AQUATIC EFFECTS TECHNOLOGY EVALUATION (AETE) PROGRAM

Quality Assurance Program for Assessing Mine-Related Effects Using Benthic Invertebrate Communities

AETE Project 2.1.4

QUALITY ASSURANCE PROGRAM FOR ASSESSING MINE-RELATED EFFECTS USING BENTHIC INVERTEBRATE COMMUNITIES

Report Prepared for:

Aquatic Effects Technology Evaluation (AETE) Program
Canada Centre for Mineral and Energy
Technology (CANMET)
Natural Resources Canada
555 Booth Street
Ottawa, Ontario
K1A 0G1

Prepared by:

BEAK INTERNATIONAL INCORPORATED
14 Abacus Road
Brampton, Ontario
L6T 5B7

February, 1999
Beak Ref: 3359.1

AQUATIC EFFECTS TECHNOLOGY EVALUATION PROGRAM

Notice to Readers

Quality Assurance Program for Assessing Mine-Related Effects Using Benthic Invertebrate Communities

The Aquatic Effects Technology Evaluation (AETE) program was established to review appropriate technologies for assessing the impacts of mine effluents on the aquatic environment. AETE is a cooperative program between the Canadian mining industry, several federal government departments and a number of provincial governments; it is coordinated by the Canada Centre for Mineral and Energy Technology (CANMET). The program was designed to be of direct benefit to the industry, and to government. Through technical and field evaluations, it identified cost-effective technologies to meet environmental monitoring requirements. The program included three main areas: acute and sublethal toxicity testing, biological monitoring in receiving waters, and water and sediment monitoring.

The technical evaluations are conducted to document certain tools selected by AETE members, and to provide the rationale for doing a field evaluation of the tools or provide specific guidance on field application of a method. In some cases, the technical evaluations included a go/no go recommendation that AETE takes into consideration before a field evaluation of a given method is conducted.

The technical evaluations are published although they do not necessarily reflect the views of the participants in the AETE Program. The technical evaluations should be considered as working documents rather than comprehensive literature reviews. The purpose of the technical evaluations was to document specific monitoring tools. AETE committee members would like to stress that no one single tool can provide all the information required for a full understanding of environmental effects in the aquatic environment.

For more information on the monitoring techniques, the results from their field application and the final recommendations from the program, please consult the AETE Synthesis Report to be published in the spring of 1999.

Any comments concerning the content of this report should be directed to:

Geneviève Béchard
Manager, Metals and the Environment Program
Mining and Mineral Sciences Laboratories - CANMET
Room 330, 555 Booth Street, Ottawa, Ontario, K1A 0G1
Tel.: (613) 992-2489 Fax: (613) 992-5172
E-mail: gbechard@nrcan.gc.ca

PROGRAMME D'ÉVALUATION DES TECHNIQUES DE MESURE D'IMPACTS EN MILIEU AQUATIQUE

Avis aux lecteurs

Un programme d'assurance de la qualité applicable à l'évaluation des effets de l'activité minière sur les communautés de macro-invertébrés benthiques

Le Programme d'évaluation des techniques de mesure d'impacts en milieu aquatique (ÉTIMA) visait à évaluer les différentes méthodes de surveillance des effets des effluents miniers sur les écosystèmes aquatiques. Il est le fruit d'une collaboration entre l'industrie minière du Canada, plusieurs ministères fédéraux et un certain nombre de ministères provinciaux. Sa coordination relève du Centre canadien de la technologie des minéraux et de l'énergie (CANMET). Le programme était conçu pour bénéficier directement aux entreprises minières ainsi qu'aux gouvernements. Par des évaluations techniques et des études de terrain, il a permis d'évaluer et de déterminer, dans une perspective coût-efficacité, les techniques qui permettent de respecter les exigences en matière de surveillance de l'environnement. Le programme comportait les trois grands volets suivants : évaluation de la toxicité aiguë et sublétale, surveillance des effets biologiques des effluents miniers en eaux réceptrices, et surveillance de la qualité de l'eau et des sédiments.

Les évaluations techniques ont été menées dans le but de documenter certains outils de surveillance sélectionnés par les membres de l'ÉTIMA et de fournir une justification pour l'évaluation sur le terrain de ces outils ou de fournir des lignes directrices quant à leur application sur le terrain. Dans certains cas, les évaluations techniques pourraient inclure des recommandations relatives à la pertinence d'effectuer une évaluation de terrain que les membres de l'ÉTIMA prennent en considération.

Les évaluations techniques sont publiées bien qu'elles ne reflètent pas nécessairement toujours l'opinion des membres de l'ÉTIMA. Les évaluations techniques devraient être considérées comme des documents de travail plutôt que des revues de littérature complètes. Les évaluations techniques visent à documenter des outils particuliers de surveillance. Toutefois, les membres de l'ÉTIMA tiennent à souligner que tout outil devrait être utilisé conjointement avec d'autres pour permettre d'obtenir l'information requise pour la compréhension intégrale des impacts environnementaux en milieu aquatique.

Pour des renseignements sur l'ensemble des outils de surveillance, les résultats de leur application sur le terrain et les recommandations finales du programme, veuillez consulter le Rapport de synthèse ÉTIMA qui sera publié au printemps 1999.

Les personnes intéressées à faire des commentaires concernant le contenu de ce rapport sont invitées à communiquer avec M^{me} Geneviève Béchard à l'adresse suivante :

Geneviève Béchard
Gestionnaire, Programme des métaux et de l'environnement
Laboratoires des mines et des sciences minérales - CANMET
Pièce 330, 555, rue Booth, Ottawa (Ontario), K1A 0G1
Tél.: (613) 992-2489 / Fax : (613) 992-5172
Courriel : gbechard@nrcan.gc.ca

EXECUTIVE SUMMARY

Benthic invertebrates are commonly used to monitor spatial and temporal impacts on aquatic systems. Similar to other aquatic monitoring tools, the methods have to be applied by adequately trained and experienced individuals to help ensure that the data generated are not of questionable quality. Unfortunately, even though benthic invertebrate community assessments have been a cornerstone of aquatic effects monitoring for many decades, there has been little attention paid towards quality assurance programs to improve data quality. Unlike other components of environmental effects monitoring programs, such as analytical chemistry and aquatic toxicity testing, very little effort has been made to standardize benthic invertebrate monitoring protocols and to implement mandatory quality control measures to ensure data quality. Quality Control associations, such as Canadian Association for Environmental Analytical Laboratories (CAEAL), have developed extensive quality assurance programs for chemistry and aquatic toxicity labs but have yet to incorporate benthic ecology labs into the program.

Quality control procedures are necessary to ensure that benthic invertebrate data achieve an acceptable minimum level of quality and that the level of quality attained is well documented. It is still very common to encounter biomonitoring programs and published literature on benthic invertebrates that have not documented any level of quality assurance.

In Canada, the most recent benthic invertebrate quality assurance program was developed by Environment Canada for the federal government's environmental effects monitoring program for the pulp and paper industry. Despite the attempts to standardize the benthic invertebrate monitoring protocols for this program and requirements for mandatory quality control components, in many cases, the data from the first cycle of monitoring were still of questionable quality. This was generally due to a lack of enforcement of the quality control requirements.

The Assessment of Aquatic Effects of Mining (AQUAMIN) is presently developing a program to evaluate the effectiveness of Canada's *Metal Mine Liquid Effluent Regulations* (MMLERs) in protecting fish, fish habitat (e.g., benthic invertebrates) and the beneficial uses of fisheries resources. This will require monitoring of benthic invertebrates on a national scale. Because the impacts from mines with treated effluents tend to be more subtle than the impacts typically associated with pulp and paper discharges, it is important to have a quality assurance program in place to ensure that the invertebrate programs are sensitive enough to detect whether or not there are impacts that are associated with a mining operation.

The objective of this document is to provide guidance and recommend a Quality Assurance Program that, when adopted by any benthic invertebrate monitoring program, will ensure that the data generated are of known quality, reproducible and comparable among studies.

Typically, most components of a quality assurance program are generic and apply to any aquatic monitoring program. These have been highlighted in this document along with those that are more specific to benthic ecology. The mandatory requirements of a quality assurance program for benthic invertebrate monitoring programs which are considered to be the minimum requirements are as follows:

- documentation of a study design and objectives;
- stipulation of data quality objectives;
- documented standard operating procedures for field and laboratory work;
- an average of 95% recovery of invertebrates from samples with no samples having less than 90% recovery;
- calculation of the error associated with any subsampling techniques;
- archiving of sorted invertebrates and bench sheets;
- compilation of a voucher collection;
- a listing of taxonomic keys used; and
- documentation of the sorters' and taxonomists' qualifications.

RÉSUMÉ

Les invertébrés benthiques sont couramment utilisés pour la surveillance des incidences spatiales et temporelles sur les écosystèmes aquatiques. Comme tout autres méthodes de surveillance, ces méthodes doivent être appliquées adéquatement par des personnes compétentes et expérimentées afin d'assurer la qualité des données. Malheureusement, bien que depuis plusieurs décennies les évaluations de la communauté d'invertébrés benthiques aient été la pierre angulaire de la surveillance de la mesure d'impacts en milieu aquatique, peu d'importance a été accordée aux programmes d'assurance de la qualité en vue d'améliorer la qualité des données. Contrairement à d'autres composantes des programmes de surveillance des effets sur l'environnement, tels la chimie analytique et les essais de toxicité, peu d'efforts ont été déployés pour normaliser les protocoles de surveillance des invertébrés benthiques et mettre en oeuvre des mesures obligatoires de contrôle de la qualité en vue d'assurer la qualité des données. Les associations de contrôle de la qualité, telle la *Canadian Association for Environmental Analytical Laboratories* (CAEAL), ont élaboré des programmes exhaustifs d'assurance de la qualité pour les laboratoires de chimie et d'analyse de la toxicité en milieu aquatique mais ont encore à intégrer les laboratoires d'écologie benthique dans leur programme.

Des procédures de contrôle de la qualité sont nécessaires pour s'assurer que les données sur les invertébrés benthiques sont d'un niveau acceptable et bien documenté. Il arrive trop souvent que les programmes de surveillance biologique et que les publications portant sur les invertébrés benthiques ne fournissent pas suffisamment d'information sur la qualité des données.

Au Canada, le plus récent programme d'assurance de la qualité des données sur les invertébrés benthiques a été élaboré par Environnement Canada dans le cadre du Programme de surveillance des incidences environnementales sur le milieu aquatique, mis en oeuvre par le gouvernement fédéral à l'intention des fabriques de pâtes et papiers. Malgré les efforts consacrés à la normalisation des protocoles de surveillance des invertébrés benthiques dans le cadre de ce programme et les exigences découlant des composantes obligatoires d'assurance de la qualité, dans bien des cas, la qualité des données compilées après le premier cycle d'étude était en général douteuse en raison du fait que les exigences liées au contrôle de la qualité n'était généralement pas appliquées.

Le Programme d'évaluation des effets de l'exploitation minière sur le milieu aquatique au Canada (AQUAMIN) élabore actuellement un programme dont le but est d'évaluer l'efficacité du *Règlement sur les effluents liquides des mines de métaux* (RELMM) visant à protéger le poisson, l'habitat du poisson (p. ex., les invertébrés benthiques) et les ressources halieutiques. Ce projet exigera la surveillance des invertébrés benthiques à l'échelle nationale. Les impacts des mines dont les effluents ont été traités ayant tendance à être plus subtils que ceux qui sont généralement associés aux rejets des fabriques de pâtes et papiers, il est important de mettre en oeuvre un programme d'assurance de la qualité afin d'assurer que les études basées sur les invertébrés soient assez sensibles pour détecter si les impacts associés à l'exploitation des mines.

Ce document vise à fournir des principes directeurs et recommande la mise en oeuvre d'un programme d'assurance de la qualité qui, une fois adopté dans le cadre de la surveillance des invertébrés benthiques, assurera des données reproductibles de qualité, comparables entre les sites d'études.

Généralement, la plupart des composantes d'un programme d'assurance de la qualité sont génériques et s'appliquent à tout programme de surveillance des impacts en milieu aquatique. Ces composantes ont été mises en évidence dans le présent document de même que les celles plus spécifiquement liées à l'écologie benthique. Les exigences obligatoires minimales découlant d'un programme d'assurance de la qualité applicable aux programmes de surveillance des invertébrés benthiques sont les suivantes :

- documentation sur la conception de l'étude et des objectifs;
- stipulation des objectifs de qualité des données;
- documentation des procédures opérationnelles standardisées relatives au méthodes de travail sur le terrain et en laboratoire;
- un taux moyen de récupération des invertébrés de 95 % à partir des échantillons et un taux de récupération minimum de 90 % pour chaque échantillon;
- calcul de l'erreur associée à toute technique de sous-échantillonnage;
- archivage des invertébrés triés et relevés des analyses en laboratoire;

- compilation de la collection de spécimens justificatifs;
- liste des clés de taxonomie utilisées; et
- documentation des compétences des trieurs et taxonomistes.

TABLE OF CONTENTS

Page

1.0 INTRODUCTION

Since the late 1950s, benthic invertebrates have been commonly used to assess impacts of industrial discharges on aquatic environments. Using benthic invertebrates to monitor impacts has long been recognized as a cost-effective monitoring tool. The advantages of using benthic invertebrates have been well documented (e.g., Klemm *et al.,* 1990; Rosenberg and Resh, 1993). Two of the main advantages of using invertebrates as biomonitors, that are always cited, are that invertebrates are easily collected and readily identified. However, the proper collection and in particular, the correct identification of benthic invertebrates is not as simple as much of the literature tends to suggest. Like any monitoring technology, if the techniques are not properly applied by trained and experienced individuals, benthic invertebrate studies often end up basing conclusions on questionable data. At a 1993 benthic invertebrate monitoring workshop sponsored by Environment Canada, Scientific Authorities hosting the workshop indicated that for many Environment Canada studies, some with up to 20 years of benthic data, sound conclusions could not be made because the quality of the data was so poor (Environment Canada, 1993).

Unlike other components of environmental effects monitoring programs, such as analytical chemistry and toxicity testing, very little effort has been made to standardize benthic invertebrate monitoring protocols and to implement mandatory quality control measures to ensure data quality. It was not until the mid-1980s that agencies, which routinely use benthic invertebrates for large-scale biomonitoring studies, initiated quality assurance programs in order to produce data that were of known quality, reproducible and comparable (e.g., U.S. EPA Puget Sound monitoring program, Tetra Tech, 1987; U.S. Geological Survey's National Water Quality Assessment Program, 1993). Quality assurance/quality control (QA/QC) procedures are necessary to ensure that environmental data achieve an acceptable minimum level of quality and that the level of quality attained is documented adequately (Tetra Tech, 1986).

Many benthic invertebrate monitoring programs now incorporate a quality management plan, however, it is still quite common to encounter programs and published literature on benthic ecology that have not documented any level of quality assurance. For groups or programs (e.g., Aquatic Effects Technology Evaluation Program (AETE), Assessment of Aquatic Effects of Mining (AQUAMIN)) that are developing habitat assessment procedures, an effort to define QA/QC procedures is essential (Davis and Simon, 1995). In Canada, the most recent quality assurance program was developed by Environment Canada for the federal government's environmental effects monitoring program for the pulp and paper industry (EC/DFO, 1991). Despite the attempts to standardize the benthic invertebrate protocols for this program and requirements for mandatory quality control components, in many cases, data were still of questionable quality. This was generally due to a lack of enforcement of the QA/QC requirements by regulatory agencies.

Studies using benthic invertebrates make comparisons of data among sites (spatial comparisons) or times (temporal comparisons) and a QA/QC program minimizes and quantifies the variability introduced by collection or processing methods, which in turn improves the sensitivity of the methods in demonstrating whether or not an effect exists. In any monitoring program, the principles of QA/QC are as critical as the use of standard sampling protocols (Gibson, 1995).

AQUAMIN is presently developing a program to evaluate the effectiveness of Canada's *Metal Mine Liquid Effluent Regulations* (MMLERs) in protecting fish, fish habitat (e.g., benthic invertebrates) and the beneficial uses of fisheries resources. CANMET, in cooperation with representatives from the Mining Association of Canada and other federal and provincial agencies through a program known as the Aquatic Effects Technology Evaluation (AETE) program, is evaluating cost-effective monitoring tools that the mining industry could use to assess the effectiveness of the regulations, as well as for routine environmental effects monitoring. It is within the mandate of the AETE Committee to recommend cost-effective environmental effects monitoring methods to assist the Canadian mining industry. The Technical Committee for the AETE program recognizes the need for a well-documented quality assurance program to accompany any of their recommendations for using benthic invertebrates for monitoring the effects of mining on aquatic systems in Canada.

Because there is always the desire to be as cost effective as possible, during routine benthic invertebrate monitoring programs it is critical that a quality management program is in place to ensure that as cost saving measures are implemented the effectiveness of the monitoring tool is not compromised. For example, sieves and flotation methods are used to reduce the costs of separating benthos from debris. However, if a large percentage of the benthos are not recovered by the cost saving measure then the effectiveness of the benthic monitoring tool is reduced. A QA/QC program to some extent limits the ability of practitioners to reduce costs because in such a program data quality takes precedence over cost reduction.

The objective of this document is to provide guidance and recommend a Quality Assurance Program (QAP) that, when implemented by any benthic invertebrate monitoring program, will ensure that the data generated are of known quality, reproducible and comparable among studies. In the event that a national environmental effects monitoring program is developed for the Canadian mining industry it will be critical that a QA/QC program is in place to ensure that the data, when entered into a national database, are comparable and of known quality. A QAP in the context of environmental effects monitoring at mine sites is particularly critical because effects on the benthic invertebrate community by low-level metal contamination or other disruptions associated with mine operations are expected to be smaller and more subtle than effects from organic enrichment (Taylor, 1996). Therefore, any increase in data variability that is introduced because of collection or processing methods will increase the effect size that can be reliably detected. Consequently, the impacts will have to be more severe in order to be detected with sufficient power. All project activities from sampling and laboratory analysis to statistical analysis and reporting are potential error sources and therefore, there needs to be a quality control check in every aspect of a biomonitoring program to minimize the occurrence of these errors.

In this document special attention has been paid to the use of the terms “should” and “must”. Quality control criteria preceded by “must” are considered to be the minimum acceptable level of quality in a benthic invertebrate monitoring program to assess mine-related impacts. Those criteria preceded by “should” are desirable components and if included in a QAP will improve the overall technical quality of the data.

In development of the QAP, it was assumed that temporal trends in benthic invertebrate community structure are important and that data generated for different mine sites will need to be comparable in order to assess the overall effectiveness of the MMLERs across Canada.

2.0 QUALITY MANAGEMENT PLAN

The first step towards ensuring high quality data for benthic invertebrate monitoring programs is the development of a Quality Management Plan (QMP). A key component of any QMP for benthic invertebrate monitoring is a Quality Assurance Policy Statement which affirms the senior management's or the proprietor's commitment to, and ultimate responsibility for, achievement of data quality objectives as a priority equal to or greater than that of schedule or budget. Although the quality assurance policy statement is a simple motherhood statement, it puts the onus on management or company owners to ensure that the data are of high quality.

The following components that should be included in a QMP are generally common practice and are often presented in proposals or detailed in the scope or statement of work for a particular benthic invertebrate monitoring project.

A glossary of the terms used in this document is provided in Appendix 1.

2.1 Study Description

The first stage in quality management is a well-documented description of the study. The study description must state the overall purpose and objectives of the study. It must also present specific data quality objectives (DQOs), and details of the project organizational structure and management process which are focused on achievement of objectives. The quality management plan should have an in-depth discussion on each of these topics with particular emphasis on the study objectives and the data quality objectives. The study description should define the schedule and general process for assessing whether each DQO is being or has been achieved, and for determining, based on this, whether the basic study design, management practice (organization) or technical operations (field, laboratory, quality assurance) should be revised.

2.1.1 Overall Scope of the Program

The description provides an overview of the main components of the study. If the benthic invertebrate study is a component of a larger multidisciplinary program then the project overview needs to elaborate on these other areas of study so that the benthic ecologists have a clear understanding on how their data fit into the bigger picture. The description should focus on data collection activities, basic design approaches, objectives in terms of questions to be resolved, data uses to address these questions, and the overall framework for decision making based on the answers to these questions.

The description should identify any blocks of data to be produced for or used by another project, either as supplementary data or for comparisons between projects or in data pooling. Any criteria that would permit or preclude such data uses should be stated.

2.1.2 Study Objectives

The objectives can be stated either as general or specific questions to be resolved. These questions usually refer to the nature and extent of effects on benthic invertebrate communities in relation to contaminant sources. They generally involve comparison to reference areas, or to baseline time periods, or they refer to gradients in community structure. They may also involve correlations between biological responses and stressor variables. General questions can be partitioned into more specific questions pertaining to individual variables or biotic indices that are testable as hypotheses.

2.1.3 Study Design

For each data collection activity, the sampling strategy (e.g., random, stratified random, systematic) and level of effort (e.g., number of replicate samples, number of stations in reference and exposure areas) must be identified. The spatial and temporal domain of the study and the bounds of interpolation or extrapolation should be clearly defined and justified. It is also important to ensure that the statistical methods that are being considered for the study are appropriate for the study design selected.

2.1.4 Data Uses

The use of the data is largely dictated by the form of the specific questions posed; however, statistical methods (including transformations) may depend on the assumptions that the user wishes to make, which in turn may depend on preliminary examination of data. Also, the variables and areas of interest may be redefined or revised after preliminary data review. This kind of flexibility is allowable and should therefore be identified in the study description.

2.1.5 Decision Framework

The logic leading from the answers to specific questions, to the ultimate selection of abatement, remedial or further investigative initiatives, should be defined as clearly as possible. Often there are significant socio-political rather than scientific inputs to these decisions, which should be acknowledged.

2.1.6 Data Collection Activities

The data collection activities must refer to the specific measurements to be made and the general methods of sampling and analysis. Reference to specific standard operating procedures (SOPs) must be documented, whenever possible.

2.2 Data Quality/Quantity Objectives

The linkage of fine scale questions to broader scale questions means that data may have to meet the needs of different users at the project or site-specific level and at regional or national program levels. For example, on a project level a mine owner may want to know if a discharge is having a detrimental effect on the receiving environment, whereas on a national level Environment Canada may want to incorporate the data from this mine with those from other mines to determine whether or not effluent regulations are sufficient to protect the environment. The data quality (and quantity) needs of these users may differ, and it is critical that project managers understand the broader context in terms of the potential use of the benthic invertebrate data.

A process for defining and assessing achievement of Data Quality Objectives (DQOs) is essential and should be described. The process of DQO development is iterative, involving both program managers and technical experts. It should lead to DQOs that are realistic and achievable. The process often involves compromise between what we would like to achieve and what we can afford, but ensures an appreciation of what is achievable within the available budget. The process also encourages cost-effective resource allocation and guards against spreading resources too thinly which in turn may result in the inability to meet the study objectives. Data quality objectives should never be defined at quality levels lower than what is considered to be a minimum acceptable level of data quality. Minimum acceptable levels for DQOs are often established by regulatory agencies (e.g., Environment Canada) or associations that focus on improving and maintaining data quality (e.g., CAEAL, Canadian Association of Environmental Analytical Laboratories).

All DQOs should be clearly stated along with the rationale on which they are based and the corrective actions that may be considered when they are not achieved. Different types of DQOs are briefly discussed below.

An example of the development of data quality objectives and corrective actions is provided in Appendix 2.

2.2.1 Power to Answer Questions with Specified Resolution

A specified power to answer the project questions with a specified resolution is the ultimate data quality objective. It is determined as an objective from consideration of the consequence of getting incorrect answers and thus making "wrong" decisions. Other data quality objectives, often called measurement quality objectives (MQOs), are designed to control components of the overall uncertainty that limits the power of the study.

2.2.2 Sensitivity, Precision and Accuracy

Sample detection limits should be sufficiently low so that unacceptable bias due to censored data is not introduced to the summary statistics that are used to answer questions (e.g., mean or variance). The detection limit for organism density in a benthic sample is determined by the sampling area of the collection device, the number of grabs pooled per sample and the subsampling practice used in the laboratory which may change the effective sampling area.

Precision is specified as a permissible limit for random error (variance or standard deviation). It may be specified at any or all levels. For example, it could be specified for a measurement process (e.g., dissolved oxygen) or for sampling and processing methods. It is determined by replication at each level.

Accuracy is specified as a permissible limit for the degree to which a quality control check sample deviates from the expected value. For benthic studies, a check sample may be a benthic sample of relevant composition that has been enumerated by a number of sorters or taxonomists so that a consensus composition has been established, or it may be a sample that has received intensive study internally so that its composition is “known”. For example, intensive examination of residual debris in a sorting check may reveal additional organisms; the total for the sample now is “known”. There was incomplete recovery previously.

It should be noted that accuracy, as defined above, based on a single sample, includes precision and bias components. The average under-recovery over many samples represents a systematic error or bias.

2.2.3 Comparability and Compatibility

Comparability is a measure of the confidence with which one data set can be compared to another (U.S. EPA, 1995). Two data sets may be considered comparable if they are both essentially unbiased or share essentially the same bias as determined by analysis of checked samples for quality control. Differences in relevant error may require special methods of comparison leading to some loss of power, but generally do not preclude comparison.

If two data sets are to be pooled for subsequent treatment as a single data set, they must be essentially the same (compatible) with respect to both mean value and variance for all variables and domains of interest. If they are not, the data likely belong to different populations and this structure will have to be recognized in the statistical treatment. Compatibility is not necessarily ensured by adherence to sampling and analytical protocols. Any requirements for data set comparability or compatibility should be described as DQOs.

In a national monitoring program, such as Environment Canada’s environmental effects monitoring program for Canadian pulp and paper mills, data comparability and compatibility are essential. This national program is designed to monitor temporal trends at a particular mill, while concurrently, developing a national database to assess the effectiveness of the pulp and paper effluent regulations in protecting the aquatic environment. Ensuring that data sets are comparable among mines does not preclude the ability to develop site-specific study objectives and methods to better delineate impacts at a mine site.

2.2.4 Representativeness

Samples may be considered to represent the area of interest if collected by an unbiased sampling plan (i.e., some type of random sampling) and if the sampling method is not unacceptably biased. If the sampling method is unbiased, the benthic invertebrates sorted from a number of pooled grabs should accurately represent the benthic community at the station where they were collected. By establishing a DQO for representativeness, the study has to demonstrate the number of grabs that have to be pooled to accurately reflect the benthic community at a station. This demonstration can involve simple plots of grabs versus number of taxa, whereby, at some point no new taxa are found with additional grabs. Assessment of representativeness in the sampling plan involves consideration of time and location and other conditions of sampling, as compared to the time/space domain that is the object of study. Representativeness is not necessarily ensured by adherence to sampling and analytical protocols.

2.2.5 Completeness

Completeness is the percentage of planned measurements that are actually obtained and useable for any measurement variable. Criteria for any declaration of data as not useable should be stated or referenced. Note that DQO exceedence does not necessarily make data unusable. Benthic invertebrate biomonitoring studies typically set DQOs for completeness at 90% (e.g., U.S. EPA, 1994a)

2.3 Study Organization

The integrated structure and process by which management ensures that objectives will be achieved for a benthic biomonitoring study must be defined in order to be effective. This includes definition of responsibilities and authorities, reporting channels and schedules, QA policy, resource allocations and strategies, document controls, procurement processes and staff training and evaluation procedures.

2.3.1 Responsibilities and Authorities

In the QMP, individuals responsible for planning, implementing and assessing each of the main field, laboratory and QA operations must be identified. Too often staff are unaware that it is their responsibility to ensure that a particular quality control check is performed to demonstrate that a data quality objective was met. An example of a typical organizational chart is provided in Figure 2.1.

The document should also define each report product, the schedule or frequency of preparation, and the individuals responsible for preparation, review and final approval. Dissemination of QA information is particularly important and the individuals responsible must be identified.

2.3.2 Resource Allocations

The QMP should briefly identify a strategy for budget and manpower allocations to specific field, laboratory and QA operations, including QA planning, and any process for cost monitoring and/or increasing allocations to meet program needs. The QMP should also document the steps to be taken that will ensure that the quality of the data is not compromised if the project is going over budget.

2.3.3 Document and Data Control

The QMP should identify the process for revision, certification and circulation of key documents such as standard operating procedures (SOPs), project or QA reports and data files, to ensure information integrity and use of the most recent versions. It is quite common to find field staff following old versions of SOPs because a system was not in place to ensure that revised SOPs were circulated to the appropriate individuals. A regular review schedule, a process for revisions and a list of changes should be attached to all documents. Older versions should be archived.

2.3.4 Staff Training and Evaluation

The QMP should describe the process for design and implementation of related staff training and proficiency evaluation programs. This is particularly critical for benthic invertebrate sorters which is generally overlooked because it is assumed that sample sorting is a simple task. The QMP should also identify the current status of staff assigned to the project in relation to the training program. This is particularly important in the laboratory operations where staff are performing detailed invertebrate identifications. Staff should only be identifying invertebrate groups which they have been trained to identify.

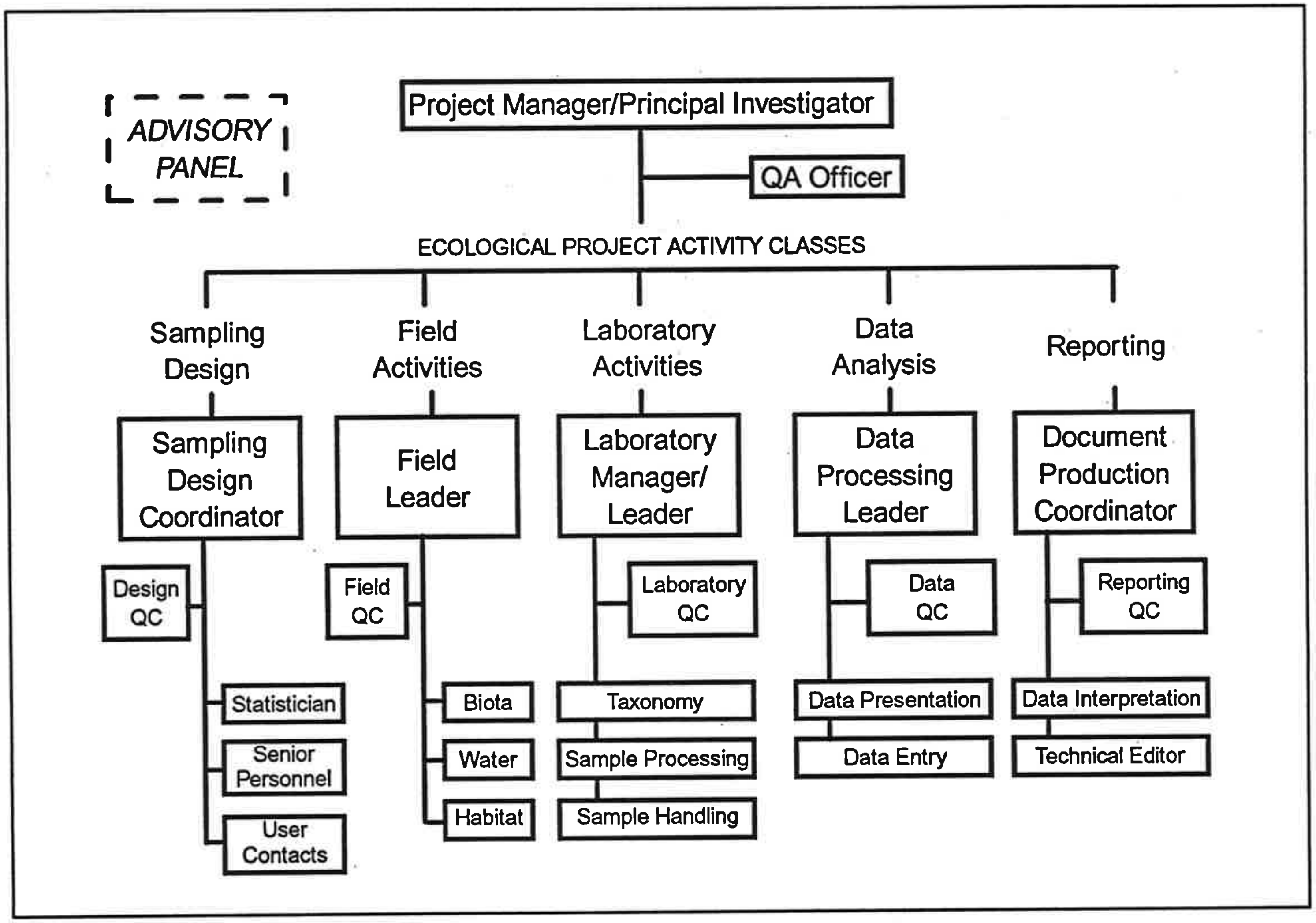

Figure 2.1: Typical Organizational Chart Showing Project Organization and Lines of Communication (taken from U.S. EPA, 1995)

3.0 FIELD OPERATIONS

Field operations include sampling design and sampling procedures. Execution of sampling designs requires good communication between managers and technical staff, and consideration of available resources. Procedures may reference or build on any pertinent SOPs that exist, rather than repeating them in detail.

There should be a health and safety plan in place prior to any field operation. For many provinces in Canada it is a mandatory requirement. Guarding the health and safety of the field crew ensures that such concerns do not interfere with the collection of quality data (Tetra Tech, 1986).

3.1 Sampling Design

Studies should be designed to meet the study and data quality objectives, based on previous reconnaissance or historical data, clearly defined questions and stated acceptable confidence levels for the answers. The design will specify the sampling time period, areas to be sampled, numbers of stations or time points in each, how to perform unbiased sampling within areas or time periods, and field and laboratory replication to check or control components of error.

3.2 Sampling Procedures

It is acceptable to reference or build on existing SOPs for sampling procedures. The field staff must have a copy of the most recent versions of the SOPs. Non-standard methods or deviations from standard methods are only allowed if described in the QMP. The methods described should include any real time quality control checks, such as repeated measurements, meter calibration logs or sampling efficiency tests. They should be organized to facilitate updating as needed, and must be communicated to field staff. Standardized forms facilitate consistent data recordings and should include places for observational (meta) data and comments, as well as places to indicate that required methods were followed, or to note exceptions. Standardized data forms also minimize the occurrence of missing data points that are generally due to forgetfulness.

Guidance should be provided on the proper selection of sampling sites. This guidance should be directed toward maximizing accuracy, minimizing uncertainty or, at least, providing a means by which variability may be reduced (U.S. EPA, 1995). This guidance should be detailed enough such that individual field crews would sample similar habitats and would also ensure that reference and exposed sites would have similar habitat features.

3.2.1 Equipment and Instrument Maintenance and Calibration

All equipment used to make supporting measurements during a benthic invertebrate survey must be accompanied by SOPs for their use and calibration. For meters used for supporting measurements of geographical coordinates, dissolved oxygen, conductivity and pH, logs tracking the calibration and precision of meters at the start and end of each field day should be kept. If a reading appears unusual (e.g., anoxic or supersaturated conditions), meter calibration should be checked and recorded.

3.2.2 Sample Collection and Preservation

Detailed SOPs for sample collection and preservation must be provided to the field staff. Quality control criteria for what constitutes an acceptable sample (e.g., depth of Ponar penetration, no sample leakage, etc.) should be developed. Collecting representative samples is crucial to subsequent decision making, and obtaining good results on non-representative samples is of no use (U.S. EPA, 1995).

Detailed instructions on how to preserve benthic samples must also be provided. It is surprising how many field biologists believe that preserving samples with 10% buffered formalin means making up a 10% formalin solution and then adding it to a 1 L jar that may have 750 mL of sample in it only to find out weeks later during sorting that the sample has deteriorated. It is important to note that the final solution after adding the formalin must be a minimum of 10% buffered formalin so that the invertebrates are adequately fixed. Samples containing large quantities of organic matter require higher percentages of formalin.

Sample labelling instructions should also be documented. As a minimum, samples must have external and internal labels.

3.2.3 Chain of Custody, Shipping, Storage and Records

It is important to have chain of custody forms which track samples and handling from the time of collection to delivery of the benthic invertebrate data. This is especially important when using contract laboratories for sample processing to ensure that all of the samples arrived at the lab in good condition.

In Canada, shipping of benthic invertebrate samples preserved with formaldehyde is covered under the transportation of dangerous goods regulations. Therefore, it is imperative that field crews have the proper documentation and are aware of the appropriate methods for shipping samples with formaldehyde.

The storage procedures should also be documented and all records maintained on file. Generally, benthic invertebrate samples must be stored in a secure and well ventilated area. The samples should not be frozen at any time as this will hamper identification of the invertebrates.

3.2.4 Staff Training and Proficiency

Training is the most important quality control element of field operations (Gibson, 1995) and is essential to reduce between-user error. Field staff must be appropriately trained and have adequate experience in sampling benthic invertebrates. All field crews must be accompanied by a crew leader who should have adequate training and experience to make decisions while in the field. It is quite common that changes to the study plan occur due to unforeseen circumstances in the field and only a qualified crew leader can make decisions that will not compromise the ability of the data to meet the objectives of the study.

3.2.5 Field Quality Control

The field quality control plan must ensure that there is consistency in sampling methods throughout the study period (Environment Canada, 1993). All sampling equipment, especially sieves and nets, must be thoroughly cleaned between stations to ensure that there is no transfer of invertebrates from one station's samples to other samples.

Standardized field data sheets must be filled out completely and accurately and these sheets should be included in the appendices of the final reports. The field data sheets should be detailed enough to allow someone else to reconstruct the field activities without relying on the memory of the field crew.

Supporting measurements should be replicated at a minimum of 10% of the stations to document the precision of the measurements. It should be documented whether or not measurements and samples met the acceptance criteria.

4.0 LABORATORY OPERATIONS

Laboratory operations include laboratory facilities and analytical procedures. Benthic invertebrate sample processing requires a considerable amount of labour and technical expertise to produce consistent high quality data in a timely manner (U.S. Geological Survey, 1993). Sample processing for benthic invertebrate monitoring programs is one of the most critical elements and yet it is only recently that quality control checks have been incorporated into benthic invertebrate studies. The problem stems from the common belief that sorting and taxonomy of benthic samples require little or no training. Analysis of benthic invertebrate samples is like any other component of environmental monitoring (e.g., chemistry, toxicity) in that it requires staff who have adequate training and experience and it requires quality control checks to demonstrate that the data are of known and high quality.

Most of the invertebrate samples from Canadian biomonitoring projects are subcontracted out to small firms (generally one to three individuals) that specialize in invertebrate taxonomy. Therefore, the following QA/QC components have been developed bearing this in mind and no quality control criteria have been recommended that a single-person operation could not meet.

In Canada, there are currently no interlaboratory or accreditation studies related to benthic invertebrate analysis. Typically these types of programs (e.g., Canadian Association of Environmental Analytical Laboratories- CAEAL) are external to the project, but the results can be pertinent, particularly the interlaboratory studies with project-relevant matrices. Variations on the interlaboratory theme that could be incorporated into a benthic invertebrate study include split sample exchanges, re-processing samples or exchange of voucher collections among participating laboratories.

System audits involve site visits, interviews and document review and are generally quite comprehensive. Currently, there are no such audits by authorities but they should be encouraged. Periodic audits should be carried out to ensure that field and laboratory staff are following SOPs and meeting QC requirements.

4.1 Facilities

The general layout of the laboratory, separation of areas and floor space should be documented. The description of the facility should emphasize means for avoidance of cross-contamination of invertebrates (e.g., invertebrates tangled in sieves used in sample washing), tracking and storage of samples, and ventilation to remove formalin and alcohol fumes. The description should also emphasize responsibilities for QA, health and safety, sample handling and data management.

4.2 Sample Processing Procedures

Sample processing procedures must be documented in meticulous detail in order to facilitate repeatability of the study. This has been one of the most common shortcomings in invertebrate studies that focus on temporal trends. For example, studies have shown that there is up to two orders of magnitude difference in benthos abundance between samples sorted with the aid of a stereomicroscope compared to samples sorted, by eye, in white enamel trays (Burt *et al.*, 1989). It was common in the 1960s and 1970s to sort invertebrates in white enamel trays and recent studies can still be frequently encountered that highlight temporal trends in invertebrate density not realizing that the increase in density is simply attributed to sorting with a microscope and not to improvements in the health of the receiving environment. Reference must be made to SOPs for standard methods, and a brief summary of the key methods must be provided in final reports. Non-standard methods or deviations from standard methods should be described in detail. Descriptions should include or reference a description of the internal quality control, data validation, data management and reporting systems. Analytical procedures should be verified and method performance criteria stated.

4.2.1 Sample Receiving and Storage

Login, pre-treatment and storage, criteria for refusing samples, chain of custody and sample tracking mechanics, and archiving procedures should be described. All relevant field data should be captured during login. These data are often helpful in confirming the identity of rare taxa.

4.2.2 Instrument Calibration and Maintenance

In a taxonomic lab there is generally very little equipment that requires extensive maintenance and calibration. Samples are generally sorted with a microscope in order to meet the minimum recovery levels specified below, however, microscopes need very little maintenance as long as the lenses are not out of alignment.

Balances are often used for weighing subsample fractions or determining the biomass of invertebrates. Balances must be calibrated on a daily basis and a calibration log should be maintained.

Electronic or manual counters used for enumerating invertebrates should be checked on a routine basis to ensure the equipment is functioning properly.

Sieves should also be checked periodically to ensure there is no separation along the mesh seams.

4.2.3 Sample Sorting

Sorting benthic invertebrates involves the segregation of individual organisms within a sample or subsample generally into some predetermined taxonomic grouping. Sorters must be trained so that they can identify invertebrates to basic levels and be familiar with the different techniques used to separate organisms from different types of debris (e.g., gravel, organic matter, sand etc.). Inexperienced sorters (considered to be those that have processed less than 500 samples) must undertake some form of proficiency testing before working independently on samples. QC checks on new sorters should be frequent during a training program until it is clear that the sorter has consistently achieved the minimum acceptable level of sorting efficiency.

For any benthic invertebrate biomonitoring project, regardless of the experience of the sorter, a portion of the samples (generally 10 to 20%) must be resorted by the QC Officer or another qualified sorting technician to identify any missed specimens. The number of animals missed must be expressed as a percentage of the total. Generally there are two recommended ways to confirm sorting recovery. The first is to randomly select 10 to 20% of the samples and have them sorted in their entirety by another sorter. The second method is to resort 10 to 20% of each sample to confirm sorting recovery. The latter method is better in the sense that it examines all samples, however, it is more costly and also introduces additional error from subsampling because only 10 to 20% of the sample is sorted and any missed animals found are extrapolated to the whole sample.

The minimum acceptable level of invertebrate recovery is generally accepted to be 90 to 95%. An achievable standard is 95% recovery of total organisms in initial sorts (Environment Canada, 1993), which corresponds to a log scale standard deviation of about 0.02 between initial and final sorts. This standard was readily achieved by most reputable labs that participated in the first-cycle pulp and paper EEM program in Canada and it was agreed to as an acceptable level by many of Canada's benthic ecologists at a 1993 benthic biomonitoring workshop (Environment Canada, 1993). It is also the standard adopted by most regulatory agencies that have ongoing bioassessment programs using benthic invertebrates (e.g., U.S. EPA Puget Sound Monitoring program, 1987; EC/DFO Pulp and Paper EEM program, 1993a; U.S. EPA Contaminated Sediment (ARCS) program, 1994; U.S. EPA EMAP, 1994a).

Typically most programs stipulate that if the standard is not met, all samples should be resorted. This stipulation is generally overly conservative. Missing organisms is generally a function of the amount of organic debris. For example, if a sample is comprised of 500 mL of organic matter with only ten organisms, the probability of missing two animals is quite high which means a sorting recovery of only 80%. For these types of samples in which one or a few organisms represent >5% of the total abundance, the standard could specify the maximum number of organisms which could be missed on the initial sort (Environment Canada, 1993). For example, recovery of 9 out of 10 organisms from a sample might be considered acceptable, even though >95% of the organisms are not recovered.

When doing recovery checks it is also not uncommon for one of the samples to have a recovery slightly less than 95% whereas all the other checked samples are much higher. In these cases having one sample with slightly less than 95% recovery does not justify resorting all of the samples.

Since some benthic invertebrate monitoring programs accept 90% as the minimum acceptable sorting recovery (e.g., Plafkin *et al.*, 1997) even though 95% is readily achievable with trained sorters, it is recommended that the QA/QC criteria for invertebrate sorting recovery for biomonitoring of mine-related effects be that the average sorting recovery of the checked samples (10 to 20% of the total number of samples for an individual project) be at least 95% and no sample should have a sorting recovery of less than 90%. If these conditions are not met, then the samples must be resorted.

If a recovery problem is identified, it is prudent to assess the efficiency of any elutriation or other facilitation techniques used to reduce sorting costs to ascertain whether non-sorter related errors are the reason for the less than 95% recovery levels.

4.2.4 Subsampling

QA/QC checks for subsampling are intended to ensure that density estimates obtained through subsampling meet minimum precision requirements.

The current subsampling requirement adopted by the pulp and paper EEM program (EC/DFO, 1993a) is based on a minimum subsampling amount of 25 percent. Klemm *et al.* (1990) also recommends that a subsample should represent at least one-quarter of the original sample.

A minimum of 10% of subsampled samples must be evaluated for subsampling precision which is estimated in terms of total organism abundance and number of taxa. There is no specified criterion for acceptable subsampling error or recommended action if the target is not met.

Environment Canada (1993) has listed four criteria that must be met when subsampling benthos for biomonitoring:

1. the fraction subsampled must be a known percentage of the total sample;
2. the subsamples must be representative of the entire sample and should not be biased towards or against certain taxa;
3. the variance associated with subsampling should be small relative to the variance among replicate samples. If this criterion is not met, then subsampling will reduce the power of any statistical tests; and
4. subsampling methods should be easy to use and have explicit SOPs.

The National Water Quality Assessment Program for benthic macroinvertebrate studies (U.S. Geological Survey, 1993) requires that consecutive 1/4-subsamples are processed until either the QA/QC criteria are met or the entire sample is processed. The two

QA/QC criteria used to evaluate subsampling are that the number of taxa the subsamples have in common must be at least 90% of the combined number of taxa and the similarity between the two communities described by the subsamples must be at least 90%, as determined using a percentage similarity coefficient (PSC). The PSC between two subsamples is calculated as:

$$PSC = 100 - 0.5*\Sigma \ |a_i-b_i|)$$

where a_i and b_i are the percentages of taxon i in the first and second subsample fractions, respectively.

An alternative approach to investigating the effects of subsampling error on the outcome of data analysis is to calculate the metrics on which the interpretation will be based (e.g., indices, principal component scores) for each of the subsamples. The subsampling error should be small relative to the relevant error term used in the statistical analysis. Elliott (1977) reports that a standard error of about 20% of the mean is an acceptable error for most benthic invertebrate samples. Therefore in routine biomonitoring surveys it is only reasonable to expect to reach this level of precision for variables such as total number of invertebrates or number of individual dominant (comprising at least 10% of the total abundance) taxa (Alberta Environment, 1990).

The criteria for acceptability of subsampling error will be determined at the final design stage considering the components of variability in the available preliminary data, and the effect of this variability on the study's ability to detect ecological differences and relationships according to the study hypotheses. This exercise is part of the process of developing meaningful Data Quality Objectives. The example of DQOs provided in Appendix 2 established that subsampling error should not exceed 20% of the within area standard deviation.

It has been suggested that subsampling fractions should comprise a minimum of 25% of the original sample or a fraction that contains 300 to 500 organisms. If the latter method is followed, the weight of the fraction containing the prescribed number of animals would be used to estimate the total number of organisms in the sample.

Regardless of which method is used, final reports must specify the error associated with the subsampling method chosen. All sorted sample debris should be represerved in 10%

buffered formalin or in 70% alcohol and stored for up to six months after the submission of final reports in the event that report reviewers or other researchers wish to re-examine the samples or sort additional subsamples.

4.2.5 Invertebrate Identification

Discrepancies in identification and quantification can take three forms: (1) simple errors in counts, (2) errors in invertebrate identification that require adjustment to counts but not to taxa lists and (3) errors in identification that require adjustments to counts and additions to the taxa list (U.S. Geological Survey, 1993). In order to establish confidence in the quality of taxonomic work, and to permit checking of this work, Environment Canada (1993) recommends that:

- identifications should be verified by an expert in the taxonomic group of interest (a working definition of a taxonomic expert is an individual who specializes in the taxonomy of a particular group or an individual who through work experience has more than ten years of taxonomic experience and has processed more than 5,000 samples (D. Zaranko, ZEAS, 1998, pers. comm.));
- persons who carry out the identifications must be identified in final reports, together with details of their qualifications and experience;
- literature and taxonomic keys used for benthos identification must be referenced; and
- details of both reference and voucher collections should be given (institutions or agencies holding the specimens; catalogue numbers of the specimens, etc.).

Reference collections consist of specimens for which confirmed identifications have been provided for each specimen, including the authorities who verified the specimens, their affiliations and the dates of the verifications. These collections are used by the laboratory staff for verifying invertebrate identifications. Voucher collections (or specimens) are site-specific representative collections from the project under evaluation. Voucher collections must be assembled for all studies, with details of their deposition (agency or institution name; catalogue or reference numbers) indicated in the final report. Voucher collections for projects that assess temporal trends must be archived for the life of the project. This was the approach adopted by the pulp and paper EEM program (EC/DFO, 1993a).

The National Water Quality Assurance Program for benthic macroinvertebrate studies (U.S. Geological Survey, 1993) requires a thorough check of 10 percent of the samples including taxonomic verification and enumeration checks. Percent difference between original and check values is calculated for total abundance and number of taxa. The Bray-Curtis dissimilarity index (Bray and Curtis, 1954) is calculated on the original and corrected values. If the degree of similarity (1-Bray Curtis index) is less than 90%, the sample fails the criterion.

In addition, if more than 10% new taxa are added as a result of the taxonomic verification, the sample fails the criterion. In the event of a QC check failure, the sample must be re-processed. If other re-checked samples fail the QC criteria, the processing techniques must be re-evaluated and, if necessary, re-adjustment of the QC criteria is considered.

An alternative approach to investigating the impact of identification and enumeration error on the outcome of an analysis is to calculate the metrics on which the interpretation will be based (e.g., indices, principal component scores) for both original and check values of the sample. The identification and enumeration error should be small relative to the relevant error term used in the statistical analysis.

The minimum level of quality assurance for taxonomic identifications is the compilation of a voucher collection and archiving of the sorted invertebrates for at least 10 years.

4.2.6 Laboratory Data Management

Standardized forms facilitate consistent data recording, reporting and tracking of samples through the processing stages. They should allow flagging of questionable data, with some indication of the reason. Automated information management systems facilitate real-time data validation and flagging.

All data entered into spreadsheets must be checked by another staff member to ensure that there are no transcription errors.

The laboratory management system should include maintenance of chain-of-custody records and sample tracking logs. Bench sheets, voucher specimens, sorted invertebrates, and calibrations logs should all be archived for at least ten years and this information should be incorporated in the laboratory's data management plan.

5.0 STATISTICAL ANALYSIS AND INTERPRETATION

5.1 Data Management

In order to ensure data integrity throughout the study process, access to the master data files must be controlled. There should be updating and archiving procedures in place. Data as entered should go through a validation procedure, involving transcription checks, comparison of associated QC data to QC criteria, and assignment of data quality flags. The date of last change to either data or quality flags should be known.

Once the data have been quality control checked, the data should be stored in read-only files to prevent accidental changes to the data.

5.2 Statistical Procedures

Selected statistical procedures must be appropriate to the study hypotheses (or questions) and the study design. They should conform to the procedures identified at the study design stage. Any deviation should be justified as necessary and power implications should be considered. Investigators should also indicate the rationale for choosing a particular method (Environment Canada, 1993).

In the final reports the statistical methods must be documented so that the analyses can be repeated by another investigator. All raw data must also be available in the report, usually in appendices.

5.3 Data Reduction

Data reduction may include selection of variables or construction of composite variables (e.g., principal components) in order to retain statistical degrees of freedom, or to ensure that meaningful variables are included in the analysis. Reduction procedures should be objective rather than subjective (i.e., there should be specific criteria). The use of ecological relevance as a basis for selection of taxa variables may be difficult in that ecological characteristics of all taxa may not be known.

5.4 Assumptions and Transformations

Data should be investigated for adherence to the assumptions that are associated with the statistical procedures used. Where assumptions are violated, transformations should be selected to rectify (or improve) the situation. If assumptions cannot be approximately satisfied then alternate (assumption free) statistical methods should be utilized. All transformations performed must be described and any remaining violations of assumptions must be documented.

All raw data files must be in an appendix and statistical data files should be kept for at least ten years especially if the project is part of a long-term monitoring program (Environment Canada, 1993).

5.5 Missing Values and Outliers

Missing values can be an important consideration because they may necessitate the dropping of important cases or parameters. Missing variables may be replaced by the mean of an appropriate subset of the data (e.g., from adjacent locations or times), however this must be clearly documented if performed. Outliers are generally identified based on the selected statistical model (i.e., by examination of residuals) using an objective criterion (e.g., 2 to 3 standard deviations). Their validity should be confirmed and if they are valid observations, the analysis should be repeated with and without these outliers to check the robustness of the conclusions.

Any missing data or outliers must be clearly identified in the final reports.

5.6 Ecological Interpretation

It is important to remember that statistical significance does not equal ecological significance. Statistically significant results should be evaluated as to their ecological significance. The ecological significance of an effect can be evaluated against the yardstick of normal temporal and/or spatial variation. If derived variables have been utilized (e.g., principal components), their ecological meaning should be examined. Both input variables and statistical results should make ecological sense.

5.7 Statistical Analysis Review

This review includes evaluation of repeatability by checking that reports include data files and clearly defined procedures and assumptions, and perhaps by having the analysis independently repeated from this documentation. The review also may include validation of statistical packages used (e.g., by comparison of results among packages). Conclusions should be checked for robustness to outliers, transformations, data reductions and alternate statistical methods.

6.0 REPORTING

At the completion of any project using benthic invertebrates to assess mine-related impacts there must be a quality assurance data review and report. This includes data validation and reporting whether or not the quality control criteria were met. Any bias correction of data (e.g., correction between studies for differences in taxonomy or selection of taxa for inclusion in subsequent statistical analyses) also needs to be reported. Validation is based on associated QC data, in relation to invalidation or flagging criteria.

Corrective actions may be taken in real-time or after quality control or project data reviews. Specific trigger conditions should be defined for each corrective action. Any such action taken must be recorded and its effectiveness assessed.

The final report must include a quality assurance report which should contain study design recommendations based on assessment of DQO achievement. The content is a summary of results from all of the quality assurance operations.

The following is a checklist of quality assurance components that have been addressed in this report and that should be contained in the quality assurance report (Table 6.1). Some of the items are considered desirable, whereas others are considered to be mandatory (i.e., minimum quality control requirements) for any benthic invertebrate monitoring program in order to provide data of known quality.

TABLE 6.1: CHECKLIST FOR A QUALITY ASSURANCE PROGRAM

	Desirable	Mandatory
Quality Management Plan		
Quality Assurance Policy Statement	√	
Study Objectives		√
Overview of Study	√	
Study Design (e.g., sampling strategy, level of effort)		√
Identification of Data Uses	√	
Decision Framework (if applicable)	√	
Data Collection Activities		√
Data Quality Objectives		√
Study Organization		√
Field Operations		
Health and Safety Plan	√	
Sampling Design		√
Sampling Procedures (SOPs)		√
Equipment Calibration		√
Chain-of-Custody and Standardized Data Collection Forms	√	
Staff Training		√
Replication of Supporting Measurements (e.g., dissolved oxygen)		√
Laboratory Operations		
Facility Description	√	
Documentation of Sample Processing		√
Standard Operating Procedures		√
Chain-of-Custody and Sample Tracking	√	
Instrument Calibration		√
Staff Training		√
Sorting Recovery		√
Subsampling Error		√
Archive Sorted Debris	√	
Taxonomy Verified by Expert	√	
Documentation of Taxonomists' and Sorters' Qualifications		√
List of Taxonomic Keys Used		√
Reference Collection	√	
Voucher Collection		√
Internal Taxonomic Verification Checks	√	
Data Management Procedures	√	

TABLE 6.1: CHECKLIST FOR A QUALITY ASSURANCE PROGRAM (cont'd)

	Desirable	Mandatory
Data Transcription Checks		√
Bench Sheets Archived	√	
Sorted Invertebrates Archived		√
Statistical Analysis		
Data Management System	√	
Documentation of Procedures		√
Presentation of Data Files		√
Independent Review	√	

7.0 REFERENCES

Alberta Environment. 1990. Selected Methods for the Monitoring of Benthic Invertebrates in Alberta Rivers.

Bray, J.R. and J.T. Curtis. 1954. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. 27: 325-349.

Burt, A.B., P.M. McKee and D. Hart. 1989. Effects of Pollution on Benthic Invertebrate Communities of the St. Mary's River, 1995. Report to the Ontario Ministry of the Environment.

Davis, W.S. and T.P. Simon (Eds.). 1995. Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. CRC Press, Inc.

Ellliott, J.M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. Freshw. Biol. Ass. Sci. Pub. No. 25. Second Edition. 160 pp.

Environment Canada and Department of Fisheries and Oceans (EC/DFO). 1993a. Technical Guidance Document for Aquatic Environmental Effects Monitoring Related to Federal Fisheries Act Requirements. Version 1. April 1993.

Environment Canada and Department of Fisheries and Oceans (EC/DFO). 1991. Aquatic Environmental Effects Monitoring Requirements, Annex 1, Aquatic Environmental Effects Monitoring at Pulp and Paper Mills and Off-Site Treatment Facilities Regulated under the Pulp and Paper Effluent Regulations. EPS 1/RM/18. Environmental Protection Service, Ottawa.

Environment Canada. 1993. Guidelines for Monitoring Benthos in Freshwater Environments. Report prepared for Environment Canada by EVS Consultants, North Vancouver, B.C.

Environment Canada. 1998. Pulp and Paper Technical Guidance for Aquatic Environmental Effects Monitoring. EEM/1998/1.

Gibson, G.R. (Editor). 1994. Biological Criteria: Technical Guidance for Streams and Small Rivers. EPA-822-B-94-001. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC. 162. pp.

Gibson, G.R., Jr. 1995. Estuarine/Near Coastal Marine Waters: Bioassessment and Biocriteria Technical Guidance. U.S. Environmental Protection Agency, Office of Science and Technology, Health and Ecological Criteria Division.

Klemm, D.J., Q.J. Stober and J.M. Lazorchak. 1993. Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters. EPA/600/R-92/111. U.S. EPA, Office of Modelling, Monitoring Systems and Quality Assurance Office of Research and Development, Cincinnati, OH.

Klemm, D.J., P.A. Lewis, F. Fulk and J.L. Lazorchak. 1990. Macroinvertebrate Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters. EPA/600/4-90/030. U.S. EPA, Office of Modelling, Monitoring Systems and Quality Assurance Office of Research and Development, Cincinnati, OH.

Plafkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross and R.M. Hughes. 1997. Revision to: Rapid Bioassessment Protocols for Use in Streams and Rivers: Periphyton, Benthic Macroinvertebrates, and Fish. EPA Internet Website.

Rosenberg, D.M. and V.H. Resh (Eds.). 1993. Freshwater Biomonitoring and Benthic Macroinvertebrates. Chapman &Hall, New York.

Taylor, B.R. 1996. Optimization of Field and Laboratory Methods for Benthic Invertebrate Biomonitoring. Second Draft Report.

Tetra Tech, Inc. 1986. General QA/QC Considerations for Collecting Environmental Samples in Puget Sound. Prepared for U.S. EPA, Seattle, WA.

Tetra Tech, Inc. 1987. Recommended Protocols for Sampling and Analyzing Subtidal Benthic Macroinvertebrate Assemblages in Puget Sound. Prepared for U.S. EPA, Seattle, WA.

U.S. Environmental Protection Agency (EPA). 1990. Quality Assurance Glossary and Acronyms, U.S. EPA, Quality Assurance Management Staff, Office of Modeling, Monitoring Systems and Quality Assurance, Office of Research and Development, Washington, D.C. July 1990.

U.S. Environmental Protection Agency (EPA). 1991. Environmental Monitoring and Assessment Program, Quality Assurance Program Plan. Environmental Monitoring Systems Laboratory, Cincinnati.

U.S. Environmental Protection Agency (EPA). 1994. Assessment and Remediation of Contaminated Sediments (ARCS) Program - Assessment Guidance Document, Chapter 7: Assessment of Benthic Macroinvertebrate Community Structure. EPA-905-B94-002.

U.S. Environmental Protection Agency (U.S. EPA). 1994a. Quality Assurance Project Plan for the southern California Bight Pilot Project. EPA Environmental Monitoring and Assessment Program.

U.S. Environmental Protection Agency (EPA). 1995. Generic Quality Assurance Project Plan Guidance for Programs Using Community Level Biological Assessment in Wadable Streams and Rivers. EPA-B-95-004. Office of Water, Washington, DC.

U.S. Geological Survey. 1993. Guidelines for the Processing and Quality Assurance of benthic Invertebrate Samples Collected as Part of the National Water-Quality Assessment Program. U.S. Geological Survey. OFR-93-407. Denver, CO.

Zar, J.H. 1974. Biostatistical Analysis. Prentice-Hall, Inc., Englewood Cliffs, N.J.

APPENDIX 1

Glossary of Terms

APPENDIX 1: GLOSSARY OF TERMS

Acceptance criteria - criteria specifying the limit above which data quality is considered satisfactory and below which it is not [Modified from USEPA (1990) "Acceptable quality level"].

Accuracy - the degree of agreement between an observed value and an accepted reference value. Accuracy includes a combination of random error (precision) and systematic error (bias) components which are due to sampling and analytical operations; a data quality indicator. The U.S. EPA recommends that this term not be used and that *precision* and *bias* be used to convey the information usually associated with accuracy [Klemm *et al.*, (1993)].

Assemblage - an association of interacting populations of organisms in a given waterbody, for example, fish assemblage or a benthic macroinvertebrate assemblage [Gibson (1994)].

Benthos - Benthos refers to invertebrates living in or on the sediments of aquatic habitats.

Bias - the systematic or persistent distortion of a measurement process which deprives the result of Representativeness (i.e., the expected sample measurement is different than the sample's true value). A data quality indicator [Klemm *et al.*, (1993)].

Biological assessment/Bioassessment - an evaluation of the condition of a waterbody using biological surveys and other direct measurements of the resident biota in surface waters [Gibson (1994), USEPA (1991)].

Biological integrity - the condition of an aquatic community inhabiting unimpaired waterbodies of a specified habitat as measured by an evaluation of multiple attributes of the aquatic biota. Three critical components of biological integrity are that the biota is (1) the product of the evolutionary process for that locality, or site, (2) inclusive of a broad range of biological and ecological characteristics such as taxonomic richness and compositions, trophic structure, and (3) is found in the biogeographic region of study [Gibson (1994)].

Biomonitoring - multiple, routine biological assessments over time using consistent sampling and analysis methods for detection of changes in biological condition.

Calibration - to determine, by measurement or comparison with a standard, the correct value of each scale reading on a meter or other device, or the correct value for each setting of a control knob. The levels of the calibration standards should bracket the range of planned measurements [USEPA (1990)].

Community - any group of organisms belonging to a number of different species that co-occur in the same habitat or area; an association of interacting assemblages in a given waterbody.

Comparability - the degree to which different methods, data sets and/or decisions agree or can be represented as similar; a data quality indicator [Klemm *et al.,* (1990; 1993)].

Completeness - the amount of valid data obtained compared to the planned amount, and usually expressed as a percentage; a data quality indicator [USEPA (1990); Klemm *et al.*, (1993)].

Confidence level - the probability, usually expressed as a percentage, that a confidence interval will include a specific population parameter; confidence levels usually range from 90 to 99 percent [USEPA (1990)].

Confidence interval - an interval that has the stated probability (e.g., 95 percent) of containing the true value of a fixed (but unknown) parameter [Gibson (1994)].

Corrective action - corrective actions are measured to correct identified problems, maintain documentation of the results of the corrective process, and continue the process until each problem is eliminated. The corrective action is the process to remediate defects.

Damaged and unusable samples - are samples that have been damaged and part of all of the sample was destroyed or not recoverable.

Damaged and usable samples - samples that have been damaged but the entire sample was salvageable (i.e., *all* organisms were saved).

Data quality objectives (DQOs) - qualitative and quantitative statements developed by data users to specify the quality of data needed to support specific decisions; statements about the level of uncertainty that a decisionmaker is willing to accept in data used to support a particular decision. Complete DQOs describe the decision to be made, what data are required, why they are needed, the calculations in which they will be used; and time and resource constraints. DQOs are used to design data collection plans [Gibson (1994)].

Data reduction - the process of transforming raw data by arithmetic or statistical calculations, standard curves, concentration factors, etc., and collation into a more useful form [USEPA (1990)].

Data validation - the process of substantiating specified performance criteria [Klemm *et al.*, (1993)].

Data verification - the ability to be proven or substantiated [Klemm *et al.*, (1993)].

Ecological integrity - the condition of an unimpaired ecosystem as measured by combined chemical, physical (including habitat), and biological attributes [Gibson (1994)].

Endpoints - a measurable ecological characteristic [Klemm *et al.*, (1993)].

Environmental monitoring - the periodic collection of data to be used to determine the condition of ecological resources [Klemm *et al.*, (1993)].

Indicator - characteristics of the environment, both abiotic and biotic, that can provide quantitative information on ecological resources [Klemm *et al.*, (1993)].

Interlaboratory - activities that occur among different laboratories [USEPA (1990)].

Intralaboratory - activities that occur within a laboratory [USEPA (1990)].

Level of effort - the amount of effort (e.g., person-hours, sampling effort per time, or sampling vigor) needed to complete a task or project.

Measurement parameters - any quantity such as a mean or standard deviation characterizing a population. Commonly misused for "variable", "characteristic" or "property" [USEPA (1990)].

Measurement quality objectives (MQO) - the QA objectives for precision, representativeness, comparability and completeness for each measurement [USEPA (1995)].

Metric - a calculated term or enumeration which represents some aspect of biological assemblage structure, function or other measurable aspect of a characteristic of the biota that changes in some predictable way with increased human influence [Gibson (1994)].

Multimetric approach - is an assessment approach that uses a combination of multiple metrics to provide synthetic assessments of the status of water resources [Gibson (1994)].

Percent recovery - for invertebrate sorting accuracy is usually calculated as "percent recovery" and is applied in the form of sample sorting checks.

Performance audit - a type of audit in which the quantitative data generated in a measurement system are obtained independently and compared with routinely obtained data to evaluate the proficiency of an analyst or laboratory [USEPA (1990)].

Pilot studies - studies implemented based on questions that require field work to evaluate indicators, sampling strategy, methods and logistics [Klemm *et al.*, (1993)].

Precision - the degree of variation among individual measurements of the same property, usually obtained under similar conditions; a data quality indicator. Precision is usually expressed as standard deviation, variance or range, in either absolute or relative terms [USEPA (1990)].

Preventive maintenance - an orderly program of activities designed to ensure against equipment failure [USEPA (1990)].

Quality Assurance (QA) - an integrated system of activities involving quality planning, quality control, quality assessment, quality reporting and quality improvement to ensure that a product or service meets defined standards of quality with a stated level of confidence [USEPA (1990)].

Quality objectives - the upper and lower limiting values of the data quality indicators as defined by the data user's acceptable error bounds [USEPA (1990)].

Quality Control (QC) - the overall system of technical activities whose purpose is to measure and control the quality of a product or service so that it meets the needs of users. The aim is to provide quality data or results that are satisfactory, adequate, dependable, and economical [USEPA (1990)].

Quality Assurance Plan (QAP) - a formal document describing the detailed quality control procedures by which the data quality requirements defined for the data and decisions in a specific project are to be achieved [USEPA (1990)].

Raw data - data that have not been manipulated; the actual measurements taken.

Reference collection - a set of biological specimens, each representing some taxonomic level and not necessarily limited to specific projects or activities.

Representativeness - the degree to which data accurately and precisely represent the frequency distribution of a specific variable in the population; a data quality indicator [USEPA (1990)].

Sensitivity - capability of method or instrument to discriminate between measurement responses of a variable of interest [USEPA (1990)].

Subsampling - a subset of a sample; subsample may be taken from any laboratory or field sample [USEPA (1990)].

Type II error - (beta error) an incorrect decision resulting from acceptance of a false null hypothesis (a false negative decision) [USEPA (1990)].

Type I error - (alpha error) an incorrect decision resulting from the rejection of a true null hypothesis (a false positive decision) [USEPA (1990)].

Uncertainty of data - a measure of the total variability associated with sampling and measuring, taking into account two major error components: systematic error (bias) and random error [USEPA (1990)].

Voucher collection - a curated collection consisting of the actual specimens collected in a survey that is maintained following identification and enumeration.

APPENDIX 2

Example of Data Quality/Quantity Objectives and Corrective Actions

APPENDIX 2: EXAMPLE OF DATA QUALITY/QUANTITY OBJECTIVES AND CORRECTIVE ACTIONS

Study Objective: To ascertain whether there are impacts on a stream benthic community from mine-related activities.

Design Approach: An area downstream of the mine site, which receives mine effluent, and is relatively homogeneous with respect to habitat and effluent exposure, will be compared to an area of similar habitat upstream of the mine site. A simple CI study design will be used where an equal number of stations will be sampled in each area.

Data Quality/Quantity Objectives:

1. The objective is to be able to detect a 50% decrease in total organism abundance, if one really exists, between upstream and downstream areas, with a small chance of missing the difference (Type II error) ($\beta<0.1$) and a detection criterion that minimizes the chance of false positive detection (Type I error) ($\alpha<0.05$).

2. Data Quantity: The number of samples required to detect a specified effect size can be determined from previous sampling data using the following power equation (Environment Canada, 1998)

$$n \geq 2\,(Z_\alpha + Z_\beta)^2\,(S_T/\Delta)^2 + 0.25Z_\alpha^{\,2}$$

where:

$\mathbf{S_T}$ = standard deviation in log abundance = **0.123384** (from Table A2.1). It is derived from the MSE (mean square error) from the ANOVA conducted on the preliminary survey results presented in Table A2.1.

Δ = -log(0.5) = 0.301 (based on the objective of detecting a 50% decrease in abundance). This is the minimum effect size (50%) which we are interested in detecting.

$\mathbf{Z_\alpha}$ = 1.645 (one-tailed) (proportion of the normal curve table, e.g., Zar, 1974, p. 412). This is the detection criterion that minimizes the chance of a false positive detection (Type I error) ($\alpha<0.05$). A one-tailed test is used where the investigator knows the direction (i.e., less than the reference value) of the difference being tested for. A two-tailed value (1.96) would be used if the investigator did not know the direction of the difference and/or if it did not matter (e.g., the number of taxa would be less if an effluent was toxic or the number of taxa could be more if an effluent causes nutrient enrichment).

$\mathbf{Z_\beta}$ = 1.282 (one-tailed) (proportion of the normal curve table). This is the detection criterion which establishes the chance of missing a difference that really exists (Type II error) ($\beta<0.1$ or 10%). In this example, assuming the above values for the power equation and using the data in Table A2.1, at least

4 stations per area would be necessary. If the Z_α for a two-tailed test were assumed (1.96), then at least 5 stations per area would be necessary.

3. Relevant Error: Standard deviation (S_T) ≤ **0.123384**. This is the value used for study design purposes in item 2 above. It is derived from the MSE (mean square error) from the ANOVA conducted on the survey results presented in Table A2.1.

4. Subsampling Error: Standard deviation between subsamples (S_1) was found to be ≤ 0.02 (generally achievable based on the data, see Table A2.1). For this study the investigator may consider setting a DQO for S_1 of ≤ 0.2 S_T (action level) on log scale. The action level is designed to be both achievable and small relative to the error (S_T) that is expected to be used in testing hypotheses.

5. Sorting Error: Standard deviation between sorts (S_2) was found to be < 0.02 (generally achievable based on the data, see Table A2.1). Similar to the DQO for subsampling error, a DQO for sorting (S_2) could be established as ≤ 0.2 S_T (action level) on log scale. The action level is designed to be both achievable and small relative to the error (S_T) that is expected to be used in testing hypotheses.

Note: S_T could increase by 3% if action levels of S_1 and S_2 come into play (i.e., new $S_T^2 = S_T^2 + S_1^2 + S_2^2$); the corresponding 3% increase in detectable Δ (on log scale) may be considered marginally acceptable. If not acceptable then corrective action would have to be taken by sorting additional subsamples or achieving a higher sorting recovery.

Corrective Actions:

1. If generally achievable errors are frequently exceeded in subsampling or sorting (i.e., in more than 50% of check samples) then investigate possible reasons and consider staff retesting/retraining.

2. If project action levels are exceeded in any check sample, then investigate impact on ability to detect differences between areas, consider restrictions on subsampling and consider routine resorting.

TABLE A2.1: BENTHIC INVERTEBRATE SURVEY DATA (taken from the AETE program)

	Invertebrate Abundance (number/m^2)		Log$_{10}$ Abundance	
Station	Reference	Exposure	Reference	Exposure
1	898	161	2.9533	2.2068
2	666	197	2.8235	2.2945
3	845	320	2.9269	2.5051
4	634	211	2.8021	2.3243
5	519	267	2.7152	2.4265
6	454	351	2.6571	2.5453
Mean	**669**	**251**	**2.8130**	**2.3838**
Standard Deviation	**175**	**74**	**0.116**	**0.131**
Difference	**418**		**0.4292**	

ANOVA: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	6	16.87792	2.81299	0.013352
Column 2	6	14.30254	2.38376	0.017095

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.552714	1	0.55271	36.30622	**0.000128**	4.96459
Within Groups	0.152237	10	0.01522			
Total	0.70495	11				

$S_T = (MSE)^{0.5} = (0.015244)^{0.5} = \mathbf{0.123384}$

Subsampling Error

Station	No. Animals Fraction 1	No. Animals Fraction 2	Standard Deviation	S_1	
HS-E1	150	142	5.65		
HS-E4	253	245	5.65	**0.017**	(maximum of the two standard deviations on log scale)

Sorting Error

Station	No. Animals Recovered	No. Animals Re-sort	Percent Recovery	S_2	
HS-R2	746	12	98.4		
HS-E5	695	7	99.0	**0.005**	(maximum of the two standard deviations on log scale)