VARIABILITY IN THE IDENTIFICATION AND ENUMERATION OF MARINE BENTHIC INVERTEBRATE SAMPLES AND ITS EFFECT ON BENTHIC ASSESSMENT MEASURES

J. ANANDA RANASINGHE[1*], DAVID E. MONTAGNE[2], STEPHEN B. WEISBERG[1]
MARY BERGEN[1,4] AND RONALD G. VELARDE[3]

[1]*Southern California Coastal Water Research Project, 7171 Fenwick Lane, Westminster, CA 92683-5218, USA;* [2]*Los Angeles County Sanitation Districts, P.O. Box 4998, Whittier, CA 90607, USA;* [3]*City of San Diego, 4918 N. Harbor Drive, Suite 101, San Diego, CA 92106, USA;* [4]*Present address: P.O. Box 729, Ojai, CA 93024-0729, USA.*

[*]*(author for correspondence, email: anandar@sccwrp.org)*

Abstract: Studies designed to measure anthropogenic impacts on marine benthic communities depend on the ability of taxonomists to consistently discriminate, identify, and count benthic organisms. To quantify errors and discrepancies in identification and enumeration, 20 samples were completely reprocessed by another one of four participating laboratories. Errors were detected in 13.0% of the data records, affecting total abundance by 2.1%, numbers of taxa by 3.4%, and identification accuracy by 4.7%. Paired t-tests were used to test for differences in the Benthic Response Index (BRI), total abundance, numbers of taxa, and the Shannon-Wiener index between the original and the reanalysis data. Differences in the BRI were statistically insignificant. Although statistically significant differences were observed for numbers of taxa, total abundance, and the Shannon-Wiener index, the differences were small in comparison to the magnitude of differences typically observed between anthropogenically affected and reference sites.

Keywords: marine benthic invertebrates, identification, enumeration, quality assurance, quality control, inter-laboratory calibration, benthic assessment

1. Introduction

Many approaches to benthic community assessment depend on accurate species identification. Measures based on pollution tolerance ratings of the species in each sample, such as those proposed by Eaton *et al*. (2001) and Smith *et al*. (2001), are most dependent. Ordination and cluster analyses based on species abundance data also require accurate species separation. Many recently developed multi-metric benthic indices include species richness (numbers of species) or diversity (Engle *et al*., 1994; Weisberg *et al*., 1997; Engle and Summers, 1999; Van Dolah *et al*., 1999; Paul *et al*., 2001). Traditional methods such as the Abundance Biomass Comparison (ABC) method (Warwick *et al*., 1987) and diversity indices (Washington, 1984) also depend on accurate species identification.

Several steps in benthic sample processing can introduce laboratory error into measurements of species richness and diversity. First, under-counts of organisms and species can result from failure to remove all organisms from the sediment during sorting. Second, misidentification of sorted organisms can cause underestimates of species richness if similar species are not distinguished, or over-counts if a single species is erroneously divided. Errors are also introduced if identified organisms are miscounted.

Environmental Monitoring and Assessment **81**: 199–206, 2003.

Most laboratories resort a subset of samples to quantify and minimize sorting error, but quality assurance practices to ensure taxonomic accuracy are typically less well developed although Ellis, (1988) identified the need. Many laboratories maintain collections of voucher specimens that are sent to outside experts for confirmation of identifications, but voucher specimens are typically limited to the best specimens for each species. There is also no guarantee that voucher materials accurately represent all specimens reported under a name. Inter-laboratory calibrations, like those of Ellis and Cross, (1981), may provide better quality assurance. Here we present the results of a calibration exercise in which four laboratories reprocessed samples to evaluate taxonomic and counting consistency and their effects on measures used in benthic assessments.

2. Methods

Samples were obtained from a regional survey of benthic infauna in the Southern California Bight (Bergen *et al*., 2000). The samples were collected in August and September 1994 with a $0.1m^2$ Van Veen grab, sieved through 1mm screens, relaxed for 30 min in $MgSO_4$ or propylene phenoxytol, fixed in sodium borate buffered 10% formalin, and transported to four laboratories for identification and enumeration analysis.

Taxonomic accuracy was assessed from twenty samples selected at random. Each sample was analyzed at one of the four laboratories and reanalyzed at another, selected at random. Taxonomists performing reanalysis had no access to original analysis results. When reanalysis was complete, we compared the original and reanalysis data and compiled a list of differences. These differences were classified as errors when they were caused by inaccurate identifications, incorrect counts, or specimens overlooked in the original analysis. They were classified as discrepancies, rather than errors, when they resulted from use of a junior synonym or other unconventional nomenclature, failure to note removal of specimens for vouchers, or differences in opinion about the taxonomic level to which an organism could be identified (e.g., *Polydora* sp. vs. *Polydora narica*). For each sample, error rates for total abundance and numbers of taxa were calculated as the ratio of the difference between the original and resolved values to the resolved value, expressed as a percentage; original values greater than resolved values resulted in negative rates. Error rates for identification accuracy were calculated as the ratio of misidentifications to resolved identifications, also expressed as a percentage. The resolved value represented "truth" by consensus agreement between the original and reanalysis taxonomists.

To assess the effects of differences in laboratory results on benthic assessments, the number of taxa per sample, total abundance, Shannon-Wiener diversity, and the Benthic Response Index (Smith *et al*., 2001) were calculated for the original and the reanalysis data. Paired t-tests were then used to test for differences between the original and the re-analyzed data.

3. Results

Differences between original and reanalysis data were detected in 25.3% of the data records (Table 1), where a record consists of a taxon and its reported abundance. The differences were nearly equally divided between errors and discrepancies. Miscounts affected 4.8% of the data records and were the most common type of error; misidentifications (4.5%) and overlooked specimens (3.3%) were almost as common. The errors yielded total abundances, numbers of taxa and identifications that differed from "truth", on average, by 2.1%, 3.4% and 4.7%, respectively (Table 2).

Total abundance, numbers of taxa, and Shannon-Wiener index values were significantly different between the original and reanalysis data (Table 3). Mean differences between the original analysis and reanalysis were 4.75 per sample for total abundance, 2.25 per sample for number of taxa, and 0.037 for the Shannon-Wiener index. Differences in the Benthic Response Index were small and not statistically significant.

Table 1. Frequencies of differences in identification and enumeration for 20 samples. The data comprised 1,715 records each consisting of a taxon and its reported abundance. Negative values indicate that the net result of the error was an understatement of the true value; positive values indicate overstatement of the true value.

	Type of Difference	*Number*	*% of Differences*	*% of Records*
Errors	Miscount	83	19.1	4.8
	Mididentification	78	18.0	4.5
	Overlooked specimen (s)	57	13.1	–3.3
	Misapplication of identificationrules	5	1.2	0.3
	Total Errors	223	51.4	13.0
Discrepancies	Judgment differences	131	30.2	7.6
	Specimen loss or unrecorded voucher removal	57	13.1	3.3
	Unconventional nomenclature	23	5.3	1.3
	Total Discrepancies	211	48.6	12.3
	Total Differences	**434**		**25.3**

Table 2. Means (and ranges) of error rates for total abundance, numbers of taxa and identification accuracy.

Original Analysis Laboratory	*Number of Reanalyzed Samples*	*Mean Error Rate (%)*		
		Total Abundance	*Number of Taxa*	*Identification Accuracy*
A	6	3.1 (2.2 – 6.1)	4.8 (2.9 – 5.9)	6.9 (4.3 – 10.5)
B	2	1.0 (0.3 – 1.5)	1.8 (1.2 – 2.3)	3.6 (2.3 – 5.0)
C	6	2.2 (0 – 3.1)	4.5 (1.0 – 9.2)	3.0 (0 – 4.3)
D	6	1.5 (–1.2 – 4.9)	1.1 (0 – 2.0)	4.6 (2.0 – 11.7)
All	20	2.1 (–1.2 – 6.1)	3.4 (0 – 9.2)	4.7 (0 – 11.7)

Table 3. Mean values for the original and reanalysis data and, based on paired t-tests, the probability that the difference > 0. NS = Not significant.

Assessment Measure	*Mean Values*			*Probability*
	Original Data	*Reanalysis Data*	*Difference*	
Total Abundance (per sample)	334.25	339.00	4.75	0.040
Number of Taxa (per sample)	80.90	83.15	2.25	0.004
Shannon-Wiener Diversity Index	5.162	5.199	0.037	0.006
Benthic Response Index	14.242	14.238	0.004	0.979^{NS}

4. Discussion

The data presented in Tables 1 and 3 provide, for the first time, rates at which errors and discrepancies occur when different taxonomists process samples and their effect on commonly used assessment measures. While the percentage of records affected was high, the overall affect on conclusions that would be reached by application of the data for assessment purposes was not large. The observed differences of 4.75, 2.25 and 0.037 in abundance, numbers of taxa and the Shannon-Wiener index were small relative to the sample means of 334.25, 80.9 and 5.162, respectively. Moreover, the differences were much smaller than the differences typically observed between anthropogenically affected and reference sites (Weisberg *et al*., 1997; Van Dolah *et al*., 1999). The Benthic Response Index (BRI), which is the abundance-weighted pollution tolerance of the species in the sample (Smith *et al*., 2001), was not affected by the observed differences. Apparently the BRI is robust to minor taxonomic errors, presumably because similar species have similar pollution tolerance values.

Two factors contributing to the small values of among-laboratory differences are the experience of the taxonomists and the extent of communication among them. Most of the taxonomists have more than two decades of experience working on the benthos of southern California. Moreover, recognizing the need for consistency, in 1982, taxonomists in southern California formed an organization, the Southern California Association of Marine Invertebrate Taxonomists (SCAMIT, http://www.scamit.org), that is dedicated to standardizing taxonomy in the region. SCAMIT publishes lists of accepted nomenclature, maintains reference collections, produces keys and other taxonomic tools, and fosters communication among its members through monthly workshops, newsletters, a web site and e-mail lists. Throughout the regional survey from which these samples were drawn, and independent of the quality assurance exercise presented here, descriptions and figures of unusual or unknown organisms were distributed using these communication mechanisms to keep other taxonomists abreast of developments as they occurred. Error rates in regions where this communication does not exist may be higher.

Another factor contributing to our low error rates was distinguishing between discrepancies and errors. One type of discrepancy was the use of unconventional

nomenclature. We chose not to count this as an error because trained taxonomists easily recognize and correct synonymous names while merging data from different sources; however, this is an issue of concern because not all data users have the expertise to recognize synonyms when merging data sets. Although participating taxonomists made efforts to minimize these problems by relying on SCAMIT lists, discrepancies still occurred suggesting that larger numbers occur in areas where there is no authoritative standardization of nomenclature.

Of greater concern was the prevalence of instances where organisms were identified to different taxonomic levels in analysis and reanalysis. Some of these discrepancies resulted from differences of opinion whether condition of a specimen was sufficient for species-level identification while others reflected differences in experience of the taxonomists. We chose to classify these instances as discrepancies, rather than errors, because data can be lumped to the higher taxonomic level before analysis. However, lumping may affect measures of species richness and diversity (Wu, 1982; Wilson and Jeffrey, 1994) and this is of concern when richness and diversity data are compared to threshold values to infer condition of the benthos, as in the case of B-IBI measures (Weisberg *et al*., 1997; Van Dolah *et al*., 1999). Interpretations of assessment results are distorted if the level of taxonomy differs from the level used while developing thresholds.

This issue is of particular concern because taxonomy improves over time and, therefore, benchmarks developed from early data should be re-evaluated over time. Our regional surveys provide an example of improving taxonomy. In the survey providing data for these analyses, it was necessary to lump 43 taxa due to taxonomic uncertainty (Table 4). In a subsequent regional survey involving the same group of taxonomists it was only necessary to lump 16 taxa (Table 4). Our list does not include taxa that were recognized from the beginning as impractical to identify to species at the current state of knowledge; rather, it lists groups that the taxonomists thought they were identifying consistently and accurately but, after reanalysis and review of species lists and specimens, it became clear that they were not. The improvement between surveys was achieved in two ways. First, "failures" stimulated production of new keys and other identification tools by SCAMIT. Second, in the subsequent survey, all four laboratories referred a few groups (ceriantharian and edwardsiid anemones and euclymeninaen and lumbrinerid polychaetes) to single "specialty taxonomists" for identification and enumeration. Despite the efforts of SCAMIT, these groups still presented obstacles to consistent treatment unless one taxonomist identified all of them. The challenge is to ensure consistency between assessment tools and levels of taxonomy when the tools are applied.

Our study explored sources of variability and error often ignored when interpreting the results of benthic assessments (Ellis, 1988). Most importantly, our results provide a standard against which subsequent efforts may be judged. By integrating reanalysis by sample exchange among external taxonomists into quality assurance plans, multi-laboratory monitoring programs can greatly increase the

Table 4. Level of taxonomic identification assigned after re-identification of specimens and inter-laboratory comparison of data. Bolded entries indicate taxonomic groups affected (all or in part) by reliance on a single taxonomist for identification in 1998.

Group	*Name Adopted after Interlaboratory Comparison*	*Level*	*Number of Taxa Combined* 1994	1998
PHYLUM CNIDARIA, Class Anthozoa	**Ceriantharia**	**Order**	**4**	
	Actiniaria	**Order**	**5**	
Order Pennatulacea	*Acanthoptilum* spp.	Genus	3	3
PHYLUM NEMERTEA, Class Anopla	Anopla	Class	7	3
	Paleonemertea	Order	2	3
Class Enopla, Order Hoplonemertea	Hoplonemertea	Order	3	
	Lineidae	Family	12	9
	Amphiporus spp.	Genus	4	4
	Tetrastemma spp.	Genus	2	
PHYLUM MOLLUSCA, Class Aplacophora				
Order Aplacophora	Chaetodermatidae	Family	5	3
Class Gastropoda, Order	*Bittium* spp.	Genus	2	
Megagastropoda	*Lirobittium* spp.	Genus		3
	Asperiscala spp.	Genus	2	
	Nitidiscala spp.	Genus	2	
	Crepidula spp.	Genus	4	
Order Neogastropoda	*Ophiodermella* spp.	Genus	3	
Class Bivalvia, Order Veneroida	*Solen* spp.	Genus	2	
Order Myoida	*Corbula* spp.	Genus	2	
Order Septibranchida	*Cardiomya* spp.	Genus	2	
PHYLUM ANNELIDA, Class Polychaeta				
Order Orbiniida	*Levinsenia* spp.	Genus	3	
	Paradoneis spp.	Genus		4
Order Cossurida	*Cossura* spp.	Genus	2	
Order Spionida	*Boccardia* spp.	Genus		2
	Protocirrineris spp.	Genus	2	
	Monticellina spp.	Genus	5	
	Mediomastus spp.	Genus	3	3
Order Capitellida	***Clymenella* spp.**	**Genus**	**3**	
	Maldanidae	**Family**	**11**	
Order Opheliida	*Ophelina* spp.	Genus	2	
Order Phyllodocida	*Eusyllis* spp.	Genus		3
	Harmothoinae	Subfamily	15	
	Sthenelais spp.	Genus	3	
	Sphaerosyllis spp.	Genus	2	
Order Eunicida	***Lumbrineris* spp.**	**Genus**	**15**	
	Drilonereis spp.	Genus	3	
	Dorvillea (S.) spp.	Genus		3
	Arabella spp.	Genus		2
	Nothria spp.	Genus		2
Order Fauveliopsida	*Fauveliopsis* spp.	Genus	3	
Order Terebellida	*Terebellides* spp.	Genus	2	
Order Sabellida	*Demonax* spp.	Genus	2	
	Bispira spp.	Genus	2	2

continued on following page

Table 4. Continued

Group	*Name Adopted after Interlaboratory Comparison*	*Level*	*Number of Taxa Combined*	
			1994	*1998*
PHYLUM ARTHROPODA, Class Malacostraca				
Order Leptostraca	*Nebalia* spp.	Genus	3	
Order Isopoda	*Edotia* spp.	Genus	2	
	Synidotea spp.	Genus	2	
Order Amphipoda	*Aorides* spp.	Genus	6	
	Corophium spp.	Genus		3
	Photis spp.	Genus	4	
	Protomedeia spp.	Genus	2	
	Synchelidium spp.	Genus		3
PHYLUM ECHINODERMATA	Holothuroidea	Class	2	
PHYLUM CHORDATA	Ascidiacea	Class	4	

likelihood of producing results that are accurate and comparable. The levels of error measured in this study provide the first available data points about variability in identification and abundance measures during multi-laboratory taxonomic analysis. As additional data about these errors are accumulated they can be incorporated as targets and limits in quality assurance and quality control programs to ensure that laboratory data quality are maintained.

Acknowledgments

The authors gratefully acknowledge the contributions of Kelvin Barwick, Cheryl Brantley, Donald Cadien, Nancy Carder, Ann Dalkey, Douglas Diener, Thomas Gerlinger, Kathy Langen, Megan Lilly, John Ljubenkov, Lawrence Lovell, Ricardo Martinez-Lara, Thomas Parker, Dean Pasko, Charles Phillips, William Power, James Roney, Rick Rowe, Timothy Stebbins and Ronald Velarde, the participating taxonomists; their efforts and specialized expertise made this study possible. The Southern California Association of Marine Invertebrate Taxonomists (SCAMIT) provides a mechanism for standardizing the names of organisms on the mainland shelf of southern California and promotes communication among taxonomists; they were an integral part of this study. We appreciate the perseverance of the participants in the many meetings that culminated in the quality assurance plan and the many more that were required to carry it through. We also thank Virginia Engle, Dr. Daniel M. Dauer, Dr. Jack Word, and Dr. Derek Ellis who reviewed our manuscript. Finally, we acknowledge Dr. Derek Ellis, who has long recognized the importance and need for formal quality control and quality assessment of biological data that are used to assess anthropogenic impacts on marine invertebrate communities.

References

Bergen, M., Cadien, D., Dalkey, A., Montagne, D. E., Smith, R. W., Stull, J. K., Velarde, R. G. and Weisberg, S. B.: 2000, 'Assessment of benthic infaunal condition on the mainland shelf of southern California', *Environmental Monitoring and Assessment* **64**, 421–434.

Eaton, L.: 2001, 'Development and validation of biocriteria using benthic macroinvertebrates for North Carolina estuarine waters', *Marine Pollution Bulletin* **42**, 23–30.

Ellis, D. V.: 1988, 'Quality control of biological surveys', *Marine Pollution Bulletin* **19**, 506–512.

Ellis, D. V., and Cross, S. F.: 1981, 'A protocol for inter-laboratory calibrations of biological species identifications (Ring Tests)', *Water Research* **15**, 1107–1108.

Engle, V. D., Summers, J. K. and Gaston, G. R.: 1994, 'A benthic index of environmental condition of Gulf of Mexico estuaries', *Estuaries* **17**, 372–384.

Engle, V. D. and Summers, J. K.: 1999, 'Refinement, validation, and application of a benthic condition index for Northern Gulf of Mexico estuaries', *Estuaries* **22**, 624–635.

Paul, J. F., Scott, K. J., Campbell, D. E., Gentile, J. H., Strobel, C. S., Valente, R., Weisberg, S. B., Holland, A. F. and Ranasinghe, J. A.: 2001, 'Developing and applying a benthic index of estuarine condition for the Virginian Biogeographic Province', *Ecological Indicators* **1**, 83–99.

Smith, R. W., Bergen, M., Weisberg, S. B., Cadien, D., Dalkey, A., Montagne, D., Stull, J. K. and Velarde, R. G.: 2001, 'Benthic response index for assessing infaunal communities on the Southern California mainland shelf', *Ecological Applications* **11**, 1073–1087.

Van Dolah, R. F., Hyland, J. L., Holland, A. F., Rosen, J. S. and Snoots, T. R.: 1999, 'A benthic index of biological integrity for assessing habitat quality in estuaries of the southeastern USA', *Marine Environmental Research* **48**, 269–283.

Warwick, R. M., Pearson, T.H. and Ruswahyuni: 1987, 'Detection of pollution effects on marine macrobenthos: further evaluation of the species abundance/biomass method', *Marine Biology* **95**, 193–200.

Washington, H. G.: 1984, 'Diversity, biotic and similarity indices: a review with special relevance to aquatic ecosystems', *Water Resources* **19**, 653–694.

Weisberg, S. B., Ranasinghe, J. A., Dauer, D. M., Schaffner, L. C., Diaz, R. J. and Frithsen, J. B.: 1997, 'An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay', *Estuaries* **20**, 149–158.

Wilson, J. G. and Jeffrey, D. W.: 1994, 'Benthic biological pollution indices in estuaries', in: Kramer, K. J. M. (ed), Biomonitoring of coastal waters and estuaries, CRC Press, Boca Raton, FL, pp. 311–327.

Wu, R. S. S.: 1982, 'Effects of taxonomic uncertainty on species diversity indices', *Marine Environmental Research* **6**, 215–225.