CHAPTER 13

Selection of benthic macroinvertebrate metrics for monitoring water quality of the Fraser River, British Columbia: implications for both multimetric approaches and multivariate models

VINCENT H. RESH[1], DAVID M. ROSENBERG[2] AND TREFOR B. REYNOLDSON[3]

[1]*Department of Environmental Science, Policy & Management, 201 Wellman Hall, University of California, Berkeley, CA 94720, USA*

[2]*Department of Fisheries & Oceans, Freshwater Institute, 501 University Crescent, Winnipeg, MB R3T 2N6, CANADA*

[3]*National Water Research Institute, Environment Canada, CCIW, 867 Lakeshore Road, Burlington, ON L7R 4A6, CANADA*

Summary

Rapid bioassessment protocols are widely used in the USA and are based on the multimetric approach in which scores for several individual metrics (or measures) are evaluated against thresholds developed from reference sites, and a composite score is then calculated. The multimetric and multivariate approaches that use benthic macroinvertebrates in assessments of water quality (e.g. RIVPACS, AUSRIVAS, the BEAST), are similar in their data collection methods, but differ in the way reference sites are selected, test sites are classified and test site assessments are made.

Forty-four metrics were calculated from collections of benthic macroinvertebrates made at seventeen sites in the Fraser River catchment, British Columbia, Canada, in 1994. They included measures of richness (17 metrics), numbers of individuals (16), functional feeding groups (10), and a biotic index (1). Richness measures had the lowest variability in mean values across the five basins (or subcatchments) examined, and the lowest coefficients of variation based on replicate samples collected at a single site within each basin. Most (59.1%) metrics could be calculated at all sites examined in this study, and most (55.3%) correctly indicated impairment when impaired and unimpaired sites were compared. However, incorrect indications of impairment were noted in 40 to 60% of the metric comparisons made between unimpaired sites located in different ecoregions, between unimpaired sites in different streams of the same ecoregions, and between unimpaired sites in the same stream.

Richness metrics consistently had the lowest error rates of all the metrics examined. Incorporation of non-richness metrics into multivariate approaches may increase incorrect indications of impairment (i.e. Type I errors).

Multimetric approaches should consider incorporating multivariate analyses for defining reference conditions and assessing impairment of test sites. Collaboration among users of multimetric and multivariate approaches can enhance both types of water quality monitoring and assessment programs.

Introduction

Various approaches employed for water quality monitoring have become closely identified with their countries of origin and development, examples being the Saprobien index from Germany (Cairns & Pratt 1993) and RIVPACS from the UK (see Chapter 1). The rapid bioassessment protocols developed in the mid-1980s (Plafkin, Barbour *et al.* 1989) are most identified with monitoring in the USA. Rapid bioassessment attempts to provide an integrated assessment of an aquatic resource, comparing habitat (e.g. physical structure, flow regime) and biological measures with empirically defined reference conditions (Barbour, Gerritsen *et al.* 1999).

Resh *et al.* (1995) have noted that in almost all its permutations (see the appendix in Resh & Jackson 1993), rapid bioassessment approaches use techniques that attempt to evaluate assemblages of benthic macroinvertebrates at reduced costs relative to those associated with traditional, more quantitatively rigorous assessments. The effort (and cost) of benthic analysis is reduced with rapid bioassessment because of four specific features. First, a single relatively large sample, covering an area several-fold larger than that in traditional quantitative collections, is taken instead of several replicate samples. Second, a standardized subsampling procedure is used (e.g. the first 100 to 300 organisms randomly sorted), which both reduces the number of organisms processed and provides a relatively consistent unit of effort for the processing of all samples. Third, identification is often only to family level. Fourth, the results of surveys can be summarized in ways that can be understood by non-specialists, such as managers, other decision makers, and the concerned public.

Rapid bioassessment protocols in the USA are based on the use of multimetrics. This approach attempts to provide an integrated analysis of the biological community at a site, by calculating various metrics (or measures) representing functional or structural aspects of the community, and summing these into a single score. Their use has not been without controversy; for example, potentially important ecological information may be lost by aggregating individual measures into an index (e.g. Suter 1993; Polls 1994). In addition, some metrics are too variable to detect impairment consistently (or may indicate impairment when it does not occur) and are based on subjective criteria (Resh & Jackson 1993; Hannaford & Resh 1995). However, supporters have argued that the advantage of a multimetric approach is that it incorporates ecological information on how aquatic organisms feed, reproduce, and exploit their habitats (Fore *et al.* 1996) into assessments of water quality. They suggest that reliance on combinations of multiple measures minimizes the weaknesses of individual metrics (Barbour, Gerritsen *et al.* 1999, among others). These debates are often "apples and oranges" in content, and the resolution of these differences seems to be far from over.

The use of the multimetric approach in biological monitoring has expanded greatly in the USA over the past decade. Currently, 47 of the 50 states use benthic macroinvertebrates as the "target assemblages" (Utah, Nevada, and South Dakota do not, probably because their wadeable streams flow intermittently); 31 states also use fish and five use periphyton in bioassessments (Barbour *et al.* 1995; Barbour, Gerritsen *et al.* 1999). Benthic macroinvertebrates are the basis for most biomonitoring programs currently in use worldwide (Rosenberg & Resh 1993b), and the reasons for this choice are clear (Rosenberg & Resh 1996). State programs in the USA are usually based on modifications of a national program developed and promulgated by the United States Environmental Protection Agency (USEPA) (Plafkin, Barbour *et al.* 1989).

The purpose of this chapter is to determine which metrics are most appropriate for examining impairment in the Fraser River catchment, Canada, and to examine how these results could influence users of multimetric and multivariate approaches. To do this we examine a range of metrics in two ways: (1) do they detect impairment when impairment occurs and (2) do they give incorrect indications of impairment when no impairment occurs (Type I errors)?

How do multimetric and multivariate approaches differ?

Both approaches involve similar methods for collecting benthic macroinvertebrates, but the range of environmental variables used in the multivariate predictive models are not generally measured when multimetric collections are made. Instead, a habitat assessment supplements the biological information collected in the multimetric approach (Plafkin, Barbour *et al.* 1989; Hannaford *et al.* 1997). The two approaches diverge further once the samples have been collected, sorted, and specimens identified. In multimetric analysis, sites are grouped *a priori* based on their geophysical attributes, and final classification is based on taxonomic composition. In multivariate approaches, sites are classified into groups using clustering methods based on the similarity of their species composition.

In multimetric analysis, selection of reference sites for comparison with a test site is based on the geographical or physical attributes of the site, whereas in multivariate approaches, selection may be based on the sites in the reference group with which the test site has the highest probability of inclusion, using a discriminant model (e.g. the BEAST, Chapters 11 and 12). Alternatively, the selection may draw on information from several reference groups, according to the weighted probabilities with which the test site would be included in those reference groups (e.g. RIVPACS and AUSRIVAS, Chapters 1 to 10). Finally, test site assessment in the multimetric approach is based on quartile distributions of additive metrics. In the multivariate approach, it may be based on a comparison of the test and reference sites in taxa ordination space, using probability ellipses constructed around reference sites (the BEAST) or a comparison of the taxa observed at the test site and those expected to be present at the site, based on weighted probabilities of taxon occurrence (RIVPACS and AUSRIVAS). The above distinctions are discussed in detail by Reynoldson, Norris *et al.* (1997).

Reynoldson, Norris *et al.* (1997) compared two multivariate predictive models (the BEAST and AUSRIVAS) with a multimetric analysis. The latter was done using two groups of metrics. First, a fixed list of metrics modified from Plafkin, Barbour *et al.* (1989) was used, which included calculation of the following: (1) number of individuals; (2) number of families; (3) percent of Ephemeroptera, Plecoptera and Trichoptera (EPT) individuals; (4) percent of Chironomidae individuals; (5) ratio of the number of EPT individuals/number of EPT + Chironomidae individuals; (6) ratio of the number of Hydropsychidae individuals/number of Trichoptera individuals; (7) percent dominance of a single taxon; (8) the Family Biotic Index (Hilsenhoff 1998). Based on input from M. T. Barbour and J. Gerritsen (Tetra Tech Inc.), a second multimetric analysis was carried out in which the number of individuals was deleted, and the ratio of the number of EPT individuals/number of EPT + Chironomidae individuals was replaced with the ratio of the number of Baetidae individuals/number of Ephemeroptera individuals. A composite score was then calculated, based on the similarity of metrics to the appropriate reference site classification. Precision (i.e. whether all replicates at a single site were consistently designated as impaired or unimpaired) and accuracy (i.e. designations of unimpaired sites as unimpaired) of multimetric assessments were estimated. The precision and accuracy of the two groups of metrics were then compared within ecoregions, stream order, and biotic classifications.

The results of this comparison indicated that the two multivariate models performed consistently better than either of the fixed metric designs (Multimetric 1 and 2 in Table 13.1). In one comparison of precision, the BEAST performed less well than AUSRIVAS. However,

AUSRIVAS failed to designate a known impaired site as impaired, which the BEAST did designate as impaired.

Table 13.1. *Comparison of precision (A) and accuracy (B) of four methods employed for assessing water quality by macroinvertebrate composition at sites in the catchment of the Fraser River, Canada. Results are expressed as percentages of correct assessments for impaired and unimpaired sites.*

(See Reynoldson, Norris *et al.* 1997 for additional details).

Method of assessment	Ecoregion	Stream order	Biotic grouping
(A). Designation of replicates at each site as either ALL impaired or ALL unimpaired			
Multimetric 1	40	80	60
Multimetric 2	60	80	80
BEAST	80	100	80
AUSRIVAS	100	100	100
(B). Designation of unimpaired sites as unimpaired			
Multimetric 1	50	38	75
Multimetric 2	69	38	88
BEAST	100	100	100
AUSRIVAS	100	100	100

Table 13.2. *Description of sampling sites in the catchment of the Fraser River, Canada, used for the analysis shown in Tables 13.3 and 13.4.*

* Sites where replicate samples were taken.

Subcatchment	(Site)	River	Order	Ecoregion	Impacts
Salmon	(01)	Salmon River	3	Thompson-Okanagan Plateau	None
	(02)	Salmon River	4		None
	(03)*	Salmon River	4		Agriculture, logging
	(04)	Salmon River	3		Agriculture, logging
Chilcotin	(04)	Cluska River	4	Fraser Plateau	None
	(05)*	Palmer Creek	4		None
	(08)	Cluska River	4		None
Clearwater	(03)	Hobson Creek	2	Southern Rocky Mt. Trench	None
	(06)*	Hemp Creek	2		None
Pitt	(01)	Pitt River	2	Pacific Ranges	None
	(02)	Pitt River	2		None
	(03)	Pitt River	3		None
	(06)*	Pitt River	3		None
	(07)*	Pitt River	4		None
	(08)	Pitt River	4		None
Stuart	(02)*	Condit Creek	2	Omeineca Mts, Fraser Basin	None
	(04)	Lion Creek Trib.	2		None
	(06)	Lion Creek	3		None

Methods

Study area

The analysis is based on benthic macroinvertebrate collections made in the Fraser River, a catchment that covers *ca* 230,000 km² or 25% of British Columbia, Canada's westernmost province (see Chapter 12 for further description). Although the Fraser is one of the last unregulated large rivers of North America, urban and industrial pressures pose a severe threat to the health of this ecosystem. The research described here attempts to address the problem of pollution in the Fraser River catchment, through the development of a biomonitoring program for assessing water quality using benthic macroinvertebrates.

Sampling methods and analysis

Eventually, the Fraser River study will involve analysis of benthic macroinvertebrate data from more than 250 sites. In the analysis presented here, we use results from second to fourth order sites in five subcatchments from five different ecoregions of the Fraser River catchment (Table 13.2).

Rosenberg *et al.* (2000) provide details of the sampling and laboratory sorting program. The data used for this present study were obtained from 18 sites located throughout the Fraser River catchment (Table 13.2). Samples were collected from riffles using a triangular kicknet sampler (38.5 cm on each side). Either one sample unit or 3–5 replicate sample units, each of 3 minutes duration, were collected on each site; subsamples of the first 200 organisms encountered were subsequently sorted and identified from each sample unit.

Benthic macroinvertebrates were identified to the lowest taxonomic level (genus and in some cases species) in the Fraser River study. In this analysis, metrics were calculated in two ways: (1) at family level, because this is the approach used in many rapid bioassessment programs; (2) at genus or species level when possible (e.g. not with the Family Biotic Index or percent Chironomidae), as used in other programs. A total of 44 metrics was examined, including the same nine metrics used by Reynoldson, Norris *et al.* (1997) in the analysis described in Table 13.1, and an additional 35 metrics chosen from Barbour, Gerritsen *et al.* (1999).

The 44 metrics can be divided into four groups: measures of richness; numbers of individuals (or enumerations); functional feeding group ratios; and a biotic index. Analysis of each metric involved calculating the coefficients of variation (CV) for sites in six streams (marked with an asterisk in Table 13.2) in the Fraser River catchment. Test sites included a known impaired site (Salmon River 04) and known unimpaired sites. These designations were determined by personnel who are knowledgeable with this area (see the description of the workshop conducted for this purpose, in Rosenberg *et al.* 2000), examination of the area around a site (by helicopter) before sampling, and examination of a site while sampling.

To determine if impairment could be detected at the known impaired sites and, conversely, to determine if impairment would be incorrectly indicated (Type I error) at known unimpaired sites, we compared: (a) known impaired sites with known unimpaired sites (Salmon 03 cf. 01, and 03 cf. 02); (b) unimpaired sites of the same stream order in different ecoregions (selected because of their geomorphic similarity, e.g. Clearwater 06 cf. Stuart 02); (c) unimpaired sites in different streams of the same order in the same ecoregion (without regard to geomorphic similarity, Clearwater 03 cf. 06, Chilcotin 04 cf. 05, 05 cf. 08, Stuart 02 cf. 04, 02 cf. 06); (d) sites of the same or ±1 order in the same stream (Pitt 01 cf. 06, 02 cf. 06, 03 cf. 06, 07 cf. 08). Statistical analysis of these comparisons involved one-tailed t-tests at $p = 0.05$, as recommended by Barbour, Gerritsen *et al.* (1999); a one-tailed test was used because the metrics change in one direction (e.g. a decrease in number of taxa present with impairment).

Because we were using known impaired or unimpaired sites, and examining metric response individually, a Bonferroni correction was not necessary.

Results

Variability of 44 metrics

The first analysis examined the variability of 44 metrics, which included both the fixed list of metrics advocated by Plafkin, Barbour *et al.* (1989), marked with an asterisk in Table 13.3, and other metrics proposed for use in Barbour, Gerritsen *et al.* (1999). Mean values and CVs (expressed as percent) for all metrics varied greatly, sometimes by over an order of magnitude at the six sites in different basins (Table 13.3).

Table 13.3. *Mean metric values/coefficients of variation (CV %) for selected metrics at six stream sites in the catchment of the Fraser River (see Table 13.2 for descriptions of sites) for which replicate samples were collected.*

* Metrics that were typically used in the fixed-metric approach of Plafkin, Barbour *et al.* (1989); for a complete description of metric calculation see Barbour, Gerritsen *et al.* (1999).

— indicates that the metric could not be calculated at that site.

N = number of Families or Taxa; values for families may be higher than values for taxa because some smaller specimens could be identified only to family. EPT = Ephemeroptera, Plecoptera and Trichoptera combined.

(1) Ratio of EPT Individuals/Chironomidae + EPT Individuals.

CHI = Chilcotin subcatchment; CLR = Clearwater subcatchment; PIT = Pitt River subcatchment; SAL = Salmon River subcatchment; STU = Stuart subcatchment (see Table 13.2).

Metrics	CHI 05	CLR 06	PIT 06	PIT 07	SAL 03	STU 02
Richness:	(n = 5)	(n = 5)	(n = 3)	(n = 5)	(n = 3)	(n = 5)
Total N Taxa	27.6/11.0	28.4/ 6.9	12.0/25.0	19.8/ 8.3	26.0/ 6.7	21.0/21.6
*Total N Families	20.0/ 8.7	16.4/14.0	10.3/14.8	14.2/12.7	14.7/10.4	13.2/12.4
N EPT Taxa	9.6/16.3	15.4/14.2	9.0/15.1	7.5/11.1	8.7/ 8.7	14.4/19.4
N EPT Families	9.6/11.9	11.2/9.8	7.7/15.1	8.6/13.3	7.7/19.9	10.2/16.1
N Ephemeroptera Taxa	3.0/62.4	5.2/21.1	3.0/0	6.2/13.5	1.3/43.3	5.6/24.0
N Ephemeroptera Families	4.0/ 0	4.4/12.5	4.0/0	3.0/0	3.7/15.8	4.0/17.7
N Trichoptera Taxa	3.6/15.2	2.8/46.6	2.0/0	1.3/43.3	3.3/17.3	1.6/55.9
N Trichoptera Families	3.2/14.0	2.6/43.9	2.0/0	1.3/43.3	3.0/33.3	1.4/39.1
N Plecoptera Taxa	6.0/ 0	7.4/12.1	4.0/0	4.0/0	4.0/0	7.2/15.2
N Plecoptera Families	2.4/47.5	4.2/20.0	3.0/0	4.6/ 9.3	1.0/0	4.8/17.4
N DipteraTaxa	8.8/23.3	13.6/13.9	5.3/39.0	9.2/11.9	14.3/10.7	7.2/26.7
N Diptera Families	3.6/34.3	3.2/26.2	2.0/ 0	3.2/13.1	3.3/34.7	2.4/22.8
N Chironomidae Taxa	5.6/32.4	11.6/13.1	3.3/62.5	6.2/17.9	11.3/18.4	5.8/37.4
N Odonata Taxa	— —	— —	— —	— —	— —	— —
N Odonata Families	— —	— —	— —	— —	— —	— —
N Coleoptera Taxa	2.4/22.8	— —	— —	— —	1.0/ 0	— —
N Coleoptera Families	1.0/0	— —	— —	— —	1.0/ 0	— —

(Table 13.3 is continued on facing page)

Metrics	CHI 05	CLR 06	PIT 06	PIT 07	SAL 03	STU 02
Number of Individuals:						
*N EPT Ind/Ch+EPT Ind[1)]	0.6/22.0	0.5/13.1	1.0/2.1	0.6/2.0	0.6/23.3	0.7/12.6
*% EPT Individuals	45.4/22.2	46.5/12.8	94.1/2.9	81.4/3.0	49.6/32.8	66.7/12.8
% Ephemeroptera	36.3/19.5	26.1/23.0	56.6/12.4	28.7/8.5	27.6/20.8	55.4/10.4
% Plecoptera	4.0/50.0	17.7/6.5	35.9/24.5	52.2/24.5	6.1/102.3	10.2/23.4
% Trichoptera	5.1/34.2	2.7/28.1	1.6/0	0.6/34.6	16.1/72.9	1.1/49.8
*% Chironomidae	25.9/38.2	49.4/12.9	2.1/91.7	11.7/13.0	31.7/25.5	32.1/26.1
% Coleoptera	2.4/31.6	— —	— —	— —	1.0/108.7	— —
% Odonata	— —	— —	— —	— —	— —	— —
% Tribe Tanytarsini	19.3/37.9	17.0/48.2	— —	0.8/47.1	18.1/31.6	26.7/30.8
% Diptera+non-insects	46.3/18.8	56.2/10.8	7.5/47.6	19.1/13.1	57.0/15.7	34.4/23.3
*% Dominant Taxon	17.9/22.6	17.5/29.2	39.3/ 6.2	36.7/ 2.6	21.6/17.6	32.7/8.7
% 2 Dominant Taxa	30.5/12.0	30.6/19.9	64.7/ 6.4	54.7/ 7.2	37.6/14.4	55.1/7.0
% Contribution of 5 Dominant Taxa	57.3/ 5.3	56.4/8.0	92.5/ 6.1	73.2/ 2.1	65.2/ 5.1	79.0/4.2
*% of Trichoptera that are Hydropsychidae	53.3/40.1	20.0/0	— —	75.0/47.1	52.2/25.3	88.9/21.7
% of Ephemeroptera that are Baetidae	38.0/42.7	31.3/9.6	59.0/14.6	61.5/14.0	16.2/57.4	48.2/15.4
*Total Abundance	11,191.6/23.3	12,904.2/20.5	324.3/41.0	414.6/44.0	4,325.0/19.1	4,216.4/23.3
Functional feeding measures:						
% Gatherers	53.6/8.6	76.4/3.3	89.1/0.6	79.0/2.3	45.6/10.8	86.1/3.4
% Gatherer Families	37.9/9.9	51.8/13.7	51.9/6.2	42.1/5.2	48.1/10.8	48.0/15.7
% Filterers	21.7/34.2	8.3/74.1	— —	1.9/60.0	10.0/63.8	— —
% Filterer Families	19.2/17.2	10.9/18.2	— —	8.4/28.9	11.2/27.2	— —
% Predators	3.3/43.7	3.8/42.2	8.7/8.7	11.7/ 9.8	10.9/50.7	5.8/24.8
% Predator Families	17.7/37.0	16.4/53.4	39.3/14.2	24.7/38.9	21.0/42.6	30.9/28.9
% Scrapers	18.3/18.4	16.1/38.1	54.4/16.9	55.3/6.0	7.8/120.8	23.6/113.5
% Scraper Families	18.0/12.5	24.9/37.4	42.0/5.3	21.4/12.8	18.0/13.2	26.8/23.6
% Shredders	3.9/38.9	21.7/30.1	— —	4.2/31.0	30.1/49.8	5.9/41.7
% Shredder Families	11.1/37.4	30.2/30.4	— —	17.1/24.1	20.6/10.8	21.2/11.8
Biotic Indices:						
*Family Biotic Index	3.3/11.1	2.9/8.9	2.2/7.9	1.6/11.8	3.0/16.3	3.0/16.9

Richness measures showed low variability in terms of either mean values or CVs (Table 13.3). Among sites, mean values ranged over less than a 2-fold difference for the following: number of families; number of Ephemeroptera, Plecoptera and Trichoptera (EPT) taxa and families; number of Ephemeroptera families; number of Plecoptera taxa; number of Diptera families; the ratio of EPT individuals to EPT+Chironomidae individuals; the contribution of the five dominant taxa; the percentage of gatherer taxa and families (Table 13.3). Coefficients of variation showed the lowest range for number of families (8.7 to 14.8%) and the highest range for percent Chironomidae individuals (12.9 to 91.0%). The lowest CV values were obtained for the following: number of taxa (≤25.0% at all six sites) and families (≤14.8%); number of EPT taxa (≤19.4%) and EPT families (≤19.9%); number of Ephemeroptera families (≤17.7%); number of Plecoptera taxa (≤15.2%); percentage of Ephemeroptera (≤23.0%), two dominant taxa (19.9%) and five dominant taxa (≤8.0%); percentage of gatherer taxa (≤10.8%) and families (≤15.7%); Family Biotic Index (≤16.9%).

Detection of impairment in twelve sites

In the second analysis, twelve among-site comparisons were made for the detection of impairment. For each site, metrics were calculated when the information required to do this was available from collections at the site; e.g. calculation of the number of Trichoptera taxa would require that Trichoptera occurred at the site. With this restriction, 6 of 17 richness metrics, 12 of 16 enumerations, 7 of 10 functional feeding group metrics, and the Family Biotic Index, were calculated in the twelve comparisons (Table 13.4).

Table 13.4. *Summary of one-tailed t-test evaluations ($p = 0.05$) comparing metrics in terms of:*
(a) how often data were available for a statistical comparison, from 12 sites;
(b) correct indications of impact ($n = 2$);
(c to e) incorrect indications of impact in terms of site comparisons in different rivers from
(c) different ecoregions ($n = 1$), (d) the same ecoregion ($n = 5$), (e) different sites in the same river ($n = 4$).

(1) Ratio of EPT Individuals/Chironomidae+EPT Individuals.
EPT = Ephemeroptera, Plecoptera and Trichoptera combined.

N = number of Families or Taxa. See text on sampling methods for sites involved in the comparisons.

Metrics	(a) No. of sites for calculation	(b) Metric indicated impairment	Did the metric incorrectly indicate impairment when sites were compared in: (c) Different ecoregions?	(d) Different rivers?	(e) The same rivers?
Richness:					
Total N Taxa	12	Yes	Yes	Yes (2/5)	Yes (1/4)
Total N Families	12	Yes	No	Yes (4/5)	No (0/4)
N EPT Taxa	12	Yes	No	Yes (1/5)	No (0/4)
N EPT Families	12	Yes	No	Yes (2/5)	No (0/4)
N Ephemeroptera Taxa	9	Yes	No	Yes (1/5)	No (0/1)
N Ephemeroptera Families	6	No	No	Yes (1/3)	—
N Trichoptera Taxa	9	Yes	No	Yes (4/5)	Yes (1/1)
N Trichoptera Families	12	No	No	Yes (4/5)	Yes (1/4)
N Plecoptera Taxa	6	No	No	No (0/3)	—
N Plecoptera Families	7	—	No	Yes (1/5)	No (0/1)
N DipteraTaxa	12	No	Yes	Yes (4/5)	Yes (1/4)
N Diptera Families	9	No	No	Yes (3/5)	No (0/1)
N Chironomidae Taxa	12	No	Yes	Yes (4/5)	Yes (1/4)
N Odonata Taxa	0	—	—	—	—
N Odonata Families	0	—	—	—	—
N Coleoptera Taxa	3	—	No	Yes (2/2)	—
N Coleoptera Families	3	—	No	Yes (2/2)	—
Number of Individuals:					
N EPT Ind/Ch+EPT Ind[1]	12	No	Yes	Yes (3/5)	Yes (1/4)
% EPT Individuals	12	No	Yes	Yes (3/5)	Yes (1/4)
% Ephemeroptera	12	Yes	Yes	Yes (4/5)	Yes (1/4)
% Plecoptera	12	No	No	Yes (3/5)	Yes (3/4)
% Trichoptera	9	No	No	Yes (3/5)	Yes (1/1)
% Chironomidae	12	Yes	Yes	Yes (3/5)	Yes (1/4)
% Coleoptera	4	No	—	Yes (2/2)	—
% Odonata	0	—	—	—	—

(Table 13.4 is continued on facing page)

Metrics	(a) No. of sites for calculation	(b) Metric indicated impairment	Did the metric incorrectly indicate impairment when sites were compared in: (c) Different ecoregions?	(d) Different rivers?	(e) The same rivers?
% Tribe Tanytarsini	9	Yes	No	Yes (3/5)	Yes (1/1)
% Diptera+non-insects	12	Yes	Yes	Yes (4/5)	Yes (1/4)
% Dominant Taxon	12	Yes	Yes	Yes (4/5)	Yes (3/4)
% 2 Dominant Taxa	12	Yes	Yes	Yes (4/5)	Yes (2/4)
% Contribution of 5 Dominant Taxa	12	Yes	Yes	Yes (3/5)	Yes (1/4)
% of Trichoptera that are Hydropsychidae	7	Yes	—	Yes (4/4)	Yes (1/1)
% of Ephemeroptera that are Baetidae	12	Yes	Yes	Yes (2/5)	Yes (2/4)
Total Abundance	12	Yes	Yes	Yes (4/5)	Yes (2/4)
Functional feeding measures:					
% Gatherers	12	Yes	Yes	Yes (4/5)	Yes (2/4)
% Gatherer Families	12	Yes	Yes	Yes (3/5)	Yes (3/4)
% Filterers	7	No	No	Yes (2/3)	No (0/1)
% Filterer Families	7	Yes	Yes	Yes (2/3)	No (0/1)
% Predators	12	No	No	Yes (3/5)	Yes (2/4)
% Predator Families	12	No	No	No (0/5)	Yes (2/4)
% Scrapers	12	Yes	No	Yes (5/5)	Yes (4/4)
% Scraper Families	12	No	Yes	Yes (3/3)	Yes (4/4)
% Shredders	12	No	Yes	Yes (1/5)	No (0/1)
% Shredder Families	9	Yes	Yes	Yes (3/5)	No (0/1)
Biotic Indices:					
Family Biotic Index	12	No	No	Yes (3/5)	Yes (2/4)

With regard to correct indications of impairment (i.e. impairment was noted when it occurred; e.g. Salmon River 01 and 02 cf. 03 comparisons), 6 of 12 richness metrics, 10 of 15 enumeration metrics, 5 of 10 functional feeding group metrics, but not the Family Biotic Index, had t values higher than expected at $p = 0.05$ (Table 13.4).

Incorrect designations of impairment when a site was not impaired (Type I error) were as follows. For unimpaired streams in different ecoregions, errors were found for 3 of 15 richness metrics that were calculable, for 10 of 13 enumerations, and for 6 of 10 functional feeding groups metrics. In different rivers of the same ecoregion, errors were found for 35 of 65 richness comparisons, for 49 of 71 enumerations, for 26 of 44 functional feeding groups metrics, and for 3 of 5 Biotic Index comparisons. Unimpaired sites in the same river had the lowest errors, with errors in richness being 5 of 32, enumerations 21 of 47, functional feeding groups 17 of 28, and Family Biotic Index 2 of 4 (Table 13.4).

Combining error rates across different scales (different ecoregions, different rivers, sites in the same river), by summing the last three columns in Table 13.4, indicated that the best performing metrics were all richness metrics: number of taxa (only 4 errors in 10 comparisons); families (4 of 10); EPT taxa (1 of 10) and families (2 of 10); Ephemeroptera taxa (1 of 7) and families (1 of 4); Plecoptera taxa (0 of 4) and families (1 of 7).

The lowest error rates occurred when unimpaired sites in the same rivers were compared, and this was especially evident for richness metrics. Fewer incorrect designations occurred (although only one site comparison was used) when sites in different ecoregions (located several hundred km apart) were picked because of physical habitat similarities, than when streams in the same ecoregion were compared (Table 13.4).

Discussion

This analysis of metrics, calculated from macroinvertebrate collections from the Fraser River, clearly shows three main points. (1) Richness metrics are the most useful of all the types of metrics tested, in terms of ability to indicate impairment when it occurs and not indicating impairment when it does not occur. (2) The fixed metric approach of Plafkin, Barbour *et al.* (1989), which attempts to include a variety of structural and functional measures of benthic communities (i.e. those marked by an asterisk in Table 13.3), would not be significantly improved by substitution of other metrics (i.e. those without an asterisk in Table 13.3), because it is mainly the richness metrics that perform well. (3) There are more classification errors (i.e. incorrect indications of impairment) with the multimetric approach when sites in different rivers are compared than when sites in the same river are compared, and two sites in different ecoregions selected because of geomorphic similarity may have fewer classification errors than sites in the same river or region.

The high variation in mean values for the six sites in Table 13.3 indicates that rather than having geographically broad-based thresholds reflecting unimpaired conditions, local thresholds must be established. Given that these streams have different underlying geology, nutrient bases, geomorphology, etc., the variability observed in mean values of benthic macroinvertebrate metrics between subcatchments is not unexpected. Consequently, even for metrics that have high CVs, finding statistically significant differences in the absence of impairment is not surprising. Other studies have examined the appropriateness of benthic macroinvertebrate metrics for different regions, and their results indicated that some metrics could be used successfully in different regions (Table 13.5) but most could not. Although these studies examined a range of metrics, it is important to note that only the richness measures seem to be reported as consistently useful across studies.

What can the multimetric approach tell us about RIVPACS-type models? Water quality monitoring agencies worldwide are considering the use of multivariate models as the basis for monitoring programs. Although reliance on the use of a few fixed metrics (e.g. percent EPT) may be appropriate for developing countries (Resh 1995; Sivaramakrishnan *et al.* 1996), this certainly goes against international trends. Because the regulatory agencies in the USA have based most of their benthic macroinvertebrate biomonitoring programs on the multimetric approach (Resh & Jackson 1993; Resh *et al.* 1995), Reynoldson, Norris *et al.* (1997) recommended that a safe, cost-effective strategy for these agencies may be to (1) supplement the multimetric biological collections, which are fundamentally the same as those used for multivariate approaches, with similar environmental measurements required for multivariate analyses, and (2) do multimetric and multivariate analysis side-by-side and base the ultimate decision of site impairment on analysis and interpretation of both approaches.

RIVPACS and other multivariate approaches (e.g. AUSRIVAS) develop predictive models using the presence or absence of species as the basis for predictions of faunal occurrence in test sites, although data on density (e.g. the BEAST) and abundance (e.g. RIVPACS) are also used in these models. From the other chapters in this volume, it is apparent that a trend in the future development and expansion of multivariate models is to include non-richness metrics that reflect other structural and functional aspects of benthic macroinvertebrate communities (e.g.

Table 13.5. *Benthic macroinvertebrate metrics found to be useful in previous analyses.*

Kerans *et al.* (1992) determined success of a metric if no differences or consistent differences were found between the sampling devices that were used, between riffles and pools, and year-to-year differences.
Barbour *et al.* (1992) used ability to distinguish classes (montane versus valley/plains).
Resh & Jackson (1993) used low variability between sites and years, and consistent patterns of difference between impacted and unimpacted sites.
Kerans & Karr (1994) used concordance with water quality and fish assemblage analyses, and variability across habitats and ecoregions.
Fore *et al.* (1996) distinguished disturbed sites from minimally disturbed sites.

Kerans *et al.* (18 metrics)	Barbour *et al.* (17 metrics)	Resh & Jackson (20 metrics)	Kerans & Karr (18 metrics)	Fore *et al.* (30 metrics)
Plecoptera richness	No. of taxa	No. of taxa	Taxa richness	Taxa richness
Intolerant snail and mussel richness	EPT richness	No. of EPT taxa	Intolerant snail & mussel richness	Ephemeroptera richness
% Individuals in two numerically dominant taxa	Pinkham-Pearson index	No. of families	Ephemeroptera richness	Plecoptera richness
% Omnivores	Quantitative similarity index	Margalef's index	Trichoptera richness	*Pteronarcys* richness
% Gatherers	Biotic index	Family biotic index	Plecoptera richness	Trichoptera richness
% Grazers	% Dominant taxa	% Scrapers	% *Corbicula*	Intolerant taxa richness
% Predators	Dominants in common for five most abundant taxa		% Oligochaeta % Omnivores	Sediment-intolerant taxa richness
% Filterers	Ratio of individuals of Hydropsychidae to total Trichoptera		% Filterers	Sediment-tolerant taxa richness
% Shredders			% Grazers	
% Chironomidae	% Scrapers		% Predators	% Tolerant species
	% Shredders		% Individuals in two numerically dominant taxa	% Sediment-tolerant species
	Quantitative similarity index for functional feeding groups		Total abundance	% Dominance of the three most abundant taxa

see Chapter 9). However, the analysis performed in this study showed that richness metrics were most accurate in detecting impairment and avoiding classification errors; it should be expected that predictive, multivariate models incorporating non-richness measures will also produce higher misclassification rates. A comparison of the BEAST and AUSRIVAS (Reynoldson, Norris *et al.* 1997) indicated that although data on presence and absence alone (in AUSRIVAS) has lower rates of classification errors, the inclusion of density (as in the BEAST) may make the model more sensitive in detecting impairment. This suggestion needs further testing; however, it does underscore the potential problems of increased misclassification but also the benefits of increased sensitivity from including non-richness metrics.

Taxon richness is the most widely used evaluation measure in benthic macroinvertebrate studies of pollution effects (Resh & McElravy 1993). Why did richness measures work better than enumerations, functional feeding group and biotic index metrics in this study? The success of using presence or absence information would suggest that with impairment, taxon abundance is not just reduced, but rather that taxa are eliminated. Furthermore, changes in density can result from either impairment-related or non-impairment-related sources. Problems with functional feeding group designations and biotic indices may involve issues of methods: the need for correct designation of feeding groups (or the concept itself), and correct designations for tolerances of individual taxa, respectively. The above questions require far more experimental study.

The present analysis also has implications for the multimetric approach. Each of the metrics in a test site that is compared to a reference site value, using percent similarity or interquartile ranges, could also be compared using a multivariate analysis. This inclusion could bring federal and local regulatory agencies in the USA more in line with worldwide trends in approaches to monitoring water quality, and lead to greater international cooperation and collaboration that can advance the performance of all water quality monitoring programs (Resh & Yamamoto 1994).

Acknowledgments

We thank M. Barbour and C. Faulkner for making a draft version of the 1999 USEPA rapid bioassessment protocols available to us, T. Pascoe for assistance with database management, and D. W. Sutcliffe, J. F. Wright and an anonymous reviewer for their comments on the manuscript of this chapter.