Freshwater Biology (2001) **46,** 109–123

APPLIED ISSUES

Assessing the performance of volunteers in monitoring streams

LESKA S. FORE,* KIT PAULSEN† and KATE O'LAUGHLIN‡
*Statistical Design, 136 NW 40th St., Seattle, WA 98107, U.S.A.
†City of Bellevue, P.O. Box 90012, Bellevue, WA 98009, U.S.A.
‡King County Land and Water Resources Division, 700 Fifth Ave., Suite 2200, Seattle, WA 98104, U.S.A.

SUMMARY

1. Citizens are concerned about the quality of water resources and many participate in monitoring activities, though doubts remain about the quality of the data volunteers collect. We trained volunteers to collect benthic macroinvertebrates using professional protocols. Of the seven stream sites sampled by volunteer crews, six sites were also sampled by professional crews.
2. In the laboratory, volunteers used morphological features to identify as many different taxa as possible within the major insect orders; their identification was approximately to family. Volunteers calculated five metrics: total taxon richness, richness of three key groups (Ephemeroptera, Plecoptera and Trichoptera), and percentage dominance of the three most abundant taxa. All metrics were strongly correlated with (a) the percentage of urbanized area in the catchment and (b) the metrics derived from a more complete taxonomic identification by a professional scientist. Taxon richness metrics declined with urban development, while percent dominance increased.
3. An overall summary multimetric index was used to compare the field and laboratory procedures of volunteers and professionals. Using an ANOVA model, we detected no significant difference between field samples collected by volunteers and professionals. The variance of index values associated with differences between crews was zero. The ability of the index to detect significant differences among sites (statistical power) improved by only 13% for assessments based on professional laboratory identification instead of volunteer laboratory identification.
4. Citizen volunteers, when properly trained, can collect reliable data and make stream assessments that are comparable to those made by professionals. Data collected by volunteers can supplement information used by government agencies to manage and protect rivers and streams.

Keywords: citizens, macroinvertebrates, power analysis, streams

Introduction

Public interest in water resources has increased dramatically in the past decade (Kerr *et al*., 1994; Firehock & West, 1995; Penrose & Call, 1995; Lathrop & Markowitz, 1995; Karr, Allan & Benke, 1999). A US intergovernmental task force on monitoring, for ex-

Correspondence: Leska S. Fore, Statistical Design, 136 NW 40th St., Seattle, WA 98107, U.S.A.
E-mail: leska@seanet.com

ample, cited more than 500 volunteer groups involved in monitoring water quality, urging that their efforts be integrated into government programmes (USGS, 1995). Although many programmes do include volunteer data in their official reports (Kerr *et al*., 1994; Mattson *et al*., 1994; Mayio, 1994; Lathrop & Markowitz, 1995; Beauchene, 1997; Carlson, 1997), volunteer involvement is not well documented in published literature (but see Reynoldson, Hampel & Martin, 1986; Dvornich, Tudor & Grue, 1995; Penrose & Call, 1995). Results from volunteer programmes are often easier to find on the Internet. The US Environmental Protection Agency (EPA) and the state of Kentucky, for example, use the Internet to connect volunteers with agencies and other resources (Kentucky Water Watch, 1998; US EPA, 1997).

In the state of Washington (U.S.A.), 11000 volunteers are involved in surface water monitoring and protection. Yet in 1996 reliable information existed for only 4% of Washington's surface waters (Washington DOE, 1997). Training volunteers to collect missing data would seem an obvious way to fill the gaps, but many managers and scientists question the quality and reliability of data from volunteers. We therefore sought to train volunteers to assess the impact of urbanization on freshwater resources and to test the reliability of volunteer efforts by comparing them with assessments made by professional scientists.

This project used a multimetric index, the benthic index of biological integrity (B-IBI), to assess the condition of invertebrate assemblages collected from streams (Kerans and Karr, 1994; Karr, 1999). Assessments based on multimetric indices are increasingly used in the U.S.A. to monitor and manage surface waters under the Clean Water Act (Ransel, 1995; Southerland & Stribling, 1995; Davis *et al*., 1996; Karr & Chu, 1999). Multimetric indexes typically combine selected biological attributes, called metrics (e.g. the number of Ephemeroptera taxa, the relative abundance of predators, and the relative abundance of tolerant organisms) to evaluate the impact of human activities on aquatic systems (Karr *et al*., 1986; Davis & Simon, 1995; Barbour *et al*., 1998). The B-IBI developed for the northwestern U.S.A. and Japan includes 10 metrics selected for their association with urbanization, timber harvest, recreation and pollution (Fore, Karr & Wisseman, 1996; Rossano, 1996; May *et al*., 1997; Karr, 1998, 1999).

Monitoring programmes in the U.K. and Australia base their assessments on a different analytical approach that uses multivariate statistical models to predict the expected invertebrate assemblage (Wright, Furse & Armitage, 1993; Parsons & Norris, 1996; Marchant *et al*., 1997; Hawkins *et al*., 2000). Although analytical methods differ, our comparisons of volunteer field and laboratory performance are relevant to those programmes because invertebrate collection methods and laboratory protocols are similar to those used in the U.S.A.

Methods

The purpose of this study was: (1) to determine whether biological information collected by volunteers was related to changes associated with urbanization and development and (2) to compare the precision of assessments made by volunteers with those made by professionals. We first tested whether volunteer assessments of biological condition correlated with an independent measure of human disturbance (the percentage of catchment area covered by urban development or by forest). We ranked stream sites according to these percentages and measured the biological condition of each site using B-IBI.

To evaluate the precision of volunteer data, we compared volunteer and professional results at two distinct stages of the assessment process: sample collection in the field and sample processing in the laboratory. Laboratory processing included sorting invertebrates from leaves, sticks and sediment; identifying them; and calculating biological metrics. Volunteers and professionals followed identical methods in the field. In the laboratory, however, volunteer and professional methods differed because volunteers identified invertebrates approximately to family, whereas, for most insects, the professionals took their taxonomic analysis to genus or species.

To compare the quality of volunteer and professional sampling in the field, we held laboratory methods constant by sending field samples of both groups to the same professional laboratory for taxonomic identification. Volunteers first sorted and identified invertebrates in their own samples; afterward, the samples were sent to the professional laboratory for more complete taxonomic identification. To compare volunteer and professional laboratory methods, both volunteers and professionals analysed the same set of

volunteer-collected field samples. Thus, we held field sampling constant in order to compare laboratory methods. The last step of our analysis estimated the statistical precision of volunteer and professional indices. Precision was compared in terms of the number of categories of biological condition that volunteers and professionals could detect with their respective assessments.

Four data sets were derived from the two sample sets collected at each site by the volunteer and professional crews. For convenience, we refer to each data set with a two-letter code, volunteer (V) or professional (P), where the code's first letter refers to the field method and the second to the laboratory method. The volunteer field samples were analysed in three ways: first, by volunteers in the laboratory (VV), second by a professional taxonomic laboratory (VP), and third, by the present authors to regroup genera and species identified by the professional laboratory into families (VP.family). Professional field samples were analysed once by the professional laboratory (PP).

These combinations gave us four data sets: The VV data set included five metrics calculated from volunteer field sampling and volunteer laboratory analysis. The VP data set included 10 metrics calculated from volunteer field samples and professional laboratory identification. Like the VV data, the VP.family data set included five metrics, which were derived from a modification of the VP data in which genera and species were combined to the taxonomic level of family. The PP data set included 10 metrics calculated from professional field samples and professional laboratory identification.

Study site descriptions and 'a priori' ranking

We selected seven stream sites in the Seattle area (King County) of the Puget Sound basin, Washington, U.S.A. We chose sites to represent a gradient of human influence, from minimally disturbed to extremely degraded. All seven catchments had been logged extensively. Present land uses ranged from scattered dwellings and farms (minimal disturbance) to dense urban development. Three of the catchments (Kelsey, Thornton and Pipers) were completely developed; their only green spaces were located in city parks. Two catchments (Evans and Soos) were mostly developed as suburbs and had little remaining native vegetation. Two catchments (Rock and Holder) were still heavily forested, though not with the original, old growth, vegetation. Human population density and development were relatively low in these catchments but increasing rapidly.

We ranked stream sites according to the percentage of the catchment that was developed (covered with buildings or roads) before evaluating the biological condition of the invertebrate assemblage. Using GIS software, we estimated land cover from 1995 satellite imagery. Catchment areas ranged from 4 to 38 km^2. We categorized most of the area in the catchments ($> 80\%$) as either forested or developed; small percentages of bare ground, bare rock and grass made up the rest. Development and forest cover were inversely correlated; our sites were ranked the same by both measures (Fig. 1). We did not distinguish among types of forest cover (e.g. among seral stages or tree type) or among levels of development (e.g. high or low density); neither did we calculate the area impervious to water, a common measure of development (Booth & Jackson, 1997).

The area of developed land only roughly approximates human impact on a stream or river; overall human influence depends on the location, intensity and type of human activity. For this study, we tried to sample sites that were typical of each stream; we generally avoided bridges, culverts, mines, or other sites of human activity not typical for the catchment. For Holder Creek, we were restricted to sampling just upstream (< 50 m) of a four-lane bridge because other access points were unsafe or inaccessible.

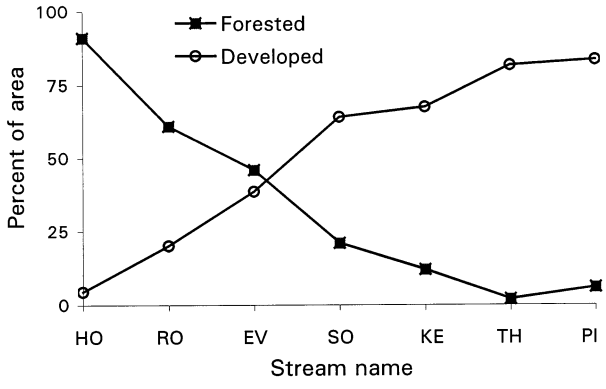

Fig. 1 Forest is replaced by urban development across these seven Puget Sound lowland catchments. Catchment area was calculated upstream of sample sites. (HO, Holder; RO, Rock; EV, Evans; SO, Soos; KE, Kelsey; TH, Thornton; and PI, Pipers).

Therefore, although Rock Creek had a slightly greater fraction of developed land in its catchment, we ranked Holder as more disturbed because of the bridge and evidence of construction at our sample site. Furthermore, our Rock Creek site had large conifers in the riparian area and large woody debris, cobble and boulders in the stream. In contrast, our Holder Creek site had buildings near the stream, brush and second-growth trees in the riparian area, and small cobbles and fine sediment in the stream.

Field methods

Six of the seven sites sampled by volunteers were also sampled by professional biologists using exactly the same sampling protocol and equipment. Pipers Creek was sampled only by volunteers. Both groups used a Surber sampler (0.3×0.3 m) with a 500 μm mesh net to collect benthic macroinvertebrates during August–October 1997. At each stream site, three Surber samples were collected 3–6 m apart near the centre of the channel. Each Surber sample was preserved separately. Samples were collected from riffles whenever possible, but glides were substituted at more degraded sites.

Volunteers and professionals collected samples within approximately 1 month of each other. The second sampling event at each site took place upstream of the first, to avoid any disturbance effects caused by the first sampling, except at Evans Creek, where the samples came from the same location. At this site, there were two problems. First, rather than simply disturbing the substratum within the sample frame, the volunteer crew shovelled sediment and cobble into the Surber sampler from the area within and surrounding the frame; this yielded an unusually large sample with greater than 6500 individuals. Second, the professional crew collected their sample at exactly the same location where the volunteers had previously disturbed the substratum and surrounding banks.

Laboratory methods

Field methods were the same for both groups whereas laboratory methods differed, primarily because professionals were more skilled in taxonomic identification. In the laboratory, both groups sorted as many invertebrates as possible from the sample detritus. For this step, volunteers used hand lenses, whereas the professionals used dissecting microscopes. For both groups, the goal was to identify all the invertebrates from each Surber sample; no subsampling was used.

After sorting, volunteers identified invertebrates to taxonomic order using pictorial (rather than dichotomous) keys. They learned which morphological features distinguish taxa within an order—gill shape and placement in Ephemeroptera, for example. Using dissecting microscopes, they worked in pairs to determine the total number of distinct taxa in each sample and the number of distinct taxa within the three insect orders Ephemeroptera, Plecoptera and Trichoptera. Volunteers also recorded the number of animals in each morphological group. Each of the 21 stream samples collected by volunteers (seven sites $\times$ three Surbers) was processed by one pair of volunteers.

The volunteers identified invertebrates approximately to family (though they did not necessarily learn the Latin names). In contrast, professional identification went to genus or species for most insects, genus for chironomids, and order or higher for non-insects (Plotnikoff & White, 1996).

Calculating B-IBI

B-IBI is composed of 10 metrics representing multiple levels of biological organization such as taxonomic composition and feeding ecology (Table 1). Because each metric has its own range of potential values (e.g. percentages for relative abundance, number of taxa for taxon richness), each metric is given a score before it is incorporated into B-IBI. A score of 5 indicates values expected in sites with minimal human disturbance; a score of 3, moderate human disturbance; and a score of 1, severe degradation. The sum of metric scores at a site is the final index. Relatively undisturbed sites thus have high B-IBIs, degraded sites low B-IBIs.

Volunteers calculated five of the 10 B-IBI metrics during their laboratory time, whereas we calculated all 10 B-IBI metrics from the data analysed by the professional laboratory. Volunteers calculated total taxon richness; taxon richness of Ephemeroptera (mayflies), Plecoptera (stoneflies) and Trichoptera (caddisflies); and percentage dominance (the number of animals in the most abundant taxon divided by the

Table 1 Biological metrics for invertebrates (Karr, 1998), response to human disturbance, and scoring criteria used to integrate metrics into a multimetric B-IBI†

Metric	Response	Scoring criteria		
		1	3	5
Taxon richness and composition				
Total number of taxa*	Decrease	[0, 20)	[20, 40]	>40
Number of Ephemeroptera taxa*	Decrease	[0, 4]	(4, 8]	>8
Number of Plecoptera taxa*	Decrease	[0, 3]	(3, 7]	>7
Number of Trichoptera taxa*	Decrease	[0, 5)	[5, 10)	≥10
Number of long-lived taxa	Decrease	[0, 2]	(2, 4]	>4
Tolerance				
Number of intolerant taxa	Decrease	[0, 2]	(2, 3]	>3
% Tolerant individuals	Increase	≥50	(19, 50)	[0, 19]
Feeding ecology				
% Predator individuals	Decrease	[0, 10)	[10, 20)	≥20
Number of clinger taxa	Decrease	[0, 10]	(10, 20]	>20
Population attributes				
% Dominance (three taxa)*	Increase	≥75	[50, 75)	[0, 50)

† Metrics used by volunteers (VV) and for family-level index (VP.family) are marked by an asterisk. Square brackets indicate a closed interval and that the value next to the bracket is included in the range; round brackets indicate an open interval and that the value is not included in the range. For example, for a sample with a total of 20 taxa, the metric score would be 3.

Table 2 Scoring criteria for volunteer (VV) and family level (VP.family) metrics*

Metric	Scoring criteria				
	1	2	3	4	5
Total number of taxa	[0, 7]	(7, 13]	(13, 19]	(19, 25]	>25
Number of Ephemeroptera taxa	[0, 1]	(1, 2]	(2, 4]	(4, 6]	>6
Number of Plecoptera taxa	0	(0, 1]	(1, 2]	(2, 3]	>3
Number of Trichoptera taxa	0	(0, 2]	(2, 4]	(4, 6]	>6
% Dominance (one taxon)	>85	[70, 85]	[55, 70)	[40, 55)	<40

* Square brackets indicate a closed interval and that the value next to the bracket is included in the range; round brackets indicate an open interval and that the value is not included in the range.

total number of animals collected in the sample). For data derived from professional taxonomic identification, we calculated percentage dominance for the three most abundant taxa, instead of just one, because the most abundant taxon identified by volunteers was typically a family, which could include multiple genera or species. We calculated five additional metrics from professionally identified data; these were related to life history (taxon richness of long-lived or semi-voltine taxa), habit (taxon richness of clingers), tolerance to human disturbance (percentage tolerant individuals and richness of intolerant taxa), and feeding group (percentage of individuals that were predators).

We defined metric scoring criteria using data from our seven sites and 15 additional sites in western Washington that were evenly distributed across a gradient of human influence, from undisturbed to intensely urban (unpublished data). Because the 21 sites were evenly distributed across this human influence gradient, we divided the range of values for each metric approximately into thirds to give three scoring categories for professionally identified data (see Table 1). B-IBI could range from 10 (poor) to 50 (excellent).

We calculated a five-metric index for the VV and VP.family data sets. For these data, our metric scoring criteria reflected the lower taxonomic resolution

(Table 2). Because there were only five metrics in this index, variability of index values at a site increased dramatically if several metric values happened to fall near a break point in the scoring criteria. For a 10-metric index this is less of a concern due to the averaging effect of more metrics. To compensate for this difference, we increased the number of scoring categories from three to five, greatly reducing index variability. (The same increase in the number of scoring categories had no effect on the variability of the 10-metric B-IBI, so we retained the typical three-category scoring.) Values for the five-metric index ranged from 5 (poor) to 25 (excellent).

Data analysis

We evaluated the quality of volunteer assessments on the basis of four comparisons using our four VV, VP, VP.family and PP data sets. First, we compared volunteer assessments of stream condition with an independent measure of human disturbance. Second, we compared volunteer with professional field efforts using B-IBI values derived from the same level of professional laboratory analysis. Third, we compared volunteer with professional laboratory methods by comparing metric values obtained by volunteers and professionals. Fourth, we compared the statistical precision of the five-metric volunteer index with the professional B-IBI.

We first tested whether volunteer assessments were associated with human disturbance by correlating the five VV metrics with a ranking of our stream sites based on intensity of human land use (Spearman's ρ, $n = 7$ sites). Second, we compared B-IBI values from volunteer (VP) and professional (PP) field data. Because both groups used the same field protocol and samples were sent to the same professional laboratory, we expected no differences. To test this hypothesis, we used Pearsons's r to compare B-IBI scores for the six sites sampled by both crews. We also used ANOVA to partition variance according to its possible sources: differences among sites (human disturbance), differences among crews (volunteer or professional), interaction of site and crew, and error (Hicks, 1982; Wiley, Kohler & Seelbach, 1997). If there were no difference in field sampling, we would expect the crew effect to equal zero. We used parametric models for these statistical tests because multimetric indexes satisfy the required assumptions (Fore, Karr & Conquest, 1994).

Third, we compared metric values based on the volunteers' morphological grouping with metric values based on professional identification to genus or species. We correlated the five VV metrics with their counterparts from the VP data set. (This comparison used only the volunteer field samples in order to limit the source of any differences to the laboratory methods.) Non-parametric testing (Spearman's ρ, $n = 6$) avoided any concerns about the metrics' underlying distributions.

For our final comparison, we estimated the relative statistical precision of the volunteer and family-level five-metric indices and the professional B-IBI. We translated statistical precision into the number of distinct categories of biological condition that each index could detect. In order to calculate precision, replicate samples are necessary to estimate the variance associated with repeat measures of the target variable, in our case, index values at a stream site. We had two ways of estimating index variance at a site, or measurement error. We could use index values calculated for each of the three repeat Surbers or we could average metric values from the three Surbers and use the two repeat visits by the volunteer and professional crews. In fact, we used both methods. Using the first approach, we compared the relative precision of volunteer (VV) and professional (VP and VP.family) laboratory methods. After we determined that field sampling by volunteers did not differ from professionals, we used the second approach to estimate the precision of the protocol actually used by King County.

The full professional protocol specifies that the metric values calculated from the three Surbers be averaged and a single B-IBI value reported for each sampled site. Averaging repeat samples is designed to reduce variability due to microhabitat differences (Kerans, Karr, & Ahlstedt, 1992). The estimates of precision we used to compare volunteer and professional laboratory methods may be somewhat conservative because metric values for each Surber were not averaged but were kept separate to estimate variance. Because we were interested in comparing precision, its possible underestimate was not a concern for this comparison.

The number of categories was based on the calculation of the minimum detectable difference (MDD) for a two-sample t-test (Zar, 1984, p. 135). Note that MDD varies depending on the statistical model cho-

sen and the questions being asked (Thomas, 1997). By setting $\alpha = 0.05$ and $\beta = 0.20$ (Peterman, 1990) for a two-sided test, we were asking: How large a difference between index values have we an 80% chance of detecting with a P-value < 0.05? We divided the possible range of each index (20 for the five-metric index and 40 for the 10-metric B-IBI) by MDD to obtain the number of distinct categories of biological condition that each index could detect.

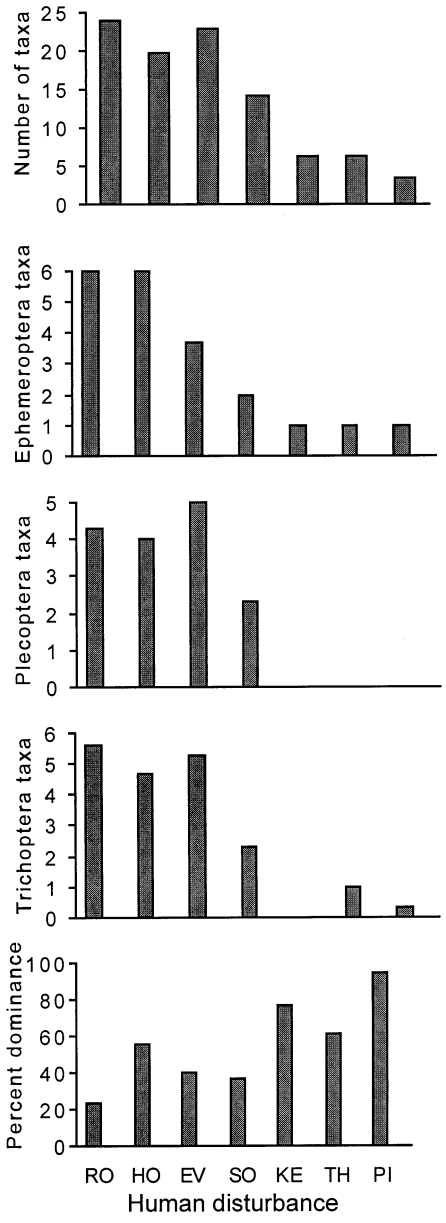

Fig. 2 Volunteer measurements (VV) of five metrics (total number of taxa; number of Ephemeroptera, Plecoptera and Trichoptera taxa; and percent dominance) were significantly correlated with human disturbance in the catchment. For statistics, see Table 3.

Table 3 Correlation (Spearman's ρ) of volunteer metrics (VV) with (1) human disturbance and (2) metrics based on professional taxonomic identification (VP)

	Correlation of VV metrics with	
Metric name	Disturbance	VP metrics
Total number of taxa	0.95**	0.92**
Number of Ephemeroptera taxa	0.95**	0.99**
Number of Plecoptera taxa	0.85*	0.97**
Number of Trichoptera taxa	0.86*	0.96**
% Dominance (1 taxon)	−0.82*	0.96**

* $P < 0.05$; ** $P \ll 0.01$.

Results

Volunteer assessments were significantly correlated both with human disturbance and professional assessments. Results from volunteer and professional field samples differed very little. Volunteer laboratory methods were less precise than professional methods because of volunteers' lower taxonomic resolution, yet the results of volunteer and professional laboratory analyses were highly correlated.

All five VV metrics showed a strong correlation with human disturbance (Fig. 2, Table 3). Total taxa and Ephemeroptera taxon richness declined steadily as forested land was replaced by development. Plecoptera taxon richness also declined with disturbance; stoneflies were not found at the most disturbed sites. Volunteers found fewer Trichoptera taxa as disturbance increased. Percentage dominance increased with human disturbance as predicted. The five-metric index calculated from VV data declined steadily with disturbance (Fig. 3).

We detected no consistent differences in B-IBI values calculated from volunteer and professional field samples, and both volunteer and professional indices were significantly correlated with human disturbance. Indices calculated from VP and PP data sets were significantly correlated (Fig. 4) and were not consistently higher or lower than one another. B-IBI values differed by less than five for all sites except Evans Creek, where the VP index exceeded the PP

index by 12. Recall that Evans was the only site where the professional crew sampled the identical spot sampled earlier by volunteers. We suspect that the removal of 6500 individuals from that spot by volunteers may have contributed to the lower B-IBI for the professional sample.

Differences between crews (professional versus volunteer) did not contribute to the variability of B-IBI. We calculated two ANOVA models because of the sampling problems at Evans Creek. One model included Evans Creek and the other did not. For both models, most of the variation (80% and 90%) in B-IBI values was attributed to differences among sites (human disturbance). Both models attributed 0% of the variance to crew differences and 10% to error. (The error term represents differences among Surber samples at a single site associated with small habitat features or time of sampling.) The models differed in the percentage of variation attributed to the interaction term, which represents an effect in addition to site and crew effects. Without Evans Creek, the interaction term went from 10% to 0% without changing the error term. This result further supports the explanation that the professional sample was biased for this particular site by the previous volunteer visit.

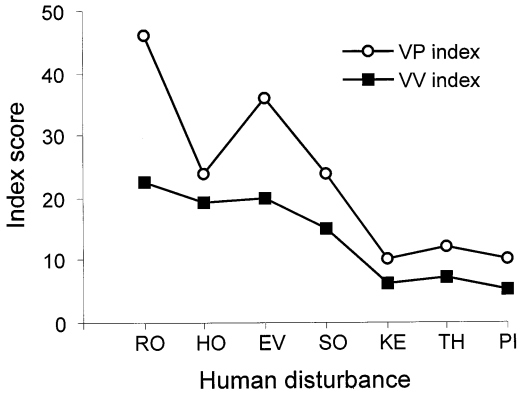

Fig. 3 The volunteer (VV) five-metric index and the professional (VP) benthic index of biological integrity (B-IBI) both declined as human disturbance increased. Indices were correlated with each other (Spearman's $\rho = 0.98$, $P \ll 0.01$; $n = 7$) and with disturbance ($\rho = 0.93$, $P < 0.01$; $\rho = 0.87$, $P < 0.05$).

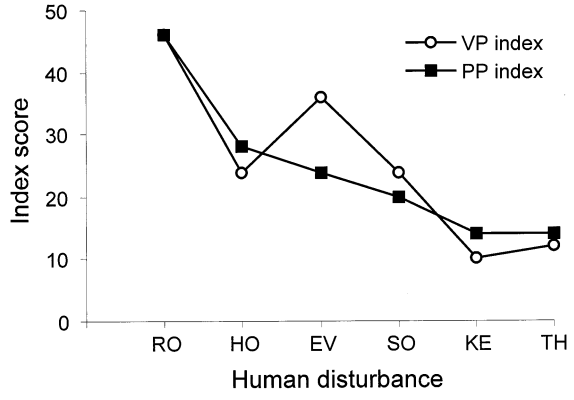

Fig. 4 B-IBI based on volunteer (VP) and professional (PP) field data with professional laboratory analysis both declined as human disturbance increased. The indices were highly correlated with each other (Pearson's $r = 0.90$, $P < 0.01$; $n = 6$) and with disturbance (Spearman's $\rho = 0.84$, $P < 0.05$; $\rho = 0.99$, $P \ll 0.01$).

Although the laboratory methods were not identical, metrics and indices calculated by volunteers (VV) and professionals (VP) were highly correlated (Fig. 5). Even though volunteers identified many fewer taxa than the professionals, the proportion of taxa they identified was relatively constant, and metric values of both groups tended to rank sites in a similar order. Volunteer estimates of taxon richness were lower for two reasons: (1) they missed small invertebrates while sorting, and (2) they typically did not recognize genus- or species-level differences. They missed small Plecoptera (Nemouridae), small Trichoptera (e.g. Hydroptilidae), and many of the Diptera, including all the Chironomidae. During identification, they had the most difficulty distinguishing among Diptera.

The strong correlation between metrics derived from volunteer and professionally identified data was matched in the indices, although the VV index had fewer metrics and, thus, a smaller range than B-IBI calculated for the VP data set (see Fig. 3). The five-metric index (VV) and 10-metric B-IBI did differ in their relative assessments of Evans and Holder Creeks. The VV index ranked the two streams similarly, whereas B-IBI ranked Evans's biological condition higher, primarily because of higher taxon richness metrics; relative abundance metrics were similar. Three times as many individuals were collected at Evans, and the higher sampling effort may have inflated taxon richness.

The professional B-IBI was more precise than the volunteer five-metric index; the precision of the VV index was equivalent to the VP.family index. The VV and VP.family indices could detect four categories of biological condition, and B-IBI based on VP data could detect 4.5 (Fig. 6). For the range of biological condition represented by our sites, professional taxo-

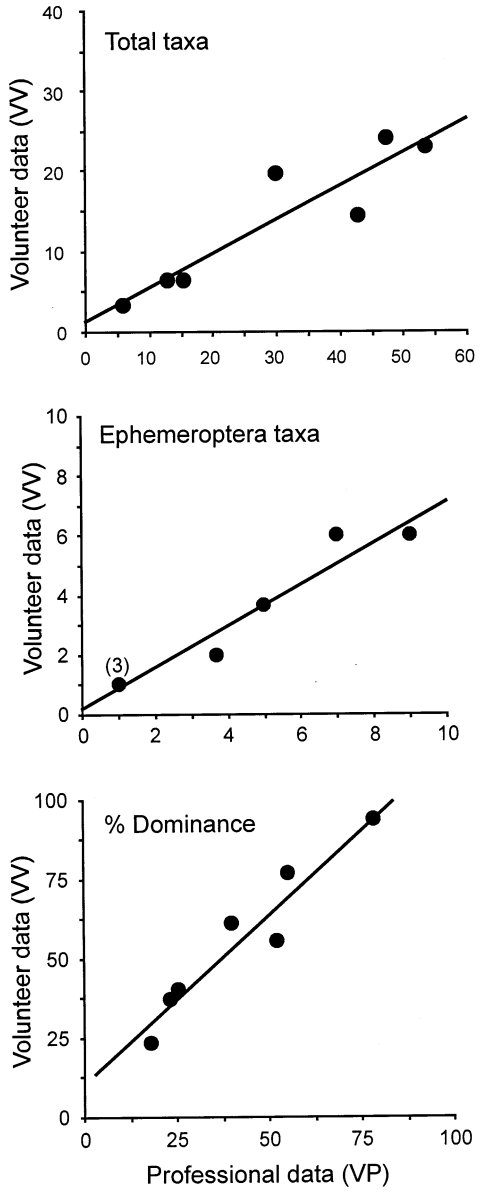

Fig. 5 Mean values for three metrics (total number of taxa, Ephemeroptera taxa, and percent dominance) from volunteer data (VV) plotted against professional data (VP). Regression lines are drawn, but relationship was tested with correlation (see Table 3).

nomic effort translated into a 13% increase in precision over volunteer effort.

The actual protocol used by King County was more precise, as expected, because the three metric values calculated for each Surber sample were averaged before calculating B-IBI. We assumed no crew differences and used the two visits by volunteers and professionals as replicates. B-IBI could detect 5.8 categories of biological condition. Evans was excluded

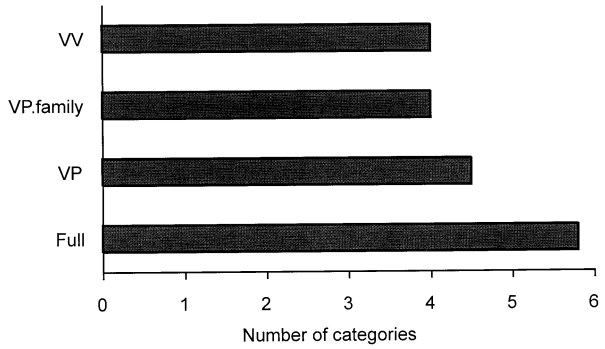

Fig. 6 Number of distinct categories of biological condition that different protocols can detect at 80% statistical power. VV, volunteer field and volunteer laboratory methods; VP.family, volunteer field methods and professional laboratory identification to family; VP, volunteer field and professional laboratory methods; FULL, full protocol (see text).

from this analysis because of possible sampling bias, as explained above.

Discussion

The purpose of this study was to determine whether volunteers could collect high quality data relevant to the management of urban catchments. We compared volunteer and professional assessments at two points in our analysis. We compared (1) volunteer versus professional field collections using the same (professional) level of laboratory analysis (VP versus PP), and (2) volunteer and professional taxonomic identification and metric calculation for the same set of field samples (VV versus VP). We found no differences between the field samples collected by professionals and volunteers. In the laboratory, professional taxonomic analysis yielded better identification of taxa, but precision of the assessment increased only slightly, by 13%.

Not only were volunteer assessments comparable to those derived from professional data, they were also strongly correlated with human disturbance, measured as developed area within the catchment. The total number of different taxa present and the number of taxa of Ephemeroptera, Plecoptera and Trichoptera all declined as forest was replaced by urban development. As invertebrate taxon richness declined, the percentage, or relative abundance, of animals belonging to the most abundant taxon (percentage dominance) increased, illustrating how single

groups can dominate a living assemblage as environmental conditions deteriorate (Pedersen & Perkins, 1986).

If biological assessments are so strongly correlated with measures derived from satellite imagery, why bother sampling invertebrates? For this study, we selected a very simple gradient based on developed area and excluded catchments with other confounding effects, such as dairy farms, channelization or mine waste. Satellite imagery is not necessarily an appropriate surrogate for biological condition where multiple human influences and cumulative effects are involved. In those cases, direct measures of the biota provide a better ranking of catchments in terms of human impact and biological condition.

Defining the statistical precision of assessment protocols is crucial to protect water resources. Analysing the full professional protocol used by King County demonstrated that B-IBI could detect between five and six categories of biotic integrity at 80% power. Thus, we know that a change in B-IBI of seven or more (across a range of 10–50) is required to represent a statistically significant and biologically meaningful change in the condition of a stream's biota. Such an analysis can be directly translated into guidelines for assessing the impact of human activities on streams and deciding whether a site is impaired within the framework of laws such as the Clean Water Act (Ransel, 1995).

Comparison of field performance

We compared the field performance of volunteers (VP) and professionals (PP) in two ways. First, we compared B-IBI values based on professional and volunteer field samples; second, we partitioned the variance of B-IBI values according to its possible sources. Laboratory effort was held constant for these comparisons; that is, invertebrate samples for both crews were sent to the same laboratory for full taxonomic identification. B-IBI values derived from the two sets of field samples (our VP and PP data sets) were highly correlated and gave very similar site assessments. The index scores based on volunteer data were neither consistently lower nor higher than those based on professional data. We attribute the small differences in B-IBI values to measurement error, e.g. differences in sample location or time of day.

The one exception was Evans Creek, whose B-IBI value based on volunteer field data (VP) was 12 points higher than that based on professional data (PP). We suspect that the taxon richness metrics for the professional sample were so low because the sample was inadvertently collected from the identical spot sampled earlier by volunteers. This result suggests that routine monitoring of the same sites may harm the resource it is designed to protect.

The largest component of variance (90%) in B-IBI values was associated with differences among streams, which were originally selected to represent different levels of human disturbance. None of the variance in B-IBI was due to differences between volunteer and professional crews (0%). A relatively small amount of variance (10%) was attributable to measurement error. Thus, we conclude that B-IBI is more sensitive to changes associated with disturbance, in this case urbanization, than to natural variability related to time of sampling or location within a stream reach.

Comparison of laboratory performance

In order to compare volunteer and professional methods in the laboratory, we held field results constant and examined only samples collected by volunteers. First, volunteers analysed the invertebrate samples and made their assessment (VV); then their field samples were sent to a professional laboratory for full identification (VP). We compared results for individual metrics and for the resulting overall indices. Five of the 10 B-IBI metrics were adapted for volunteers because they did not require identifying specimens to species. Even though volunteers distinguished many fewer taxa than professionals, volunteer metrics were strongly correlated with the professional metrics and with human disturbance. Volunteers had particular difficulty distinguishing among the Diptera, which is reflected in the slightly lower (though highly significant) correlation between volunteer and professional values for total taxon richness.

Volunteer-measured taxon richness of mayflies, stoneflies and caddisflies, though generally lower, was still very highly correlated with professional values. On average, volunteers separated about 85% of the invertebrates present from the inorganic matrix and missed many of the small invertebrates. Nevertheless, the only metric based on number of individu-

als (percentage dominance) was also highly correlated with professional values. High correlation between volunteer and professional metrics meant that the overall index values composed of those metrics were also highly correlated.

Although highly correlated, professional (VP) and volunteer (VV) metrics and index values were not equal. In particular, the range of possible values for metrics of taxon richness was smaller because volunteers identified many fewer taxa. Differences in metric values could be adjusted with scoring criteria to make indices comparable, but such an adjustment would mask the fact that volunteers do miss small taxonomic distinctions and important biological differences. We expect, for example, that volunteer metrics based on simple, morphological sorting or family-level identification will be less likely to distinguish among sites with relatively little human disturbance, because the differences among minimally disturbed sites typically appear at the genus and species levels.

In the laboratory, volunteers sorted invertebrates using many of the same morphological features scientists use to distinguish families. They did not have to learn exact taxonomic features or jargon; requiring that they do so would probably have prevented their participation (Oliver & Beattie, 1997). We expect that, as volunteers gain experience, they will learn family names and exact taxonomic features. A field guide similar to those available for amateur naturalists (e.g. Robbins, Bruun & Zim, 1983; Stokes & Stokes, 1996) would be very helpful with an emphasis on overall shape rather than the specific body parts used in dichotomous keys (e.g. Lehmkuhl, 1979; Merritt & Cummins, 1996).

The role of statistical power

Statistical power is defined as the probability of detecting a difference, or change, when a difference truly exists (Peterman, 1990). Low statistical power means that only extreme differences will be statistically significant and that resource condition must decline dramatically before managers can detect a change (Peterman & M'Gonigle, 1992; Dayton, 1998). We therefore recommend that monitoring programmes report the statistical precision of the measurement tools used to assess resource condition (Humphrey, Faith & Dostine, 1995; Barbour *et al*., 1998; Hughes *et al*., 1998; Carlisle & Clements, 1999).

The number of distinct categories detected by an index, i.e. its statistical power, depends on the experimental design; for our comparisons, we chose a two-sample *t*-test with three replicates because of its general nature and broad applicability to resource management questions (Peterman, 1990; Resh & McElravy, 1993; Fore *et al*., 1994). Models that estimate precision assume that replicate samples are independent. Samples collected within a stream are not independent because they experience similar flow conditions and water chemistry (Heffner, Butler & Reilly, 1996). Our calculations of precision were all equally affected by this limitation; as a consequence, our results only apply to similar situations where repeat samples are collected within a stream reach.

A multimetric index based entirely on volunteer effort (VV) could distinguish four categories of biological condition, exactly the same number as the professional analysis at family level (VP.family). For these protocols in our study, the index ranged from 5 to 25. Thus, a change in an index score of 5 or more probably represents a real biological change. The protocol based on full taxonomic identification (VP) could detect 4.5 categories of biological condition. The gain in precision from a professional taxonomic analysis (about 13%) was thus quite small.

We demonstrated that B-IBI following the full professional protocol used by King County (PP) can detect between five and six categories of biotic condition at 80% power. This result is nearly identical to that reported for the fish IBI used in Ohio (Fore *et al*., 1994), and it is similar to Ohio's invertebrate index (DeShon, 1995), suggesting that multimetric indices comprising similar metrics but derived for different taxonomic groups or geographic areas have similar intrinsic statistical properties. The practical application of this result means that a change in B-IBI of seven or more (for a range from 10 to 50) probably represents a real change in biological condition. These values can be translated into biological criteria used to define whether streams are biologically impaired under the Clean Water Act. The state of Washington is in the process of defining criteria for streams as required by the US Environmental Protection Agency (Karr, 1991).

Not only must an index for biological assessment be relatively immune to measurement error associated with natural variability (e.g. weather or location within the riffle), it must also correlate with an inde-

pendent measure of human disturbance. Index values may be *precise*, in that they provide similar values for repeated measurements, but precision does not necessarily mean that they provide *accurate*, or meaningful, measures of biological condition. For this study, the two indices derived from volunteer efforts (VV and VP) and the professional (PP) index showed a strong association with an independent measure of human disturbance, declining significantly as forest was replaced by urban area.

The role of volunteers

Volunteers and amateurs have added to scientific knowledge for centuries; fields such as astronomy and ornithology encourage volunteers to collect data on stars or bird migration (Root & Alpert, 1994; Mims, 1999). In the U.S.A., many states include water chemistry data collected by volunteers in their biennial reports to the US Environmental Protection Agency (Kerr *et al*., 1994; Mayio, 1994; Lathrop & Markowitz, 1995; Penrose & Call, 1995). Results of this study indicate that trained volunteers can also collect reliable biological data to supplement stream assessments. This project demonstrated a very strong interest in stream health by local citizens (77 people participated) and a willingness to spend long hours sorting and identifying very small animals ($>$ 880 h). Most volunteers said they would participate again.

This project focused on data collection and analysis, but a robust biological monitoring programme involves more than simple data collection (Yoder, 1995; Yoder & Rankin, 1995; Maher, Cullen & Norris, 1994). The role that volunteers play depends on a project's purpose and the questions asked (Cairns, McCormick & Niederlehner, 1993; Mattson *et al*., 1994; Carlson, 1997). For this project, the overall sampling design and final interpretation of the data were done by professional biologists. Because sample sites were carefully selected by biologists familiar with the region, volunteers successfully demonstrated the effects of human influence on stream invertebrates. Volunteer programmes will be most effective when guided by experienced researchers.

There are two stages at which volunteers can participate in biological monitoring: field collection and laboratory analysis. When carefully trained, volunteers collect field samples of similar quality to those of professionals. Because similar data are collected for assessments of surface waters in the U.K. (Resh, Norris & Barbour, 1995), Australia (Chessman, 1995; Marchant *et al*., 1997) and other countries (Graca & Coimbra, 1998; Barton & Metcalfe-Smith, 1992), our results are relevant for those monitoring programmes as well.

On the other hand, laboratory analysis by volunteers will never equal that of professional taxonomists. Family level identification is sufficient when large differences exist in biological condition or intensity of human disturbance (Thorne & Williams, 1997). To make distinctions among less disturbed catchments, for example, undisturbed versus catchments harvested for timber, additional metrics derived from genus and species identification would probably be necessary (see Table 1; Fore *et al*., 1996). Our study sites were located across a broad gradient of disturbance from suburban to extremely urban; consequently, the morphological sorting of volunteers to the approximate level of family was able to detect large differences in biological condition. As for multimetric indices, studies that derive assessments from multivariate models have found that extreme differences in biological condition can be detected at the family level of identification (Chessman, 1995; Growns *et al*., 1995; Wright *et al*., 1995; Marchant *et al*., 1997), but that smaller differences in resources condition require identification to the level of genus or species for most insects (Hawkins & Norris, 2000; Hawkins *et al*., 2000).

Volunteer groups may be unable to pay for professional taxonomic analysis ($\sim$ US$250 per site for this protocol). In this case, we recommend that they collect samples, make their assessments and then archive their samples. Sample jars are small, easy to store and their contents can be kept for years if properly preserved. If, over time, volunteers observe a change and their assessment is contested, they could send their archived samples to a professional taxonomist for a more complete analysis.

This project demonstrated that citizen volunteers are very interested in participating in biological monitoring, they are capable of collecting meaningful data, and their assessments are comparable to those made by professionals. Properly trained volunteers can extend our knowledge of current stream conditions by sampling more sites than professionals may have the resources for. Because they have a personal interest in their local catchments, volunteers are ideal candidates to monitor streams and watch for changes.

Acknowledgments

Although they do not appear as co-authors, the many volunteers involved in this project made this study possible. Special recognition goes to J. Hobbs and C. Schulze. We would also like to recognize the work of J. Fairchild, D. Frostholm, J. Jackman, A. Kimpo, N. and L. Malmgren, K.C., J. and K. Miedema, S. Morley, B.D. and E. Nelson, J. and K. Nelson, E. O'Connor, R. Oslin, B. Rockenbeck, S. Still, P. Szewczykowski, L. and J. Theriault, J. Wall and M. Windus. Special thanks to J. Anderson, C. Anderson and J. White for project support. E.W. Chu, A.G. Hildrew and an anonymous reviewer provided many helpful suggestions. Funding was provided by a King County Water Quality Block Grant (No. 421347).

References

Barbour M.T., Gerritsen J., Snyder B.D. & Stribling J.B. (1998) *Rapid bioassessment protocols for use in wadeable streams and rivers*: *periphyton*, *benthic macroinvertebrates*, *and fish*.EPA 841-D-97-002. Office of Water, USEPA, Washington, DC.

Barton D.R. & Metcalfe-Smith J.L. (1992) A comparison of sampling techniques and summary indices for assessment of water quality in the Yamask River, Quebec, based on benthic macroinvertebrates. *Environmental Monitoring and Assessment*, **21,** 225–244.

Beauchene M. (1997) Wanted! Preferably alive! *Volunteer Monitor*, **9,** 1–4.

Booth D.B. & Jackson C.R. (1997) Urbanization of aquatic systems: degradation thresholds, stormwater detection, and the limits of mitigation. *Journal of the American Water Resources Association*, **33,** 1077–1090.

Cairns Jr. J., McCormick P.V. & Niederlehner B.R. (1993) A proposed framework for developing indicators of ecosystem health. *Hydrobiologia*, **263,** 1–44.

Carlisle D.M. & Clements W.H. (1999) Sensitivity and variability of metrics used in biological assessments of running waters. *Environmental Toxicology and Chemistry*, **18,** 285–291.

Carlson R.E. (1997) The 1995 and 1996 great American Secchi dip-in: a report to the volunteers. *Lakeline*, **17** (Suppl.).

Chessman B.C. (1995) Rapid assessment of rivers using macroinvertebrates: a procedure based on habitat-specific sampling, family level identification and a biotic index. *Australian Journal of Ecology*, **20,** 122–129.

Dayton P.K. (1998) Reversal of the burden of proof in fisheries management. *Science*, **279,** 821–822.

Davis W.S. & Simon T.P. (1995) *Biological Assessment and Criteria*, *Tools for Water Resource Planning and Decision Making*. Lewis Publishers, Boca Raton, FL.

Davis W.S., Snyder B.D., Stribling J.B. & Stoughton C. (1996) *Summary of state biological assessment programs for streams and rivers*.EPA 230-R-96-007. Office of Policy, Planning, and Evaluation, US Environmental Protection Agency, Washington, DC.

DeShon J.E. (1995) Development and application of the invertebrate community index (ICI). In: *Biological Assessment and Criteria*, *Tools for Water Resource Planning and Decision Making* (eds W.S. Davis & T.P. Simon), pp. 217–243. Lewis Publishers, Boca Raton, FL.

Dvornich K.M., Tudor M. & Grue C.E. (1995) NatureMapping: improving management of natural resources through public education and participation. *Wildlife Society Bulletin*, **23,** 609–614.

Firehock K. & West J. (1995) A brief history of volunteer biological water monitoring using macroinvertebrates. *Journal of the North American Benthological Society*, **14,** 197–202.

Fore L.S., Karr J.R. & Conquest L.L. (1994) Statistical properties of an index of biotic integrity used to evaluate water resources. *Canadian Journal of Fisheries and Aquatic Sciences*, **51,** 1077–1087.

Fore L.S., Karr J.R. & Wisseman R.W. (1996) Assessing invertebrate responses to human activities: evaluating alternative approaches. *Journal of the North American Benthological Society*, **15,** 212–231.

Graca M.A.S. & Coimbra C.N. (1998) The elaboration of indices to assess biological water quality. A case study. *Water Resources*, **32**(2), 380–390.

Growns J.E., Chessman B.C., McEvoy P.K. & Wright I.A. (1995) Rapid assessment of rivers using macroinvertebrates: case studies in the Nepean River and Blue Mountains, NSW. *Australian Journal of Ecology*, **20,** 13–141.

Hawkins C.P. & Norris R.H. (2000) Effects of taxonomic resolution and exclusion of taxa on the sensitivity of RIVPACS models used to predict biotic conditions in streams. In: *Assessing the Biological Quality of Freshwaters*. *RIVPACS and Other Techniques* (eds J.F. Wright, D.W. Sutcliffe & M.T.

Furse), pp. 217–228. Freshwater Biological Association, Ambleside.

Hawkins C.P., Norris R.H., Hogue J.N. & Feminella J.W. (2000) Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications.*, **10,** 1456–1477.

Heffner R.A., Butler M.J. IV & Reilly C.K. (1996) Pseudoreplication revisited. *Ecology*, **77,** 2558–2562.

Hicks C.R. (1982) *Fundamental Concepts in the Design of Experiments*. Saunders College Publishing, Orlando, FL.

Hughes R.M., Kaufmann P.R., Herlihy A.T., Kincaid T.M., Reynolds L. & Larsen D.P. (1998) A process for developing and evaluating indices of fish assemblage integrity. *Canadian Journal of Fisheries and Aquatic Sciences*, **55,** 1618–1631.

Humphrey C.L., Faith D.P. & Dostine P.L. (1995) Baseline requirements for assessment of mining impacts using biological monitoring. *Australian Journal of Ecology*, **20,** 150–166.

Karr J.R. (1991) Biological integrity: a long-neglected aspect of water resource management. *Ecological Applications*, **1,** 66–84.

Karr J.R. (1998) Rivers as sentinels: using the biology of rivers to guide landscape management. In: *River Ecology and Management: Lessons from the Pacific Coastal Ecosystem* (eds R.J. Naiman & R.E. Bilby), pp. 502–528. Springer, New York, NY.

Karr J.R. (1999) Defining and measuring river health. *Freshwater Biology*, **41,** 221–234.

Karr J.R., Allan J.D. & Benke A.C. (1999) River conservation in the United States and Canada: science, policy, and practice. In: *River Conservation: Science, Policy, Practice* (eds P.J. Boon, B.R. Davies & G.E. Petts). J. Wiley, London.

Karr J.R. & Chu E.W. (1999) *Restoring Life in Running Waters: Better Biological Monitoring*. Island Press, Washington, DC.

Karr J.R., Fausch K.D., Angermeier P.L., Yant P.R. & Schlosser I.J. (1986) Assessment of biological integrity in running water: a method and its rationale. *In*: *Illinois Natural History Survey Special Publication, No. 5.* Champaign, IL.

Kentucky Water Watch (1998) *http://www.state.ky.us/nrepc/water/vm.htm*.

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

Kerans B.L., Karr J.R. & Ahlstedt S.A. (1992) Aquatic invertebrate assemblages: spatial and temporal differences among sampling protocols. *Journal of the North American Benthological Society*, **11,** 377–390.

Kerr M., Ely E., Lee V. & Mayio A. (1994) A profile of volunteer environmental monitoring: national survey results. *Lake and Reservoir Management*, **9,** 1–4.

Lathrop J.E. & Markowitz A. (1995) Monitoring water resource quality using volunteers. In: *Biological Assessment and Criteria, Tools for Water Resource Planning and Decision Making* (eds W.S. Davis & T.P. Simon), pp. 303–314. Lewis Publishers, Boca Raton, FL.

Lehmkuhl D.M. (1979) *How to Know the Aquatic Insects*. Brown Company Publishers, Dubuque, IA.

Maher W.A., Cullen P.W. & Norris R.H.. (1994) Framework for designing sampling programs. *Environmental Monitoring and Assessment*, **30,** 139–162.

Marchant R., Hirst A., Norris R.H., Butcher R., Metzeling L.& & Tiller D. (1997) Classification and prediction of macroinvertebrate assemblages from running waters in Victoria, Australia. *Journal of the North American Benthological Society*, **16,** 664–681.

Mattson M.D., Walk M., Kerr P.A., Slepski A.M., Zajicek O.T. & Godfrey P.J. (1994) Quality assurance testing for a large scale volunteer monitoring program: the acid rain monitoring project. *Lake and Reservoir Management*, **9,** 10–13.

May C.W., Horner R.R., Karr J.R., Mar B.W. & Welch E.B. (1997) Effects of urbanization on small streams in the Puget Sound lowland ecoregion. *Watershed Protection Techniques*, **2,** 483–494.

Mayio A. (1994) Volunteer data in the 305(b) Report. *The Volunteer Monitor*, **6,** 13.

Merritt R.W. & Cummins K.W. (Eds) (1996) *An Introduction to the Aquatic Insects of North America*, 3rd edn. Kendall/Hunt Publishing Company, Dubuque, IA.

Mims F.M. (1999) Amateur science — strong tradition, bright future. *Science*, **284,** 55–56.

Oliver I. & Beattie A.J. (1997) Future taxonomic partnerships: reply to Goldstein. *Conservation Biology*, **11,** 575–576.

Parsons M. & Norris R.H. (1996) The effect of habitat-specific sampling on biological assessment of water quality using a predictive model. *Freshwater Biology*, **36,** 419–434.

Pedersen E.R. & Perkins M.A. (1986) The use of benthic invertebrate data for evaluating impacts of urban runoff. *Hydrobiologia*, **139,** 13–22.

Penrose D. & Call S.M. (1995) Volunteer monitoring of benthic macroinvertebrates: regulatory biologists' perspectives. *Journal of the North American Benthological Society*, **14,** 203–209.

Peterman R.M. (1990) Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Sciences*, **47,** 2–15.

Peterman R.M. & M'Gonigle M. (1992) Statistical power analysis and the precautionary principle. *Marine Pollution Bulletin*, **24,** 231–234.

Plotnikoff R.W. & White J.S.. (1996) *Taxonomic laboratory protocol for stream macroinvertebrates collected by the Washington State Department of Ecology*. Publication No. 96-323, Washington State Department of Ecology, Olympia, WA.

Ransel K.P. (1995) The sleeping giant awakes: PUD No. 1 of Jefferson County v. Washington Department of Ecology. *Environmental Law*, **25,** 255–283.

Resh V.H. & McElravy E.P. (1993) Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. In: *Freshwater Biomonitoring and Benthic Macroinvertebrates* (eds D.M. Rosenberg & V.H. Resh), pp. 159–194. Chapman & Hall, New York, NY.

Resh V.H., Norris R.H. & Barbour M.T. (1995) Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Australian Journal of Ecology*, **20,** 108–121.

Reynoldson T., Hampel L. & Martin J. (1986) Biomonitoring networks operated by school children. *Environmental Pollution*, **41,** 363–380.

Robbins C.S., Bruun B. & Zim H.S. (1983) *Birds of North America, a Guild to Field Identification*. Golden Press, New York, NY.

Root T. & Alpert P. (1994) Volunteers and the NBS [letter]. *Science*, **263,** 1205.

Rossano E.M. (1996) *Diagnosis of Stream Environments with Index of Biotic Integrity*. Museum of Lakes and Streams, Sankaido, Tokyo, Japan(in Japanese and English).

Southerland M.T. & Stribling J.B. (1995) Status of biological criteria development and implementation. In: *Biological Assessment and Criteria, Tools for Water Resource Planning and Decision Making* (eds W.S. Davis & T.P. Simon), pp. 81–96. Lewis Publishers, Boca Raton, FL.

Stokes D. & Stokes L. (1996) *Stokes Field Guide to Birds: Western Region*. Little, Brown and Company, New York, NY.

Thomas L. (1997) Retrospective power analysis. *Conservation Biology*, **11,** 276–280.

Thorne R., St. J. & Williams W.P. (1997) The response of benthic macroinvertebrates to pollution in developing countries: a multimetric system of bioassessment. *Freshwater Biology*, **37,** 671–686.

USGS (US Geological Survey) (1995) *The Strategy for Improving Water-Quality Monitoring in the United States—Final Report of the Intergovernmental Task Force on Monitoring Water Quality*. Open-File Report 95-742, http://water.usgs.gov/wicp/itfm.html.

US EPA (Environmental Protection Agency) (1997) *What is volunteer monitoring?* http://www.epa.gov/owow/monitoring/volunteer/epavm.html.

Washington DOE (Department of Ecology) (1997) *What's watch over Washington?* http://www.wa.gov/ecology/wq/wow/wdw/survey.html.

Wiley M.J., Kohler S.L. & Seelbach P.W. (1997) Reconciling landscape and local views of aquatic communities: lessons from Michigan trout streams. *Freshwater Biology*, **37,** 133–148.

Wright I.A., Chessman B.C., Fairweather P.G. & Benson L.J. (1995) Measuring the impact of sewage effluent on the macroinvertebrate community of an upland stream: the effect of different levels of taxonomic resolution and quantification. *Australian Journal of Ecology*, **20,** 142–149.

Wright J.F., Furse M.T. & Armitage P.D. (1993) RIVPACS: a technique for evaluating the biological water quality of rivers in the UK. *European Water Pollution Control*, **3,** 15–25.

Yoder C.O. (1995) Policy issues and management applications for biological criteria. In: *Biological Assessment and Criteria, Tools for Water Resource Planning and Decision Making* (eds W.S. Davis & T.P. Simon), pp. 327–343. Lewis Publishers, Boca Raton, FL.

Yoder C.O. & Rankin E.T.. (1995) Biological criteria program development and implementation in Ohio. In: *Biological Assessment and Criteria, Tools for Water Resource Planning and Decision Making* (eds W.S. Davis & T.P. Simon), pp. 109–144. Lewis Publishers, Boca Raton, FL.

Zar J.H. (1984) *Biostatistical Analysis*, 2nd edn. Prentice-Hall, Englewood Cliffs, NJ.

(*Manuscript accepted 6 March 2000*)