MAHI score, and other information. This application is available at https://www.fs.usda.gov/research/srs/products/dataandtools/tools/mahi and https://www.americanrivers.org/MAHI.

DISCUSSION

Our goal was to develop an index that accurately and objectively reflects the health of mussel assemblages. The final version of MAHI based on 4 metrics discriminated between healthy and degraded assemblages almost perfectly. Scores of only 2 assemblages overlapped between the groups, and these exceptions illustrate the potential inaccuracies of qualitative perceptions about assemblage health. The Minnesota River mussel assemblage is considered highly degraded, mainly because of its high species loss (42%; Sietman 2007). However, all remaining species are recruiting, and this assemblage had the highest recruitment values in our dataset, resulting in an intermediate MAHI score (6.4). This score indicates that despite species loss, remaining mussel populations are healthy. In contrast, the Elk River supports the most diverse remaining mussel assemblage in West Virginia (USFWS 2012) and retains most of its historical richness, but recruitment is low, and populations appear to be declining, resulting in only an intermediate score (5.4). MAHI

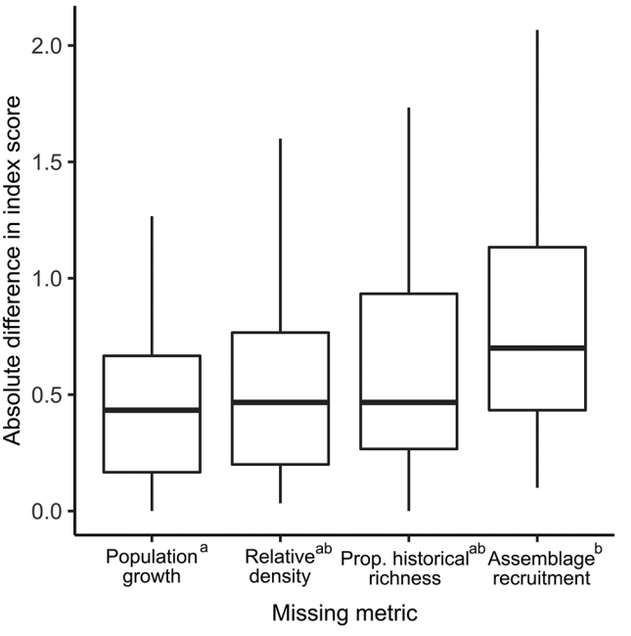

Figure 4. Absolute differences between mussel assemblage health index scores calculated with all 4 metrics and scores calculated with 1 missing metric. Horizontal lines are medians, boxes are the 25th and 75th percentiles, and whiskers are 1.5× the interquartile range. Metrics with different superscripted lowercase letters differed from each other (all $p < 0.009$). Prop. = proportion.

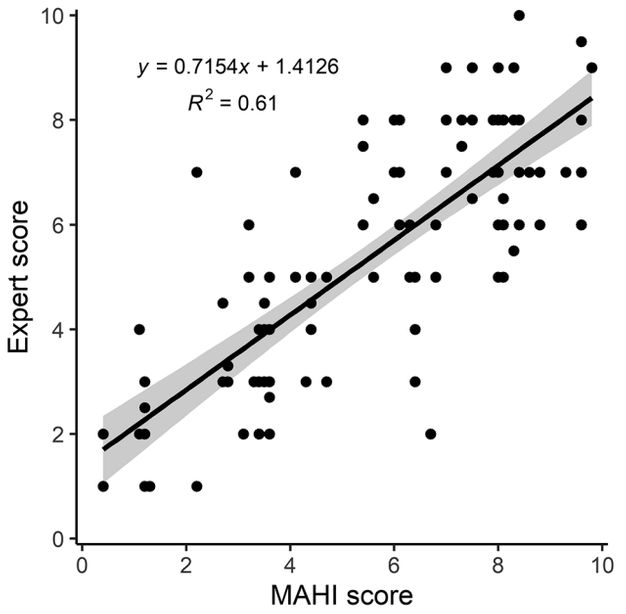

Figure 5. Comparison of mussel assemblage health index (MAHI) scores and independent expert scores for 50 mussel assemblages in stream reaches across eastern North America.

provides objective assessments that are independent of regional context and other factors that can influence individual perceptions about assemblage health.

Both the overall accuracy and objectivity of MAHI scores were illustrated by comparison of those scores with subjective expert assessments. MAHI scores generally reflected expert scores, which were, on average, within 1.3 points of MAHI scores, but the difference was as high as 4.8. The intercept and slope of the regression of expert scores on MAHI scores showed that expert scores tended to be slightly higher than MAHI scores for degraded assemblages and slightly lower for healthy assemblages. Notably, there were substantial differences of opinion among experts for specific streams. Variation among experts in scores for the same reach was of similar magnitude (mean difference = 1.5, maximum = 6.0) as differences between MAHI and mean expert scores. By eliminating individual subjectivity, MAHI provides objective health assessments that can be compared across time and space.

Selection of MAHI metrics

We omitted from MAHI 5 metrics for which scores differed between healthy and degraded assemblages. Two of these, mean species recruitment strength and proportion of species with recruitment, are redundant to assemblage recruitment strength, which integrates both measures, but both probably would be informative metrics. The other 3 metrics imply value judgements about assemblage characteristics that limit their comparability among streams. The proportion of Anodontini was higher in healthy assemblages, contrary to predictions, but Anodontini may dominate assemblages in some biogeographic regions and habitat types regardless of stream health (Haag 2012). H' was higher in healthy assemblages, as expected, but H' and species richness often are naturally low in small streams and in all Atlantic Coast streams in contrast to larger streams in more species-rich regions

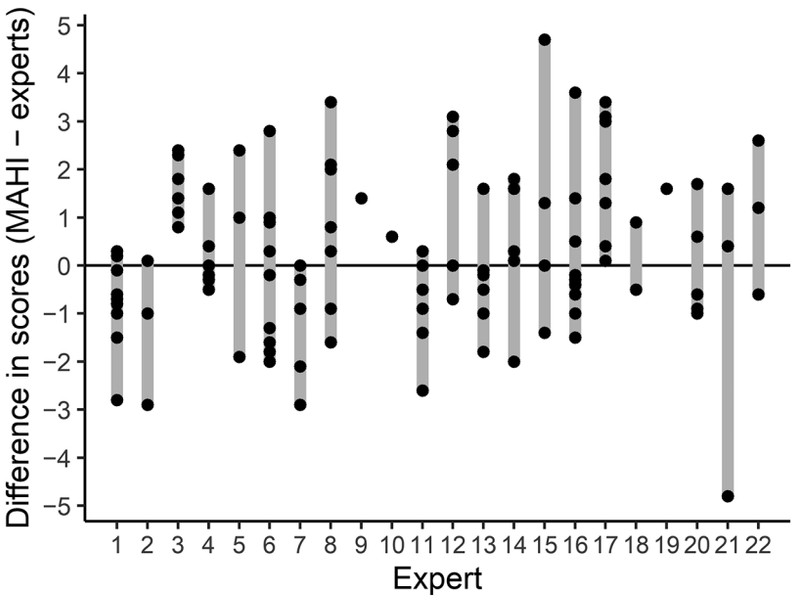

Figure 6. Differences between mussel assemblage health index (MAHI) scores and scores from 22 experts for 50 mussel assemblages in stream reaches across eastern North America. Black dots within a gray line represent multiple assemblages scored by the same expert. The horizontal line at 0 represents no difference between MAHI and expert scores.

(Haag 2012). For example, the Delaware River supports high mussel abundance and retains 100% of its historical species richness, resulting in a high MAHI score. However, only 7 species are native to that reach, and mussel assemblages are dominated by *Elliptio complanata*, resulting in low scores for current richness and *H′*. Omission of these 3 metrics avoids penalizing or rewarding assemblages based on factors unrelated to stream or assemblage health.

All previous mussel indices include metrics related to species tolerance or conservation status, but such metrics were not discriminatory in our dataset. This lack of discrimination was due, in part, to variation in the distribution of imperiled species among regions and habitat types. For several assemblages that received high MAHI scores (e.g., Delaware, Hudson, Neversink, Grand, South Thames, and Chippewa rivers), few or no imperiled species are native to those streams, or assemblages are dominated by nonimperiled *E. complanata*, both because of biogeographic or habitat factors.

The lack of discrimination by these metrics also illustrates the dangers of basing assemblage health assessments on untested assumptions about the relative tolerance of imperiled species. Copper Creek and the Tellico River both had a high proportion of individuals of imperiled species (threatened and endangered + proposed = 0.381 and 1.000, respectively) despite receiving low MAHI scores and uniformly low metric scores indicative of highly degraded assemblages. The assumption that imperiled mussel species are less tolerant than nonimperiled species has not been evaluated widely and is not supported by available information. For example, effective and lethal concentrations for acute and chronic effects of Cu and NH_3 did not differ consistently or substantially between imperiled and nonimperiled mussel species (Augspurger et al. 2003, Wang et al. 2007a, b). In another study, proportional occupancy loss (i.e., decrease in distributional range) of state-imperiled and nonimperiled mussel species over time in Illinois ranged from 0 to 1 for both groups, and the difference between the groups was small (Cao et al. 2017; mean proportional loss, imperiled = 0.81, nonimperiled = 0.62; Appendix S5). Metrics related to imperiled species could be useful for identifying streams of high conservation importance, but such metrics are not useful indicators of assemblage health across a broad area.

The 4 metrics retained in MAHI provide a comprehensive assessment of assemblages based on fundamental and largely nonredundant aspects of assemblage or population health. The metrics relative density and population growth both relate to the broad metric group abundance, but they depict different aspects of assemblage health. For example, an assemblage could be declining rapidly but still have relatively high density if the decline is in its early stages. Conversely, an assemblage with low density could be experiencing rapid population growth, which may indicate improvement in assemblage health and habitat conditions. Assemblage recruitment strength and proportion of historical richness both represent aspects of assemblage health that are largely independent of abundance, and they may reveal important features of assemblages. For example, without estimates of proportion of historical richness, the Minnesota River would receive a higher MAHI score that would not reflect the loss of nearly ½ of its fauna. Population growth is somewhat redundant to assemblage recruitment strength because rapidly growing populations are expected to have high recruitment. However, high recruitment might not result in high population growth if adult mortality is high, and by including both metrics, MAHI may provide evidence of such a scenario. Furthermore, because of its dependence on long-term monitoring data, population growth is the metric most likely to be unavailable to users of MAHI, and when unavailable, recruitment may provide a proxy for population growth. However, the ability to use CPUE data to estimate population growth is an important feature of MAHI that provides flexibility for users.

Considerations for using MAHI

Because the MAHI composite score is robust to missing metric values, useful information about assemblage health can be obtained even with incomplete data. There were no substantial differences among metrics in the effects on composite MAHI scores when each metric was omitted, which indicates that all 4 metrics have similar effects on the composite score. It is particularly noteworthy that omission of the population growth metric did not have an inordinate effect on the composite MAHI score, which is a desirable feature of MAHI because population growth often is unavailable until long-term monitoring has been conducted.

In addition to the composite MAHI score, more detailed information about assemblages can be obtained by examining individual metric scores, as described previously for the Minnesota and Elk rivers. Another example is the Little Tennessee River (after the appearance of *Corbicula*): although that stream retains most of its historical richness (see Results, Evaluation of MAHI performance), the absence of recruitment portends continued declines in abundance and species loss. Conversely, although most species have been lost from the North Fork Holston River and density is low, its intermediate recruitment score may indicate modest recovery of populations of remaining species. Raw metric values also provide useful information about assemblages. For example, the raw values for assemblage recruitment strength and population growth provide the actual magnitude of those measures, unlike the scaled index scores. Individual metric scores may provide valuable clues about the causes of mussel declines, which can inform conservation efforts. In addition, the variability among individual metrics may provide insights into assemblage health. CVs were low across metrics for most healthy assemblages, showing that all 4 fundamental aspects of those assemblages were uniformly healthy. In contrast, degraded assemblages showed higher variation across metric

scores, which may indicate that degraded assemblages were affected by a wide range of stressors that affected different aspects of assemblage health.

Information about all 4 MAHI metrics can be obtained from standard mussel surveys and other sources. However, there are considerations for each metric that inform their interpretation. First, obtaining historical estimates of species richness may be challenging for previously unsurveyed streams, but in many cases, reasonable inference can be made about the historical fauna (see Methods). Because MAHI is robust to missing metric values, proportion of historical richness could be omitted if necessary. However, we recommend making every effort to obtain reasonable estimates of historical species richness because this metric captures a fundamental aspect of assemblage health not reflected by other metrics.

Detecting changes in mussel abundance over time can be difficult because of high error inherent in many mussel sampling methods (Strayer and Smith 2003). Our method of estimating population growth purposely avoids statistical inference about differences in abundance over time, but sampling error could result in erroneous conclusions about population growth, particularly in cases where abundance estimates are available for only 2 time periods. However, only very large sampling error is likely to unduly affect the metric and composite scores, and longer time series of monitoring data will further reduce chances for erroneous conclusions about population growth.

Our shell length criterion for identifying recruits and calculating recruitment strength was meant to be applicable to as many species as possible and, being based on maximum observed size for the population, it accounts for variation in size among streams. However, the accuracy of this method varies among species and datasets. This warrants 3 cautionary notes about our recruitment metric. First, recruitment strength estimates will be most accurate for longer-lived species (more than ~15 y). Recruitment may be grossly underestimated for small species with rapid growth and a short lifespan (e.g., *Simpsonaias ambigua*, *Toxolasma parvum*). For example, *T. parvum* may reach maturity in its 1st y and live only 5 y (Haag 2012). Therefore, all individuals in a population will represent recent recruitment by our definition (<10 y old), but few or no individuals of <50% maximum size may be encountered, in which case recruitment strength will erroneously equal 0. For similar reasons, recruitment may also be underestimated for larger species with short life spans (e.g., *Potamilus fragilis*, *Utterbackia imbecillis*), but the bias will be less severe because of the greater size range of individuals in the population. Because these species rarely dominate mussel assemblages and our metric uses mean recruitment strength across all species, this bias will rarely have an inordinate effect on the metric or composite score. However, this potential bias should be considered in the rare cases when such species dominate an assemblage.

The 2nd cautionary note relates to use of mean species recruitment strength. This metric is not used by itself in MAHI, but it can be easily calculated, and our Excel application provides values for that metric and for each species. Estimates of recruitment strength for a species can be compared across years at a site, but they should not be compared among species to evaluate relative recruitment strength. This caution is necessary because fast-growing species that approach L_{max} quickly (i.e., higher values of the von Bertalanffy growth parameter K) will tend to exceed 50% of L_{max} sooner and have lower values of recruitment strength than slower-growing species with lower K. These types of within-assemblage comparisons are beyond the scope and purpose of the index and would require species-specific maximum length criteria developed from length-at-age data. Differences between assemblages in the proportion of fast- vs slow-growing species could bias MAHI recruitment scores, but such bias is unlikely to have inordinate effects on the composite score.

The final cautionary note about our recruitment metric pertains to datasets with very low sample sizes due to low mussel abundance. Recruitment strength estimates for such datasets may be either underestimated or inflated because of low sample sizes or impossible to calculate if no species have $n \geq 5$. Our method of correcting for nondetection of recruits addresses this issue to an extent, but it remains a concern of which users should be aware. This potential source of error will be an issue primarily in severely degraded assemblages, but other metrics unrelated to recruitment should more robustly reflect assemblage health in those streams.

MAHI was developed with data from streams in eastern North America, but it could be modified for application in streams in other parts of the world or in lakes. The metrics proportion of historical species and population growth could be applied unmodified in any water body. Our recruitment metric would require evaluation and perhaps modification. Of the 2 measures used to calculate assemblage recruitment strength, proportion of species with recruitment could be applied unmodified. The accuracy of our maximum length criterion for recruits used to calculate mean species recruitment strength would need to be evaluated based on length-at-age data for species in a new region of application. Relative density may be most challenging to modify for other areas. To our knowledge, relationships between mussel density and lake size or lake watershed area have not been evaluated, but it seems unlikely that those factors would be predictive of mussel density. If not, raw mussel density could be an informative metric if scaled within the context of observed density in lakes in the region of interest. In streams, it would be necessary to evaluate whether watershed area is predictive of mussel density. If not, raw density could be scaled as proposed for lakes. In addition, at least 1 special case may be particularly challenging for adapting MAHI and may affect its applicability in some geographic locations. In small streams in eastern and northwestern North America, Europe, and

elsewhere, *Margaritifera* spp. can achieve high densities (>300/m^2) that greatly exceed mussel densities typically seen in small streams (e.g., Johnson and Brown 1998, Geist 2010). The density metric of MAHI will not be applicable in such situations.

MAHI can be calculated using data readily obtained from standard mussel surveys, and the Excel application we developed can further streamline calculation of MAHI scores. MAHI effectively discriminated healthy and degraded assemblages from a diverse array of biogeographic regions and habitat types. It provided scores that generally reflected expert opinion, but it eliminates variation due to differences in individual perception and experience. The scaling of metric scores allows intuitive assessment of 4 fundamental aspects of assemblage health that can be used to diagnose specific symptoms of assemblage degradation. MAHI is free of untested assumptions about species sensitivity to anthropogenic factors, and it accounts for expected differences in mussel assemblages due to stream size, biogeographic region, and other intrinsic factors. MAHI will be a valuable tool for early detection of mussel declines, better understanding the causes of declines, and evaluating the effectiveness of conservation strategies and management actions.

ACKNOWLEDGEMENTS

Author contributions: WRH conceived the index and compiled survey data with input from AKB, TPJ, and SJP. WRH and AKB developed candidate metrics, and WRH, AKB, and TPJ analyzed the data. WRH wrote the manuscript with input from AKB, TPJ, and SJP.

We thank the following people for contributing survey and monitoring data, providing scores for streams, and providing constructive input or other assistance: David Aldridge, Carla Atkinson, Caitlin Carey, Janet Clayton, Mike Compton, Andrea Darracq, Mike Davis, Gerry Dinkins, Kristi Dobra, Karel Douda, Ani Escobar, Luke Etchison, Kevin Eliason, Brant Fisher, Juergen Geist, Lilagrace Haag, Bill Henley, Rachael Hoch, Dan Hornbach, Mark Hove, Julieann Jacobs, Paul Johnson, Jess Jones, Tim Lane, Manuel Lopes-Lima, Heather Malcom, Monte McGregor, Todd Morris, Kelly McNichols-O'Rourke, Bernard Sietman, David Smith, Ronaldo Sousa, David Strayer, Jason Wisniewski, and Kristin Womble.

This work was funded by the BAND Foundation and supported by American Rivers, the United States Forest Service Southern Research Station, Oak Ridge Institute for Science and Education, and the University of Kentucky Department of Forestry and Natural Resources.

LITERATURE CITED

Aldridge, D. C., Y. V. Bespalaya, I. N. Bolotov, K. Douda, J. Geist, W. R. Haag, M. W. Klunzinger, M. Lopes-Lima, M. C. Mlambo, N. Riccardi, R. Sousa, D. L. Strayer, S. H. Torres, C. C. Vaughn, T. Zając, A. Zieritz, and I. S. Ollard. 2022. Freshwater mussel conservation: A global horizon scan of emerging threats and opportunities. Global Change Biology 29:575–589.

Augspurger, T., A. E. Keller, M. C. Black, W. G. Cope, and F. J. Dwyer. 2003. Water quality guidance for protection of freshwater mussels (Unionidae) from ammonia exposure. Environmental Toxicology and Chemistry 22:2569–2575.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: Periphyton, benthic macroinvertebrates and fish. 2nd edition. EPA841-B-99–002. United States Environmental Protection Agency, Washington, DC. (Available from: https://www3.epa.gov/region1/npdes/merrimackstation/pdfs/ar/AR-1164.pdf, Accessed February 21, 2023)

Cai, Y., J. Xu, M. Zhang, J. Wang, and J. Heino. 2019. Different roles for geography, energy, and environment in determining three facets of freshwater molluscan beta diversity at broad spatial scales. Science of the Total Environment 659:451–462.

Cao, Y., K. Cummings, L. Hinz, S. A. Douglass, A. P. Stodola, and A. M. Holtrop. 2017. Reconstructing the natural distribution of individual unionid mussel species and species diversity in wadeable streams of Illinois, USA, with reference to stream bioassessment. Freshwater Science 36:669–682.

Carboni, J., D. Gatelli, R. Liska, and A. Saltelli. 2007. The role of sensitivity analysis in ecological modelling. Ecological Modelling 203:167–182.

Daniel, W. M., K. M. Brown, and M. D. Kaller. 2014. A tiered aquatic life unit bioassessmentmodel for Gulf of Mexico coastal streams. Fisheries Management and Ecology 21:491–502.

Dinno, A. 2017. *Dunn.test*: Dunn's test of multiple comparisons using rank sums. (Available from: https://CRAN.R-project.org/package=dunn.test)

Dunn, H. L., S. Zigler, and T. Newton. 2020. Mussel community assessment tool for the upper Mississippi River. Freshwater Mollusk Conservation and Biology 23:109–123.

Geist, J. 2010. Strategies for the conservation of endangered freshwater pearl mussels (*Margaritifera margaritifera* L.): A synthesis of conservation genetics and ecology. Hydrobiologia 644:69–88.

Greco, S., A. Ishizaka, M. Tasiou, and G. Torrisi. 2019. On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. Social Indicators Research 141:61–94.

Haag, W. R. 2009. Extreme longevity in freshwater mussels revisited: Sources of bias in age estimates derived from mark-recapture experiments. Freshwater Biology 54:1474–1486.

Haag, W. R. 2012. North American freshwater mussels: Natural history, ecology, and conservation. Cambridge University Press, New York, New York.

Haag, W. R., and A. L. Rypel. 2011. Growth and longevity in freshwater mussels: Evolutionary and conservation implications. Biological Reviews 86:225–247.

Haag, W. R., and M. L. Warren. 2010. Diversity, abundance, and size structure of bivalve assemblages in the Sipsey River, Alabama. Aquatic Conservation: Marine and Freshwater Ecosystems 20:655–667.

Haag, W. R., and J. D. Williams. 2014. Biodiversity on the brink: An assessment of conservation strategies for North American freshwater mussels. Hydrobiologia 735:45–60.

Hawkins, C. P. 2006. Quantifying biological integrity by taxonomic completeness: Its utility in regional and global assessments. Ecological Applications 16:1277–1294.

Henley, W. F., J. J. Schmerfeld, S. A. Budischak, C. M. Hall, R. J. Neves, S. Ciparis, and J. W. Jones. 2016. Freshwater mussel

(Unionidae) abundance and diversity upstream and downstream of a Superfund site on the North Fork Holston River, Saltville, Virginia. Journal of Shellfish Research 35:875–883.

Herman, B. D., T. W. Slack, and T. S. Swannack. 2021. Developing conceptual models for assessing benefits and impacts of USACE activities on freshwater mussel communities. Technical Note ERDC/TN EMRRP-EBA-25. United States Army Corps of Engineers, Engineer Research and Development Center, Ecosystem Management and Restoration Research Program, Vicksburg, Mississippi. (Available from: https://erdc-library.erdc.dren.mil/jspui/handle/11681/42161, Accessed November 28, 2022)

Hill, B. H., A. T. Herlihy, P. R. Kaufmann, R. J. Stevenson, F. H. McCormick, and C. B. Johnson. 2000. Use of periphyton assemblage data as an index of biotic integrity. Journal of the North American Benthological Society 19:50–67.

Hoggarth, M. A., and M. D. Grumney. 2013. A report on a mussel survey of Big Walnut Creek from Hoover Dam to its mouth, Franklin and Pickaway counties, OH. Unpublished report, Otterbein University, Ohio. (Available from: https://www.yumpu.com/en/document/view/51684654/mussel-report-friends-of-big-walnut-creek, Accessed November 28, 2022)

Hornbach, D. J., D. C. Allen, M. C. Hove, and K. R. MacGregor. 2018. Long-term decline of native freshwater mussel assemblages in a federally protected river. Freshwater Biology 63:243–263.

Hornbach, D. J., M. C. Hove, K. R. MacGregor, J. L. Kozarek, B. E. Sietman, and M. Davis. 2019. A comparison of freshwater mussel assemblages along a land-use gradient in Minnesota. Aquatic Conservation: Marine and Freshwater Ecosystems 29:1826–1838.

Illari, M., A. T. Souza, L. Amorim, and R. Sousa. 2019. Decay and persistence of empty bivalve shells in a temperate riverine system. Science of the Total Environment 683:185–192.

Iooss, B., S. Da Veiga, A. Janon, G. Pujol, B. Broto, K. Boumhaout, T. Delage, R. El Amri, J. Fruth, L. Gilquin, J. Guillaume, M. Herin, M. Il Idrissi, L. Le Gratiet, P. Lemaitre, A. Marrel, A. Meynaoui, B. L. Nelson, F. Monari, R. Oomen, O. Rakovec, B. Ramos, O. Roustant, G. Sarazin, E. Song, J. Staum, R. Sueur, T. Touati, V. Verges, and F. Weber. 2022. *sensitivity*: Global sensitivity analysis of model outputs. (Available from: https://cran.r-project.org/src/contrib/Archive/sensitivity/sensitivity_1.28.0.tar.gz)

Jarvis, J. D. 2011. Water quality in the upper Little Tennessee River and its potential effects on the Appalachian Elktoe mussel (*Alasmidonta raveneliana*). MS Thesis. Western Carolina University, Cullowhee, North Carolina.

Johnson, P. D., and K. M. Brown. 1998. Intraspecific life history variation in the threatened Louisiana pearlshell mussel, *Margartifera hembeli*. Freshwater Biology 40:317–329.

Jones, J., T. Lane, B. Ostby, B. Beaty, S. Ahlstedt, R. Butler, D. Hubbs, and C. Walker. 2018. Collapse of the Pendleton Island mussel fauna in the Clinch River, Virginia: Setting baseline conditions to guide recovery and monitoring. Freshwater Mollusk Biology and Conservation 21:36–56.

Kerans, B. L., and J. R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. Ecological Applications 4:768–785.

Lenat, D. R. 1993. A biotic index for the southeastern United States: Derivation and list of tolerance values, with criteria for assigning water-quality ratings. Journal of the North American Benthological Society 12:279–290.

Lydy, M. J., C. G. Crawford, and J. W. Frey. 2000. A comparison of selected diversity, similarity, and biotic indices for detecting changes in benthic-invertebrate community structure and stream quality. Archives of Environmental Contamination and Toxicology 39:469–479.

Ma, M., X. Wang, and L. Cui. 2020. Development of a periphyton-based index for biological integrity for lakes and reservoirs in Baiyangdian Watershed, China. Aquatic Ecosystem Health and Management 23:288–301.

Metcalfe-Smith, J. L., S. K. Staton, G. L. Mackie, and N. M. Lane. 1998. Changes in the biodiversity of freshwater mussels in the Canadian waters of the lower Great Lakes drainage basin over the past 140 years. Journal of Great Lakes Research 24:845–858.

Miller, A. C., and B. S. Payne. 1988. The need for quantitative sampling to characterize size demography and density of freshwater mussel communities. American Malacological Bulletin 6: 49–54.

Morris, T. J., and L. D. Corkum. 1996. Assemblage structure of freshwater mussels (Bivalvia:Unionidae) in rivers with grassy and forested riparian zones. Journal of the North American Benthological Society 15:576–586.

Payne, B. S., and A. C. Miller. 2000. Recruitment of *Fusconaia ebena* (Bivalvia:Unionidae) in relation to discharge of the lower Ohio River. American Midland Naturalist 144:328–341.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, EISPACK authors, S. Heisterkamp, B. Van Willigen, J. Ranke, and R Core Team. 2021. *nlme*: Linear and nonlinear mixed effects models. (Available from: https://CRAN.R-project.org/package = nlme)

Richard, J. C., E. Leis, C. D. Dunn, R. Agbalog, D. Waller, S. Knowles, J. Putnam, and T. L. Goldberg. 2020. Mass mortality in freshwater mussels (*Actinonaias pectorosa*) in the Clinch River, USA, linked to a novel densovirus. Scientific Reports 10:14498.

Saisana, M., A. Saltelli, and S. Tarantola. 2005. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Journal of the Royal Statistical Society Series A: Statistics in Society 168:307–323.

Sietman, B. E. 2007. Freshwater mussels of the Minnesota River valley counties. Pages 5.32–5.43 *in* Minnesota county biological survey. Native plant communities and rare species of the Minnesota River valley counties. Biological Report 89. Minnesota Department of Natural Resources, St Paul, Minnesota. (Available from: https://files.dnr.state.mn.us/eco/mcbs/mn_river_report.pdf, Accessed February 14, 2023)

Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquino. 2008. A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 27:878–891.

Stoeckl, K., M. Denis, and J. Geist. 2020. Conservation status of two endangered freshwater mussel species in Bavaria, Germany: Habitat quality, threats, and implications for conservation management. Aquatic Conservation: Marine and Freshwater Ecosystems 30:647–661.

Strayer, D. L., S. Claypool, and S. J. Sprague. 1997. Assessing unionid populations with quadrats and timed searches. Pages 163–199 *in* K. S. Cummings, A. C. Buchanan, C. A. Mayer, and T. J. Naimo (editors). Conservation and management of freshwater mussels II.

Initiatives for the future. Upper Mississippi River Conservation Committee, Rock Island, Illinois.

Strayer, D. L., and H. M. Malcom. 2007a. Effects of zebra mussels (*Dreissena polymorpha*) on native bivalves: The beginning of the end or the end of the beginning? Journal of the North American Benthological Society 26:111–122.

Strayer, D. L., and H. M. Malcom. 2007b. Shell decay rates of native and alien freshwater bivalves and implications for habitat engineering. Freshwater Biology 52:1611–1617.

Strayer, D. L., and D. R. Smith. 2003. A guide to sampling freshwater mussel populations. American Fisheries Society Monograph 8, Bethesda, Maryland.

Szafoni, R. E. 2002. Freshwater mussel classification index: Identifying mussel assemblages of state-wide significance. Ellipsaria 4:20–21.

USFWS (United States Fish and Wildlife Service). 2012. Improving a WV river's health by bringing back endangered mussels. United States Fish and Wildlife Service, Northeast Region. (Available from: https://usfwsnortheast.wordpress.com/2012/10/23/improving-a-wv-rivers-health-by-bringing-back-endangered-mussels/, Accessed February 14, 2023)

Vaughn, C. C., C. M. Taylor, and K. J. Eberhard. 1997. A comparison of the effectiveness of timed searches vs. quadrat sampling in mussel surveys. Pages 157–162 *in* K. S. Cummings, A. C. Buchanan, C. A. Mayer, and T. J. Naimo (editors). Conservation and management of freshwater mussels II. Initiatives for the future. Upper Mississippi River Conservation Committee, Rock Island, Illinois.

Villella, R. F., D. R. Smith, and D. P. Lemarie. 2004. Estimating survival and recruitment in a freshwater mussel population using mark-recapture techniques. American Midland Naturalist 151:114–133.

Vowell, J. L. 2001. Using stream bioassessment to monitor best management practice effectiveness. Forest Ecology and Management 143:237–244.

Wang, L., and J. Lyons. 2003. Fish and benthic macroinvertebrates as indicators of stream degradation in urbanizing watersheds. Pages 227–249 in T. P. Simon (editor). Biological response signatures: Indicator patterns using aquatic communities. CRC Press, Boca Raton, Florida.

Wang, N., C. G. Ingersoll, I. E. Greer, D. K. Hardesty, C. D. Ivey, J. L. Kunz, W. G. Brumbaugh, F. J. Dwyer, A. D. Roberts, T. Augspurger, C. M. Kane, R. J. Neves, and M. C. Barnhart. 2007a. Chronic toxicity of copper and ammonia to juvenile freshwater mussels (Unionidae). Environmental Toxicology and Chemistry 26:2048–2056.

Wang, N., C. G. Ingersoll, D. K. Hardesty, C. D. Ivey, J. L. Kunz, T. W. May, F. J. Dwyer, A. D. Roberts, T. Augspurger, C. M. Kane, R. J. Neves, and M. C. Barnhart. 2007b. Acute toxicity of copper, ammonia, and chloride to glochidia and juveniles of freshwater mussels (Unionidae). Environmental Toxicology and Chemistry 26:2036–2047.