Documentation of Method Performance Characteristics

For the
Anne Arundel County
Biological Monitoring Program

Prepared for:

Anne Arundel County
Department of Public Works
Watershed, Ecosystem, and Restoration Services
2662 Riva Road
Annapolis, MD 21401

Originally Prepared by:

Tetra Tech, Inc.
400 Red Brook Blvd, Suite 200
Owings Mills, MD 21117-5159

Revised by:

KCI Technologies
936 Ridgebrook Rd.
Sparks, MD 21152

August 2011

ACKNOWLEDGEMENTS

The principal authors of the original 2005 report are Colin R. Hill, James B. Stribling, and Carolina Gallardo, all of Tetra Tech. The principal authors of this revision are Colin R. Hill and Michael J. Pieper of KCI Technologies. This report presents results of a performance characteristic evaluation study to set data quality and measurement quality objectives for the Anne Arundel County Biological Monitoring and Assessment Program.

The appropriate citation for this report is:

Hill, C.R., and M. J. Pieper. 2010. *Documentation of Method Performance Characteristics for the Anne Arundel County Biological Monitoring Program*. Revised, December 2010. Prepared by KCI Technologies, Sparks, MD for Anne Arundel County, Department of Public Works, Watershed, Ecosystem, and Restoration Services. Annapolis, MD.

For more information, please contact:

Mr. Christopher J. Victoria
Anne Arundel County
Department of Public Works
Watershed, Ecosystem, and Restoration Services
2662 Riva Road
Annapolis, Maryland 21401
410-222-4240
PWVICT16@aacounty.org

TABLE OF CONTENTS

ABSTRACT

A small-scale study was conducted in 2005 to demonstrate the overall performance and capability of the Anne Arundel County biological monitoring program, to allow it to be directly compared to other monitoring programs and datasets, and to define overall data quality expectations. The small-scale study consisted of a total of six sites, which were sampled for benthic macroinvertebrates on Cox Branch, just south of Crofton, MD. From the resulting bioassessment data, five performance quality characteristics (precision, accuracy, bias, representativeness, and completeness) were evaluated either quantitatively or qualitatively. These characteristics were evaluated for each of six components making up the biological assessment protocol for Anne Arundel County: field sampling, laboratory sorting and subsampling, taxonomic identification and enumeration, data entry, metric calculation, and site assessment.

From the results of the initial performance characteristic evaluation, quantitative measurement quality objectives (MQOs) were set for each of the six biological assessment components. Numeric MQOs were established for several different measures of precision including relative percent difference (RPD), root mean square error (RMSE), and coefficient of variability (CV) for benthic macroinvertebrate metrics and overall index scores. Performance characteristics calculated from annual sampling results during Round One (2005 – 2008), show that MQOs for benthic macroinvertebrate metrics and overall index scores have not been consistently achievable. As a result, performance characteristics were calculated using a larger data set (n = 24 sample pairs) collected throughout the County from 2004 – 2008. Numeric MQOs for metric and index scores were revised to reflect the increased variability encountered when sampling across a larger scale and among sites exhibiting a broader range of impairment.

Since MQOs for laboratory subsampling and taxonomic components were demonstrated to be readily achievable, these remained unchanged from the original report (i.e., percent sort efficiency (PSE; ≥90%), percent taxonomic disagreement (PTD; ≤15%), specimen enumeration (≤5%), and completeness of taxonomic identifications (≥90%)). MQOs for both data entry and metric calculation accuracy also remained unchanged (i.e., 100% following re-checks and corrections). Minor changes were made to site assessment MQOs with mRPD <20% and/or RMSE >0.6 for B-IBI scores, the 90% confidence interval of the B-IBI for site assessments was set at <0.96, and completeness remained unchanged at ≥95%. While these MQOs are intended for use as indicators of data quality, exceedence of thresholds does not imply the need to exclude the sample(s) from further analysis. Rather, data for the sample(s) should be examined in detail to determine causes of the exceedence. Decisions and corrective actions made for identified error sources should be a guide for potential dataset-wide modifications.

1. INTRODUCTION

As a supplement to the Anne Arundel County biological monitoring and assessment program, and to provide a measure of data quality (i.e., the reliability of these assessments), this report presents performance characteristics for the various field and laboratory standard operating procedures (SOPs) that, as a whole, make up the biological stream assessment protocol. This document is intended to provide a guideline for ongoing data quality assessments associated with the County's routine biological monitoring and to help enhance defensibility of data and assessments. This report presents revised performance characteristics and measurement quality objectives for benthic macroinvertebrate metric and index scores, based on calculations performed using a larger data set of replicate samples (n = 24 sample pairs compared to n = 6 sample pairs in previous study) collected throughout the County during Round One (2004-2008) of the Countywide Biological Monitoring and Assessment Program.

Overall variability of data is the combined result of accumulated error from multiple sources within any measurement system (Taylor 1988, Clark and Whitfield 1994, Taylor and Kuyatt 1994, and Diamond et al. 1996). As a general rule, error can be divided into two categories: systematic and random. The error associated with a particular method is known as systematic error, which can, to a certain extent, be controlled by using an appropriate quality assurance program. The error that results from the sample itself or the population from which it is derived is known as random error. Unlike systematic error, random error can only partly be controlled through a careful sampling design. However, it is often impossible to separate the effects of the two types of error, and they can directly influence each other (Taylor 1988). The overall degree of combined error associated with a dataset is known as data quality. Statements of data quality are critical for data users and decision makers to properly evaluate the extent to which they should rely on technical, scientific information (Costanza et al. 1992).

A major goal of the County is to produce biological assessments using objective and defensible data. Consequently, a comprehensive Quality Assurance Project Plan (QAPP) for ensuring the collection of such data was developed simultaneously with the countywide monitoring program (Anne Arundel County 2004a,b), and later revised to reflect updates and changes to the sampling and analysis procedures (Anne Arundel County 2010). The QAPP followed U. S. Environmental Protection Agency requirements for developing project plans (USEPA 1995) and describes the biological stream assessment protocol including documented standard operating procedures (SOPs) for data collection, the technical rationale behind the procedures, and the series of activities and reporting procedures that are used to document and communicate data quality, with an inherent goal of minimizing systematic error.

Beyond the requirements associated with the QAPP, it is necessary to determine if the processes and procedures detailed in the QAPP are being executed during day-to-day program activities such that data collected by these methods meets DQOs detailed in the program design document (Hill and Stribling 2004). The development of MQOs is one way of ensuring data of sufficient high quality to meet program objectives are collected

by the program, thereby reducing systematic error. MQOs represent specific numeric and statistical targets for various aspects of the data collection operations of the program. The data collected in a given sampling period are examined to see if these targets are met. Data that do not reach the MQOs are not necessarily discarded, but non-compliant data should be examined more closely, resulting in a more objective evaluation of data quality. If it is clear that MQOs were not met due to improper data collection procedures or faulty equipment, data would likely be discarded. On the other hand, if it appears that MQOs were not met due to unusual site conditions or other circumstances attributed to random variability, the data do not need to be discarded. This performance characteristics framework is based on that being developed by the Methods and Data Comparability Board of the National Water Quality Monitoring Council (MDCB 2004 [draft]).

During the original performance characteristic evaluation study, two adjacent reaches with overall similar physical habitat and drainage area characteristics (i.e. duplicate sites) were sampled in each of three stream segments of Cox Branch, and the Maryland benthic index of biological integrity (B-IBI) (Stribling et al. 1998) calculated for each. Using metric and index results, performance characteristics were documented, and from them, measurement quality objectives (MQO) were recommended for judging the acceptability of data for the use in Anne Arundel County. However, performance characteristics calculated from annual sampling results during Round One (2005 – 2008), show that MQO's for benthic macroinvertebrate metrics and overall index scores have not been consistently achievable. As a result, performance characteristics were re-calculated using a larger data set (n = 24 sample pairs) collected throughout Anne Arundel County from 2004 – 2008. Numeric MQOs for metric and index scores were revised to reflect the increased variability encountered when sampling across a larger scale and among sites exhibiting a broader range of impairment. Since MQOs for laboratory subsampling and taxonomic components were demonstrated to be readily achievable, performance characteristics were not re-evaluated using the Round One data set.

The purpose of this revision is twofold: 1) to document MQOs for B-IBI metrics that have been revised for the current Coastal Plain B-IBI as developed by Southerland et al. (2005); and 2) to revisit the performance characteristics for field sampling precision using a larger, more comprehensive (i.e., Countywide scale) dataset of duplicate (replicate) samples following the completion of Round One of the Countywide Biological Monitoring and Assessment Program in 2008. This dataset provides a more robust estimate of the variability encountered throughout the County, whereas the original dataset was both spatially and temporally homogeneous and lacked a range of impairment that is more representative of the population at the Countywide scale.

It should also be emphasized that sections 2.2 through 2.4 have not been revised, and all results discussed in these sections are based on the findings of the initial 2005 study.

Organization of this report

This report addresses the documentation of data quality, by describing the data in terms of five quality characteristics: 1) precision, or nearness of multiple measures to one

another; 2) accuracy, or nearness of a measure to a known, specified analytical truth; 3) bias, or tendency to systematically favor one outcome over another; 4) representativeness, or ability to collect a representative sample of a population; and 5) completeness, or wholeness of a dataset. Some of these were assessed quantitatively and others were qualitatively evaluated, while others were not applicable to a particular method. These characteristics are presented for six components of the biological assessment process: field sampling, laboratory sorting and subsampling, taxonomic identification and enumeration, data entry, metric calculation, and site assessment. Because detailed descriptions of methods are provided in the QAPP and in SOPs, only specific critical methodological information is presented here. Following the introduction, this report is divided into two sections. Section 2 discusses the calculation and documentation of performance characteristics and is organized according to Table 1. If indicated as not applicable (na), a performance characteristic is not listed in this section, unless an explanation is required. Section 3 provides recommendations for quantitative and qualitative MQOs. Metric and index values and scores for each sample pair are displayed in Appendix A.

Table 1. Error partitioning framework for biological assessment protocols. Performance characteristics may be quantitative (QN), qualitative (QL), or not applicable (na).

	Performance Characteristics				
Component Method or Activity	**Precision**	**Accuracy**	**Bias**	**Representativeness**	**Completeness**
1. Field sampling	QN	na	QL	QL	QN
2. Laboratory sorting/subsampling	na	na	QN	QL	na
3. Taxonomy	QN	QL	QL	na	na
4. Enumeration	QN	QL	QL	na	na
5. Data entry	na	QN	na	na	na
6. Metric calculation (e.g., data reduction)	na	QN	na	na	na
7. Site assessment	QN/QL	QN	QL	QL	QN

Brief overviews of all methods are given in Section 2 for each protocol component (Table 1). Field sampling methods are those used by the Maryland Biological Stream Survey (MBSS) (Kazyak 2001, DNR 2007). Sorting/subsampling and taxonomy are similar to MBSS (Boward and Friedman 2000) but with several differences as outlined in the County's QAPP (Anne Arundel County 2004a). Data management and metric calculation were done primarily using the Ecological Data Application System (EDAS,

Tetra Tech 1999) relational database (MS Access 2000 platform); and the metrics calculated are those developed by the MBSS (Southerland et al. 2005).

2. PERFORMANCE CHARACTERISTICS

2.1. Field Sampling

Method Overview. Benthic macroinvertebrates were collected using a 600-µm mesh D-frame net to make 20 1-ft^2 sweeps, or "jabs" through multiple best available habitats within a 75-meter stream reach (Kazyak 2001, DNR 2007). Field duplicate samples were collected from adjacent 75-m reaches, which were randomly selected prior to the start of each sampling period. A total of 48 samples (24 duplicate sample pairs) were collected from 23 different PSUs throughout the County from 2004 to 2008 to assess intra-team variability of field sampling. It should be noted, that although benthic macroinvertebrate sampling methods have remained consistent since 2004, consistency in field crews were not maintained throughout the entire Round One sampling effort. However, while potential inter-team variability was not evaluated, it is not thought to contribute substantially to the overall variability in the data set.

2.1.1 Precision

Field sampling precision, defined as the nearness of two or more repeated measures, was calculated using sample results from adjacent reaches (i.e., duplicate samples). Precision of benthic metric values and index scores was calculated from pooled data collected from 24 sample pairs and included relative percent difference (RPD), root mean square error (RMSE), and coefficient of variability (CV).

2.1.1.1 Relative Percent Difference (RPD)

RPD was calculated using the equation:

$$RPD = \left(\frac{|A - B|}{A + B} \times 2 \right) \times 100$$

where, A is the metric or index value of the first sample and B is the metric or index value of the second sample (Berger et al. 1996). In general, low RPD values indicate better precision and repeatability. However, when calculating RPD using low metric values, the results must be interpreted with caution. For example, any comparison of a sample parameter value of 0 with one of >0 will automatically result in an RPD of 200. Also, comparing metrics between samples that both lack individuals of a particular taxon or functional feeding group (e.g., Ephemeroptera taxa, scraper taxa) will result in an RPD of 0, such as is shown in (Table 2) for percent Ephemeroptera, Ephemeroptera taxa, and scraper taxa. In those instances, RPD may not necessarily be an appropriate form of

precision, and care should be taken when including these values when calculating median RPD for a dataset.

Table 2. Metric and index precision represented by median relative percent difference, root mean square error (RMSE), and coefficient of variation (CV) for 24 sample pairs collected between 2004 and 2008.

	Measures of Precision		
Attribute	**Median RPD**	**RMSE**	**CV**
Total Taxa	17.7	4.29	19.5
No. EPT Taxa	28.6	1.67	46.8
% Ephemeroptera	0.0	2.79	138.3
No. Ephemeroptera Taxa	0.0	0.45	114.5
% Intolerant Urban	81.0	15.93	82.3
Number of Scraper Taxa	0.0	0.85	123.0
% Climbers	28.6	6.88	65.7
B-IBI	12.5	0.59	22.1

2.1.1.2 Root Mean Square Error (RMSE)

Root mean square error (RMSE) (or standard error of estimate), is a pooled standard error for a set of k group means, typically associated with a one-way ANOVA, and is calculated by:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij}-\overline{y_j})^2}{\sum df_{1\ldots k}}}$$

where y_{ij} is the i^{th} individual observation in group j.$j = 1\ldots k$ (Zar 1999). It is important to note that the denominator in this operation is the sum of degrees of freedom (df) for each group of replicated samples. Similar to RPD, RMSE decreases as precision increases. However, unlike RPD, RMSE is scale-dependent; therefore, metric and index RMSE values that are on different scales cannot be directly compared. For example, RMSE values calculated for percentage metrics (%Ephemeroptera, %climbers, %Intolerant Urban), which are on a 100-point scale, cannot be directly compared to RMSE of taxa richness metrics (total taxa, EPT taxa, Ephemeroptera taxa, scraper taxa) because these metrics are not on a definitive scale. Further, none of the metric RMSE values can be compared to the B-IBI RMSE because of this dependence on scale. Unlike RPD values, however, RMSE values are not distorted by metric values of zero.

2.1.1.3 Coefficient of Variation (CV)

To standardize the scale-dependent RMSE values, the coefficient of variation (CV) was calculated for the individual metric and B-IBI scores. CV is calculated from RMSE by:

$$CV = \frac{RMSE}{\overline{Y}} \times 100$$

where $\overline{Y}$ is the mean of the dependent variable (e.g., metric, index) of the duplicate sample pair population (Zar 1999). The CV is expressed as a percentage and allows direct comparison of the standard deviations among metrics and indices. However, it should be noted that regardless of RMSE, CV values typically tend to be higher for impaired sites with low scoring metrics and B-IBI scores due to the fact that the RMSE is divided by the sample mean, and the lower the mean value of the denominator the higher the resulting CV. For example, two groups of sample pairs that have identical RMSE values of 0.60 but different mean B-IBI scores of 4.10 and 2.40 would have CV s of 14.6 and 25.0, respectively. In other words, CV can be exaggerated somewhat for severely impaired streams, as compared to minimally impaired streams, due to lower values for both metrics and index scores.

Lower values for CV indicates better precision; however, only one metric, total taxa, had a CV <20% (Table 2). Three metrics (% Ephemeroptera, Ephemeroptera taxa, and scraper taxa) had values exceeding 100%, which occurs when the value of the RMSE exceeds the sample mean. Not surprisingly, these three metrics had the lowest mean values due to their rare nature in Coastal Plain streams. For these metrics, CV does not appear to be a useful indicator of precision, especially when the sample mean approaches zero.

2.1.2 Accuracy

Accuracy is not applicable to field sampling. Since, it would require knowledge of all invertebrates present at a sampling location as the analytical truth. Such an evaluation is not feasible or necessary for routine biological monitoring activities.

2.1.3 Bias

The benthic macroinvertebrate sampling approach is to proportionately allocate samples (jabs) among available habitats in order of preference, with an emphasis on best available habitat. Because this method puts an emphasis on sampling the best available habitats first, it could be considered biased in favor of habitats given preference, when present. Although often more abundant, lower quality habitat types such as undercut banks, snags, or sandy bottoms, can often be ‘undersampled’ or systematically excluded from the sample even though they may be present, since the allocation of samples is somewhat subjective and left to the discretion of the individual sampler.

The County’s sampling procedure was developed to be consistent with MBSS protocols current at the time of program development (Kazyak, 2001); however, the ordering of

preferred habitats was modified slightly between MBSS Sampling Manuals revisions Kazyak, 2001 and DNR 2007, which occurred between years three and four of the County's Round One sampling. The County's SOP (Anne Arundel County, 2004a) for macroinvertebrate collection governed the collection procedure during Round One and was as follows:

1) Riffles
2) Gravel, broken peat, and/or clay lumps in a run area
3) Snags/logs that create a partial dam or are in a run habitat
4) Undercut banks and associate root mats in running water
5) Submerged aquatic vegetation (SAV) and associated bottom substrates in moving water
6) Detrital/sand areas in moving water

2.1.4 Representativeness

This method is intended to provide a sample of the benthic macroinvertebrate assemblage the best available habitats in a sampling reach are able to support in rough proportion of their occurrence in the sampling reach (Kazyak 2001). It is not intended to be representative of all habitat types present in the sampling reach, especially if they comprise less than 5% of the stable habitat in the reach.

2.1.5 Completeness

One hundred percent of the sampling effort was utilized in each of the streams sampled; therefore field sampling data are complete.

2.2 Laboratory Sorting and Subsampling

Method Overview. The subsampling method involved using a 30-square Caton gridded screen (Caton 1991), which allows isolation of physically-defined amounts of sample material (leaf litter detritus, substrate particles) in the full sample, and then separation/removal of the organisms from that material (Anne Arundel County 2004a). Gridded squares of material were removed and sorted until the target number of organisms (100) was reached, with the final grid being sorted to completion. Once sorting was completed, experienced laboratory personnel[1] examined the remaining sort residue to ensure that all organisms were found. If missed specimens were found, they were counted and recorded on the subsampling bench sheets. Each sample resulted in three "post-sorting" containers: 1) the subsample destined for identification, 2) the unsorted sample remains, and 3) the sort residue.

2.2.1 Precision

[1] Preferably from independent laboratories, though not done in this evaluation.

Precision of sorting and subsampling is considered not applicable. This performance characteristic was not specifically evaluated.

2.2.2 Bias

Bias of subsampling is evaluated using a performance characteristic similar to percent recovery used in analytical chemistry laboratories, called percent sorting efficiency (PSE) (Barbour et al. 1999, Anne Arundel County 2004a). After the initial sorting effort, all sort residue was rechecked by qualified/experienced sorters (Table 3). The number of missed organisms recovered in the sort residue from the initial sorting was used to calculate sorting efficiency, using the following formula:

$$PSE = \left(\frac{a}{a+b} \right) \times 100$$

where *a* is the number of organisms originally sorted and *b* is the number of organisms recovered in the QC check (Barbour et al. 1999, Anne Arundel County 2004a).

Table 3. Percent sorting efficiencies per sample.

Station #	# of organisms originally sorted	# of organisms found in QC	Total # of organisms	PSE
CxB-1A	97	6	103	94.2
CxB-1B	105	10	115	91.3
CxB-2A	121	10	131	92.4
CxB-2B	128	16	144	88.9
CxB-3A	100	2	102	98.0
CxB-3B	99	14	113	87.6

2.2.3 Representativeness

Two aspects of the sample handling and laboratory processing method are designed to enhance representativeness. First, in the laboratory, all samples are spread evenly in the Caton tray allowing for a more representative subsample to be selected. Second, the grids that are sorted are randomly selected and sorted in their entirety.

2.3. Taxonomic Identification and Enumeration

Method Overview. Identification and enumeration (direct counts of taxa and total samples) was performed by a taxonomic laboratory using the most appropriate, up-to-date, and accepted taxonomic keys. Taxonomy was performed primarily to genus level, some to species, and others to higher levels (i.e., tribe, subfamily, family, order, or class). Target taxonomic levels (Table 4) are used to define the necessary level of effort for identification. Three of the samples were randomly selected for re-identification by an independent laboratory/taxonomist. Once the primary identifications were completed for all three samples, the vials and slides were shipped to that lab for re-identification.

Samples were sent with site information only (i.e., without identifications), thus representing blind samples.

2.3.1. Precision

Results from each lab were compared and precision estimates for enumeration and identification were calculated.

Table 4. Hierarchical targets for taxonomic identifications.

PHYLUM/Class	Order/*Family*	Taxonomic target
PHYLUM ANNELIDA		
Class Branchiobdellida		Identify to genus
Class Hirudinea		Identify to genus
Class Oligochaeta		Identify to genus
Class Polychaeta		Identify to genus
PHYLUM ARTHROPODA		
Class Arachnoidea	*Acari*	Identify to genus
Class Insecta	Coleoptera	Identify to genus
	Diptera	*Identify all to genus except for the following taxa:*
	Chironomidae	Identify to genus (this may not be possible for some groups, which should be identified to at least tribe or subfamily)
	Dolichopodidae	Identify to family
	Phoridae	Identify to family
	Scathophagidae	Identify to family
	Syrphidae	Identify to family
	Ephemeroptera	Identify to genus
	Hemiptera	Identify to genus
	Lepidoptera	Identify to genus
	Megaloptera	Identify to genus
	Odonata	Identify to genus
	Plecoptera	Identify to genus
	Trichoptera	Identify to genus
Class Malacostraca	Amphipoda	Identify to genus
	Decapoda	Identify to genus
	Isopoda	Identify to genus
	Mysidacea	Identify to genus
Class Ostracoda		Identify to genus
PHYLUM COELENTERATA		
PHYLUM MOLLUSCA		
Class Bivalvia		Identify to genus
Class Gastropoda		Identify to genus except in the following cases:
	Hydrobiidae	Identify to family
PHYLUM NEMERTEA		Identify to class
PLATYHELMINTHES PLANARIIDAE		Identify to genus

2.3.1.1 Enumeration

Final specimen counts for samples are dependent on the taxonomic identifications, not the rough counts obtained during the initial sorting activity. Comparison of counts uses Percent Difference in Enumeration (PDE) (Stribling et al. 2003), calculated as:

$$PDE = \left(\frac{|Lab1 - Lab2|}{Lab1 + Lab2} \right) \times 100$$

Identical numbers of individuals were counted by each lab for all three samples, therefore, the PDE was 0 for all samples. This was within the MQO of ≤5% for the overall dataset. It should be noted that experience has shown us that it is extremely rare for duplicate counts (with independent labs) to be identical.

2.3.1.2 Taxonomic Identifications

Side-by-side comparison between the taxonomic results delivered by the two labs was performed (Appendix A). The process entailed examination of the taxa list for each sample and the number of organisms each lab identified for each taxon (Appendix B). For each sample, the number of disagreements was determined, divided by the number of comparisons, and multiplied by 100 to give percent taxonomic disagreement (PTD) (Stribling et al. 2003). PTD was calculated as:

$$PTD = \left[1 - \left(\frac{comp_{pos}}{N}\right)\right] \times 100$$

where $comp_{pos}$ is the number of agreements, and N is the total number of organisms in the larger of the two counts. The lower the PTD value, the more similar are sample taxonomic results, and the greater is the overall taxonomic precision.

Table 5. Percent taxonomic disagreement for (PTD) re-identified samples.

Sample	N	$comp_{pos}$	PTD
CxB-1A	102	97	4.9
CxB-2B	123	114	7.3
CxB-3A	101	97	4

2.3.2 Accuracy

Definition of accuracy requires specification of an analytical truth (Taylor 1988, Clark and Whitfield 1994). For taxonomy, that is 1) the most up-to-date technical literature/keys, 2) an identified reference collection verified by specialists in different taxonomic groups, or 3) specimen by specimen comparison with museum-based type specimens. All taxonomy in this project was completed using technical literature specified in the QAPP (Anne Arundel County 2004a). There is currently no reference collection for the County, but it is recommended that one be assembled. Option 3 is not feasible, nor considered necessary, for routine monitoring programs.

2.3.3 Bias

This type of error in taxonomy would be problematic if there were consistent misinterpretation of technical keys, misunderstanding of morphological features, or poor processing of samples (including slide mounts of Chironomidae and Oligochaeta). However, none of these problems were identified in this project.

2.3.4 Completeness

Completeness of taxonomic analyses depends on the ability of the taxonomist to identify individual specimens and the rate at which the targeted hierarchical level of identification is met. For example, identifying an organism to family level when the QAPP required a genus level identification would be considered an incomplete identification. Complete identification may not be possible due to early instar organisms with underdeveloped morphological features, damaged specimens, or poorly assembled slide mounts. This aspect of the taxonomy was not evaluated.

PTD quantifies the precision with which the taxonomic database is developed. The comparison resulted in a mean PTD of 5.4%. Individual sample PTD ranged from 4 – 7.3 (Table 7). Most of the disagreements were over worms (Oligochaeta) and midges (Insecta: Diptera: Chironomidae).

2.4 Data Entry

Method Overview. All data from the six 2005 study sites were entered into EDAS (Ecological Data Application System, version 3.2, MS Access 2000) (Tetra Tech 1999). Data types entered included header information, comments, physical characterization, water quality, physical habitat assessment, and taxonomic data.

2.4.1 Accuracy

The accuracy of the data entry was checked by direct comparison of original datasheets (handwritten in the field or laboratory) with printouts from the database. All data entries (100%) were checked by an individual other than the primary data entry technician. There were no data entry errors identified (Table 6); however, if there had been, notations of errors would have been made on the initial printouts. All errors would have been corrected in the database and marked on the initial printouts when corrections were made.

Table 6. Data entry QC statistics by data types.

Data Type	No. of Entries	No. of Errors	Percent Correct
Header Info	30	0	100
Habitat	78	0	100
Water Chemistry	24	0	100
Benthos	220	0	100

2.5 Metric Calculation

Method Overview. In structuring the biological portion of the database, it was necessary to relate several sources of secondary data to each taxon. Three tables were developed in EDAS that organized tolerance values, functional feeding groups, and habit, which were

taken primarily from Merritt and Cummins (1996) and Barbour et al. (1999). Seven metrics and one index were calculated for each site based on the MBSS B-IBI calibrated for Coastal Plain streams (Southerland et al. 2005). They are:

1. *Total number of taxa*. This metric is a sum of all taxa present in the sample. The taxa richness of a community is commonly used as a quantitative measure of stream water and habitat quality. Stream degradation generally causes a decrease in the total number of taxa (Resh and Grodhaus 1983).
2. *Number of EPT taxa*. This metric is a sum of all taxa in the insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies). These taxa are generally sensitive to degraded stream conditions. A low number of insects within these orders is indicative of stream degradation (Lenat 1988).
3. *Number of Ephemeroptera taxa.* This metric is a sum of all taxa in the insect order Ephemeroptera (mayflies). Ephemeroptera are generally considered pollution sensitive, thus communities dominated by Ephemeroptera usually indicate lower disturbances in water quality.
4. *Percent Intolerant Urban.* This metric measures the percentage of the sample that is considered intolerant to urbanization. It is equal to the percentage of individuals in the sample with a designated tolerance value between 0-3. As impairment increases the percent of intolerant taxa decreases.
5. *Percent Ephemeroptera*. This metric measures the percentage of mayfly nymphs present in the sample. Mayflies are generally sensitive to pollution, so the degree to which a sample is composed of this order of insects can be an indicator of stream conditions, generally decreasing in value with increasing stress.
6. *Number of scraper taxa*. This metric is a sum of all taxa that are in the functional feeding group of scrapers. Scrapers feed on pollution intolerant microfauna, therefore, stream conditions that affect the food source can also affect scraper populations. This metric is expected to decrease with increasing stressors.
7. *Percent climbers*. This metric is a sum of all taxa that are adapted to living on stem type surfaces. Higher percentages of climbers typically represent a decrease in stressors and overall better water quality.

Resulting metric values were compared to criteria and scored on a 5 (closest to reference), 3 (neutral), 1 (farthest from reference) basis.

2.5.1 Accuracy

A subset of metric values was hand-calculated using only the taxonomic and enumeration data, and then compared to those that resulted from the EDAS queries. For one randomly selected site, all metrics were checked. Additionally, one metric was checked across all

sites. The purpose of this QC activity was to ensure that the metric calculation queries were performed correctly. All hand calculations (13) resulted in the exact same metric values as were generated by the database, suggesting queries were working correctly.

2.6 Site Assessment

Method Overview. The final site assessment is based on the Coastal Plain B-IBI score (Southerland et al. 2005), which is an aggregate of seven metrics, calculated for each site. The B-IBI can be used to designate thresholds and/or assign narrative assessments (e.g., good, fair, poor, very poor) to a given waterbody/watershed.

2.6.1 Precision

Table 7 shows the results of repeat sampling on metric and overall index precision. All but one metric (% Intolerant Urban) had a median RPD value <30%. However it should be noted that three metrics (% Ephemeroptera, Ephemeroptera Taxa, and Scraper Taxa) had median RPD values of zero due to the large number of sample pairs completely lacking Ephemeroptera and Scraper Taxa (n = 13 for both % Ephemeroptera &Ephemeroptera Taxa; n = 11 for Scraper Taxa), which results in zero RPD for each sample pair and effectively skews the median towards zero due to its high frequency in the data set. Because these taxa are infrequently found in Coastal Plain streams, the median may not be the best measure of central tendency. A comparison of median RPD and average RPD shows that there is far more variability in the data for these three metrics than the median RPD would indicate (Table 9).

Table 7. Comparison of median and average relative percent difference (RPD) values.

	Median RPD	Average RPD
Total Taxa	17.7	23.1
No. EPT Taxa	28.6	54.8
% Ephemeroptera	0.00	34.6
No. Ephemeroptera Taxa	0.00	29.4
% Intolerant Urban	81.0	92.7
Number of Scraper Taxa	0.00	48.3
% Climbers	28.6	49.1
B-IBI	12.5	16.2

Only one metric demonstrated good precision with CV<30% (Total Taxa), and the overall index had a CV of 22.09%. It should be noted, however, that these values were obtained from duplicate samples collected from adjacent stream reaches that had been selected at random prior to the site visit, which resulted in some duplicate samples being collected from sites where overall habitat and condition were dissimilar (Roberts et al., 2006). It is likely that they could be lower when duplicate samples are collected from reaches that are highly similar in habitat and complexity as well as presence/absence of stressor sources.

2.6.2 Accuracy

The accuracy of a biological indicator, such as the B-IBI, is characterized as its capacity to correctly identify stressor conditions (physical, chemical, and hydrologic). It is quantified as discrimination efficiency (DE) using the formula:

$$DE = \frac{a}{b} \times 100$$

where *a* is the number of correctly identified stressor sites, and *b* is the total number of stressor sites. Southerland et al. (2005) found the MBSS Coastal Plain B-IBI to have a DE of 96%.

2.6.3 Bias

Bias in site assessments is associated with DE. Thus, the four percent of stressor sites that are not indicated as degraded by the B-IBI (Southerland et al. 2005), can be attributed to bias.

2.6.4 Representativeness

Since the countywide program is a probability based stratified-random design (Anne Arundel County 2004b), the stream assessments are considered to be representative of a broader area than the individual site. Collectively, BIBI scores from multiple sites can then be averaged to obtain a mean condition, which is representative of the overall sampling unit.

The index score is not calculated if the final count for a subsample was <80 organisms and all 30 grids are sorted. This is intended to minimize the bias that may be associated with performing assessments with inadequate samples and data.

2.6.5 Completeness

Biological assessments were completed (100%) for all streams sampled.

2.6.6 Sensitivity

Sensitivity refers to the amount of change an indicator can detect relative to an independent variable (Flotemersch et al. 2006). The sensitivity of biological metrics or index scores used for site assessment can be determined by calculating the confidence interval. The confidence interval (CI), also known as detectable difference, is the likely range around the observed mean that the true mean is likely to fall and is determined using measures of precision. For this analysis the CI was chosen to be 90%, that is, the range within which the true mean is likely fall 90% of the time. A 90% CI (i.e., p = 0.10) of a single observation (i.e., only one replicate) is calculated from RMSE using the equation:

$$CI = \pm RMSE \times t_{\alpha}$$

where t_{α} is the t- value multiplier (i.e., 1.64) derived from a standard t-table (Zar 1999). With additional replicate samples, the confidence interval is divided by the square root of the number of replicates (two-tailed 90% CI = [1.64 RMSE]/√n). In other words, the confidence interval gets tighter as the number of replicates increases. Once the confidence interval is applied to the metric and B-IBI scores, it is possible to calculate the range of values that contain the true mean the majority (i.e., 90%) of the time. The 90% confidence interval calculated from this data set (n = 24 sample pairs) was 0.96, which means 90% of the time, the true population mean will fall within ± 0.96 of the observed mean.

3. RECOMMENDED MEASUREMENT QUALITY OBJECTIVES

Field Sampling

To account for differences in data, MQOs were established for several different measures of precision (RPD, RMSE, CV) for benthic metrics and overall index scores (Table 11). However, when calculating these measures, it is important to understand the limitations of each, as described above, to determine which are the most appropriate to use for evaluating acceptability of the data since not all measures will be applicable for a given data set. For instance, if a sample pair results in a comparison of zero with a value greater than zero for a given metric, RPD would not provide a useful measure of precision and RMSE or CV would be the preferred measures. Likewise, if the sample mean approaches zero, which is often the case for Ephemeroptera taxa and/or scraper taxa, CV would not be a useful measure of precision, but RMSE would be appropriate to evaluate those metrics. In other words, the most appropriate measure, or measures, of precision should be chosen for comparison to MQOs, and not all measures are necessary.

Values exceeding MQOs should be investigated for potential error, and corrective actions or adjustments, as required. Corrective action often requires the use of best professional judgment based on weight of evidence. For example, if it is determined that the MQOs are not being met for samples collected by certain personnel, it may require further investigation to determine the specific cause of that error. Then, if it is determined that it is due to error in applying the sample methodology, it may be necessary to re-train personnel in proper sampling techniques. Or, if it is determined to be a result of equipment failure (e.g., hole in net) or use of improper equipment (e.g., incorrect mesh size on net), a different type of corrective action may be required. Even if the data are not easy to fix to meet the MQO, it does not necessarily mean that they should be thrown out or redone. This depends entirely on how much uncertainty one is willing to accept in the data set, and hence, overall site assessment. For each field season, duplicate sampling should occur at a minimum of 10% of the locations, preferably one duplicate sample pair per primary sampling unit.

MQO: For a sampling event (field season, watershed, or other strata) the B-IBI mRPD should be <20%, the RMSE should be <0.6 and/or the coefficient of variability (CV) should be <22%.

Laboratory Sorting and Subsampling

Laboratory sorting/subsampling MQO for this program is to have a dataset where ≤ 10% of the samples have a percent sort efficiency (PSE) of ≤ 90%. It is recommended that outside, independent laboratories perform the sort residue re-checks on a randomly selected 10% of the samples. For internal QC, individual sorters will be trained to consistently attain a 90% or greater sorting efficiency. QC checks would be performed on every sample until a sorter is able to demonstrate the ability to attain a 90% or greater sorting efficiency on five samples in a row, and after that, samples should be checked at a rate of 10%. If, for example, the outside laboratories determines that greater than 10% of the samples they check failed, then corrective action should be taken, such as re-checking another 10% of samples to see if they meet the established MQO. If they continue to fail, further corrective action may be necessary, such as re-sorting all samples until the MQO is met.

MQO_{PSE}: Less than or equal to 10% of externally QC'd sort residues should have a PSE ≤90%.

Table 8. Measurement quality objectives for metric and index scores.

	MQO		
Attribute	**Median RPD**	**RMSE**	**CV**
Total Number of Taxa	20	4.3	20
Number of EPT Taxa	30	1.7	50
% Ephemeroptera	30	2.8	100
% Intolerant to Urban	80	15.9	80
Number of Ephemeroptera Taxa	30	0.5	100
Number of Scraper Taxa	30	0.9	100
% Climbers	30	6.9	70
B-IBI	20	0.6	22

Taxonomic Identification and Enumeration

For re-identified samples, percent taxonomic disagreement (PTD) between laboratories should be ≤ 15%. Individual sample PTD should be evaluated to determine the cause of disagreements, however, corrective actions are generally not necessary if mean PTD for the dataset is ≤ 15%. However, differences in individual sample comparisons should still be evaluated for patterns of disagreement. The MQO for specimen enumeration is 5%, where samples having greater than 5% count difference are examined for sample integrity and reasons for the differences. The MQO for the overall dataset is ≤ 5%. For percent

taxonomic completeness (PTC), the MQO should be ≥ 90%. That is, on average, ≥ 90% of the organisms in the subsample should be identified to the target level (see Table 4).

MQO_{PTD}: Mean PTD ≤ 15%; Samples with PTD ≥ 15% should be further examined for patterns of error.

MQO_{PDE}: Mean PDE ≤ 5%; Samples with PDE ≥ 5% should be further examined for patterns of error.

MQO_{PTC}: Mean PTC ≥ 90%; Samples with PTC ≤ 90% should be examined and those taxa not meeting targets isolated.

Data Entry

The MQO for data entry accuracy should be 100% after QC checks have occurred and before data is used for subsequent analysis. After data is checked by QC personnel, all errors should be noted and corrected in the database. This is a very important step for tracking the consistency of data in the database.

MQO: Entered data are 100% correct.

Metric Calculation

The MQO for metric calculation accuracy should be 100% after QC checks have occurred. Hand calculations of each individual metric should be compared to the values calculated by the database or spreadsheets. If any errors are discovered in the metric calculations, they should first be noted and recorded, then corrected in the database to ensure the metric scores used in the site assessment are accurate.

MQO: Metric calculations are 100% correct.

Site Assessments

Since the B-IBI score is the primary measure used for the site assessments, the median RPD should not exceed 20% and/or the RMSE should not exceed 0.6. Since both RPD and CV tend to become exaggerated with lower index scores, RMSE was selected over CV since it does fluctuate based on the absolute value of the sample mean. All efforts should be made to ensure that, whenever possible, all samples are collected, processed, and identified, and the resulting the data is entered into the database and metrics calculated to achieve a final site assessment. Valid assessments should be completed for ≥ 95% of sites selected for sampling during a given sampling period.

MQO: The mRPD calculated from the replicate samples should not exceed 20% and/or the RMSE should not exceed 0.6 for the B-IBI score.

MQO_{SENS}: The 90% confidence interval calculated from the replicate samples should not exceed 0.96 for the B-IBI.

MQO_{COMP}: Assessments should be completed for ≥ 95% of sites selected for sampling.

4. SUMMARY

This report presents qualitative and quantitative information on the quality of data underlying, and acceptable for, Anne Arundel County biological stream assessments. These results are intended to demonstrate the overall performance and capability of the monitoring program (Anne Arundel County 2004b), to allow it to be directly compared to other monitoring programs and datasets, and to define data quality expectations. In combination with results from subsequent years of monitoring, they also, in part, form the foundation for defining criteria for acceptable data quality. In general, this report reflects a dataset acceptable for its stated purposes. We emphasize that if a data point or sample fails to meet an MQO, it should not automatically be considered bad data. Rather, it should be evaluated in detail to determine reasons for the exceedence and that the appropriate measure is used for the given data set. Quality issues identified in field sampling, laboratory sorting and subsampling, taxonomy and data entry should continue to be addressed through training, field and laboratory audits, rigorous senior oversight and external review, with very specific corrective actions detailed as necessary in all appropriate QA/QC documents.

5. LITERATURE CITED

Anne Arundel County. 2010. *Anne Arundel County Biological Monitoring and Assessment Program: Quality Assurance Project Plan*. Revised December 2010. Prepared by KCI Technologies, Inc. for Anne Arundel County Department of Public Works, Watershed Ecosystem and Restoration Services. Annapolis, MD. For additional information, contact Mr. Chris Victoria (*410-222-4240*, <PWVICT16@aacounty.org>)

Anne Arundel County. 2004a. *Anne Arundel County Biological Monitoring and Assessment Program: Quality Assurance Project Plan*. Prepared by Tetra Tech, Inc. for Anne Arundel County Office of Environmental & Cultural Resources. January 2004, 44pp., Annapolis, MD. For additional information, contact Mr. Chris Victoria (*410-222-4240*, <PWVICT16@aacounty.org>)

Anne Arundel County. 2004b. *Design of the Biological Monitoring and Assessment Program for Anne Arundel County, Maryland*. Prepared by Tetra Tech, Inc. for Anne Arundel County Office of Environmental & Cultural Resources. January 2004, 18pp., Annapolis, MD. For additional information, contact Mr. Chris Victoria (*410-222-4240*, <PWVICT16@aacounty.org>)

Barbour, M.T., J. Gerritsen, ,B.D. Snyder, and J.B. Stribling. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish. Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Berger, W., H. McCarty, and R.K. Smith. 1996. *Environmental Laboratory Data Evaluation*. Genium Publishing Corp. Amsterdam, NY.

Boward, D. and E. Friedman. 2000. *Laboratory methods for benthic macroinvertebrate processing and taxonomy.* Maryland Biological Stream Survey, Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division. CBWP-MANTA-EA-00-6. Annapolis, Maryland. November.

Caton, L.W. 1991. Improving subsampling methods for the EPA "Rapid Bioassessment" benthic protocols. Bulletin of the North American Benthological Society 8(3):317-319.

Clark, M.J.R. and P.H. Whitfield. 1994. Conflicting perspectives about detection limits and about the censoring of environmental data. Water Resources Bulletin 30(6): 1063-1079.

Costanza, R., S.O. Funtowicz, and J.R. Ravetz. 1992. Assessing and communicating data quality in policy-relevant research. Environmental Management 16(1):121-131.

Diamond, J.M., M.T. Barbour, J.B. Stribling. 1996. Characterizing and comparing bioassessment methods and their results: a perspective. Journal of the North American Benthological Society 15(4): 713-727.

Flotemersch, J.E., J.B. Stribling, and M.J. Paul. 2006. Concepts and Approaches for the Bioassessment of Non-Wadeable Streams and Rivers. EPA/600/R-06/127. U. S. EPA, Office of Research and Development, Cincinnati, OH.

Kazyak, P.F. 2001. *Maryland Biological Stream Survey Sampling Manual.* Maryland Department of Natural Resources, Chesapeake Bay Research and Monitoring Division, Annapolis, MD.

Lenat, D. R. 1988. Water quality assessment of streams using a qualitative collection method for benthic macroinvertebrates. Journal of North American Benthological Society 7: 222-233.

Maryland Department of Natural Resources (DNR). 2007. Maryland Biological Stream Survey Sampling Manual: Field Protocols. CBWP-MANTA-EA-07-01. Published by the Maryland Department of Natural Resources, Annapolis, MD. Publication # 12-2162007-190.

Methods and Data Comparability Board (MDCB). 2004 (draft). Guidance on Bioassessment Method and Data Comparability Evaluation. Draft September 30, 2004. http://wi.water.usgs.gov/methodsboard/loveland_co_04/bioassessment%20comparability%20guidance%20report%20v.1.doc

Merritt, R.W. and K.W. Cummins. 1996. *An Introduction to the Aquatic Insects of North America.* Third Edition. Kendall/Hunt Publishing, Dubuque Iowa.

Resh, V. H. and G. Grodhaus. 1983. Aquatic insects in urban environments. In: G. W. Frankie and C. S. Koehler (editors). Urban Entomology: Interdisciplinary Perspectives. Praeger Publishers, New York. pp. 247-276.

Roberts, M., C. Smith, and C. Victoria. 2006. Aquatic Biological Assessment of the Watersheds of Anne Arundel County, Maryland: 2005. Anne Arundel County, Office of Environmental and Cultural Resources, Annapolis, Maryland.

Stribling, J.B., S.R. Moulton, and G.T. Lester. 2003. Determining the quality of taxonomic data. Journal of the North American Benthological Society 22(4): 621-631.

Stribling, J. B., B. K. Jessup, J. S. White, D. Boward, and M. Hurd. 1998. *Development of a Benthic Index of Biotic Integrity for Maryland Streams.* Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division. CBWP-EA-98-3. Annapolis, Maryland.

Southerland, M.T., G.M. Rogers, M.J. Kline, R.P. Morgan, D.M. Boward, P.F. Kazyak, R.J. Klauda, S.A. Stranko. 2005. New Biological Indicators to Better Assess the Condition of Maryland Streams. DNR-12-0305-0100. Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division. Annapolis, MD.

Taylor, J.K. 1988. Defining the Accuracy, Precision, and Confidence Limits of Sample Data. Chapter 6, pages 102-107, *In* Lawrence H. Keith (editor), *Principles of Environmental Sampling*. ACS Professional Reference Book. American Chemical Society. Columbus, Ohio.

Taylor, B.N. and C.E. Kuyatt. 1994. *Guidelines for evaluating and expressing the uncertainty of NIST measurement results.* NIST Technical Note 1297. National Institute of Standards and Technology, U.S. Department of Commerce, Washington, DC. 24 pp.

Tetra Tech, Inc., 1999. Ecological Data Application System (EDAS). A user's guide. November. For further information, contact Erik Leppo at (410-356-8993), <*Erik.Leppo@tetratech.com*>).

U.S. Environmental Protection Agency. 1995. *EPA Requirements for Quality Assurance Project Plans*. EPA QA/R-5. U.S. Environmental Protection Agency, Quality Staff, Washington, D.C.

Zar, J.H. 1999. *Biostatistical Analysis.* Fourth Edition. Prentice Hall, Upper Saddle River, NJ. 663 pp.

APPENDIX A. SAMPLE PAIR RESULTS

Table A-1. Metric values for sample pairs (n = 24) collected throughout Anne Arundel County from 2004 to 2008. Samples denoted by FR indicate the Field Replicate sample.

Station	Year	No. Taxa	No. EPT Taxa	% Ephem.	No. Ephem. Taxa	% Intolerant Urban	No. Scraper Taxa	% Climbers
03-05 FR	2004	21	4	0.00	0	13.86	2	20.79
03-05	2004	30	7	0.00	0	20.18	4	14.91
09-10 FR	2004	24	8	5.21	1	6.25	4	35.42
09-10	2004	26	5	2.59	1	16.38	6	16.38
10-09 FR	2004	18	4	3.74	1	2.80	1	12.10
10-09	2004	23	4	5.00	1	20.00	2	10.00
18-11A FR	2004	29	6	0.94	1	9.43	0	21.70
18-11A	2004	23	6	1.00	1	8.00	0	18.00
21-09 FR	2004	18	6	5.71	1	10.48	0	36.19
21-09	2004	19	5	6.59	1	7.69	0	39.56
11-14A FR	2005	12	0	0.00	0	3.80	1	9.40
11-14A	2005	21	1	3.03	1	23.23	4	8.10
12-04 FR	2005	18	2	0.00	0	66.70	0	8.10
12-04	2005	22	3	1.04	1	22.92	0	33.00
15-19A FR	2005	19	5	18.60	2	17.70	1	14.70
15-19A	2005	22	8	12.12	1	20.20	2	12.10
19-10 FR	2005	30	9	9.70	2	24.80	1	16.80
19-10	2005	26	9	20.41	3	28.57	0	8.16
22-03 FR	2005	22	2	0.00	0	15.20	1	22.90
22-03	2005	12	0	0.00	0	1.00	0	24.00
05-08 FR	2006	33	4	0.00	0	7.14	1	3.57
05-08	2006	29	5	0.00	0	9.48	1	6.90
06-05 FR	2006	32	7	0.00	0	26.05	0	15.13
06-05	2006	23	6	0.00	0	35.90	0	18.80

Station	Year	No. Taxa	No. EPT Taxa	% Ephem.	No. Ephem. Taxa	% Intolerant Urban	No. Scraper Taxa	% Climbers
07-10 FR	2006	21	4	0.00	0	0.00	0	4.08
07-10	2006	26	3	0.97	1	4.85	1	6.80
24-04 FR	2006	18	3	0.00	0	9.91	0	0.90
24-04	2006	20	3	0.00	0	14.91	0	0.88
01-08 FR	2007	30	2	0.00	0	0.00	0	8.91
01-08	2007	23	0	0.00	0	0.93	0	12.96
02-03 FR	2007	22	2	0.00	0	3.70	0	0.93
02-03	2007	25	4	0.00	0	19.61	0	0.00
16-02 FR	2007	18	3	0.00	0	75.45	0	1.82
16-02	2007	21	0	0.00	0	50.59	0	3.53
16-12A FR	2007	31	0	0.00	0	2.88	1	1.92
16-12A	2007	22	0	0.00	0	0.00	0	5.41
17-05 FR	2007	12	1	0.00	0	1.72	0	0.00
17-05	2007	12	1	0.00	0	3.81	0	0.00
04-08 FR	2008	26	1	0.00	0	0.00	0	10.68
04-08	2008	24	1	0.00	0	1.80	0	7.21
13-05 FR	2008	20	4	0.00	0	49.56	0	0.88
13-05	2008	14	3	0.00	0	86.79	0	0.00
14-02 FR	2008	8	1	0.00	0	38.74	0	0.00
14-02	2008	17	4	0.00	0	48.70	0	0.00
20-05 FR	2008	17	4	0.00	0	20.19	0	0.00
20-05	2008	15	4	0.00	0	51.00	0	0.00
23-02 FR	2008	32	3	0.00	0	22.55	0	3.92
23-02	2008	31	4	0.00	0	4.00	0	5.00

Table A-2. Metric and index scores for sample pairs (n = 24) collected throughout Anne Arundel County from 2004 to 2008. Samples denoted by FR indicate the Field Replicate sample.

Station	Year	No. Taxa	No. EPT Taxa	% Ephem.	No. Ephem. Taxa	% Intolerant Urban	No. Scraper Taxa	% Climbers	BIBI	Rating
03-05 FR	2004	3	3	1	1	3	5	5	3.00	Fair
03-05	2004	5	5	1	1	3	5	5	3.57	Fair
09-10 FR	2004	5	5	3	3	1	5	5	3.86	Fair
09-10	2004	5	5	3	3	3	5	5	4.14	Good
10-09 FR	2004	5	3	3	3	3	5	5	3.86	Fair
10-09	2004	3	3	3	3	1	3	5	3.00	Fair
18-11A FR	2004	5	5	3	3	1	1	5	3.29	Fair
18-11A	2004	5	5	3	3	1	1	5	3.29	Fair
21-09 FR	2004	3	5	3	3	3	1	5	3.29	Fair
21-09	2004	3	5	3	3	1	1	5	3.00	Fair
11-14A FR	2005	1	1	1	1	1	3	5	1.86	Very Poor
11-14A	2005	3	1	3	3	3	5	5	3.29	Fair
12-04 FR	2005	3	3	1	1	5	1	5	2.71	Poor
12-04	2005	5	3	3	3	3	1	5	3.29	Fair
15-19A FR	2005	3	5	5	5	3	3	5	4.14	Good
15-19A	2005	5	5	5	3	3	5	5	4.43	Good
19-10 FR	2005	5	5	3	5	3	3	5	4.14	Good
19-10	2005	5	5	5	5	5	1	5	4.43	Good
22-03 FR	2005	5	3	1	1	3	3	5	3.00	Fair
22-03	2005	1	1	1	1	1	1	5	1.57	Very Poor
05-08 FR	2006	5	3	1	1	1	3	3	2.43	Poor
05-08	2006	5	5	1	1	1	3	3	2.71	Poor
06-05 FR	2006	5	5	1	1	3	1	5	3.00	Fair
06-05	2006	5	5	1	1	5	1	5	3.29	Fair
07-10 FR	2006	3	3	1	1	1	1	3	1.86	Very Poor

Station	Year	No. Taxa	No. EPT Taxa	% Ephem.	No. Ephem. Taxa	% Intolerant Urban	No. Scraper Taxa	% Climbers	BIBI	Rating
07-10	2006	5	3	3	3	1	3	3	3.00	Fair
24-04 FR	2006	3	3	1	1	1	1	3	1.86	Very Poor
24-04	2006	3	3	1	1	3	1	1	1.86	Very Poor
01-08 FR	2007	5	3	1	1	1	1	5	2.43	Poor
01-08	2007	5	1	1	1	1	5	1	2.14	Poor
02-03 FR	2007	5	3	1	1	1	1	3	2.14	Poor
02-03	2007	5	3	1	1	3	1	1	2.14	Poor
16-02 FR	2007	3	3	1	1	5	1	3	2.43	Poor
16-02	2007	3	1	1	1	5	1	3	2.14	Poor
16-12A FR	2007	5	1	1	1	1	3	3	2.14	Poor
16-12A	2007	5	1	1	1	1	1	3	1.86	Very Poor
17-05 FR	2007	1	1	1	1	1	1	1	1.00	Very Poor
17-05	2007	1	1	1	1	1	1	1	1.00	Very Poor
04-08 FR	2008	5	1	1	1	1	1	5	2.14	Poor
04-08	2008	5	1	1	1	1	1	3	1.86	Very Poor
13-05 FR	2008	3	3	1	1	5	1	1	2.14	Poor
13-05	2008	3	3	1	1	5	1	1	2.14	Poor
14-02 FR	2008	1	1	1	1	5	1	1	1.57	Very Poor
14-02	2008	3	3	1	1	5	1	1	2.14	Poor
20-05 FR	2008	3	3	1	1	3	1	1	1.86	Very Poor
20-05	2008	3	3	1	1	5	1	1	2.14	Poor
23-02 FR	2008	5	3	1	1	3	1	3	2.43	Poor
23-02	2008	5	3	1	1	1	1	3	2.14	Poor

APPENDIX B. TAXONOMIC COMPARISONS

Table B-1. Taxonomic comparisons between two independent laboratories, anonymously indicated as T1 for the primary, or original, taxonomist and T2 as the QC, or re-identification taxonomist.

Sample ID	Order	Family	Final ID	T1	T2	# of Agreements
CxB-1A	Lumbricina	Lumbricidae	Unid. Genus	1		
CxB-1A	Haplotaxida	Sparganophilidae	Sparganophilus		1	
CxB-1A	Tubificida	Naididae	Nais communis		2	2
CxB-1A	Tubificida	Naididae	Nais	2		
CxB-1A	Tubificida	Tubificidae	Limnodrilus hoffmeisteri		2	2
CxB-1A	Tubificida	Tubificidae	Limnodrilus	2		
CxB-1A	Amphipoda	Crangonyctidae	Synurella	1	1	1
CxB-1A	Isopoda	Asellidae	Caecidotea	1	1	1
CxB-1A	Plecoptera	Leuctridae	Unid. Leuctrid	1		
CxB-1A	Plecoptera	Nemouridae	Amphinemura	21	21	21
CxB-1A	Plecoptera	Nemouridae	Nemouridae		1	
CxB-1A	Plecoptera	Perlodiade	Isoperla	1	1	1
CxB-1A	Plecoptera	Perlodiade	Cultus	1	1	1
CxB-1A	Megaloptera	Corydalidae	Nigronia	1	1	1
CxB-1A	Trichoptera	Hydropsychidae	Diplectrona	2	2	2
CxB-1A	Trichoptera	Lepidostomatidae	Lepidostoma	1	1	1
CxB-1A	Trichoptera	Limnephilidae	Pycnopsyche	5	5	5
CxB-1A	Trichoptera	Limnephilidae	Ironoquia	2	2	2
CxB-1A	Coleoptera	Dryopidae	Helichus	3	3	3
CxB-1A	Coleoptera	Ptilodactylidae	Anchytarsus	2	2	2
CxB-1A	Diptera	Ceratopogonidae	Ceratopogon?	1		
CxB-1A	Diptera	Ceratopogonidae	Stilobezzia		1	
CxB-1A	Diptera	Chironomidae	Ablabesmyia	1	1	1
CxB-1A	Diptera	Chironomidae	Brillia	2	2	2
CxB-1A	Diptera	Chironomidae	Chaetocladius	3	3	3
CxB-1A	Diptera	Chironomidae	Corynoneura	2	2	2
CxB-1A	Diptera	Chironomidae	Cricotopus/Orthocladius	2	2	
CxB-1A	Diptera	Chironomidae	Diplocladius	1	1	1
CxB-1A	Diptera	Chironomidae	Heterotrissocladius	1	1	1
CxB-1A	Diptera	Chironomidae	Hydrobaenus	4	4	4
CxB-1A	Diptera	Chironomidae	Micropsectra	2	2	2
CxB-1A	Diptera	Chironomidae	Parametriocnemus	6	6	6
CxB-1A	Diptera	Chironomidae	Paratanytarsus	1	1	1
CxB-1A	Diptera	Chironomidae	Polypedilum	8	8	8
CxB-1A	Diptera	Chironomidae	Pseudorthocladius	3	3	3
CxB-1A	Diptera	Chironomidae	Rheocricotopus	7	7	7
CxB-1A	Diptera	Chironomidae	Rheotanytarsus	1	1	1
CxB-1A	Diptera	Chironomidae	Thienemannimyia group	4	4	4
CxB-1A	Diptera	Chironomidae	Tvetenia	2	2	2
CxB-1A	Diptera	Chironomidae	Zavrelimyia	2	2	2
CxB-1A	Diptera	Tipulidae	Tipula	2	2	2
			Number of individuals	102	102	97
			PTD	**4.9**		
			PDE	**0.0**		

Sample ID	Order	Family	Final ID	T1	T2	# of Agreements
CxB-2B	Lumbricina	Lumbricidae	Lumbricidae Unid. Genus	2		2
CxB-2B	Lumbricina	Lumbricidae	Lumbricidae		2	
CxB-2B	Eulamellibranchia	Sphaeriidae	Pisidium	1	1	1
CxB-2B	Eulamellibranchia	Sphaeriidae	Unid. sphaeriid (immature)	1		1
CxB-2B	Eulamellibranchia	Sphaeriidae	Sphaeriidae		1	
CxB-2B	Ephemeroptera	Leptophlebiidae	Leptophlebiidae		1	1
CxB-2B	Ephemeroptera	Leptophlebiidae	Unid. leptophlebiid	1		
CxB-2B	Plecoptera	Capniidae/Leuctridae	Capniidae/Leuctridae		1	1
CxB-2B	Plecoptera	Leuctridae	Leuctra	1		
CxB-2B	Plecoptera	Nemouridae	Amphinemura	25	26	26
CxB-2B	Plecoptera	Nemouridae	Nemouridae		1	
CxB-2B	Plecoptera	Taeniopterygidae	Unid. taeniopterygid	1		
CxB-2B	Plecoptera	Unid. plecoptera	Unid. plecoptera	1		
CxB-2B	Trichoptera	Hydropsychidae	Diplectrona	4	4	4
CxB-2B	Trichoptera	Lepidostomatidae	Lepidostoma	1	1	1
CxB-2B	Trichoptera	Limnephilidae	Pycnopsyche	1	1	1
CxB-2B	Coleoptera	Hydrophilidae	Hydrobius	1	1	1
CxB-2B	Diptera	Chironomidae	Apsectrotanypus	1		
CxB-2B	Diptera	Chironomidae	Bethbilbeckia		1	
CxB-2B	Diptera	Chironomidae	Brillia	1	1	1
CxB-2B	Diptera	Ceratopogonidae	Ceratopogon?	2	2	2
CxB-2B	Diptera	Chironomidae	Chaetocladius	3	3	3
CxB-2B	Diptera	Chironomidae	Constempellina	1		
CxB-2B	Diptera	Chironomidae	Cricotopus/Orthocladius	2	2	
CxB-2B	Diptera	Chironomidae	Diamesa	13	13	13
CxB-2B	Diptera	Chironomidae	Eukiefferiella	1	1	1
CxB-2B	Diptera	Chironomidae	Heterotrissocladius	3	3	3
CxB-2B	Diptera	Chironomidae	Hydrobaenus	1	1	1
CxB-2B	Diptera	Chironomidae	Larsia	1		
CxB-2B	Diptera	Chironomidae	Natarsia	3	1	
CxB-2B	Diptera	Chironomidae	Parametriocnemus	9	9	9
CxB-2B	Diptera	Chironomidae	Polypedilum	24	24	24
CxB-2B	Diptera	Chironomidae	Potthastia	3	3	3
CxB-2B	Diptera	Chironomidae	Pseudorthocladius	1	1	1
CxB-2B	Diptera	Chironomidae	Stempellina		1	
CxB-2B	Diptera	Chironomidae	Thienemannimyia group	4	4	4
CxB-2B	Diptera	Chironomidae	Zavrelimyia	2	5	2
CxB-2B	Diptera	Simuliidae	Simulium	6	6	6
CxB-2B	Diptera	Tipulidae	Tipula	2	2	2
			Number of individuals	123	123	114
			PTD	**7.3**		
			PDE	**0.0**		

Sample ID	Order	Family	Final ID	T1	T2	# of Agreements
CxB-3A	Tubificida	Enchytraeidae	Enchytraeidae		2	2
CxB-3A	Tubificida	Enchytraeidae	Enchytraeidae Unid. Genus	2		
CxB-3A	Lumbricina	Lumbricidae	Lumbricidae		1	1
CxB-3A	Lumbricina	Lumbricidae	Lumbricidae Unid. Genus	1		
CxB-3A	Tubificida	Naididae	Nais communis		1	1
CxB-3A	Tubificida	Naididae	Nais	1		
CxB-3A	Odonata	Aeshnidae	Boyeria	1	1	1
CxB-3A	Plecoptera	Chloroperlidae	Haploperla	7	6	6
CxB-3A	Plecoptera	Neumoridae	Amphinemura	12	12	12
CxB-3A	Plecoptera	Perlodidae	Isoperla	4	4	4
CxB-3A	Trichoptera	Hydropsychidae	Diplectrona	2	2	2
CxB-3A	Trichoptera	Hydropsychidae	Hydropsyche	1	1	1
CxB-3A	Trichoptera	Limnephilidae	Limnephilidae		3	2
CxB-3A	Trichoptera	Limnephilidae	Unid. limnephilid	2		
CxB-3A	Trichoptera	Limnephilidae	Pycnopsyche	2	2	2
CxB-3A	Trichoptera	Lepidostomatidae	Lepidostoma	2	1	1
CxB-3A	Coleoptera	Ptilodactylidae	Anchytarsus	3	3	3
CxB-3A	Diptera	Chironomidae	Chaetocladius	9	9	9
CxB-3A	Diptera	Chironomidae	Diamesa	2	2	2
CxB-3A	Diptera	Chironomidae	Eukiefferiella	5	5	5
CxB-3A	Diptera	Chironomidae	Krenosmittia	1		
CxB-3A	Diptera	Chironomidae	Limnophyes	2	2	2
CxB-3A	Diptera	Chironomidae	Micropsectra	4	4	4
CxB-3A	Diptera	Chironomidae	Natarsia	1	1	1
CxB-3A	Diptera	Chironomidae	Orthocladiinae		2	
CxB-3A	Diptera	Chironomidae	Unid. Orthocladiinae	1		1
CxB-3A	Diptera	Chironomidae	Orthocladius	1		
CxB-3A	Diptera	Chironomidae	Parametriocnemus	2	4	2
CxB-3A	Diptera	Chironomidae	Paratendipes	1	1	1
CxB-3A	Diptera	Chironomidae	Polypedilum	12	12	12
CxB-3A	Diptera	Chironomidae	Pseudorthocladius	2	2	2
CxB-3A	Diptera	Chironomidae	Rheocricotopus	2	2	2
CxB-3A	Diptera	Chironomidae	Rheotanytarsus	2	2	2
CxB-3A	Diptera	Chironomidae	Thienemannimyia group	5	5	5
CxB-3A	Diptera	Chironomidae	Tvetenia	6	6	6
CxB-3A	Diptera	Chironomidae	Zavrelimyia	1	1	1
CxB-3A	Diptera	Tipulidae	Tipula	2	2	2
			Number of individuals	101	101	97
			PTD			**4.0**
			PDE			**0.0**