

# Comparison of bioassessment results and costs between preserved and unpreserved macroinvertebrate samples from streams

Hanneke E. Keizer-Vlek · Paul W. Goedhart ·  
Piet F. M. Verdonshot

Received: 5 February 2010 / Accepted: 25 May 2010 / Published online: 12 June 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** The choice to use or not use a preservative before sorting macroinvertebrate samples (i.e., dead specimens vs. living specimens) is based on studies not solely focused on the effects of preservation. Using identical sample processing protocols, we compared preserved and unpreserved samples for the following parameters: (1) the number of taxa and individuals for each major macroinvertebrate group, (2) ecological quality classes calculated with a multimetric index developed for the assessment of small Dutch lowland streams, and (3) costs of sample processing. We collected macroinvertebrate samples from three lowland streams in the Netherlands. At each site, we collected six replicate samples, of which three samples were preserved and three were not. Significantly different numbers of Ephemeroptera individuals and Hydracarina taxa and individuals were collected from preserved samples compared to unpreserved samples. In assessments based on these individual metrics, standardiza-

tion of sample processing will be required. In streams with Ephemeroptera, the preservation of samples is necessary to optimize the number of Ephemeroptera individuals collected. In streams that contain Hydracarina, the preservation of samples will result in an underestimation of the number of Hydracarina taxa and individuals present. In only one instance there was a difference in ecological quality between preserved and unpreserved samples, indicating that assessing small Dutch lowland streams does not require standardization of sample preservation as part of the sample processing protocol. We detected no significant differences in sample processing costs between preserved and unpreserved samples.

**Keywords** Preservative · Costs · Macroinvertebrates · Bioassessment · Streams

## Introduction

Macroinvertebrates are the most commonly used organisms to assess the biological quality of streams in monitoring programs (Hawkes 1979; Hellawell 1986; Chessmann 1995). Biological monitoring usually has two purposes: (1) to estimate variables of interest at a site and (2) to make comparisons among sites or time intervals. Variables of interest in biological monitoring are primarily metric values (e.g., the number of

---

H. E. Keizer-Vlek (✉) · P. F. M. Verdonshot  
Alterra, Centre for Ecosystem Studies, Wageningen  
University and Research Centre, P.O. Box 47, 6700  
AA Wageningen, The Netherlands  
e-mail: hanneke.keizer-vlek@wur.nl

P. W. Goedhart  
Biometris, Wageningen University and Research  
Centre, P.O. Box 100, 6700 AC Wageningen,  
The Netherlands

taxa, average score per taxon values, and saprobic index values) and ecological quality classes resulting from biological assessment systems. Metric values and ecological quality classes are calculated based on the macroinvertebrate community composition. Various methods have been developed to collect macroinvertebrates from streams and to process macroinvertebrate samples. These sampling and sample processing methods can vary in terms of sampled area, mesh size of sampling gear, sampled habitats, intensity of sorting, and taxonomic resolution of identification, among other parameters. The methodology applied influences the accuracy and variability of bioassessment results (expressed as metric values and/or ecological quality classes) (e.g., Barbour and Gerritsen 1996; Diamond et al. 1996; Haase et al. 2004). Also, each method can be selective for certain species or groups of species that vary in their exposure and sensitivity to anthropogenic stress (Barton and Metcalfe-Smith 1992).

Accuracy and variability are both important aspects of bioassessment. Accuracy refers to the closeness of a measurement to its true value (Norris et al. 1992). Differences in accuracy between methods may, therefore, result in different bioassessment results. Differences in accuracy depend on the spatial and temporal scale at which the true value is defined—a method might be accurate at representing the organisms present in a sample, but less accurate at representing the biota at a site. Variability is important in making comparisons because the validity of conclusions depends on data variability (Norris et al. 1992); higher variability increases the probability of incorrect bioassessment results. An increase in accuracy or a reduction in variability is not always possible because associated costs are often high. When assessing ecological quality for biological monitoring purposes, however, it is not necessary to catch all organisms or taxa present at a site (Barbour and Gerritsen 1996). Standardization of sampling is required, though, for valid comparisons among sites and points in time (Courtemanch 1996; Vinson and Hawkins 1996). The question then focuses on which steps to standardize in sampling and sample processing. After all, when two methods are equally variable and give comparable bioassessment results, stan-

dardization is not necessary. Apart from accuracy and variability, costs play an important role in decision-making related to the standardization of methods. The costs for collection and processing of macroinvertebrate samples are high and (can) depend strongly on the sampling technique used (e.g., Barbour and Gerritsen 1996; Metzeling et al. 2003; Vlek et al. 2006).

Many studies have focused on variability, accuracy, and/or costs in terms of sampled area (e.g., Metzeling and Miller 2001; Vlek et al. 2006), number of samples (e.g., Canton and Chadwick 1988), sampling device (e.g., Drake and Elliott 1982; Mackey et al. 1984; Barton and Metcalfe-Smith 1992; Cheal et al. 1993), sampled habitats (e.g., Kerans et al. 1992), intensity of sorting (e.g., Barbour and Gerritsen 1996; Courtemanch 1996; Grown et al. 1997), and taxonomic resolution of identification (e.g., Nijboer and Verdonschot 2000; Bailey et al. 2001; Lenat and Resh 2001). An important aspect of sample processing, which has only been the subject of a few studies, is the preservation (or not) of samples immediately after collection. Many sampling protocols recommend 'live sorting' in which organisms are collected from the sample while still alive. Live sorting is frequently applied in the Netherlands, Southern European countries (Buffagni, CNR-IRSA, personal communication), and Germany (Braukmann 2000). Live sorting is also commonly applied in Australia for the rapid biological assessment of rivers (Metzeling et al. 2003), either for set periods (Chessman and Robinson 1987) or until a fixed number of specimens is collected (Chessman 1995).

In the few studies comparing sorting results between preserved and unpreserved samples, sorting of the unpreserved samples has been performed in the field, and sorting of the preserved samples has been performed in the laboratory (e.g., Humphrey et al. 2000; Metzeling et al. 2003; Haase et al. 2004; Nichols and Norris 2006). In these studies, other aspects of the sample processing protocol also differed between preserved and unpreserved samples. Humphrey et al. (2000) state that the live-sort procedure results in poor recovery of small and cryptic taxa. Metzeling et al. (2003) found that Oligochaeta were underrepresented in unpreserved/field samples compared to

preserved/laboratory samples. In our view, these findings are the result of field sorting and other sample processing aspects, rather than live sorting. In fact, live sorting in the laboratory might increase accuracy and reduce variability and costs. Sorting in the Netherlands is commonly performed in the laboratory to avoid (1) the high variability associated with field sorting (Haase et al. 2004), arising from differences in weather conditions and illumination at the sampling site (Carter and Resh 2001; Rawer-Joost 2001), and (2) loss of small organisms.

People who prefer using preservatives often mention the following disadvantages of live sorting: (1) specimens may be eaten by others before sorting is completed; (2) specimens may disintegrate before sorting is completed; (3) removing fast-moving taxa (like *Gammarus* sp.) from a sample may be time consuming; and (4) as a consequence of arguments 1 and 2, samples have to be sorted as soon as possible (within 5 days) after collection, making it impossible to collect a large number of samples at the same time. People in favor of live sorting often mention the following disadvantages of using preservatives: (1) it is more difficult to spot dead than living specimens because of the lack of movement and (2) it is not possible to use different preservatives depending on macroinvertebrate group, i.e., identification of Chironomidae and Bivalvia is less time consuming when they are preserved in ethanol compared to formaldehyde, while Oligochaeta are easier to identify when preserved in formaldehyde. The question is whether these disadvantages will significantly influence bioassessment results and/or the costs of sample processing. The aim of this study was (1) to compare bioassessment results between preserved (i.e., sorting dead specimens) and unpreserved samples (i.e., sorting living specimens), and (2) to compare sample processing costs between preserved and unpreserved samples.

**Methods**

**Study site and data collection**

For this study, we used data collected from three streams in the Netherlands: the Springendalse

beek, the Tongerensche beek, and the Swalm. Catchment areas of all streams are smaller than 100 km<sup>2</sup>, with all sites located between 0 and 200 m above sea level. We sampled the Springendalse beek in September 2002, the Tongerensche beek in June 2003, and the Swalm in April 2003. In each stream, a uniform 100-m stretch of the stream was selected for sampling. At each site, we collected six replicate composite samples, each consisting of sampling units from different habitats. In each stream, three habitats were sampled, and sample size varied between streams (Table 1); replicate samples collected from the same stream did not differ in sample size. To ensure collection of most species present in the habitat (expert judgment), we sampled each habitat that represented at least 5% of the total surface area over a set distance. Prior to sampling, the surface area covered by the different habitats was estimated at each site (Table 1). The samples were collected by pushing a pond net (25 cm × 25 cm, 500-µm mesh) through the upper part (2–5 cm) of the substrate. The sampling units from the different habitats were stored separately in buckets. Three out of six sampling units from each habitat were preserved in 4% formaldehyde directly after sampling. The buckets were transported to the laboratory, where the sampling units without formaldehyde were stored in a refrigerator, oxygenated, until sorting. All sampling units were kept separately during sample processing, which began with units being washed through 1,000- and 250-µm-mesh sieves. Live sorting was performed

**Table 1** Habitat coverage and sampled length of each habitat for the three streams sampled in this study

Stream	Habitat	Sampled length (m)	Coverage (%)
Tongerensche beek	Mud	0.5	50
	Sand	0.5	20
	Submerged vegetation	0.5	30
Swalm	Mud/detritus	0.25	5
	Gravel	0.75	75
	Sand	0.75	20
Springendalse beek	Gravel	0.5	5
	Sand	0.5	95
	Submerged vegetation	0.25	5

for the three unpreserved replicate sampling units (per stream and habitat). From the remaining three preserved sampling units, we collected dead organisms. After washing, the sampling units were poured into transparent trays and placed on a light box. According to Dutch common practice, units were sorted in their entirety and organisms picked from the trays using unaided visual guidance. Organisms were preserved in 70% ethanol, except for live Oligochaeta and Hydracarina. Live Oligochaeta were preserved in 4% formaldehyde and live Hydracarina in Koenike fluid (20% acetic acid, 50% glycerol, and 30% demineralized water). Organisms were identified to the lowest taxonomic level possible, i.e., at the species level for almost all specimens. Time spent on sorting and identification of all specimens in each sampling unit was recorded.

#### Data analysis

In total, 18 composite samples were collected from three different streams. The number of taxa and the number of individuals for each major macroinvertebrate group (e.g., Diptera, Ephemeroptera, Plecoptera) were evaluated to determine whether (potential) differences between preserved and unpreserved samples varied depending on the macroinvertebrate group. We refer to the number of taxa and the number of individuals for each major macroinvertebrate group as a metric.

The number of individuals per taxon was standardized to a total sampled length of 5 m according to formula 1.

$$T_x = \sum_{i=1}^h a_{xi} \left( \frac{l_i}{5 (c_i/100)} \right) \quad (1)$$

where  $T_x$  is the total number of individuals of taxon  $x$ ,  $a_{xi}$  is the abundance of taxon  $x$  for habitat  $i$ ,  $l_i$  is the sampled length (m) of habitat  $i$ ,  $c_i$  is the habitat coverage (%) of habitat  $i$ , and  $h$  is the total number of habitats sampled.

For the composite samples from each of the three streams, we calculated ecological quality classes. For this purpose, we used a revised version of the multimetric index described by Vlek et al. (2004). The multimetric index consists of 11 metrics and has been developed to assess the

ecological quality of small Dutch lowland streams (Vlek et al. 2004). The multimetric index assigns samples to an ecological quality class that can range from 1 (bad ecological quality) to 5 (high ecological quality or reference situation) based on a macroinvertebrate species list. The ecological quality classes were calculated with the program ASTERICS.

An ANOVA with blocks (streams) ( $\alpha = 0.05$ ) was applied to assess differences in metric values between preserved and unpreserved samples. Prior to statistical analysis, abundance data were  $\log_{10}(x + 1)$  transformed according to Brinkman and Duffy (1996) and Gowns et al. (1997). Taxa counts were not transformed, according to Kerans et al. (1992).

For the macroinvertebrate groups Gastropoda, Heteroptera, Hirudinea, Megaloptera, Odonata, Plecoptera, and Turbellaria, low numbers of specimens were collected from the samples. In some samples, these macroinvertebrate groups were not present at all. Performing a statistical test in these cases would be misleading because significant differences will not be observed simply because of few or no specimens in the samples. To avoid conclusions based on very low numbers of specimens, for analyses we used only macroinvertebrate groups with abundances higher than 0 in 17 out of 18 samples.

To determine whether non-significant results were the result of inadequate power of the study design, we performed an *a posteriori* power analysis (Peterman 1990). Power is defined as  $1 - \beta$ , or the inverse probability of committing a type II error in a statistical test. Low power indicates that little confidence should be placed in a conclusion based on a failure to reject  $H_0$ , i.e., no difference in metric values between preserved and unpreserved samples. Minimum detectable differences (MDDs) were calculated given the experimental design applied in this study using an alpha of 0.05 and power of 0.80 as commonly accepted values for significance level and power (Peterman 1990; Carlisle and Clements 1999). MDD is the effect size (expressed as the difference in metric values between preserved and unpreserved samples) that is necessary to generate acceptably high power (Rotenberry and Wiens 1985; Cohen 1988), which was considered to be 0.8 in this study.

Sample processing time (time spent on sorting and identification) was recorded for each sample. Costs of a person-hour vary, so we used the time required for sample processing as a measure of sample processing costs for preserved and unpreserved samples. Because the time required for sorting and identification strongly depends on the number of individuals sorted and identified (Barbour and Gerritsen 1996), differences between replicate samples in the number of individuals could confound results. Therefore, the recorded time was divided by the number of specimens in a sample and multiplied by the average number of individuals for all six samples from the respective stream. Data on recorded times (corrected for the number of individuals) were  $\ln(x)$  transformed according to Grows et al. (1997) prior to analysis. To test for differences in sample processing time between preserved and unpreserved samples, we performed an ANOVA with blocks (streams) ( $\alpha = 0.05$ ). Residuals were plotted against predicted values to check for normality in sample processing time (sorting and identification), and no deviations from normality were found.

**Results**

**Metrics**

In total, four of the 16 metrics showed differences ( $p < 0.05$ ) between preserved and unpreserved samples (Table 2). The number of Ephemeroptera individuals and Trichoptera taxa was consistently higher in preserved than in unpreserved samples (Table 3). The number of Hydracarina taxa and individuals was consistently lower in preserved samples (Table 3).

Power analysis revealed large differences between metrics in the required MDD (Table 2). Most metrics required an MDD of less than 50% (MDD/overall mean) to reach a power of 0.8 (Table 2). Only the number of Coleoptera taxa and Coleoptera individuals required MDDs of more than 50% (Table 2).

**Multimetric index**

The only difference in ecological quality class between preserved and unpreserved samples was detected in samples from the Swalm. One preserved

**Table 2** Summary of ANOVA results (stream = blocking factor) for comparison of preserved and unpreserved samples on three streams ( $n = 18, \alpha = 0.05$ )

Acronym	Metric description	Mean unpreserved	Mean preserved	Effect	MDD	MDD (%)	<i>p</i> value
OL-taxa	Number of Oligochaeta taxa	14.44	13.78	0.67	2.40	17.0	0.417
OL	Number of Oligochaeta individuals	2.72	2.88	0.16	0.36	13.0	0.198
BIVAL-taxa	Number of Bivalvia taxa	2.89	3.00	0.11	0.94	31.9	0.727
BIVAL	Number of Bivalvia individuals	1.81	2.21	0.39	0.68	33.7	0.104
CRUS-taxa	Number of Crustacea taxa	3.44	3.33	0.11	0.99	29.2	0.740
CRUS	Number of Crustacea individuals	3.2	3.02	0.18	0.37	11.7	0.152
EPHE-taxa	Number of Ephemeroptera taxa	3.22	3.00	0.22	1.28	41.3	0.610
EPHE	Number of Ephemeroptera individuals	1.89	1.42	0.47	0.36	21.6	0.002*
TRIC-taxa	Number of Trichoptera taxa	8.11	6.33	1.78	2.11	29.2	0.023*
TRIC	Number of Trichoptera individuals	2.15	2.27	0.12	0.26	11.7	0.194
COL-taxa	Number of Coleoptera taxa	2.33	2.11	0.22	1.56	70.0	0.673
COL	Number of Coleoptera individuals	1.07	1.21	0.14	0.65	56.8	0.528
DIP-taxa	Number of Diptera taxa	33.22	31.44	1.78	6.29	19.5	0.409
DIP	Number of Diptera individuals	3.39	3.39	0.01	0.26	7.8	0.912
HYD-taxa	Number of Hydracarina taxa	3.89	5.67	1.78	1.56	32.5	0.004*
HYD	Number of Hydracarina individuals	1.38	2.05	0.68	0.56	32.7	0.003*

Asterisks indicate significant differences between preserved and unpreserved samples. Effect = difference between the mean metric score of the unpreserved samples and the preserved samples in this study, MDD = minimum detectable difference ( $\alpha = 0.05, \text{power} = 0.8$ ), MDD (%) = minimum detectable difference/overall mean metric value

**Table 3** Mean metric values (and standard deviations) for preserved and unpreserved samples from the Springendalse beek, Swalm, and Tongerensche beek

Metric	Springendalse beek		Swalm		Tongerensche beek	
	Preserved	Unpreserved	Preserved	Unpreserved	Preserved	Unpreserved
EPHE	20 (9)	17 (6)	168 (40)	66 (13)	152 (64)	17 (7)
HYD	4 (3)	10 (5)	17 (21)	90 (39)	419 (124)	1,670 (434)
HYD-taxa	2 (2)	3 (1)	2 (2)	3 (1)	7 (1)	10 (1)
TRIC-taxa	8 (1)	6 (1)	7 (2)	6 (1)	9 (1)	7 (2)

Only metrics that showed differences ( $p < 0.05$ ) between preserved and unpreserved samples are incorporated in the table. *EPHE* = number of Ephemeroptera individuals, *HYD* = number of Hydracarina individuals, *HYD-taxa* = number of Hydracarina taxa, *TRIC-taxa* = number of Trichoptera taxa

sample indicated good ecological quality, while all unpreserved samples indicated poor ecological quality. All samples from the Tongerensche beek indicated poor ecological quality. Two preserved and two unpreserved samples from the Springendalse beek indicated good ecological quality, while one preserved and one unpreserved sample indicated high ecological quality.

#### Sample processing costs

We detected no significant difference between preserved and unpreserved samples in the total time required for sample processing ( $F = 1.64$ ,  $p = 0.221$ ). When comparing the time required for sorting and identification separately, we also detected no significant differences between preserved and unpreserved samples ( $F = 0.25$ ,  $p = 0.626$  and  $F = 1.11$ ,  $p = 0.310$ ).

#### Discussion

For most major macroinvertebrate groups (five out of eight), we detected no significant differences in the number of taxa or individuals between preserved and unpreserved samples. The required MDD for the number of Coleoptera taxa and individuals suggests our study was not adequately designed to detect significant differences with acceptable power. Carlisle and Clements (1999), however, suggested that metrics requiring MDDs of more than 50% to reach a power of 0.8 cannot possibly detect ecologically relevant changes given realistic sampling efforts. Therefore, we conclude that the low power of our study design is not relevant. As a result of

the high within-site variability in the number of Coleoptera taxa and individuals, these metrics are per definition not suited for biological assessment purposes in the case of the studied streams.

The metrics that required MDDs of less than 50% and showed no significant differences between preserved and unpreserved samples necessitate closer consideration. The question is whether these non-significant results should be considered as (1) the true absence of ecologically relevant differences between preserved and unpreserved samples or (2) as a reflection of inadequate power. To answer this question, the degree of change that is considered ecologically relevant for bioassessment purposes must be determined. This degree of change will vary depending on the method used for bioassessment and is also not an entirely scientific decision (Carlisle and Clements 1999).

For metrics that did show significant differences in values between methods, values were not always higher for the same method. Instead, the method that resulted in higher values depended on the organism group. Significantly higher numbers of Ephemeroptera individuals and Trichoptera taxa were collected from the preserved samples. The lower number of Ephemeroptera individuals collected from the unpreserved samples might have been caused by disintegration during transportation, storage, and sorting because of a lack of oxygen in the samples. Supporting this suggestion is the fact that during sorting, we often found only parts instead of complete Ephemeroptera specimens. The difference in Trichoptera taxa collected between both methods is considered an artifact. When we counted the number of species instead of the number of taxa,

two out of three preserved samples contained five species and one preserved sample contained four species. All three unpreserved samples contained four species. Significantly higher numbers of Hydracarina individuals and taxa were collected from the unpreserved samples. This finding supports the suggestion that small organisms, like Hydracarina, are easier to detect when they are moving.

Our results seem to contradict those of studies by Humphrey et al. (2000), Metzeling et al. (2003), and Nichols and Norris (2006), who found that small and cryptic taxa such as Oligochaeta, Diptera, and Hydracarina were often overlooked in unpreserved samples. However, in these studies, sample processing procedures varied for preserved and unpreserved samples, making it impossible to identify the exact cause of overlooking small and cryptic taxa in unpreserved samples. Nichols and Norris (2006) suggest that the small taxa were missed because operators sorted the unpreserved samples unaided by a microscope. However, Grows et al. (2006) showed that using magnification did not improve the efficiency of collection of small and cryptic taxa.

Some macroinvertebrate groups were not included in the analyses because they were absent from some samples. These macroinvertebrate groups may show significant differences in the number of individuals and the number of taxa between preserved and unpreserved samples in streams where they are more abundant.

In only one instance, we identified a difference in ecological quality class between preserved and unpreserved samples. This difference was the result of higher values for the metric EPT-taxa (%). Although values of individual metrics may vary between preserved and unpreserved samples, the final assessment result will not necessarily also differ between sample processing methods. Indeed, Fore et al. (2001) and Lorenz et al. (2004) showed that differences in metric values will not necessarily result in differences in the final assessment result. Differences in the final assessment result develop when metric values happen to fall near a break point in the scoring criteria (Fore et al. 2001), as observed for the metric EPT-taxa (%).

In addition, we found no significant differences between preserved and unpreserved samples in

the time required for sorting. Two possible explanations for this finding are that (1) the advantage of easier detection of moving organisms is cancelled out by the disadvantage of their being more difficult to catch, and/or that (2) differences between replicates are so large that they efface statistically significant differences between methods. The results show that differences between replicates are large, possibly because of differences in macroinvertebrate community composition between exact sampling locations or other sources of variation resulting from differences in sample processing (e.g., differences in refrigerator storage time).

Considering the results of this study, two things should be kept in mind. First, we used formaldehyde to preserve the samples, leaving the question of whether using ethanol as a preservative would have resulted in the same findings. Second, all samples were sorted in the laboratory, as is common practice in the Netherlands. The results of sorting unpreserved samples in the laboratory cannot be compared to the results of sorting samples in the field, especially given that circumstances for sorting in the field can be far from optimal (Carter and Resh 2001; Rawer-Joost 2001).

In some cases, we found a significant difference between preserved and unpreserved samples for individual metrics. When assessment is based on these individual metrics, the choice to use a preservative or not becomes relevant. This study indicates that in streams with Ephemeroptera, the preservation of samples is necessary to optimize the number of Ephemeroptera individuals collected. In streams that contain Hydracarina, the preservation of samples will result in underestimation of the number of Hydracarina taxa and individuals present. Problems arise when both groups are likely to be present in a stream and a sample processing method has to be chosen. The decision should always be made based on the system/metric(s) used for assessment. Additionally, in this study, ecological quality classes did not depend on the sample processing method used. This finding indicates that for the assessment of small Dutch lowland streams, the sample processing protocol does not require standardization in terms of sample preservation. However,

standardization of sampling and sample processing methods, including sample preservation, remains essential in case of (long-term) routine monitoring programs. Since there are limits to standardization, e.g., among different agencies and water types, we agree with Diamond et al. (1996) that it is important to document method performance characteristics through monitoring of data quality to make comparisons between monitoring programs possible.

## Conclusions

Significantly different numbers of Ephemeroptera individuals and Hydracarina taxa and individuals were collected from preserved samples compared to unpreserved samples. In assessments based on these individual metrics, standardization of sample processing will be required. In streams with Ephemeroptera, the preservation of samples is necessary to optimize the number of Ephemeroptera individuals collected. In streams that contain Hydracarina, the preservation of samples will result in an underestimation of the number of Hydracarina taxa and individuals present. In only one instance was there a difference in ecological quality between preserved and unpreserved samples, indicating that assessing small Dutch lowland streams does not require standardization of sample preservation as part of the sample processing protocol. We detected no significant differences in sample processing costs between preserved and unpreserved samples.

**Acknowledgements** This study was carried out within the STAR project funded by the European Commission, 5th Framework Program, Energy Environment and Sustainable Development, Key Action Water, Contract no. EVK1-CT-2001-00089. We are grateful for the helpful comments from Ralf Verdonschot, Karin Didderen, and anonymous reviewers. We would like to thank Tjeerd-Harm van den Hoek, Martin van den Hoorn, and Rink Wiggers for their efforts in collecting the data on which this study was based.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Bailey, R. C., Norris, R. H., & Reynoldson, T. B. (2001). Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. *Journal of the North American Benthological Society*, 20, 280–286.
- Barbour, M. T., & Gerritsen, J. (1996). Subsampling of benthic samples: A defense of the fixed-count method. *Journal of the North American Benthological Society*, 15, 386–391.
- Barton, D. R., & Metcalfe-Smith, J. L. (1992). A comparison of sampling techniques and summary indices for assessment of water quality in the Yamaska River Québec, based on benthic macroinvertebrates. *Environmental Monitoring and Assessment*, 21, 225–244.
- Braukmann, U. (2000). Hydrochemische und biologische merkmale regionaler bachtypen in Baden-Württemberg. *Oberirdische Gewässer, Gewässerökologie*, 56, 1–501.
- Brinkman, M. A., & Duffy, W. G. (1996). Evaluation of four wetland aquatic invertebrate samplers and four sample sorting methods. *Journal of Freshwater Ecology*, 11, 193–200.
- Canton, S. P., & Chadwick, J. W. (1988). Variability in benthic invertebrate density estimates from stream samples. *Journal of Freshwater Ecology*, 4, 291–297.
- Carlisle, M. D., & Clements, W. H. (1999). Sensitivity and variability of metrics used in biological assessments of running waters. *Environmental Toxicology and Chemistry*, 18, 285–291.
- Carter, J. L., & Resh, V. H. (2001). After site selection and before data analysis: Sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. *Journal of the North American Benthological Society*, 20, 658–682.
- Cheal, F., Davis, A., Grouns, J. E., Bradley, J. S., & Whittles, F. H. (1993). The influence of sampling method on the classification of wetland macroinvertebrate communities. *Hydrobiologia*, 257, 47–56.
- Chessman, B. C. (1995). Rapid river assessment using macroinvertebrates: A procedure based on habitat-specific family level identification and a biotic index. *Australian Journal of Ecology*, 20, 122–129.
- Chessman, B. C., & Robinson, D. P. (1987). Some effects of the 1982–83 drought on water quality and macroinvertebrate fauna in the lower La Trobe River, Victoria. *Australian Journal of Marine and Freshwater Research*, 38, 289–299.
- Cohen, J. (1988). *Statistical analysis for the behavioural sciences* (2nd ed.). Hillsdale: Erlbaum.
- Courtemanch, D. L. (1996). Commentary on the subsampling procedures used for rapid bioassessments. *Journal of the North American Benthological Society*, 15, 381–385.
- Diamond, J. M., Barbour, M. T., & Stribling, J. B. (1996). Characterizing and comparing bioassessment methods and their results: A perspective. *Journal of the North American Benthological Society*, 15, 713–727.



- Drake, C. M., & Elliott, J. M. (1982). A comparative study of three air-lift samplers used for sampling benthic macro-invertebrates in rivers. *Freshwater Biology*, *12*, 511–533.
- Fore, L. S., Paulsen, K., & O' Laughlin, K. (2001). Assessing the performance of volunteers in monitoring streams. *Freshwater Biology*, *46*, 109–123.
- Growns, J. E., Chessman, C. C., Jackson, J. E., & Ross, D. G. (1997). Rapid assessment of Australian rivers using macroinvertebrates: Cost and efficiency of 6 methods of sample processing. *Journal of the North American Benthological Society*, *16*, 682–693.
- Growns, I., Schiller, C., O'Conner, N., Cameron, A., & Gray, B. (2006). Evaluation of four live-sorting methods for use in rapid biological assessments using macroinvertebrates. *Environmental Monitoring and Assessment*, *117*, 173–192.
- Haase, P., Pauls, S., Sundermann, A., & Zenker, A. (2004). Testing different sorting techniques in macroinvertebrate samples from running water. *Limnologia*, *34*, 366–378.
- Hawkes, H. A. (1979). Invertebrates as indicators of river water quality. In A. James & L. Evison (Eds.), *Biological indicators of water quality* (pp. 2-1–2-45). New York: Wiley.
- Hellawell, J. M. (1986). *Biological indicators of freshwater pollution and environmental management*. London: Elsevier.
- Humphrey, C. L., Storey, A. W., & Thurtell, L. (2000). AUSRIVAS: Operator sample processing errors and temporal variability—implications for model sensitivity. In J. F. Wright, D. W. Sutcliffe, & M. T. Furse (Eds.), *Assessing the biological quality of freshwaters: RIVPACS and other techniques* (pp. 143–146). Cumbria: Freshwater Biological Association.
- 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.
- Lenat, D. R., & Resh, V. H. (2001). Taxonomy and stream ecology—the benefits of genus and species-level identification. *Journal of the North American Benthological Society*, *20*, 287–298.
- Lorenz, A., Kirchner, L., & Hering, D. (2004). “Electronic subsampling” of macrobenthic samples: How many individuals are needed for a valid assessment result? *Hydrobiologia*, *16*, 299–312.
- Mackey, A. P., Cooling, D. A., & Berrie, A. D. (1984). An evaluation of sampling strategies for qualitative surveys of macro-invertebrates in rivers, using pond nets. *Journal of Applied Ecology*, *21*, 515–534.
- Metzeling, L., & Miller, J. (2001). Evaluation of the sample size used for the rapid bioassessment of rivers using macroinvertebrates. *Hydrobiologia*, *444*, 159–170.
- Metzeling, L., Chessman, B., Hardwick, R., & Wong, V. (2003). Rapid assessment of rivers using macroinvertebrates: The role of experience, and comparisons with quantitative methods. *Hydrobiologia*, *510*, 39–52.
- Nichols, S. J., & Norris, R. H. (2006). River condition assessment may depend on the sub-sampling method: Field live-sort versus laboratory sub-sampling of invertebrates for bioassessment. *Hydrobiologia*, *572*, 195–213. doi:10.1007/s10750-006-0253-6.
- Nijboer, R. C., & Verdonschot, P. F. M. (2000). Taxonomic adjustment affects data analysis: An often forgotten error. *Verhandlungen Internationale Vereinigung für Theoretische und Angewandte Limnologie*, *27*, 2546–2549.
- Norris, R. H., McElravy, E. P., & Resh, V. H. (1992). The sampling problem. In P. Calow & G. E. Petts (Eds.), *Rivers handbook* (pp. 282–306). Oxford: Scientific Publications.
- Peterman, R. M. (1990). The importance of reporting statistical power: The forest decline and acidic deposition example. *Ecology*, *71*, 2024–2027.
- Rawer-Joost, C. (2001). *Eignung und Variabilität von Verfahren zur ökologischen Bewertung von Fließgewässern im Mittelgebirge auf der Basis autökologischer Kenngrößen des Makrozoobenthos*. PhD thesis, University of Hohenheim.
- Rotenberry, J. T., & Wiens, J. A. (1985). Statistical power analysis and community-wide patterns. *American Naturalist*, *125*, 164–168.
- Vinson, M. R., & Hawkins, C. P. (1996). Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. *Journal of the North American Benthological Society*, *15*, 392–399.
- Vlek, H. E., Šporka, F., & Krno, I. (2006). Influence of macroinvertebrate sample size on bioassessment of streams. *Hydrobiologia*, *566*, 523–542. doi:10.1007/s10750-006-0074-7.
- Vlek, H. E., Verdonschot, P. F. M., & Nijboer, R. C. (2004). Towards a multimetric index for the assessment of Dutch streams using benthic macroinvertebrates. *Hydrobiologia*, *516*, 173–189.