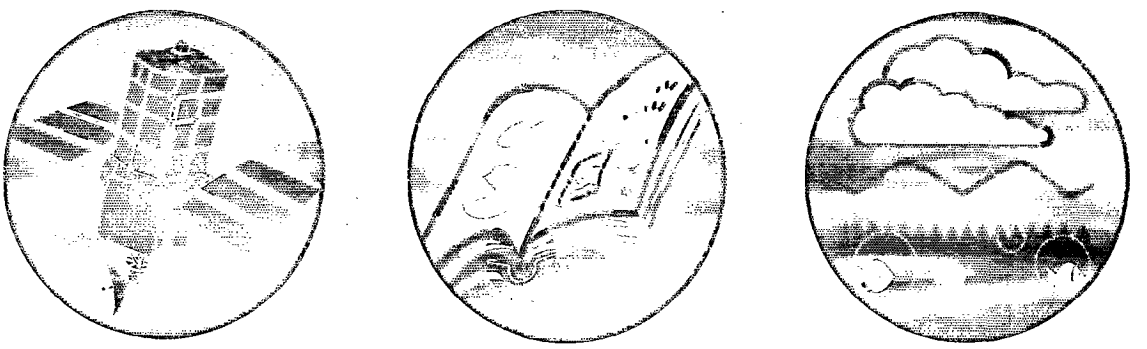
A Proposed Scheme to Ensure the Quality of Data Generated by Laboratories Undertaking Regulatory Ecotoxicological Testing



Research and Development

Technical Report
P166

ENVIRONMENT AGENCY

All pulps used in production of this paper is sourced from sustainable managed forests and are elemental chlorine free and wood free

A Proposed Scheme to Ensure the Quality of Data Generated by Laboratories Undertaking Regulatory Ecotoxicological Testing

Recommended Procedures for Laboratories Involved in Generating Regulatory Ecotoxicological Data for Direct Toxicity Assessment

R&D Technical Report P166

P Whitehouse

Research Contractor:
WRc Plc

Further copies of this report are available from:
Environment Agency R&D Dissemination Centre, c/o WRc, Frankland Road, Swindon, Wilts SN5 8YF

tel: 01793-865000 fax: 01793-514562 e-mail: publications@wrcplc.co.uk

Publishing Organisation:
Environment Agency
Rio House
Waterside Drive
Aztec West
Almondsbury
Bristol BS32 4UD

Tel: 01454 624400 Fax: 01454 624409
ISBN:1 85705 168 8

This report is the result of work jointly funded by the Environment Agency and the Scotland and Northern Ireland Forum for Environmental Research (SNIFFER).

Dissemination status
Internal: Released to Regions
External: Released to the Public Domain

Statement of use
This report represents collated documents describing proposals developed to ensure that good quality information is produced by any laboratory generating ecotoxicological data for regulatory use. Although the scheme was developed for use in the Direct Toxicity Assessment (DTA) whole effluent toxicity-based control initiative, it has relevance to all ecotoxicological testing programmes both within and outside the Agency.

Research contractor
This document was produced under R&D Project 493 by:

WRc plc
Henley Road
Medmenham
Marlow
Buckinghamshire
SL7 2DH

Tel: 01491 636500 Fax: 01491 636501

Environment Agency Project Leader
The Environment Agency's Project Leader for R&D Project 493 was Dr Jim Wharfe - Ecotoxicology and Hazardous Substances National Centre, Wallingford.

The Environment Agency's Manager for this package of work was Dr Derek Tinsley - Ecotoxicology and Hazardous Substances National Centre, Wallingford.

SNIFFER's Project Manager was Mr Dave Campbell, SEPA East Region, Clearwater House, Herriot-Watt Research Park, Avenue North, Riccarton, Edinburgh EH14 4AP

R&D Technical Report W84

CONTENTS

LIST OF FIGURES

EXECUTIVE SUMMARY

Assessing the biological effects of whole samples (e.g. effluents, receiving waters, soils and sediments) are set to play a more prominent role in controlling and monitoring the release of toxic wastes and in classifying the hazards of complex materials. Although Direct Toxicity Assessment (DTA) has many potential applications, much of the effort to date has been directed toward developing DTA as a tool alongside traditional chemical and biological survey techniques, for water quality management. Under the proposals for the control of effluents by DTA, regulatory decisions about the acceptability of effluent toxicity will be made on the basis of test data obtained from ecotoxicity tests. Consequently, the data generated in these tests play a critical part in sound regulatory decision making.

Like all biological and chemical measurements, determinations of toxicity exhibit variability which in turn can introduce uncertainty into decisions made on the basis of test results. As a result, this project makes recommendations for a comprehensive scheme aimed at promoting the quality of ecotoxicity test data used for regulatory purposes. These are based on the combination of two related components: (1) standardised test methods, described in 'Methods Guidelines' produced under R&D Project EMA 003 and (2) a Regulatory Ecotoxicology Testing Quality Scheme (RETQS, aka Register of Approved Laboratories, RAL) specifying Quality Assurance (QA) procedures for laboratories undertaking such tests and also an external Quality Control (QC) scheme, based on research carried out under R&D Project I550 ('Performance Standards for Ecotoxicity Tests'). Relevant outputs relating to the development of the RETQS have been compiled in this report.

The QA component is designed to ensure the integrity and auditability of ecotoxicity test data and is realised by requiring test laboratories to either be compliant with the Principles of Good Laboratory Practice (GLP) or to be accredited under the United Kingdom Accreditation Service (UKAS, formerly NAMAS). The QC procedures involve minimum standards for the performance of regulatory testing, particularly the amount of bias and variability that is permissible using reference toxicants. Compliance with those limits is then assessed using results obtained from regular tests carried out by participating laboratories with one of these reference toxicants. By combining QA and QC measures, the intention is to ensure that ecotoxicity tests are performed to a high standard and that the validity of the resulting data are beyond question.

When DTA is employed as a regulatory tool, laboratories generating test data will be required to participate in the RETQS. However, the RETQS is first to be piloted as part of the DTA Demonstration Programme (R&D Project P2-094). This will also provide an opportunity to compare a number of QC options (described in Annex IIId) so that a final recommendation can be made which provides an acceptable balance between cost and the ability to make valid assessments of test performance.

This report provides an overview of the components of the proposed Regulatory Ecotoxicology Testing Quality Scheme and collates documents produced during the development of this scheme.

KEY WORDS

Register of Approved Laboratories, RETQS, RAL, DTA, ecotoxicity, Quality Assurance, QA, Quality Control, QC, 'Methods Guidelines'

1. INTRODUCTION

1.1 Biological effects measures as a regulatory tool

Assessment of the biological effects of whole samples (e.g. effluents, receiving waters, soils and sediments) are set to play a more prominent role in controlling and monitoring the release of toxic wastes and in classifying the hazards of complex materials. Assessments based on ecotoxicity measures have the advantage of providing cost-effective summary information which relates directly to the impact of chemical mixtures on exposed organisms and can provide early warning of potential adverse impacts. When coupled with procedures to identify toxic substances in complex mixtures, ecotoxicological testing also provides a better link between cause and effect. The Environment Agency National Centre for Ecotoxicology and Hazardous Substances (NCEHS) Direct Toxicity Assessment (DTA) unit, in collaboration with other UK and international agencies and industry, is formulating procedures and developing appropriate methods for the ecotoxicological assessment of whole samples to meet the Agency's regulatory needs.

1.2 The current Direct Toxicity Assessment initiative

Although DTA has many potential applications, much of the effort to date has been directed toward developing DTA as a tool alongside traditional chemical and biological survey techniques, for water quality management. This work was progressed through a collaborative venture between the Environment Agency, Scottish Environmental Protection Agency (SEPA) and Department of the Environment, Northern Ireland (DoE NI) between 1993 and 1996. In 1996, following a pilot study, the UK regulators put forward proposals for using toxicity-based criteria for the regulatory control of wastewater discharges as part of a consultation exercise with industry and other interested parties. The process began with the release of a consultation document in July 1996 at a Society for Environmental Toxicology and Chemistry (SETAC) meeting in Luton and was followed by a seminar/workshop in Torquay in October 1996 at which the main issues raised through consultation were discussed. This consultation exercise was very successful and there was general support for the concept of DTA. As a result of the consultation, it was agreed that the regulators' DTA objectives could be achieved through application of action and trigger levels within discharge licences as opposed to pass/fail discharge limits, use of improvement plans, consideration of costs and benefits at each major stage of the process, and by selecting sites based on environmental need in the first instance.

It was also decided at Torquay that DTA should be further assessed in a demonstration programme in which a revised protocol would be tested at selected UK sites. A joint regulator/industry steering group was set up in January 1997 to plan the programme. The group consists of representatives of the Environment Agency, Scotland and Northern Ireland Forum for Environmental Research (SNIFFER), the UK Water Industries Association (UKWIA), the Confederation of British Industry (CBI) and the Chemical Industries Association (CIA). Three study sites were selected and the two-year study is due to be completed at the end of 1999. The key outputs from the programme will be recommendations for the implementation of DTA and also on testing methods and procedures. Working groups

have been set up by the steering group to advise on the legal aspects of introducing DTA as a regulatory tool, the technical aspects of tracing sources of toxicity in sewerage systems, DTA methods and associated Quality Assurance (QA) and Quality Control (QC) procedures.

1.3 The importance of data quality

Under the proposals for the control of effluents by DTA, regulatory decisions about the acceptability of effluent toxicity will be made on the basis of test data obtained from ecotoxicity tests. Consequently, the data generated in these tests play a critical part in sound regulatory decision-making. However, like any form of measurement, results of ecotoxicity testing are liable to variability. To be useful, ecotoxicity tests should generate data which are both accurate and precise. It is important to recognise the difference between these concepts and that they apply just as much to ecotoxicity testing as they do to a manufacturing process or to chemical analyses. These terms are defined below:

Accuracy: The degree of agreement between a measured value and the 'true' value (or specified value in the case of the manufacturing process). Accuracy may be regarded as a lack of bias.

Precision: The amount of agreement between repeated measurements (or repeated products) made under specified conditions. Precision is subject to both random error and bias.

Unlike chemical analysis, the accuracy of an ecotoxicity test is virtually impossible to assess because we never usually know the 'true' toxicity of a substance. In practice, it may be approximated from the consensus mean arising from much repeat testing of the same substance in different laboratories (Figure 1.1). Precision and accuracy can vary independently so that, for example, it is possible to have a very precise but biased test which could lead to a false conclusion. Similarly, we may estimate the toxicity of a substance with great accuracy, but not very precisely, i.e. the average of a large number of measurements is close to the 'true' population mean but the standard deviation is large.

In the context of DTA, we are primarily concerned with:

- within-laboratory variability (the variability occurring when a test is performed on different occasions within the same laboratory);
- between-laboratory variability (the variability that is evident when results from the same test are compared between laboratories)

Figure 1.1 illustrates the variability that can occur with two widely used ecotoxicity test methods when repeat tests are carried out by laboratories with the same test substance. These data show that variability (as indicated by the spread of toxicity estimates) is a feature of both test methods but that it can vary between methods. In the case of the *Daphnia* tests with 3,4-dichloroaniline, there is also evidence of bias in one laboratory which generated toxicity estimates which were consistently lower than those obtained in other laboratories.

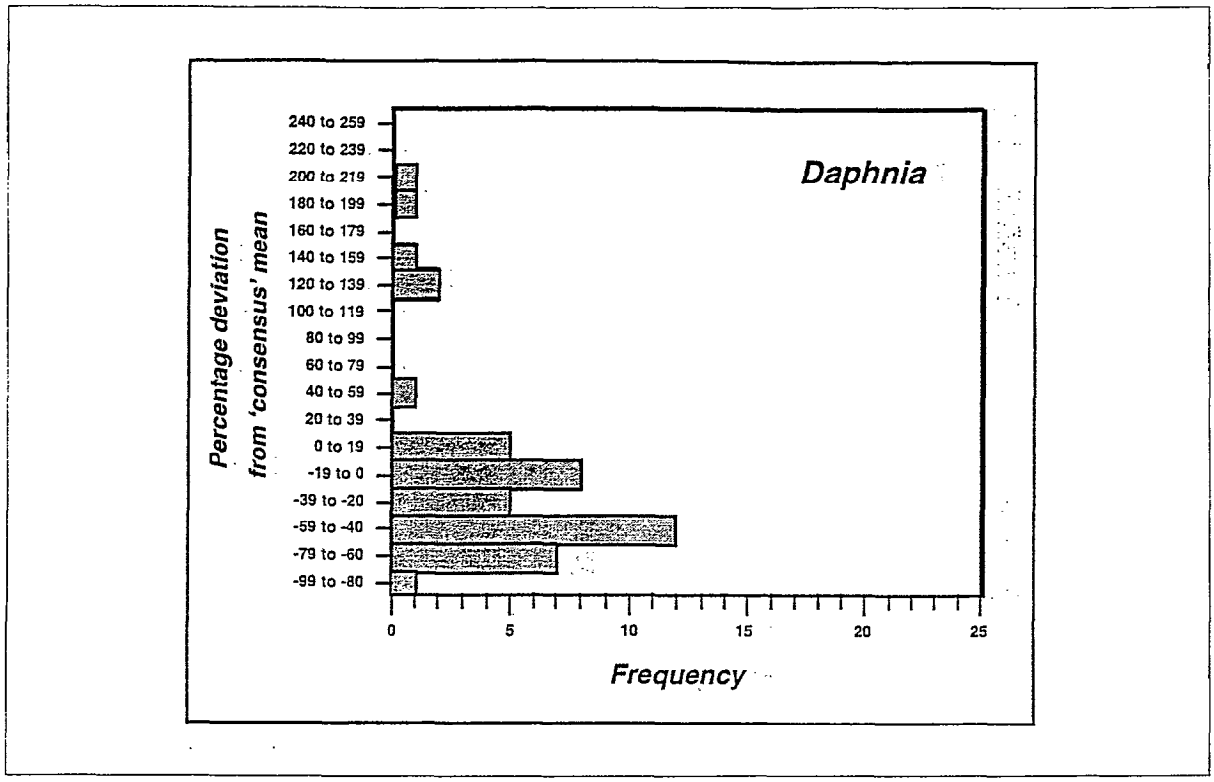


Figure 1.1 Variability in EC_{50} estimates for 3,4-dichloroaniline obtained by different laboratories

1.3.1 Regulatory implications of variability and bias

The regulatory implications of variability and bias may be profound, as indicated here:

(a) Deriving toxicity-based criteria

Variability between laboratories becomes important when toxicity-based criteria are being derived, because it may affect whether or not the criteria are derived equitably between locations. This can be illustrated by an example in which the environmental impacts of truly identical discharges (with respect to toxicity, dilution in the receiving water etc.) are investigated. If the toxicity evaluation of the effluents is conducted by different laboratories which typically differ in the sensitivity of the test methods used (because of systematic error perhaps), it is quite possible that the observed toxicities of the two discharges will differ. As a result, one discharger could be faced with a consent condition including a toxicity component whilst the other is required to both comply with a toxicity limit and embark on a programme of effluent toxicity reduction.

(b) Monitoring compliance with a toxicity-based limit

The inevitable variability involved in the measurement of the toxicity of effluents during compliance monitoring means that some effluent samples may be wrongly classified if that variability is substantial. The sample may be regarded as breaching the consent when, in fact, its true toxicity is actually compliant (Type I error or 'false positive') or the sample might

appear to comply with the consent but its true toxicity has been underestimated and it actually breaches the consent condition (Type II error or 'false negative').

The effects of variability and bias are most likely to give rise to mis-classification when 'borderline' toxicities are in evidence e.g. when the measured toxicity is close to a pass/fail threshold. Clearly, both Type I and Type II errors are to be avoided as far as possible, the first because of possible adverse commercial implications, and the second because of possible environmental impacts which go undetected.

1.3.2 The need for Quality Assurance and Quality Control measures

Clearly, false conclusions of the type indicated above should be avoided wherever possible and the Environment Agency and SNIFFER have made a commitment to promote high quality data in regulatory decision-making. The quality of test data was also highlighted as a major issue in the DTA industry consultation process in 1996 and was emphasised as an important component of successful regulatory schemes in North America. These concerns led to initiatives to select appropriate ecotoxicological test methods and to establish Quality Assurance (QA) and Quality Control (QC) procedures for DTA testing.

QA procedures for ecotoxicity testing have been established for some time but, whilst QC procedures are well established for chemical analysis, they have not been adopted to any significant extent for ecotoxicity testing. Consequently, proposals for QC for ecotoxicity testing have been developed in this project which aim to constrain the variability and bias that can occur in ecotoxicity testing and thereby strengthen the quality of decisions based on the use of these data for regulatory purposes.

1.4 Scope of this report

This report describes specific recommendations to promote the quality of ecotoxicity test data for regulatory decision-making, particularly in the context of DTA and effluent control. The recommendations are to integrate the following components into a Regulatory Ecotoxicology Testing Quality Scheme with which laboratories carrying out regulatory DTA testing must comply:

- Standardisation of test methods;
- Quality Assurance (QA) procedures;
- Quality Control (QC) procedures.

This report highlights the main recommendations and includes several Annexes each addressing the key components in more detail.

2. PROJECT OVERVIEW

2.1 Aim of project

The aim of the project was to identify and describe procedures to ensure the quality of ecotoxicity test data used in the derivation of Toxicity-Based Criteria for effluent control and for subsequent regulatory monitoring against those criteria. A 'certification' scheme was envisaged for laboratories undertaking this type of testing which latterly became known as the 'Register of Approved Laboratories' (RAL) and has subsequently been renamed the Regulatory Ecotoxicology Testing Quality Scheme (RETQS).

2.2 Outputs

This project drew heavily on other projects funded by the Environment Agency and carried out by WRc, particularly the development of 'Methods Guidelines' (R&D Project EMA 003) and the 'Performance Standards' (R&D Project i550) projects. Indeed, the RETQS may be thought of as the practical implementation of the scientific work carried out under these projects.

Specific outputs of relevance are listed below along with a brief summary of what each document covers.

2.2.1 Standardisation of test methods ('Methods Guidelines')

Variability and bias in estimates of ecotoxicity can be constrained by ensuring that all laboratories undertaking such testing adopt the same procedures. Following consultation with a Regulator/Industry DTA Methods Working Group with a particular interest in DTA test methods, a suite of predominantly short-term methods with lethal and sub-lethal endpoints has been selected for use and detailed guidelines for these methods have been prepared.

These guidelines set out procedures for the culturing and maintenance of test organisms, collection and storage of samples for testing, design and practice of several freshwater and marine toxicity tests, including the data to be collected and their subsequent analysis. These guidelines, by definition, impose some constraints on the way in which test data are generated and so should reduce variability within and between laboratories. An introduction to the 'Methods Guidelines' is reported in Annex I and the individual methods are compiled in a manual which may be updated as new methods are added or existing methods are modified.

Annex I: 'Direct Toxicity Assessment (DTA) Demonstration Programme - Methods Guidelines for Effluent and Receiving Water'

Currently, the Manual provides detailed guidance on the following test methods:

- algal growth inhibition test (*Selenastrum capricornutum, Raphidocelis subcapitata* - freshwater and *Skeletonema costatum* - marine);

- juvenile *Daphnia magna* immobilization test;
- Pacific oyster (*Crassostrea gigas*) embryo-larval development test;
- juvenile fish lethality test (rainbow trout, *Oncorhynchus mykiss* - freshwater; turbot, *Scophthalmus maximus* - marine).

2.2.2 Quality Assurance (QA)

Integrity of test data is a key element underpinning the generation of quality data and this may be achieved by incorporating Quality Assurance (QA) procedures to ensure the auditability and integrity of test data generated in regulatory testing. This need has also been recognised by the USEPA and the Canadian Association for Environmental and Analytical laboratories and these bodies also require accreditation to a formal QA system in ecotoxicology testing laboratories.

The Environment Agency and SNIFFER recognised that a Quality System which included a strong Quality Assurance (QA) component would be essential if test data were to withstand close scrutiny, possibly in a court of law. Data which may be scientifically sound but cannot be fully audited from source to the final report are unlikely to be useable for regulatory purposes. The following report (shown in Annex II) reviews possible schemes for Quality Assurance of ecotoxicity data.

Annex II: Quality System Definition for Toxicity-based Monitoring, WRc, September 1995 (NR 3993; R&D 493/9/S)

This report reviews different Quality Systems (Good Laboratory Practice, GLP; United Kingdom Accreditation Service, UKAS) and development of an independent certification scheme similar to one operating in North Carolina) and concludes that the QA aims of the RETQS could be met by requiring participating laboratories to either submit for inspection under the UK GLP Monitoring Programme or become accredited under UKAS for the test methods that they wish to offer. This flexibility would allow test laboratories to comply with the Quality System which more closely matched their core activities. The development of a new accreditation body taking responsibility for carrying out site inspections and monitoring compliance with a new scheme was considered unnecessary.

Both the UK GLP Monitoring Authority and UKAS have agreed that the QA aims of the RETQS are consistent with compliance with GLP or accreditation under NAMAS. Both organisations would be willing to inspect laboratories seeking accreditation for either scheme.

2.2.3 Quality Control (QC)

An external Quality Control (QC) scheme aimed at constraining bias and variability in the results obtained from ecotoxicity tests has been developed. This is intended to ensure that laboratories can perform tests in a repeatable way and to promote consistency between laboratories.

QA vs QC

It is important to make the distinction between Quality Assurance (QA) and Quality Control (QC). For the purposes of this document (and a widely held definition), QA is primarily concerned with the management of data, its auditability and integrity. QC, on the other hand, is primarily concerned with the performance characteristics of testing. QC features strongly in accreditation schemes for chemical analysis but, to date, has not featured in such schemes for ecotoxicity testing.

Internal vs External QC

QC may take the form of 'internal' monitoring of performance or 'external' monitoring. Conventionally, internal QC uses Shewart control charts to monitor responses to a reference toxicant over time. Based on previous experience, control limits for variability may be calculated and each new set of data compared with the 'norm' for that laboratory. Whilst this helps to maintain the status quo, it can actually reinforce bias in a laboratory because it is solely based on typical results obtained in that laboratory. For the same reason, the control limits in one laboratory may be very different to those in another and the effect can be to penalise (more stringent control limits) laboratories which typically exercise good control over variability.

External QC, on the other hand, involves the use of performance criteria (e.g. control limits on bias and variability) which are imposed externally and with which all laboratories should comply. It thereby avoids some of the potential shortcomings associated with internal QC although it requires an infrastructure to maintain it and a commitment from laboratories to participate in it. Such a scheme would set these control criteria, monitor performance and require action in the event of unacceptable performance. In this respect, it is akin to a proficiency scheme which may be used to monitor performance in chemical analytical laboratories.

In summary, external QC is the most technically sound approach to regulating the performance of tests and this is best brought about as part of a 'certification' scheme to which laboratories are invited to participate.

Proposals for a Regulatory Ecotoxicology Testing Quality Scheme (RETQS)

Several overseas agencies have developed formal 'accreditation' or 'certification' schemes which emphasise external QC and to which laboratories must belong if they are involved in regulatory wastewater testing. Laboratory certification takes on even greater importance if there is to be an emphasis on 'self-monitoring' as is proposed in the UK. Examples of schemes developed in Canada and in North Carolina are summarised in Annex III in the following reports:

Annex IIIa: Canadian Toxicological Testing Laboratory Accreditation Program: Program Description, CAEAL, December 1993.

Annex IIIb: North Carolina Biological Laboratory Certification/Criteria Procedures Document, North Carolina Department of the Environment, Health and Natural Resources, Division of Environmental Management, Water Quality Section, 1994

Proposals for a similar scheme were developed for regulatory DTA testing in the UK and a series of drafts were submitted for comment through a consultation process with test laboratories, DoH and UKAS during 1996 and 1997. A 'Regulatory Ecotoxicology Testing Quality Scheme' (RETQS) was envisaged which was based on standardised test methods (Annex I), the Quality Systems recommendations described in Annex II and, in particular, a new external QC scheme for ecotoxicity testing. The final recommendations are to be found in the following document in Annex III:

Annex IIIc: A Register of Approved Laboratories undertaking Toxicity Testing, July 1997

The external QC scheme described requires participating laboratories to carry out regular reference toxicant testing and comparison of the EC_{50} values obtained with control limits calculated from the results obtained with the same toxicants in a ring-test carried out in 1995 under the 'Performance Standards' project (R&D Project i550). Limits for bias and variability in repeated tests within a laboratory are specified for bacterial bioluminescence, *Daphnia* acute and Pacific oyster tests using two reference toxicants: zinc sulphate and 3,4-dichloroaniline. The report recommends that monitoring of performance is carried out by the Environment Agency, or a body acting on their behalf, and provides guidance on what constitutes unacceptable performance (i.e. excessive bias or variability) and suitable actions in the event of unacceptable performance.

Clearly, some expense is incurred by laboratories participating in the RETQS. However, the optimum number of tests and concentrations which are required to permit meaningful judgements of performance without incurring unnecessary costs has not been established. To address this question, a trial of different options (frequency of reference toxicant testing, number of test concentrations) is currently being undertaken in parallel with the DTA Demonstration Programme. Any changes which are deemed necessary to the scheme will be highlighted by this trial. It should also enable a cost-effective approach to external QC to be identified. These options are set out in the following documents, also shown in Annex III:

Annex IIId: Options for Trialling Quality Control Procedures for Ecotoxicity Tests, WRc 1998

Annex IIIe: Options for Trialling Quality Control Procedures for Ecotoxicity Tests - Outstanding Issues. WRc 1998

When toxicity-based criteria are employed as a regulatory tool, it is anticipated that all DTA testing concerned with either the characterisation of effluent toxicity or monitoring compliance with toxicity-based criteria will be undertaken only by laboratories participating in the RETQS.

3. CONCLUSIONS

1. A comprehensive scheme aimed at promoting the quality of ecotoxicity test data used for regulatory purposes has been developed. It is based on standardised test methods, described in 'Methods Guidelines' produced under R&D Project EMA 003 and these are underpinned by recommendations for a 'Regulatory Ecotoxicity Testing Quality Scheme' (RETQS).

2. The RETQS specifies Quality Assurance (QA) procedures for laboratories undertaking such tests and also procedures for operating an external Quality Control (QC) scheme, based on research carried out under R&D Project i550. By combining QA and QC measures, the intention is to ensure that ecotoxicity tests are performed to a high standard and that the validity of the resulting data are beyond question.

3. The QA component is designed to ensure the integrity and auditability of ecotoxicity test data and is realised by requiring test laboratories to be either GLP-compliant or accredited under NAMAS.

4. The QC procedures involve minimum standards for the performance of tests conducted for regulatory purposes, particularly the amount of bias and variability that is permissible using reference toxicants. Compliance with limits is then assessed using results obtained from regular tests carried out by participating laboratories with one of these toxicants.

5. When DTA is employed as a regulatory tool, laboratories generating test data will be required to participate in the RETQS. However, the RETQS is first being piloted as part of the DTA demonstration programme. This will also provide an opportunity to compare a number of QC options so that a final recommendation can be made which provides an acceptable balance between cost and the ability to make valid assessments of test performance.

ANNEX I STANDARDISATION OF TEST METHODS

DIRECT TOXICITY ASSESSMENT (DTA) DEMONSTRATION PROGRAMME

Methods Guidelines for Effluent and Receiving Water

CONTENTS

1. INTRODUCTION

The use of ecotoxicological methods to provide data for deriving and monitoring toxicity-based limits and for assessing receiving water column toxicity has to be carried out using standardised procedures to ensure the quality and integrity of the data generated in the course of the study (Gawadi 1990). A project was, therefore, initiated with the aim of producing Direct Toxicity Assessment (DTA) Methods Guidelines to meet the requirements of the Environment Agency and organisations within SNIFFER.

The DTA Methods Guidelines are to be used initially in the DTA Demonstration Programme to screen and characterise effluent toxicity and to assess receiving water column toxicity. The guidance given in the current document has been prepared on the basis of comments received on a previous version during a consultation exercise (Environment Agency 1997) and the output from a workshop on the DTA Methods Guidelines organised by the DTA Methods Working Group. The Workshop was held at Sundridge Park, Bromley on the 22-23 July 1997 and was attended by representatives of the regulators, industry, consultancies, testing houses and academia (DTA Demonstration Programme 1998).

In the revised guidelines considerable emphasis has been placed on the culture or maintenance of test organisms since the DTA Methods Guidelines Workshop emphasised the importance of conducting test procedures with 'healthy' organisms. Guidance is given on algal growth inhibition tests, the *Daphnia magna* immobilisation test, the oyster embryo larval development test and juvenile fish lethality tests.

The test guidelines and associated guidance on culture or maintenance methods differentiate critical steps which must be followed from those where a procedure is recommended but other approaches are allowed. They are accompanied by a glossary of terms and an appendix giving a list of suppliers of test organisms and equipment.

The guidelines given in the manual will be updated in due course following consideration of the results of the Demonstration Programme.

REFERENCES

DTA Demonstration Programme (1998) Direct Toxicity Assessment (DTA) Demonstration Programme: Proceedings of the DTA Methods Workshop: Method Guidelines and Register of Approved Laboratories for Effluents and Receiving Water Assessment. 22-23 July 1997, Sundridge Park, Bromley, pp 45.

Gawadi, N. (1990) Standard operating procedures. In: *Good laboratory and clinical practices. Techniques for the quality assurance professional* (Eds P.A. Carson and N.J. Dent), Heinemann Newnes, pp 67-83.

Environment Agency (1997) The application of toxicity-based criteria for the regulatory control of wastewater discharges - Response Compendium, pp 52.

ANNEX II QUALITY ASSURANCE (QA) PROCEDURES

Interim Report R&D Project 493

Quality System Definition for Toxicity-Based Consent Monitoring

WRc plc

September 1995

R&D 493/9/S

QUALITY SYSTEM DEFINITION FOR TOXICITY-BASED CONSENT MONITORING

D A Winnard

Research Contractor:
WRc plc
Henley Rd Medmenham
Marlow
SL7 2HD

National Rivers Authority
Rivers House, Waterside Drive,
Aztec West, Almondsbury,
Bristol BS12 4UD

493/9/S

National Rivers Authority
Rivers House, Waterside Drive
Aztec West, Almondsbury
Bristol BS12 4UD

Tel: 01454 624400
Fax: 01454 624409

Dissemination status
Internal: Restricted
External: Restricted

Research contractor
This document was produced under R&D Contract by:

WRc plc
Henley Rd Medmenham
Marlow
SL7 2HD

Tel: 01491 571531
Fax: 01491 579094

WRc Report N° NR 3993/9265

NRA Project Leader
J Wharfe, Southern Region

Additional copies
Further copies of this document may be obtained from Regional R&D Co-ordinators or the R&D Section of NRA Head Office.

493/9/S

CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

This report describes the results of the first phase of an NRA contract to devise and recommend a set of quality criteria (the 'specification') for laboratory systems, organisation, and activities, required to support the data quality objectives for toxicity-based consent (TBC) testing.

The contract has been undertaken by WRc's Quality Assurance Department, in close collaboration with the Ecotoxicology Group, who have been responsible for developing many of the scientific protocols in support of the NRA's future 'Certification' scheme for TBC testing laboratories.

KEYWORDS

Quality system, quality assurance, toxicity-based consents, certification criteria.

1. INTRODUCTION

The NRA is currently conducting a number of projects related to the direct toxicity assessment of effluents (and receiving waters), with a view to implementing toxicity-based consents (TBCs). The data used to derive these regulatory consents, and the data produced during the monitoring programme, should be generated in such a way that their validity and integrity is assured.

A formal requirement for testing facilities to establish and maintain a suitable quality system (ultimately inspectable by the Authority), would provide a mechanism for ensuring that risk decisions taken by regulatory staff are based upon sound facts. Furthermore, if routine monitoring indicates a breach of consent, enforcement actions taken by the Authority will be strengthened and supported, particularly in the case of prosecutions.

This project is to develop and document a set of fundamental guiding principles and organisational quality directives, to be used as the primary reference point both for Inspectors, and organisations wishing to enter the programme.

Stage 1 of the contract is now complete. The results of this research, including WRc's conclusions and recommendations, are given in this report. Key criteria for the scheme are identified and the extent to which these may be met by three existing quality systems have been investigated. These systems are: GLP, NAMAS, and the US system for permit and data compliance (NPDES), particularly as applied in North Carolina biological laboratories. The other system commonly applied within UK organisations, BS EN ISO 9000, was not considered because of its generalised nature of application, which does not necessarily provide the high level of control required in a laboratory testing situation.

2. PRIMARY QUALITY OBJECTIVES

In order to identify the fundamental quality requirements for a TBC specification for testing laboratories, the overall process of sampling, testing and reporting consent data was examined. A summary of the process is as follows:

- a sample of the discharge is taken by NRA officials, Consentee, or perhaps by an authorised independent contractor, as required under the conditions of the consent (end-of-pipe or other specified location);
- the appropriate paperwork is completed and the labelled sample (which must remain 'in custody' at all times) would then be sealed and tagged before transportation if necessary;
- conditions during transportation will be defined in the appropriate test method;
- upon receipt in the laboratory, there must be procedures in place to inspect seals and measure temperatures etc;
- the sample is entered into the laboratory's testing schedule;
- the test conducted;
- the results reported to the regulatory authority, and the original documentation and raw data archived at the test facility.

Due to the relatively small number of test methods to which a TBC quality requirement is to apply, and because of the number of samples and hence limited finance involved in UK testing (at least initially), one of the main considerations is that of economy of effort. To this end, systematic laboratory requirements will as far as possible be focused upon achieving improved technical performance. By the nature of all quality systems however, some aspects (e.g. the legal expectation for full documentation, sample custody, auditable raw data etc), will involve an administrative overhead for laboratories when running the system. This will be ameliorated in a number of laboratories for whom such systems are already in place, due to their GLP compliance or NAMAS accreditation. Tables 2.1 and 2.2 give details of the quality systems status of all laboratories currently involved in ring testing several of the TBC test methods (Performance Standards Project).

The broad specifications for a quality system which can accommodate TBC testing can be derived from consideration of these processes are shown in Table 2.3. The categories are simplified and refer to generally 'what' is required, and at this stage are not concerned with 'how' these would be implemented in laboratories, or 'who' would be responsible for the different elements of the programme. A comparison has then been performed to determine which categories would be satisfied by existing systems of GLP or NAMAS, and in either case, which extra requirements would be required to satisfy all the requirements for TBC 'certification'.

Table 2.1 Quality systems applied in participating (ring-test) laboratories.

ORGANISATION	QUALITY SYSTEM		SCOPE OF ACCREDITATION
	GLP	NAMAS	(Test method)
ACER Environmental	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
ABC Laboratories	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
ADAS	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
AEA	N	N	Daphnia Oyster- embryo larval Acartia Microtox
AgrEvo	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
Binnie Environmental	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
Clyde RPB	N	N N N Y	Daphnia Oyster- embryo larval Acartia Microtox
ERTL	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
Eurolabs	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
Forbairt, Eire	N	N	Daphnia Oyster- embryo larval Acartia Microtox

ORGANISATION	QUALITY SYSTEM		SCOPE OF ACCREDITATION
	GLP	NAMAS	(Test method)
Guernsey Sea Farms	N	N	Daphnia Oyster- embryo larval Acartia Microtox
Hamilton Garrod	N	N	Daphnia Oyster- embryo larval Acartia Microtox
Hazleton	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
HRC	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
Industrial Science Centre	N	Y N N Y	Daphnia Oyster- embryo larval Acartia Microtox
Inveresk Research	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
Pharmaco-LSR	Y	N	Daphnia Oyster- embryo larval Acartia Microtox
MAFF, Burnham-on-Crouch	N	N	Daphnia Oyster- embryo larval Acartia Microtox
North East RPB	N	N	Daphnia Oyster- embryo larval Acartia Microtox
NRA Leeds	N	N N N Y	Daphnia Oyster- embryo larval Acartia Microtox
NRA Waterlooville	N	N (Analytical only)	Daphnia Oyster- embryo larval Acartia Microtox

ORGANISATION	QUALITY SYSTEM		SCOPE OF ACCREDITATION
	GLP	NAMAS	(Test method)
NRA Llanelli	N	N	Daphnia
		N	Oyster- embryo larval
		N	Acartia
		Y	Microtox
Safepharm	Y	N	Daphnia
			Oyster- embryo larval
			Acartia
			Microtox
Shell Research	Y	N	Daphnia
			Oyster- embryo larval
			Acartia
			Microtox
Southern Science	N	N (Analytical only)	Daphnia
			Oyster- embryo larval
			Acartia
			Microtox
Tay RPB	N	N	Daphnia
			Oyster- embryo larval
			Acartia
			Microtox
Unilever	Y	N	Daphnia
			Oyster- embryo larval
			Acartia
			Microtox
WRc	Y	N (Analytical only)	Daphnia
			Oyster- embryo larval
			Acartia
			Microtox
Yorkshire Water	N	N	Daphnia
			Oyster- embryo larval
			Acartia
			Microtox
Zeneca (Brixham)	Y	N	Daphnia
			Oyster- embryo larval
			Acartia
			Microtox
Zeneca (Jealott's Hill)	Y	N	Daphnia
			Oyster- embryo larval
			Acartia
			Microtox

* Subject to inspection

Table 2.2 Summary of laboratory quality status

Laboratory type	GLP	NAMAS	NONE	TOTAL
Regulatory	0	3	3	6
Contract Research	11	1	4	16
Dischargers	5	0	2	7
Others	0	0	2	2

Table 2.3 TBC quality specification (see Appendices for further explanation)

	GLP	NAMAS
Organisation and Personnel	Y	Y
Facilities	Y	Y
Test schedule	Y	Y
Test sample custody	Extra requirements for TBC	Extra requirements for TBC
Equipment	Y	Y
Test organisms	Y	Extra requirements for TBC
Written procedures	Y	Y
Controlled documentation	Y	Y
Data recording	Y	Y
Archiving	Y	Y
Control charts	Extra requirements for TBC	Y
Systems Audit	Y	Y
TBC customised protocols:		
TBC methods	Compatible	Compatible
Proficiency testing	Compatible	Required
Reporting system	?	?

These specifications will be developed in more detail during the next phase of the contract, in order to ensure that at an operational level, systems are complementary and not inconsistent with GLP or NAMAS. Duplication will be resolved as far as possible by a TBC system that accepts the formats of either of these. The extra requirements identified in Table 2.3 would be included (or referred to) within the full specification, and developed as supplementary requirements to GLP and/or NAMAS, depending upon the requirements of, and in negotiation with, the respective Inspectorates.

In the initial stages of implementation of the UK programme of inspections, it is unlikely that the adequacy of sampling procedures or the competence of personnel taking samples would be included within the certification criteria, and inspections would be limited to laboratory facilities only. Furthermore, dischargers' in-house procedures for effluent treatment, flow monitoring, and toxicity reduction and identification evaluations, may also not be included. However, the inspectional remit could be broadened to include all such procedures, when resources become available as the consenting programme is extended.

Proficiency testing should form part of the requirements for laboratory QA, whether in a GLP compliant, NAMAS accredited, or independently certified laboratory for TBC testing. However, if an inter-laboratory sample check scheme will be relying on **consensus data** to derive meaningful pass/fail criteria (e.g. on a distribution by distribution basis), there should be a minimum number of laboratories taking part. Depending upon test method precision and individual performance, this is likely to be around 6-10 laboratories.

3. COMPARISON

Detailed theoretical descriptions of the principles, history and purpose of three different quality systems used in US and UK ecotoxicology laboratories are given in the Appendices, and the main elements of these are directly compared in Table 3.1. The systems are GLP, NAMAS, and the US system for permit and data compliance (NPDES), particularly as applied in North Carolina biological laboratories.

This section considers options for a UK scheme for TBC monitoring. The ramifications of application of extra protocols to GLP compliant or NAMAS accredited laboratories, and/or a new, self-contained NRA quality specification are discussed. Comments on GLP and NAMAS in this context are not directed at the intrinsic value of these systems as currently applied but to provide a means of comparison in light of requirements highlighted above.

3.1 Good Laboratory Practice (GLP)

General comment:

Communication with the DoH GLP Monitoring Authority has established that TBC testing might be incorporated under GLP, but the mechanism by which this could be established is unclear at this time. If TBC monitoring is not required by statute to come under GLP (e.g. under EC 88/320), then there are a number of documentary 'advisories' which might justify such an incorporation (David Moore, pers. comm.).

'Compliance' is facility based, with whole laboratory areas dedicated to GLP. Work flow requires the design of a study-based quality assurance (QA) system, with Study Directors appointed to each study, and monitoring by the QA unit of each study. *Separate* studies (and final reports) are required for each 'test substance'. The concept of 'an effluent sample' does not appear to be consistent with the GLP definition of a test substance.

Positive points:

GLP was originally developed for implementation within the toxicology environment, is currently applied in a number of ecotoxicology laboratories world-wide and in the UK, and would be well suited to the technology of TBC testing. GLP compliance is widely recognised throughout the UK, is well established in the testing industry, and is likely to continue to expand in product licensing and regulatory applications.

System requirements are well defined, and criteria for assessing compliance are established at national and international level (OECD), and within industry (trade and professional QA groups).

A number of contract laboratories involved in the ring-test studies developing TBC methodology are already compliant with GLP (16 from a total of 31).

Existing GLP compliant laboratories could readily develop SOPs to accommodate the additional testing and quality control (QC) requirements. The requirement for GLP in TBC testing would create low investment opportunities for compliant laboratories, thus allowing regulatory access to this source of quality testing expertise.

Negative points:

There are no specific requirements for analytical quality control (AQC) or interlaboratory tests, and there is no method-based 'accreditation' (or 'certification'), although the Memorandum of Understanding (MoU) with NAMAS could facilitate accreditation of GLP laboratories for TBC tests.

Existing GLP compliant laboratories could readily develop standard operating procedures (SOPs) to accommodate the additional testing and QC requirements. GLP is resource intensive, however, and would probably not be cost effective for laboratories, unless implementing GLP for other businesses as well as TBC testing. Generating a system from the beginning can be particularly difficult unless the laboratory has experience in QA. Development of a GLP system requires approximately 1.5 man years for a 'typical' laboratory function consisting of about 20 staff.

A formal requirement for "Good Laboratory Practice" for all TBC testing laboratories may require a new GLP Inspectorate to be commissioned by NRA. TBC requirements would be mainly additive to, but may also necessitate the amendment of, existing GLP regulations. Inspectors would have to be trained in both GLP as well as the technical and administrative procedures surrounding toxicity-based consents. The principles of GLP are governed by OECD, and as such the international acceptability, and possibly the legal aspects of a new GLP would need to be considered. It is considered unlikely that the option of a new UK GLP Inspectorate with all that this entails, would be particularly attractive to NRA.

3.2 NAMAS accreditation scheme

Positive points:

NAMAS accreditation for TBC would integrate with potential future requirements for flow measurement, self-monitoring schedules and data reporting requirements for chemical consents. Harmonisation with these requirements would allow consent setting systems to be framed within similar administrative (and inspectional) processes. The recommendation for NAMAS under some HMIP regulated IPC processes may also be a factor to support NAMAS accreditation as a suitable option for TBC Quality Systems.

The NAMAS Executive is to be incorporated within a private limited company providing the administrative systems and inspectors, to undertake Accreditation of UK laboratories. From discussion with NAMAS it has been noted that accreditation through the newly formed company will be positively received, as this will increase the business base.

NAMAS accreditation has been developed for chemical testing, and a large number of analytical laboratories are aware of, or have accreditation against NAMAS requirements. Some ecotoxicology (and some microbiological) laboratories have been accredited for tests such as *Daphnia*, oyster embryo-larval, (and Microtox™).

Negative points:

NAMAS was not designed for laboratory toxicity environments and accreditation criteria are still evolving.

Unlike GLP there is no requirement to appoint a senior member of staff to supervise each test individually. The system is not as intensive with regard to the QA monitoring of any one particular 'study' or test. Auditing procedures are developed on a system basis, and focus on the more mechanical aspects of laboratory operation, i.e. calibration and traceability.

There is no routine mechanism for procedures to be documented on a test by test basis, for example to adapt to the particular characteristics of a test material or to instruct the operator on the required concentrations of solutions to be tested for toxicity.

A typical estimate for developing a NAMAS system would be approximately one man year in effort (in a laboratory of around 20 staff), with the accreditation process taking approximately six months after this (from the time of first application).

NAMAS Inspectors ('Technical Assessors') may have to be trained and approved specifically with respect to TBC accreditation.

3.3 Development of an NRA independent certification scheme

General comment:

The final option would be to develop a completely new self-contained system, based partly on the EPA and North Carolina Certification standard. As there is no formal certification body in existence within the UK, NRA would need to establish a Certification body with an inspectorial remit. This Inspectorate would need to maintain close administrative links with the NRA consents data base(s) and an interlaboratory sample check scheme, in order that effective responsibility for requiring prompt corrective action, and authority for certification and decertification can be delegated.

It should be recognised that the type of proficiency data generated by inter-laboratory sample distribution schemes (using prepared reference materials) will not take into account important variables such as sampling. Administrative activities (e.g. records management, sample custody and archiving) may or may not contribute directly towards particular instances of inaccurate data, but will certainly impact upon the 'integrity' of the data, and should therefore be included within the monitoring activities of the Certification body.

Positive points:

The scheme could be tailored to meet the requirements of TBC testing and could be developed in the future to meet changing needs or specific data problems. Independent laboratories who do not have any other quality certification would perhaps be able to develop laboratory systems more economically than for GLP or NAMAS. However it is felt that this assumption may not always be borne out in practice, due to the possible problems and uncertainties of implementing a new and largely unknown system.

The development of an NRA Inspectorate would create an opportunity for intimate information and data on testing laboratories and consentees to be obtained by NRA officials, thus allowing certification/decertification decisions to be taken on the basis of first-hand experience.

Negative points:

A non-governmental Inspectorate would need to be commercially viable and organisationally independent from any certified testing laboratory. The creation of a (third-party) certification body would incur considerable expense particularly during set up, i.e. development of operating procedure (perhaps to EN standards), recruiting personnel, marketing and selling the scheme.

Conversely, there may be a need for a Certification body to maintain a Code of Practice demonstrating how financial incentive to accept laboratories onto the scheme would not influence Inspectors' decisions on certification. The body should be able to offer impartial advice, but could not offer contracts to act as a consultancy on setting up systems.

Laboratories which already have either GLP compliance or NAMAS accreditation would incur some costs in developing secondary systems and documentation to satisfy NRA inspectional requirements, but the extent of this duplication/overlap would depend on the TBC specification and the conditions of the respective Memorandum of Understanding.

At present the UK's use of self-monitoring is fairly limited, and therefore NRA itself could be the largest sponsor for this type of work, as well as the regulatory body ultimately responsible for laboratory certification and decertification. Therefore, if commercial laboratories were required by NRA to maintain a new quality system created solely for TBC testing, they would seek a satisfactory return for this investment and a reasonable share of the available business. This could affect the NRA's sub-contractor selection procedures, in order to avoid laboratories becoming less 'favoured', and hence less inclined to implement or maintain a quality system.

A more nebulous comment relates to enthusiasm; industry is already generally bemused and even confused, with the existing proliferation of quality systems. This often requires re-editing of quality documentation and some restructuring of organisational responsibility (etc). The bringing into existence of yet another system would certainly be unpopular, particularly in complex businesses already requiring 'QA' for a number of different purposes. Although quality systems are not generally incompatible, they are not, however, complementary to each other. Despite agreements between Inspectorates it is not usually possible to (just) add 'bolt-

on' elements to one system, in order to satisfy another. Access to GLP or NAMAS archived records (e.g. training records, instrument records, SOPs) may be refused, although it is not known to what extent this might be a problem.

Table 3.1 **Comparison of GLP, NAMAS, and NC/NPDES quality standards**

Constituent	GLP	NAMAS	NC/NPDES
Legal basis in UK	EC 88/320 UK legislation expected	No (voluntary only)	No (US only)
Inspected by	Government body (DoH)	Independent body (NAMAS)	Government body (EPA) or State Agency
Financing of Inspectorate	Indirect (no links between inspectorial time and recovery from inspected organisation)	Based on number of inspections and extent of accreditation of lab	Government ($300 000/yr + 10 man yr/yr for EPA). North Carolina State 2+ man yr/yr.
Formally stated relationships with other UK Inspectorates ('Memoranda of Understanding')	With NAMAS (for routine chemical and physical tests only)	With GLP With BS EN ISO 9000 With DWI (** DWTS) With MAFF/DoH (food control measures directive 93/99).	No
Primary scientific areas of application	Toxicology, oncology, physico-chemical measurement, pharmacology, ecotoxicology	Analytical chemistry	Analytical chemistry, ecotoxicology, flow measurement
Application to *any* industry, product or service?	No	No	No
Categories of inspections performed by Inspectorate	(i) Facility Inspection (ii) Study Inspection	(i) Horizontal Audit (ii) Study Inspection	(i) Compliance Evaluation (ii) Performance Audit
Design control (to specify job activities, scheduling, staff involved etc.)	Yes ('study plans')	No - pre-existing test methodology only. No separate job design stage	Strategic only ('QA Project Plans')
'Quality Manual' required	No (all procedures documented in SOPs)	Yes	Yes ('QA manual' or 'QA plan')
Statement of 'scope' (index of tests accredited)	No	Yes	Yes - detailed application procedures include listing of testing capabilities

Constituent	GLP	NAMAS	NC/NPDES
Process control	SOPs	SOPs	SOPs
Controlled documentation (individually numbered, distribution recorded etc)	Yes	Implied	No
Instrument maintenance/service schedules defined (in SOP)	Yes	Of greatest concern as testing is mostly using equipment. Calibrations traceable to national standards. Access to equipment strictly controlled	No
Data generation	Normally operator based (observations)	Normally instrument based (output)	Both
Raw data recording	No pencil, correction fluid, or obscuration of original entries. Corrections must be initialled and dated	No pencil, correction fluid, or obscuration of original entries. Corrections must be initialled and dated	Corrections must be initialled and dated
Identity of test item	Detailed labelling of test substance	Correct identity of sample	Detailed chain of custody record
Formal role of test item 'Custodian'	Yes	No	Yes
Sampling procedures inspectable?	Yes	No	Yes
***All* activities must comply at all times?**	Yes - by laboratory area	Yes - by procedures (accredited test methods only)	No - deliberate policy to allow flexibility in research, e.g. for TRE and TIE investigations
Documented training records (evidence of competent staff)	Yes	Yes	Yes (specific qualifications defined for supervisory and technical staff)
How identified in reports	Signed Study Director GLP and QA statements	Logo and specific format	Standard pro-forma report forms
Discussion/interpretation of data in reports	Yes	No	No

Constituent	GLP	NAMAS	NC/NPDES
Data tolerance (confidence limits) reported	No (unless integral to statistical analysis of study data)	No (unless requested by customer)	No (unless integral to statistical analysis of study data)
Content of reports	Detailed in paragraph 43 of UK GLP	Detailed	Standard pro-forma report forms
Data content of reports	*All* study data and activities must be included	Test data only	Test data only
General in-house QA requirements	Extensive - to inspect organisation of quality system and conduct of all studies at design, process and reporting stages	Focuses on the quality control of the product	No in-house independent QA function. Many specific 'QA' requirements, but not separated from other activities. Quality assurance is an *operational* process, rather than a monitoring one.
QA INSPECTION AND AUDIT PROCEDURES:			
QA review of job design	Yes	No	No in-house independent QA function
QA process inspection	Yes (all studies)	No	No in-house independent QA function
QA review of final report (includes data audit)	Yes (all studies)	No	No in-house independent QA function
QA general inspections of facility	Yes	Yes	No in-house independent QA function
Quality Control procedures	Control of data tolerance not specified (the **accuracy** of toxicity tests cannot be determined (*))	Use of control standards and control charts to monitor accuracy and precision	Use of control standards and control charts to monitor accuracy (analysis only) and precision
Peer review (auditor expertise in science)	Not required	Yes	No in-house QA function
Test validity criteria (e.g. survival rate of control organisms)	Not included in GLP regulations. However, can and do become part of company's SOPs (many test guidelines such as OECD will include such validity criteria)	No (not appropriate to science base)	As in GLP, not specified within 'quality system', but included in test methods. Test reports checked by regulator for 'required QA'

Constituent	GLP	NAMAS	NC/NPDES
Performance Evaluation (interlaboratory comparison)	No	Can be required (depends on test type)	Yes - major Permittees
Assessment (inspection) of (sub) contractors required	Yes	Yes (but use is discouraged)	No - EPA or State Agency may inspect, but Permittee is responsible for any deficiencies found. In North Carolina, certified labs only may be used.
Secure archive	Yes	Yes	No
Archivist appointed	Yes	No	No
Archiving schedule	Not defined in GLP, but in practice at least 10 years, or for duration of product licence	6 years	3 years
Concern for 'Organisational standards'	Primary (design, control and conduct of study paramount)	Primary	Secondary
Concern for 'Performance standards'	Secondary - set by individual facility and reviewed by Regulator. (A)QC not mandatory	Secondary - accreditation standards set nationally, but variations according to *purpose of data*	Primary (QC mandatory)

* EPA effluent toxicity testing methods manual
** UK Drinking Water Testing Specification

4. CONCLUSIONS AND RECOMMENDATIONS

- A non-statutory Inspectorate would need to be commercially viable and organisationally independent from any certified testing laboratory.

- A non-statutory Inspectorate would need to maintain close administrative links with NRA and an interlaboratory sample check scheme, in order that effective responsibility for requiring prompt corrective action, and authority for certification and decertification can be maintained.

- NAMAS accreditation for TBC could integrate with the administrative and inspectional requirements for flow measurement, consent setting and data reporting requirements for chemical consents.

- Proficiency testing should form part of the requirements for laboratory QA, whether in a GLP compliant, NAMAS accredited, or independently certified laboratory for TBC testing. However, if an inter-laboratory sample check scheme will be relying on **consensus data** to derive meaningful pass/fail criteria (e.g. on a distribution by distribution basis), there should be a minimum number of laboratories taking part. Depending upon test method precision and individual performance, this is likely to be around 6-10 laboratories.

- In order for data to be considered as having 'integrity', or as being 'valid', a quality specification should contain mechanisms by which a laboratory is more likely to produce a complete 'audit trail'. Data which may be sound scientifically, but cannot be fully audited from source, through each stage of its transformation, to the final report, are unlikely to be usable for regulatory or enforcement purposes.

- Any NRA policy change in relation to increased 'self-monitoring' will have enormous impact upon the overall importance of laboratory certification. However, the implications of using consentees' own data in enforcement actions may need to be considered, unless NRA 'compliance monitoring' data only would be used for this purpose. Clearly then, NRA regional ecotoxicological laboratories would still be required to provide such monitoring data and hence come into the QA programme.

- If a long-term commitment to NAMAS accreditation for TBC testing was considered appropriate by NRA, the NAMAS Executive has expressed interest and willingness to take on board the inspectorial process for TBC certification (Jane Beaumont, meeting notes 14 July 1995).

APPENDIX A GOOD LABORATORY PRACTICE (GLP)

A.1 BACKGROUND AND ORIGINS

GLP originated in the mid 1970's, as a result of investigations in the United States of pharmaceutical and contract testing organisations, by the Food and Drug Administration (FDA). In 1975 G D Searle submitted a study which was thought 'too neat and tidy' to fully reflect all the results of a toxicology study. FDA official Adrian Gross decided to visit the laboratory, and for the 2-year rodent oncogenicity study on the anti-hypertensive drug Aldactone, discovered three separate sets of pathology diagnoses, two of which had been sub-contracted. The submission had included only one data set, the one with the most favourable results.

In order to resolve the resulting allegations made during hearings of the Senate (Subcommittee on Labour and Public Welfare), the FDA undertook a full investigation of Searle Laboratories and one of its sub-contractors, Hazleton Laboratories (USA). Searle accepted the inspection, which involved as many as 20 investigators on-site at any one time. Although only a few discrepancies were found, it was enough to cast some doubt upon laboratory practices in the industry as a whole. Following the investigation, Searle submitted a document to FDA defining the organisational aspects of laboratory management, entitled *'Good Laboratory Practice'*; it was in this paper that the GLP concepts of a 'Quality Assurance Unit' (QAU) (see A.4.6), and the 'Study Director' (see A.4.3) first originated.

At this time however, FDA had little information on the way toxicology studies were performed, and had assumed the very highest level of commitment to public safety by product testing laboratories. The *conduct* of studies was not questioned, the only focus of discussion being the conclusions and interpretation that could be drawn from the reported data. In 1976 Congress allocated to FDA new resources to create positions for Inspectors, and this programme was known as the Bioresearch Monitoring Program. Subsequently, amid much trepidation within industry, FDA published proposed GLP regulations, based closely upon the Searle document. A pilot program of inspections, comprising 96 laboratories (including some from outside the USA) was instigated. These first inspections, along with industry's comments on the proposed GLP, allowed FDA officials to establish 'benchmark' quality, against which the GLP regulations were finalised (Brisson, 1987).

A.2 EARLY FDA INSPECTIONS

A.2.1 Observations

A major concern of FDA was of uncovering an industry wide scandal, but the first GLP inspections led FDA to conclude that the majority of studies were valid, and that the GLP regulations were workable. A number of shortcomings were brought to light however, these are summarised below:

- studies were poorly conceived, designed and planned 'on the hoof';

- carelessly conducted studies; study staff unaware of the need to perform work assiduously according to the protocol;
- original observations (raw data) not signed or dated;
- the manner and records of dosing did not assure that animals had received the required dosage;
- no confirmation of dosage solutions/media by analysis or other means;
- routine measurements such as animal weights were not fully kept;
- animals not identified (by clipping etc) and so mix-ups between control and test animals could occur;
- facilities treated with pesticides (to remove infestations) while test animals remained in the area;
- long delays between sacrifice and sectioning;
- absence of SOPs or study plans (documented instructions);
- the qualifications and/or experience of staff in supervisory positions was highly incompatible with the responsibilities of the task;
- staff employed to perform post-mortem and tissue sampling were not trained to perform the tasks;
- records of food and water consumption, appearance of animals etc were available, despite evidence that individuals were in fact dead (see A.2.2);
- reports inconsistent or incomplete compared with original records, employers unable to account for discrepancies;
- whilst a number of expert reviews were commissioned, the most favourable (least alarming) reports only were included in study reports;
- no Sponsor (customer) monitoring of sub-contracted studies or parts of studies;
- careless, obvious errors in reports (wrong data from wrong animals) discovered by the regulatory authority. This type of simple auditing error is the responsibility of company management (absence of in-house quality assurance procedures);
- original records not available;
- data transcribed to new records many years after work completed.

A.2.2 Non-compliant laboratories

Five of the laboratories inspected were subject to regulatory proceedings, but the most salutary example of inadequacies in laboratory practice was discovered in a company called Industrial Bio-Test (IBT). Throughout the late sixties and early seventies IBT had a bright future, was one of the largest commercial testing facilities in the US, and had performed many thousands of studies on hundreds of pharmaceutical and agrochemical products. The laboratory had been able to offer prices at around 25% less than its competitors. Prior to GLP, customer inspection of suppliers or subcontractors had not been required, and for IBT, business was booming.

A routine facility inspection in 1976 led the FDA to raise questions on IBT's procedures, and subsequently proposed to withdraw its approval of naproxen, a drug manufactured by Syntex. Further investigations followed, and irregularities included:

i) falsification and fabrication of study data;

ii) material defects in studies, by replacement of animals dying under test with healthy individuals;

iii) non-reporting of unfavourable data.

Eventually 74% of IBT's studies were invalidated, necessitating a program of re-testing at huge cost to industry, whilst products were temporarily removed from the market. IBT ceased trading in 1978, a small group of staff remaining at its headquarters dealing with the flurry of suits from its former customers. The suit against IBT by Syntex stockbrokers alone was settled in 1979 for $2.75 million. The former Company President and three ex-employees (Manager and Assistant Manager of Toxicology, and Section Head of Rat Toxicology) were accused of making fraudulent statements to federal agencies, and defrauding the Government by knowingly submitting false test data. The three had appeals rejected by the Supreme Court in 1986 and were sentenced to 20 years imprisonment.

Throughout the trial which dealt with just four specific tests, it is highly significant that it was not the *facts* of toxicity that were particularly in dispute (re-testing confirmed the safety of the products), but the carelessness in which the company ran its business in the production of regulatory data. This reflects a theme central to the principles and purpose of GLP, and has many implications in its mechanisms and job responsibilities (see also A.6).

A.3 IMPLEMENTATION

A.3.1 International

Establishing a balanced enforcement policy was important for FDA, as it was beneficial to the progression of GLP to encourage a supportive attitude within industry. However, it was necessary to establish the proper authority, and develop a forceful response to situations of fraud or recalcitrance. The legal status of GLP was not critical to its establishment, as FDA adopted a policy of non-acceptance of 'non-GLP' studies, even from foreign organisations. As the pharmaceutical industry assumes a world market for the development of most of its products, GLP rapidly became the industry standard.

In the case of non-US laboratories, there were at first no national GLP regulations, and these laboratories had to become registered with the FDA and subject to its inspectional programme. This was difficult for laboratories, expensive for the FDA, and politically embarrassing to the host nation's regulatory authorities. In 1979 the Organisation for Economic Co-operation and Development (OECD) established an expert group to devise and agree the Principles of GLP. Documents were produced covering:

(i) The OECD Principles of Good Laboratory Practice;

(ii) Implementation of the OECD Principles of GLP;

(iii) OECD Guidelines for National GLP Inspections and Study Audits.

Member States could develop national GLP regulations which were compatible, thus helping to prevent barriers to trade. Along with the quality system requirements of GLP, data produced using test methodology in accordance with the OECD Test Guidelines encourages 'mutual acceptance of data'.

Mechanisms of maintaining international compatibility include regular meetings between national compliance authorities, where GLP implementation problems are discussed, and working groups commissioned as required. Information on the training of inspectors is exchanged, and regular training courses arranged.

A.3.2 UK GLP

Prior to 1982 many UK toxicology laboratories had found it necessary to implement GLP, and were registered with the US FDA. Japanese GLPs soon followed, and the complexity of Japanese registrations further precipitated the necessity for the development of GLP within the UK. In 1982 the Health and Safety Executive produced a Code of Practice on GLP and created a GLP Inspectorate to monitor laboratories performing tests on industrial chemicals. The Department of Health and Social Security also formed an Inspectorate in 1983, extending the range of UK GLP facilities to include human health and environment studies on pharmaceuticals, agrochemicals, cosmetics and food additives.

The current GLP Monitoring Authority operates under the auspices of the now Department of Health, and took over responsibility from HSE GLP in 1984. The two UK Inspectorates were merged in 1986.

GLP is still at present 'voluntary' in the UK, although the Monitoring Authority does have powers of entry into facilities in its programme. It is expected that this year, a motion will be laid before Parliament to effect full legal status upon GLP.

A.4 COMPONENTS OF GLP

A.4.1 Introduction

GLP is defined as:

> **"the organisational processes and the conditions under which laboratory studies are planned, performed, monitored, recorded and reported."**

The purpose of GLP is to assure the quality of laboratory data through clear documentation of the control, conduct and recording of the study, such that the work may be completely reconstructed at any time in the future. In order to meet this requirement, the complete study data (e.g. study plans, raw data, reports) and all supporting non-study documentation (e.g. SOPs, instrument records, training records) must be readily available from facility archives.

GLP thus provides a means by which the *integrity* of data may be verified. It does not prescribe specific test methodology, technical consideration of the work being the responsibility of the scientific staff, and the receiving regulatory authority. GLP is based upon 'principles' or 'aims' and not 'standards', because of its wide range of application. However, skills in the interpretation of GLP will be essential when developing company standards (SOP).

Although not specifically stated in the regulations, GLP tends to be applied in such a way so as to apply to all activities, in order that the compliance of the laboratory is maintained. As GLP is a set of management principles, and impacts significantly upon the way work is conducted, it is generally not possible to administer 'non-GLP' processes within a single department. 'Part-time GLP' will attract inspectorial questioning of the level of commitment to (and understanding of) GLP by the facility.

The areas covered by GLP include the following summary list:

Organisation and Personnel:

- testing facility management
- Study Director
- Quality Assurance unit
- competence of staff
- training programme

Facilities:

- test system housing
- test system supply facilities (food, water)
- test and reference substance handling facilities
- laboratory operation areas
- specimen storage facilities

Equipment:

- correct type
- maintenance and calibration

Testing Operation:

- standard operating procedures
- reagents and solutions

Protocol and conduct of studies:

- study plans
- conduct

Records and reports:

- raw data
- contents of final reports
- archiving of data

A.4.2 The role of Management

Management must ensure that a facility operates in compliance with the Principles of GLP. This includes the provision of adequate laboratory space, and resources for instrument maintenance, administrative procedures, staff training, QA Units etc. Management must be committed to GLP as a fundamental way of working, not just as a means of demonstrating quality to customers and regulators.

Management is responsible for the organisation of a sufficient number of qualified and experienced staff within the facility. Detailed organisation charts, and policies and procedures for the appointment and training of Study Directors must be maintained. The system for monitoring a person's workload (part of the 'master schedule'), must be referred to when making this appointment. Management must ensure that the system for performing and documenting training results in a clear record of the types of study individuals are competent to direct.

A.4.3 Studies and Study Directors

A study is defined as 'an experiment or set of experiments in which a test substance is examined to obtain data on its properties and/or its safety with respect to human health and the environment' (OECD).

Studies are well defined work packages/projects, and provide the structure for the administrative control of activities within the GLP facility. For each study a 'Study Director' must be appointed by Management. The GLP regulations charge the Study Director with a great deal of responsibility. Study Directors are ultimately responsible for the technical validity, administrative control, and compliance with GLP, of the study. This overall control of the

study by one person is an essential requirement; these responsibilities must be supported by the management structure within the organisation, and cannot be delegated.

A.4.4 Study plans

A study plan is a document defining the entire scope of the study. Before a study plan can be authorised, the Study Director should ensure that Management have provided adequate staff and resources to do the work. The Study Director should also ensure that equipment has been correctly maintained, that SOPs are available, and that the test animals have been ordered or allocated from stock. The study plan is complete, and can be issued, when the Study Director signs it. The purpose of a study plan is to inform all participating staff of the purpose, methods and timing of a study. It is the key working document for the study, and in the case of contract work, forms the basis of the technical agreement between the customer (Sponsor) and the testing facility. In some cases it may also serve as the contract itself, although it is more usual for business arrangements to remain outside the GLP agreement.

The study plan should contain sufficient detail to describe the main processes, materials and solutions to be used at each stage of the study, and can therefore be used as the reference point for intention, to be measured against performance, during and after the study. Other factors such as the applicability of existing SOPs and complexity of study, will affect the content of a study plan. For routine 'base set' studies, study plans could be quite succinct.

The signature of the Study Director on the study plan is the agreement to conduct the study in compliance with the study plan and all relevant SOPs. The date of this signature is the formal 'commencement date' of the study.

A.4.5 Standard Operating Procedures (SOPs)

GLP requires test, administrative, and routine laboratory procedures to be documented and followed, in the form of SOPs. SOPs are a basic item in the reconstruction of the activities in and around a study. SOPs have to be approved for use (by Management), and should normally be reviewed by QA for auditability. There should be an index of SOPs, and this can include issue, review and withdrawal dates. SOPs are 'controlled documents', and are therefore individually numbered and recorded. SOPs should not be photocopied or annotated, and upon the issue of a revised version, all copies of the previous version should be withdrawn (apart from archive copies).

A large GLP facility may have some thousands of SOPs, a very small facility around 50 to 100. A common perception for the need for writing SOPs is that of an onerous burden, and excessive bureaucracy. However, good SOPs are a useful management tool; poor or careless SOPs are a source of inefficiency and confusion for those who are required to perform or monitor against them. SOPs are a way of maintaining processes when staff turnover occurs. New staff can easily be initiated into corporate procedures, and the operation of facilities is less dependent upon specific skills or knowledge.

SOPs should be usable by scientists, Management, QA, regulators, and possibly customers. Writing formats can vary from long informational commentary texts to short well defined auditable instructions. The best form of SOP is a clear, concise, up-to-date set of instructions,

that enable duplication of results by qualified staff and an understanding of what was/is done by non-technical personnel.

The attention to detail by which a facility maintains its SOPs is a good yardstick by which outside Inspectors can develop a feel for compliance levels. Companies for whom SOPs have become management tools will have well written SOPs that are used on a daily basis, and this will be motivating to operational staff.

A.4.6 Quality Assurance (QA)

The use of the term 'quality assurance' and other associated terms such as 'quality assessment' 'quality control', or even 'glp' (sic) is now so widespread that the impression of a common understanding has emerged. 'QA' is sometimes thought of as a single discipline with its own parameters, i.e. independent of the definitions and purpose of particular quality systems. However, when there is an attempt to ask specific operational questions it can be surprising to discover how diverse is its application between different industries, scientific disciplines, and quality systems. Furthermore, it is critical to the administration of any quality system, to maintain a common understanding of the function and remit of its quality assurance 'component'.

GLP places much emphasis on quality assurance, the responsibility for quality assurance being assigned to the Quality Assurance Unit (QAU). The QAU must be organisationally independent of the testing groups. There must be an individual appointed as 'Head of QA', reporting directly to Senior Management. Management should provide the conditions whereby the QAU can establish itself as the authority within the company on matters of GLP compliance.

One GLP requirement which encourages these conditions, is for an individually signed QA Statement to be included in every study report. This statement describes the QA monitoring of the study, and supports the Study Directors' Statement of GLP compliance. The QA statement can only be signed by QA personnel (normally the Head of QA). This reflects the authority of the company's QA function, which has a professional (and legal) obligation, and also acts as the contact point between the company and Inspectorate(s).

Specific QA activities such as inspectional and audit programmes are also well characterised within GLP. Through membership, attendance at meetings and training courses organised by the British Association of Research Quality Assurance (BARQA), QA Managers are able to maintain a harmonised approach, and communicate the expectations of industry and GLP Inspectorates. Most OECD countries have equivalent groups, and there is an international focus in the Federation of European Research Quality Assurance Societies (FERQAS).

A.4.7 The Master Schedule

The GLPs all require the maintenance of a system of logging all current, planned and completed activities within the organisation. The 'Master Schedule' can be designed and used in different ways in order to interface with other corporate resource planning tools, but must contain certain information. This includes the name of the Study Director, the identity of the 'test substance', the type of 'test system', the nature of the study (acute/chronic etc.), initiation date, current status, identity of Sponsor, completion date, and study identification number.

The Master Schedule can be a useful reference point for the planning of studies, the appointment of Study Directors, and for QA to schedule its inspectional programme. It is used by external Inspectors to identify work burdens and select studies for audit. It can also be used to identify trends and future facility needs in terms of instrumentation, staffing, method development; it can also serve as part of the archive index.

A.4.8 Archives and Records Management

The decision to operate to GLP means that a company will need to apply corporate standards for the collation and use of data, facility documentation, and also items such as fixed specimens, slides etc. These requirements will normally be in addition to the traditional needs of individual daily users, and in order to meet these regulatory requirements a rigorous records management policy will have to be implemented.

At its most basic level, data control SOPs will need to consider for example the acceptability of original raw data being taken off-site, security outside normal working hours, and the special problems of field working. The ownership of data should be considered - if product licences are contingent upon the continued existence of archived original data, then all such documentation (not just the final report) belongs to the sponsor for whom the work was performed.

GLP requires an individual to be appointed as responsible for the archives, for the movement of all data in and out of the archives to be recorded and signed for, and a detailed archive index to be maintained. Access to the archive is restricted to archive staff, QA, and Senior Management. Loans can only be authorised by Management in each case, or at the discretion of archive staff.

All facility documentation must find its way into the archive. Study files should be archived promptly (~2 weeks) upon study completion.

A.5 EFFECT OF GLP UPON ECOTOXICOLOGY LABORATORIES

GLP originated using terminology and concepts derived from mammalian toxicology, and it is perhaps surprising to consider how little it has changed despite wider application to the fields of environmental and clinical chemistry, ecotoxicology, biodegradability etc. This is partly due to the primary focus upon the organisational aspects of the work process, thus allowing the general principles to be widely applicable. Aquatic toxicology began in an atmosphere of

chemical spills, fish kills, lake algal blooms, and declining fish-eating bird populations. This was in the 1950s, in a time of pickle jar biology. By the 1980s it had become a profession in its own right, and in 1983 a paper was given at the 3rd International Meeting on GLP, discussing the application of GLP to ecotoxicology.

Ecotoxicology is the complete study of the complex effects of chemicals in the environment, and it is usual to allow experimental conditions to simulate the environment as far as possible. Research may include speculative investigation of the effects of low solubility, degradation, and the adsorptive behaviour of the test substance, and protocols need to be quite flexible. Often the data produced will be comparative (rather than absolute), and only relate to processes and effects studied in a particular ecosystem or type of receiving water.

For most GLP studies though, risk assessment by regulatory authorities requires the existence of toxicity data produced under standard conditions, to give an EC50 which is a fixed and reproducible value for the test system. In principle, toxicity equals a constant for single species laboratory tests. A different approach to these types of work is required, and it is essential to ensure that staff are fully aware of the purpose of studies and act accordingly. Performing procedures 'assiduously according to protocol' will not be an approach which is familiar to those normally involved in speculative research work. GLP field testing however is not concerned particularly with 'base-set' toxicity, as this data already exists from laboratory studies. Field studies are more concerned with the totality of effects on environmental concentrations, the fate of chemicals, interspecies activity and so on. Studies such as overspray studies in ponds are therefore at the forefront of both GLP and method development, and normally require an intensive QA involvement.

Careful attention to the identity of test organisms, particularly when collected from the field, is important, and this is achieved by using the proper expertise, recognised taxonomic keys, and of course good documentation of the examination of the batches of test organisms. Although the term 'dosage' is still sometimes used, organisms are not (usually) injected or fed. The internal concentrations of chemicals in aquatic organisms are seldom the same as the external concentrations, so the dose is unknown. The GLP study plan requirement for the planned 'route of administration' (of the dose) therefore doesn't apply, but study plans still need to consider very carefully how solutions of test substances are to be prepared and handled. In GLP there are normally at least three mechanisms for checking that the correct dosage has been given:

i) monitoring SOP/study plan compliance and operator competence (by Study Director and/or QA);

ii) 'Reconciliation' - detailed accountability of the usage of test substance provides additional evidence of correct dosage;

iii) analysis of the 'dosage vehicle' (feed, inoculum, test medium etc).

In ecotoxicology, (i) and (iii) apply, but (ii) won't provide much evidence of test solution concentrations, as these solutions are usually sub-diluted volumetrically. For most GLP studies, the protocol will require concentrations to be maintained throughout the experiment, and laboratory studies may be invalidated if concentrations are not maintained within ±20%

(OECD test guidelines). It is therefore important to have prior knowledge of stability in deciding upon study designs; in a static study, reliance upon analytical data is unacceptable when a flow-through or semi-static type of study could/should have been performed in an attempt to meet these criteria. However, there are problems with complex formulations (e.g. pesticides) as extensive analysis may be required to measure purities or monitor the stability of (all) the ecotoxicologically active ingredients. For highly unstable effluents of unknown or uncertain chemical composition as would be the test material in TBC monitoring, the presumptions behind the GLP requirements in this respect do not apply.

A.6 GLP AND MATTERS OF SCIENCE

It has already been indicated in Sections A.2.2 and A.4.1 that GLP does not prescribe test methodology or even quality control. GLP is an organisational standard, and its primary concern is the demonstration of full accountability of the conduct, and full reconstruction of the work, leading to the final reported data. Scientific evaluation of the work is the responsibility of the receiving regulatory authorities, and not the GLP Inspectorate. This important distinction (between 'compliance and 'science'), although somewhat divisive, has great impact upon the day-to day operation of the QA function. It is sometimes difficult to convince well-meaning scientific staff of the value of various administrative controls, and GLP can be viewed by some as being restrictive to scientific freedom. On the other hand it is often stated that it is perfectly possible to perform a scientifically meaningless study in complete compliance with GLP.

These truisms represent extreme cases, but reflect the need for management to ensure that the proper authority and responsibility is invested in the organisation's groups, if QA and scientific staff are not to become liabilities to one another. Sometimes known in GLP circles as 'the pure audit function' of QA, this is the objective check on the correct recording, transcription, and transformation of data, in complete compliance with the study plan and SOPs. The QA Auditor should appreciate the broad scientific objectives, but may well not be an expert in the particular science of the study. However, the QA audit is nonetheless powerful and intensive and includes *all* 'critical data' for all regulatory studies.

REFERENCES

Berny, C. and Long, D. (1989) *J. Amer. Coll. Toxicol.* **8**(2), pp419-428.

Brisson, E.L.. (1987) Good Laboratory Practices: 10 years. *J. Am. Coll. Toxicol.* **6**(2), pp221-225.

Carson, P.A. and Dent, N.J. (Eds.) (1990). Good Laboratory and Clinical Practices. Heinemann Newnes, Oxford, England.

Dewoskin, R.S., Schoeder, R.S. and Barkalow, R.E. (1989) *J. Amer. Coll. Toxicol.* **8**(2), p433.

Falconi, G., Goeke, J.E. and Morris, C. (1989) *J. Amer. Coll. Toxicol.* **8**(2), pp429-431.

Freeburg, F.E. and King, W.R. (1987) *J. Amer. Coll. Toxicol.* **6**(2), pp227-231.

Lepore, P.D. (1986) *J. Amer. Coll. Toxicol.* **5**(4), pp289-292.

McKay, D. (1987) *J. Amer. Coll. Toxicol.* **6**(2), pp213-220.

Morris, C.R (1986) *J. Amer. Coll. Toxicol.* **5**(4), pp293-296.

Morris ,C.R. (1988) Proceedings of the 6th Int. Congress on GLP, Stratford-upon-Avon, April 1988.

Royal, P.D. (1990) Society of Environmental Toxicology and Chemistry, 11th Annual Meeting, Virginia 1990, pp1-11.

Waddell, A. (1986) Proceedings of the 5th Int. Meeting on GLP, Frankfurt, May 1986, pp119-129.

APPENDIX B NAMAS ACCREDITATION

B.1 INTRODUCTION

Accreditation is the formal recognition that a laboratory is competent to carry out specific calibrations or tests, or specific types of calibrations or tests. The process of accreditation is normally administered by a national accrediting body. In the UK this body is NAMAS, the National Measurement Accreditation Service. The purpose of accreditation is to provide assurance to users of laboratory services that tests are carried out in a reputable and competent manner and to promote the acceptance of test data by producers and customers of goods and services.

Many other countries have similar accrediting bodies. There is a degree of agreement about the requirements under which accreditation can be granted to a laboratory and many different National bodies have reciprocal agreements whereby accreditation granted in one county is recognised in others. This system of mutual recognition is based on the fact that the approach to accreditation and the operation of accrediting bodies are based on the same International Standards.

B.2 APPROACHES TO QUALITY ASSURANCE

Several International Standards have been produced which define the concepts of and approaches to Quality Assurance. The most general of these is ISO 9000 - Quality Systems (European Standard 29000). This provides the basis for quality systems and quality management in many different contexts, from manufacturing to the provision of a wide range of different services (of which chemical analysis is an example).

Other more specific Standards give details of how to implement the principles of quality assurance in different situations. The Standard of principal concern is ISO Guide 25 - General Requirements for the Technical Competence of Testing Laboratories. This guide has achieved wide acceptance and has become the generic standard relating to Laboratory Accreditation. The guidance given in ISO Guide 25 is expanded upon in a series of European Standards, EN45001-3.

These standards define the important aspects of the quality system which would be required in order to ensure that results are fit for their intended purpose. These criteria also act as the basis on which to identify a competent laboratory. The standard of competence is defined partly by having a clear specification of the laboratory's organisational and record keeping requirements (quality system) and partly by ensuring that the standard of performance for each test type is adequate for the intended application. To be certain that fitness for purpose is achieved (rather than merely assumed), there is a requirement for accredited laboratories to take steps to determine their customers' needs.

B.3 QUALITY ASSURANCE AND QUALITY CONTROL

There is often some confusion concerning the terms Quality Assurance, Quality Control and even Quality System. The definitions below reflect the principles of QA as applied in these accreditation standards.

Quality Assurance: the process by which the producer or user of results is provided with the assurance that the results meet defined standards of quality.

Quality Control: steps which are taken to define and control the uncertainty associated with data and to demonstrate fitness for purpose. Quality control activities include method validation, routine checks on accuracy and precision, and periodic participation in interlaboratory tests.

Quality System: the system which is set up in the laboratory to implement the approach to quality assurance. This involves specification of how to organise laboratory work e.g. how samples are identified and handled, how methods are chosen and recorded, how problems are identified and responded to, etc. Quality control activity is one of the most important activities which is established within the quality system. Another key activity is that of audit and review - how the operation of the quality system is monitored, and how the resulting corrective actions are performed and monitored.

B.4 MEETING NAMAS REQUIREMENTS

NAMAS has produced its own 'Accreditation Standard', designated M10. This is an elaboration of the more general criteria set out in the European Standard EN 45001. The accreditation standard is supported by a comprehensive set of guidance documents, each of which deals with a different aspect of quality assurance. A laboratory which is applying for accreditation receives a set of these documents at the start of the process of gaining accreditation. The Accreditation Standard provides criteria to be adhered to by laboratories which carry out measurement or analyses - usually referred to as 'tests'.

The requirements of the Standard are defined under the following headings:

General Requirements

In summary, the requirement is that the laboratory establishes a Quality System (see above). The two key features are: the Quality Manual which is a description of the Quality System and the Scope of accreditation which lists the test for which accreditation is sought (NB a laboratory does not have to be accredited for all the tests it undertakes).

Organisation and Management

This section outlines the need for a well-defined structure within the laboratory with clear lines of responsibility. Key roles in Quality Management are identified - Technical Manager and Quality Manager - with a summary of their functions. The need for a system of authorisation

for the release of test results is emphasised. The Technical Manager has overall responsibility for the technical operation of the laboratory and for ensuring that the requirements of the Quality System are met. The Quality Manager has responsibility for ensuring that the requirements of the Quality System are met on a day to day basis. He also is responsible for organising the programme of auditing.

Quality System

This defines the basic requirements for a quality system (see above) in a testing laboratory. The Quality Manual is introduced as the reference document for the Quality System of the individual laboratory. The Quality Manual acts as an index to the functions of the Quality System, either referring directly to the key issues affecting quality or indicating how they are addressed.

The issue of documentation is crucial. The operation of all important laboratory functions must be adequately documented and such documentation needs to be subject to control and monitoring by management.

Quality Audit and Review

'Audit' is the process by which the system is checked, to ascertain whether or not the defined requirements are being complied with. 'Review' is the process of examining those requirements to ensure that Quality System meets the overall objectives of quality in the laboratory's work. Audits which encompass all aspects of the Quality System are carried out by laboratory personnel in accordance with a predetermined annual schedule. The Quality Manager plans and organises these audits. Staff are not permitted to audit their own activities.

Staff

This section specifies the need to use staff who have the appropriate combination of academic and/or professional qualifications, training, experience and skill. It is necessary to provide adequate specific training for each test, to maintain records of training and to indicate who is authorised to undertake each test.

Equipment

The requirements for test equipment are defined. Only equipment suited to the task in hand and capable of achieving the required accuracy should be used. The laboratory is required to have a system by which the fitness for purpose of equipment is demonstrated (either by test on the equipment itself or by tests on the equipment as part of the overall analytical system).

A system of equipment records is required as a means of demonstrating that the equipment used is adequately maintained. There must be a formal system to control the use of equipment and, in particular, to ensure that equipment which is malfunctioning is withdrawn from use, until it can be shown that satisfactory operation has been restored.

Measurement Traceability and Calibration

This is the means by which the laboratory's testing activity can be linked to national and international standards of measurement. The principal approach is to ensure traceability for fundamental quantities (e.g. mass, volume, temperature, time). These quantities can be linked to (checked against) their corresponding standards via calibrated masses, certified thermometers etc. This provides the fundamental assurance that the comparative tests carried out by a laboratory have a sound basis. In practice, this requirement leads to periodic checks on the accuracy of e.g. balances, ovens, incubators, refrigerators, pipettes, stop-watches.

For some measurements, the concepts of strict traceability to a fundamental standard may be difficult to apply. For example, it has not been possible to establish traceability for the parameter 'concentration' since it depends on establishing adequate criteria for the purity of standard materials. The approach to traceability which is being developed for analysis involves (a) most importantly - the traceability of all fundamental, physical aspects of analysis (especially mass); and (b) independent checks on the overall accuracy of the analytical system via analyses of reference materials and participation in appropriate interlaboratory tests.

Methods

This section defines the need to use well-recognised procedures and to have such procedures adequately documented. There is a requirement to maintain a record of all procedures which have been used and of any modifications which have been made to them. NAMAS provides guidance on the way laboratories should document test methods.

Accommodation and Environment

The need to ensure adequate facilities is stressed. If the measurement of interest demands the maintenance of specific environmental conditions e.g. temperature controlled room, steps should be taken to demonstrate that adequate conditions are maintained.

Handling of Calibration and Test Items

This outlines the requirements for an effective documented system for identifying test items (samples). It is essential to ensure that test samples should not be confused, physically or when referred to in records or other documents.

Records

Laboratories should have and maintain a systematic record of all information of practical relevance to the tests performed.

Test Reports

The necessary requirements to ensure demonstration of adequate quality for the client are set out. All information relevant to the validity and application of the test results should be recorded. Specific details of the appropriate form of test reports are set out.

Handling of Complaints and Anomalies

The Standard defines the way in which complaints and anomalies should be handled. There is a need to make sure that complaints are handled in a way which is open and documented.

B.4 PROCESS OF ACCREDITATION

The application of these criteria to the work of the laboratory is intended to provide the laboratory and its clients with confidence in the quality of the tests and in the technical and commercial integrity of the laboratory's operations.

The process of gaining accreditation involves three main stages:

(a) The laboratory applies and submits its Quality Manual and list of parameters to the accrediting body;

(b) The accrediting body appoints one of its Technical Officers and an independent Assessor or Assessors (the selection of the Assessor is agreed with the laboratory);

(c) The assessment team (Technical Officer and Assessor) visit the laboratory to make a preliminary evaluation and any informal recommendations. The visit is followed by the formal assessment at which recommendations regarding the granting of accreditation are made. Any non-compliances with the Quality Standard are dealt with and Accreditation is issued.

REFERENCES

EN45001 - General Criteria for the Operation of Testing Laboratories

EN45002 - General Criteria for the Assessment of Testing Laboratories

EN45003 - General Criteria for Laboratory Accreditation Bodies

APPENDIX C NPDES CERTIFICATION

C.1 REGULATORY FRAMEWORK

Through the mechanism of law and executive orders, US Congress empowers the Environmental Protection Agency (EPA) with primary authority to improve the nation's environment. In the area of water quality, the Clean Water Act (CWA) requires EPA to devise and maintain defensible water quality criteria, including priority pollutants listings. This research leads to criteria which take into consideration current testing capability, known health effects, and effects on aquatic life. Specific Water Quality Standards and effluent limits are then normally set by State Authorities using these criteria, while EPA oversees these activities. Most US States are authorised to issue discharge permits and take responsibility for the monitoring of laboratories, although for 25% of States this is conducted by EPA Regional Offices.

Although a number of industry specific effluent (end-of-pipe) limits are standardised (e.g. pharmaceutical manufacture), when setting effluent limits and controls consideration is given to toxicity, degradation, fate, ambient water quality and uses. These are then issued as National Pollutant Discharge Elimination System (NPDES) discharge permits, which all dischargers, commercial or state-owned, must apply for. The permit holder is responsible for all compliance conditions, and any failure is hence a violation of the CWA. EPA and State Authorities (as exemplified by the North Carolina State system) are responsible for monitoring NPDES compliance by imposing sampling frequencies, testing requirements, and document control and data reporting requirements upon the permittee. Permits must also include sampling locations and procedures, and refer to or include test methodology and data transformation and averaging. In 1984 EPA issued policy recommending the use of toxicity data to monitor water quality through the NPDES programme. Technical support documents, a permit writer's guide, and a methods manual were first published soon after.

Due in part to the policy of permittee self-monitoring, EPA in the early 1970s started to develop its 'QA programme'. Early activities included the validation and publication of many test methods, including technical procedures for quality control, calibration etc. None of these QA elements were mandatory, and data could not be adequately quantified (e.g. by confidence intervals). In 1979 efforts were centralised and co-ordinated under the 'Quality Assurance Management Staff', and Regional Laboratories and non-EPA organisations generating NPDES data were required to develop satisfactory quality systems.

Primary objectives of the overall QA programme are to ensure that data are not only technically sound and quantifiable, but are also fully and properly documented. The EPA 'Technical Support Document for Water Quality-based Toxics Control' states:

> **"Since most of the routine information gathered in compliance monitoring results from permittee self-monitoring, quality assurance (QA) is as important as compliance with limits."**

This section summarises EPA's NPDES requirements, including the State of North Carolina's certification criteria for biological laboratories. The NC scheme complies with national QA

requirements for effluent testing, and therefore relies upon EPA protocols and performance measures in a number of ways. NC State has issued a number of documents which amend or act in addition to the primary NPDES directives. The NC system is held up as one of the best examples of state compliance, and includes elements such as the Performance Evaluation (PE) programme for biological laboratories, which is a powerful mechanism of accreditation not found in many other US states.

C.2 NATIONAL DMR QA PROGRAMME

Analytical laboratories are required to take part in the EPA's national 'Discharge Monitoring Report Quality Assurance' (DMR QA) programme, which is operated annually through EPA Regional Offices. This evolved from the practice of requesting laboratories to perform analyses on control samples during EPA inspections. Pilot studies were conducted in 1979 by EMSL Cincinnati and EPA Regional Offices, and these received a favourable response.

The current programme, in which all major permittees must take part, is found to provide more effective data when assessing laboratory performance. Typical wastewater samples are distributed and are analysed by laboratories using their normal methods for NPDES self-monitoring, for determinands in their own permits. Results must be reported within 30 days, must follow a specific format, and a certification statement must be signed by laboratory management in accordance with NPDES regulations. A report on the performance of each laboratory is issued, and any results which are unacceptable are commented upon, with suggestions as to corrective action. Laboratories unable to resolve analytical problems will find that their data will no longer be accepted for effluent monitoring purposes.

A number of states operate certification schemes using site and sampling inspections (see C.4), and also DMR QA data. These systems evaluate a facility's staff, equipment, SOPs, data handling etc in general terms, and adjudicate on the laboratory's *fitness* to provide adequate data. In the absence of in-house QA however (see C.6), and as a consequence of the relative infrequency of DMR QA samples, and the fact that not all laboratories are required to take part, a number of States take the view that a more intensive system of providing regular results oriented data is required. This has led to State requirements for laboratories to measure accuracy and precision internally (control charts) for all test types, and to make this information available to regulatory authorities on request. This approach is found to be more cost-effective, and provides a closer estimate of the quality of monitoring data.

C.3 PERFORMANCE EVALUATION AND CERTIFICATION

Although a full certification programme for (analytical or) biological laboratories has not been established nationally, EPA has developed and documented the overall requirements for the various types of inspections of permittee and testing laboratories, to be performed by NPDES 'Audit Inspections' (PAIs) (see C.4) of ecotoxicology laboratories have also been established (EPA 1991).

NC officials perform inspections based upon the requirements of compliance evaluation inspections (CEIs) and performance audit inspections (PAIs) (see C.4.1 and C.4.2) in order to

ascertain permittee self-monitoring compliance. The data from these inspections are then used to decide on the frequency of inspections involving State sampling and testing of the discharge, although resource considerations means that these are mainly restricted to enforcement actions.

NC and a number of other State Agencies have developed state-wide interlaboratory sample check schemes, in order to provide more data for laboratory certification. The NC scheme involves samples ('Performance Evaluation samples') distributed at least once a year to over 500 laboratories, principally for *Ceriodaphnia, Daphnia pulex* and *Pimephales promelas* tests. The results are evaluated by NC's Certification Group personnel, who plot the data on control charts and determine the control limits for each distribution. For acute tests, ±2 standard deviations around the average LC50 is used, whereas for chronic tests the limits are set at ± one concentration level around the median NOEC value. Laboratories with results falling outside the limits are sent a further sample of the same material for re-testing; two consecutive failures automatically means a minimum 30 day decertification period.

In-house laboratory QC also forms an important part of the overall QA requirements. Reference toxicants must be tested every two weeks for acute tests (monthly for chronic tests) or based on workload, but must be at least quarterly to maintain certification for each test. Such controls are required for each test organism and test type. At least five valid controls must be entered onto the control chart of each newly certified method. Whilst laboratories are required to maintain control charts and closely monitor the coefficient of variation (CV) of test data, these are no set requirements or minimum standards for CVs. The results of national studies have shown intra-laboratory precision to range from 3% to 86%, although biological data has been shown to compare favourably with chemical data.

Discharges are normally sampled and tested for toxicity on a monthly or quarterly basis, and a single result outside the Permit limit is regarded as a non-compliance requiring corrective action. If any test fails to meet the 'required QA', then immediate re-testing is required.

C.4 TYPES OF REGULATORY INSPECTIONS

Under NPDES regulations there are a total of eight different types of site inspections performed by EPA or State Inspecting Authorities. Within these types, a number of inspectional procedures are in common, but there are specific activities or particular emphases, depending upon the purpose of the inspection. Regional EPA and State Agency approaches to inspections vary, and this can be quite a problem to those inspected. The two most frequent types of inspection performed are modelled on inspections first developed by EPA in the 1970s:

C.4.1 Compliance Evaluation Inspections (CEIs)

In this inspection, the overall management and operation of the facility is evaluated. If the inspected facility is a self-monitoring permittee, then compliance with both chemical and biological NPDES permit schedules and conditions is assessed, but no samples are taken. The inspection must include in-depth examinations of:

- the background and past performance of the facility (pre-inspection);

- the reports and data submitted by the facility (pre-inspection);
- the management and organisation of records;
- existence of SOPs reflecting up-to-date NPDES methodology;
- compliance with Permit for location and number of discharges;
- maintenance of all laboratory raw data generated during testing;
- maintenance of all records of laboratory cultures, instrumentation etc;
- dates, times, and location of sampling in accordance with Permit conditions;
- method(s) of sampling, preservation techniques, labelling, identification and tracking techniques, custody records, sample container requirements etc. in compliance with published EPA methods;
- auditability and consistency of archived raw data with previous reports.

Following the inspection, the NPDES Inspector is required to fill in an NPDES Compliance Inspection Report form and attach any supplementary narrative information, copies of other completed checklists used during the inspection, and any documentary support for important non-compliance observations. Information from the first section of this form (summarising the results of the inspection) is then entered into the Permit Compliance System (PCS), which is a centralised system for recording all national NPDES permit data.

C.4.2 Performance Audit Inspections (PAIs)

PAIs include all the in-depth inspectional requirements of the CEI, but also inspection of the process of self-monitoring. The Inspector(s) must accompany facility staff during sampling, flow measurement, laboratory testing, data manipulation and report preparation procedures. A check sample may be left by the Inspector for the laboratory to analyse. The evaluation of the Permittee's flow measurement procedures, and protocols for toxicity-based measurements are considered in greater depth in this type of inspection, but again, no samples are taken. The PAI normally takes one day, and the inspection team consists of an engineer and a chemist.

Laboratory QA is examined in more detail, including the procedures of the sample custodian and all laboratory QC methodology (reference standards, control charts). Specific checklists have been developed for use in the inspection of all EPA recommended toxicity tests.

C.4.3 Compliance Biomonitoring Inspections (CBIs)

Performed less frequently due to resource constraints (a CBI may take several days), the CBI is often conducted to support enforcement actions. It involves the collection of an effluent sample, preservation and transportation to EPA or State laboratories, and prompt testing using approved methodology. The Inspection should also include a cursory assessment of the activities in a CEI and an evaluation of permittee self-monitoring, but is less concerned with in-

house laboratory QA procedures, as the data generated by the Inspectors own samples will provide the required evidence.

C.5 SOPs

Laboratories performing ecotoxicological tests are required to 'develop and maintain' SOPs. There will be many SOPs in a large facility, and these should cover, but not be limited to, the following activities:

- all toxicity test methods;
- all culturing procedures;
- collection/purchase of test organisms;
- taxonomic examination of test organisms;
- cleaning of laboratory apparatus and glassware;
- maintenance and servicing of equipment;
- sample collection and custody;
- chemical water quality measurements;
- quality control;
- data manipulation, statistical analysis;
- reporting.

C.6 QUALITY ASSURANCE (QA)

In this quality system, 'QA' refers to any and all aspects of a method or compliance condition which is contained in an SOP, QA plan or permit. Test methods contain extra sections on specific QA conditions, such as test validity criteria (e.g. control survival). QA includes routine quality control (QC) procedures, such as the use of control charts. The complete programme for Quality Assurance is defined by EPA as 'the total programme for assuring the reliability of monitoring data' (EPA 1991). QA practices '.....must address all activities that affect the quality of the final effluent toxicity data, such as:

(1) effluent sampling and handling;

(2) the source and condition of the test organisms;

(3) condition and operation of equipment;

(4) test conditions;

(5) instrument calibration;

(6) replication;

(7) use of reference toxicants;

(8) record keeping;

(9) data evaluation.' (EPA 1993)

QA is not a *separate* function (or group of staff), independent from the testing activities, but is seen as an integral part of those activities. Appointment of a 'QA/QC officer' is not required for certification under the North Carolina State Programme for commercial testing laboratories (Matt Matthews, pers. comm.), and there are no EPA protocols defining the activities of a QA function, in relation to internal auditing. More recent EPA documentation (EPA 1993) has amended and expanded the section on quality assurance to include, as item 1, the 'appointment of a laboratory quality assurance officer with the responsibility and authority to develop and maintain a QA program', but how this is to be implemented within individual organisations is unclear. Nevertheless this may well reflect the EPA's current strategy for future NPDES programme development.

Within EPA's own regional testing laboratories however, the requirements differ from the above, and a more rigorous and independent approach to QA is necessary. The QA plan (see C.10) (which must be submitted for EPA HQ approval) must include organisation charts demonstrating the interrelationships between functional, testing, and data management units, as well as the relationship to the national program managers. There must be an office identified with overall QA responsibility, and this QA office must be organisationally independent from the testing and data management and archive functions. QA must review all SOPs, perform Performance and Systems Audits, and report directly to Regional Administrators.

C.7 STAFF COMPETENCE

Both the NPDES system principles, as documented by EPA (EPA 1991), and the NC State certification criteria (DEHN 1992) require records of training and qualifications to be maintained, as is common to all quality systems.

The NC system, however, includes a number of highly specific requirements for the qualifications of various personnel. These qualifications are not particularly unusual, but such a level of specification is unique to this system, and these are rigidly enforced:

Laboratory Manager (supervisor)	Minimum BSc (biology or biological science) + three years full-time laboratory experience of aquatic toxicity testing, or MSc + one year full-time experience.
Biologist/analyst (senior technician)	Minimum BSc + two weeks on-the-job training in culturing and toxicity testing of effluents, at a federal agency or college. One year practical (unsupervised) experience in testing protocols used in the NPDES programme.

Biological Technician	Minimum full secondary school education + two years of college. One week training in toxicological testing, and one week in culturing, at federal or academic institutions. One year practical experience.

C.8 FACILITIES

Biological testing laboratories must be well appointed, and criteria for the adequacy of the facility are well defined:

- temperature control (20-25 °C) and lighting (100 ft-candles);
- workspace 150 sq ft./person and clear benchspace 10 ft./person;
- waste treatment plant;
- vented exhaust hoods and contamination free area;
- cold storage and secure storage areas;
- separate areas for culturing and use of test systems (organisms), separate areas for chemical analysis and handling of toxic materials or samples;
- viable cultures of all test organisms included in the scope of the certification;
- tank temperature control facilities (e.g. heat exchangers, water baths);
- natural or deionised/distilled waters used in media preparation must be tested at least monthly for conductivity, pH, total hardness, total chlorine, heavy metals and organics.

Further QA requirements as required for NC certification:

- all instruments must be calibrated daily or with each usage, and all records kept;
- use of control charts for each test type;
- quarterly taxonomic examination (to species level) of all cultured organisms, and specimens archived for a minimum of one year;
- culturing and testing activities not allowed within a single incubator.

C.9 SAMPLING PROCEDURES

A detailed discussion of the technical considerations when sampling for different test types and/or differing discharge conditions is beyond the scope of this review. NPDES requirements for QA inspection and evaluation of the Permittee's sampling program are well developed. Proper sampling and custody procedures are an essential prerequisite of all test methodology, and represent an important element of data quality that cannot be evaluated using quality control techniques such as control standards or interlaboratory check samples. When

inspecting these procedures, NPDES Officials must refer to sampling methodology as detailed in the Acute and Chronic EPA Methods Manuals, as well as those required by individual Permits or in Administrative letters.

The sampling point and type of sample to be collected (spot or composite) will depend upon the consistency or flux in effluent composition and the purpose of the test, and should be defined in the Permit. Those laboratories authorised to collect samples must maintain written procedures (SOPs), and individuals must be formally trained and declared competent by the regulatory authority to conduct the procedures.

Samples of effluent must normally be tested within 36 hr of collection if for off-site testing, or 24 hr if tested on-site (EPA 1993). For chronic tests of a number of days duration, fresh samples should be collected for solution renewal. If it can be shown that the toxicity of effluents is not reduced through physico-chemical changes, a request for an extension to no greater than 72 hr, may be made to the EPA. Non-compliance with this requirement will invalidate data for NPDES purposes. All samples, except for those to be tested immediately on-site, must be cooled (to 4 °C) and shipped in iced containers. Temperatures should be recorded at dispatch and upon receipt.

Data to be recorded in field records at the time of sampling includes the exact location, date and time, signature of sampler, identity of sample, sampling procedure (ref. to SOP), preservative used, test required, and seal codes. Samples must also be fully labelled. There must be a complete audit trail for the custody of samples at all times. 'Custody' is defined as being in someone's physical possession (within sight) or in a locked store with access restricted to the sample custodian(s). Samples transported from field collection must be locked in a vehicle when not in the collector's possession. On receipt at the laboratory the codes on these seals should be confirmed. The custody record must accompany the sample(s) at all times, and must be signed by all persons when relinquishing or accepting possession.

C.10 QUALITY MANUAL

Facilities are required to maintain a 'quality manual', known as a Quality Assurance (QA) Plan, or QA Program Plan. A QA Plan is an organised collection of management policies, organisational responsibilities and general processes by which a laboratory intends to produce quality data to meet defined objectives. It should include a list of SOPs, a schedule of the tests for which approval/certification is sought, and also document Data Quality Indicators (the required precision of control/reference standards for all tests), and Data Quality Objectives (DQOs). DQOs are qualitative and quantitative statements which set the target precision and accuracy (if applicable) for all types of sample data, for all laboratory effluent tests. DQOs are derived directly from Permit conditions and by discussion with the users of the data (regulatory agencies).

REFERENCES

Department of Environment, Health, and Natural Resources (1992). Biological Laboratory Certification Guidance Manual for Effluent Toxicity Testing. DEHNR, NC, January, 1992.

EPA (1988). NPDES Compliance Inspection Manual, EPA Washington, PB88-221098.

EPA (1991). Manual for the Evaluation of Laboratories Performing Aquatic Toxicity Tests. EPA Washington, EPA/600/4-90/031.

EPA (1989). Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms. Environmental Monitoring Systems Laboratory (EMSL) - Cincinnati.

EPA (1993). Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms (Fourth Edition). EPA Washington, EPA/600/4-90/027F.

EPA (1991). Technical Support Document For Water Quality-based Toxics Control. EPA Washington, EPA/505/2-90-001.

ANNEX III EXTERNAL QUALITY CONTROL

Annex IIIa

Canadian Toxicological Testing Laboratory Accreditation Program: Program Description, CAEAL, December 1993

Canadian Toxicological Testing

Laboratory Accreditation Program

Program Description

December 1993

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 Glossary

The following definitions apply:

Acceptable Deviation (AD) Value: Concentration limits defining the acceptable deviation of a reported value from the reference value. Acceptable Deviations are based on inter-laboratory 95% confidence limits or other appropriate criteria.

Accreditation: The formal recognition of the competence of an Environmental Analytical Laboratory to carry out specified tests. Formal recognition is based on an evaluation of laboratory capability (and performance); site inspections are utilized in the evaluation of capability.

Batch or Lot Number: A specific quantity or lot of a test, control, or reference substance; (GLP)[1].

Environmental Analytical Laboratory: A laboratory engaged in the physical, chemical or biological measurement of either the receiving environment or discharges to the receiving environment.

Can: is used to mean "is (are) able to".

Certification: The formal recognition by the Canadian Association for Environmental Analytical Laboratories of the proficiency of an Environmental Analytical Laboratory to carry out specified tests. Formal recognition is based on a screening of laboratory capability and an evaluation of laboratory performance.

Control Substance: Any chemical substance or mixture, other than the test substance, feed, or water, that is added to the test system for the purpose of comparing the test substance for either a chemical or biological response with a control material; (GLP).

Good Laboratory Practice (GLP): A regulation specifying data collection and study activities for studies conducted to support the registration of a product (test substance), either a drug, pesticide, or chemical. Regulations are specified under the Organization for Economic and Cooperative Development, the U.S. Environmental Protection Act and the U.S. Food, Drug and Cosmetic Act.

Inter-laboratory Variability: Variability between laboratories.

Intra-laboratory Variability: Variability within a laboratory.

Master Schedule: Formal listing of all regulated studies conducted at a facility. A requirement of GLP laboratories; (GLP).

May: is used to mean "is (are) allowed to".

Must: is used to mean an absolute requirement.

---

[1](GLP) designates definitions and requirements mandated by the Good Laboratory Practice (GLP) Regulations.

Performance Evaluation (PE) Sample: A characterized sample, having designated reference values, which is used in the evaluation of laboratory performance.

Pre-Audit Form: A form used to update laboratory organization, testing, equipment etc. to be completed by the laboratory just prior to the Site Evaluation.

Protocol: A written study plan required for GLP studies; (GLP).

Protocol Amendment: An authorized permanent change in a protocol. Protocol Amendments are generally planned changes to the protocol; (GLP).

Protocol Deviation: A temporary deviation from protocol specifications; (GLP).

Publicity Statement: Interim recognition by CAEAL that a laboratory has successfully completed the toxicological laboratory accreditation program application form, has submitted specific required documentation, and has enrolled in the toxicological laboratory Performance Evaluation testing program

Quality Assurance: An evaluation program internal to the laboratory, but external to testing activities, employed to monitor and evaluate study activities for the test facility management. Quality Assurance programs are required for GLP studies;

Quality Control: An evaluation program internal to the laboratory which is integrated into the testing program to monitor and evaluate the accuracy and precision of recorded measurements.

Quality Manual: A document stating the quality policy, quality system and quality practices of an organization (ISO).

Raw Data: Any laboratory worksheets, records, memoranda, notes, or exact copies thereof, that are the result of original observations and activities of a study and are necessary for the reconstruction and evaluation of the report of that study; (GLP).

Recommended Action: Corrective action specified by a laboratory auditor if the consequences of an observed laboratory deficiency are not so severe as to compromise the integrity of the testing. Although neither specific implementation dates nor follow-up written confirmation are required, the laboratory should act upon such recommendations as expediously as possible.

Required Action: Corrective action specified by a laboratory auditor if an observed laboratory of deficiency is deemed to compromise the integrity of the testing (ie: absence of procedures, documentation, faulty equipment, inadequate staff performance etc.) or test requirement is not met. The laboratory must carry out these actions within a specified period of time and provide written confirmation to CAEAL that the actions have been carried out before the laboratory can receive accreditation.

Reference Toxicant Testing Program: Internal laboratory testing conducted as part of Quality Control activities to evaluate the laboratory's ability to routinely produce consistent results. Reference toxicant tests are conducted periodically and plotted on a graph to establish the laboratory's historical performance. Results are used to assess the sensitivity of an organism over time, and to assess the precision of results obtained by a laboratory for a particular toxicant.

Reference Value: Concentration value, assigned to a Performance Evaluation Sample. This value may be based on any appropriate combination of design value, inter-laboratory consensus value or value provided by a laboratory with demonsrated accuracy.

Should: is used to state that the specified condition or procedure is recommended and ought to be met if possible.

Site Evaluation: Formal on-site inspection of laboratory operations.

Sponsor: Organization which initiates and financially supports an investigation. An Organization which submits reports to federal authority to support registration of a product; (GLP).

Standard Operating Procedure (SOP): Written documents which describe routine laboratory activities for analytical testing. Both procedural and equipment maintenance and calibration SOPs are required for GLP studies.

Study Director: Principle scientist in a GLP study.

Test Standard: Work plan or technical test methods specified by government authorities. Canadian biological testing methods for the six initial accreditation tests are listed in Section 13.

Test Substance : Chemical under investigation in a GLP study.

Test System: Organism used in a GLP study.

Warning Charts: A plot of the results of reference toxicant data, sometimes referred to as "Control Charts". Methods for developing Warning Charts can be found in "Guidance Document on Control of Toxicity Test Precision Using Reference Toxicants"; Environment Canada EPS Report 1/RM/12.

2.0 Introduction

Environmental studies conducted under the Good Laboratory Practice (GLP) Regulations and non-GLP toxicological testing programs assess the effect and fate of effluents, chemical substances and chemical mixtures on the environment. Aquatic environmental studies often utilize similar testing procedures, species, water quality and environmental conditions, yet, regulatory mandates and accreditation processes for these two programs have often evolved along separate pathways. This duality is expensive and does not promote standard requirements for data integrity and data quality. The Canadian Association for Environmental Analytical Laboratories (CAEAL) has entertained the concept of providing an accreditation program which would provide either an all inclusive accreditation program or selective accreditation for laboratories conducting these types of studies. The purpose of this document is to present a framework for accrediting laboratories conducting either non-GLP environmental toxicological testing and/or studies conducted under GLP Regulations, by CAEAL.

CAEAL was formed in 1989 on the initiative of a number of public and private sector laboratories and is incorporated as a non-profit association. A principal objective of the association is to upgrade and maintain a strong environmental analytical and toxicological service capable of consistently producing high quality scientific data. It is the intent of the program described in this document to provide a vehicle for meeting this objective by offering the opportunity for member laboratories to participate in a national accreditation program. The standards and criteria applied in the accreditation process have been established by consensus among member laboratories and the CAEAL Board of Directors. Laboratories that successfully complete the accreditation process will have achieved a recognized level of competence that meets these national standards.

The overall design of the toxicological testing accreditation program was developed initially by the Toxicological Testing Laboratory QA/QC Committee, and revised further based on comments received from the general membership and the CAEAL Board of Directors. Compatibility with the existing CAEAL analytical program was a major consideration during the development of the Toxicological Testing Accreditation Program. The CAEAL chemical and toxicological programs have been designed to meet the requirements of the International Standards Organization guide entitled: "General requirements for the competence of calibration and testing laboratories" (ISO Guide 25, 1990 Edition). Although GLP accreditation is not offered at this point in time, it is anticipated that CAEAL will offer accreditation for GLP as well as non-GLP work in the future. CAEAL is currently discussing GLP program details with Environment Canada and a Memorandum of Understanding between the two parties which outlines Canadian GLP program elements that CAEAL could deliver is being developed.

3.0 Scope

The scope of the CAEAL Toxicological Testing Laboratory Accreditation Program is multi-fold in purpose. It currently offers the opportunity for laboratories to select accreditation for specific tests conducted for toxicological testing of environmental samples and will eventually offer GLP accreditation. The difference between the two programs lies in the degree of implementation of the Quality Assurance program, additional documentation and other required functions, such as specific Standard Operating Procedures (SOPs), required protocols, Master Schedule Sheets, and archival of data, which are required by the GLPs.

The CAEAL Toxicological Testing Laboratory Accreditation Program consists of 3 operational

components; the Performance Evaluation (PE), the Site Evaluation (SE) and the annual submission of Warning Charts produced monthly by the laboratory to demonstrate on-going precision and consistency in laboratory activities. The PE component utilizes unknown samples to evaluate the laboratory's performance in determining a known toxicant value, while the Site-Evaluation investigates laboratory facilities and operational activities, and assesses whether or not the laboratory meets toxicological testing and/or GLP standards. The submission of Warning Charts will be used to evaluate the continuity of laboratory procedures and activities to produce data of consistent quality. The granting of Laboratory Accreditation is test specific, and is based on participation in all components; (ie; the PE program, Site Evaluation, and annual submission of intra-laboratory Warning Charts). For those tests where suitable PE samples are not available, a laboratory may receive a Certificate of Competence based on the results of the Site Evaluation and submission of Warning Charts to CAEAL every twelve months, as specified in this document. Both types of recognition (i.e. Laboratory Accreditation and Certificate of Competence) may be referred to as Accreditation in this program. However, it is emphasized that the Certificate of Competence will only be granted for tests where PE samples are not yet available, and to laboratories which satisfy all other evaluating criteria.

Recognizing that the program is early in the implementation stage and the offering of formal recognition is about a year away, CAEAL has offered interim recognition in the form of a Publicity Statement. CAEAL will grant this recognition statement upon successful completion of the Documentation Registration Phase which includes completion of an application form, submission of specific required documentation (as outlined in this document), and enrolment in the Toxicological Laboratory Performance Evaluation testing program.

Accreditation is available to any member laboratory holding an institutional membership and will be granted by the CAEAL Board of Directors on the recommendation of the CAEAL Advisory Panel which consists of the CAEAL toxicologist and other specific government members of CAEAL. The CAEAL Advisory Panel will make recommendations to the Board of Directors, based on the results of the laboratory PE tests and the Site Evaluation as well as the submission of monthly Warning Charts.

The Site Evaluations will be carried out by a two person team of trained auditors, experienced in toxicological testing. This team will normally consist of two formally trained toxicologists, drawn from the membership. CAEAL and the applicant laboratory will discuss potential assessors to avoid potential conflict of interest.

4.0 Program Description

The major thrust of the program is to provide accreditation for laboratories conducting toxicological testing for environmental testing programs and eventually for GLP studies. Laboratories will have the opportunity of choosing which program (ie: environmental testing and/or GLP) they wish to apply for in the Application Form. The difference between accreditation for environmental testing and GLP activities will be based primarily on the requirements of the Site Evaluation and the degree of implementation of recommended procedures.

The accreditation process will include a series of document submissions to CAEAL, participation in PE sample analyses every six months, Site Evaluations every two years and successful maintenance of Warning Charts, submitted to CAEAL once a year. Accreditation will be renewed and updated annually.

As part of the first step in obtaining accreditation, laboratories may apply for accreditation by forwarding a completed Application Form to CAEAL. As discussed above, for the first 12 months of the program

(October 1992 - October 1993), a laboratory may seek recognition in the form of an interim *Publicity Statement* upon completion of the Documentation Registration Stage. Participating laboratories should begin developing Warning Charts, for tests where accreditation is desired. A list of recommended reference toxicants is provided in Table 1. Reference toxicant testing must be conducted monthly for those organisms actively cultured at the laboratory, and must be conducted as described in Environment Canada test methods when new batches of organisms are received at the laboratory.

TABLE 1 Reference Toxicants Recommended For Toxicological Testing

TEST	*REFERENCE TOXICANT*
Rainbow Trout Lethality	zinc sulphate phenol, potassium chloride
Daphnia magna Lethality	zinc sulphate, sodium chloride, potassium chloride
Bacterial luminescence inhibition	zinc sulphate, phenol
Ceriodaphnia reproduction inhibition	sodium chloride, potassium chloride
Fathead minnow growth inhibition	zinc sulphate, sodium chloride, potassium chloride
Selenastrum growth inhibition	zinc sulphate, sodium chloride, potassium chloride

The Reference Toxicants listed in Table 1 are not mandatory for intra-laboratory Warning Chart development, but are recommended for use by participating laboratories. The list was developed through membership consensus opinion and following guidance provided in Environment Canada test method documents. Intra-laboratory variability will be monitored by submitting internal laboratory Warning Charts to CAEAL annually, and inter-laboratory variability will be monitored by statistical analyses of the results of PE testing of unknown substances. An integral part of the CAEAL accreditation process is the laboratory's continued demonstration of precision and the production of data of consistent quality in a timely fashion. Thus, at the time of yearly renewal, copies of internal on-going Warning Charts will be forwarded to CAEAL for verification of completion. Participation in twice yearly PE testing rounds is mandatory for tests where accreditation is being sought.

The Site Evaluation inspection is a comprehensive review of on-going laboratory procedures and activities. During these inspections, the organization and facility, study procedures and study conduct will be reviewed in depth to verify credentials, qualifications and capabilities of the candidate laboratory.

If a laboratory wishes to receive GLP accreditation as well, an additional half-day will be needed to complete the Site Evaluation. The extra half-day will be used to evaluate those items specific to GLP accreditation, such as the Master Schedule, the Quality Assurance Unit, the Archives, and expanded Standard Operating Procedures.

5.0 Application for Accreditation

Laboratories may apply for accreditation by forwarding a completed Laboratory Accreditation Program Application Form to CAEAL. As part of the application process, applicant laboratories must agree to the terms and conditions of accreditation. The application process will also include submission of specific required documentation, outlined in Table 2, and enrolment in the toxicological laboratory performance evaluation testing program . This information will be reviewed by the CAEAL Advisory Panel, and upon successful completion of the application process, the applicant laboratory will be enrolled in the accreditation program. Until formal accreditation is available, a laboratory can obtain interim recognition in the form of a Publicity Statement upon completion of the application process (See section 8.0).

Subsequent to completing the original application, application for accreditation of additional tests can be sought at any time by either forwarding a new Application Form to CAEAL or appending the original Application Form.

TABLE 2 Documents Required to Fulfil Documentation Registration Requirements

---

- A list of staff, academic qualifications, years of experience and organization chart identifying reporting responsibility
- A schematic lay out of the testing laboratory identifying the locations and area for organism holding/culture and testing areas
- A description of the source and treatment of the primary water supply.
- A chemical description of the treated primary water supply including hardness, alkalinity, pH, TRC, ammonia, nitrite, copper, zinc; the laboratory conducting the measurements and the frequency of analysis must be identified; the laboratory supervisor must identify that measured levels are acceptable according to a recognized authority (e.g., Environment Canada, U.S., EPA, ASTM, Standard Methods);
- A list of equipment (manufacturer and model) used in the laboratory for culturing of organisms and the conducting of tests.
- A list of the standard operating procedures (with preparation date) used by the laboratory for conducting toxicity tests (e.g., trout, *Daphnia*, fathead minnow, algae, *Ceriodaphnia* and *Photobacterium*), analytical tests required of toxicity tests, procedures for organism holding and culturing and reference toxicant testing.
- The laboratory QA/QC manual or standard operating procedures identifying the Quality Management Plan, reference toxicant testing program (including toxicants used, organisms tested and frequency of testing), sample tracking system used in the lab, and data reporting format.
- Standard operating procedures for storage of test data for a period of at least five years.
- A signed statement by the laboratory supervisor that all the information describing the laboratory, its staff and operations is true.

6.0 Evaluation Process

The evaluation process includes a continuous comprehensive evaluation of laboratory activities, including a Site Evaluation and submission of Warning Charts, and the successful completion of PE samples.

6.1 Warning Chart Submission

The laboratory must submit two copies of test-specific Warning Charts each year at the time of accreditation renewal. However the results of these charts will not enter into the rating process. They are submitted only to demonstrate on-going proficiency and acceptable corrective action on the part of the laboratory, as required. Results of all reference toxicant testing, including those falling beyond the acceptance criteria in the Warning Charts, must be reported.

6.2 Performance Evaluation

Performance Evaluation (PE) testing is often used to evaluate a laboratory's ability to analyze, quantitate and report a value of an unknown substance. PE samples can be used to provide information on inter-laboratory variability. Firstly, they provide quantitative and qualitative assessment of laboratory activities and analytical capabilities; secondly, they provide useful information in understanding normal variability between laboratories; and thirdly, by increasing the reliability of data produced by laboratories, they assist regulators in making decisions relating to environmental change. PE testing can be established in several ways. However, there are presently insufficient data to select one approach for toxicity testing laboratories. As agreed at the general meeting of the Canadian toxicological laboratories in Edmonton (October 7, 1992) and by the CAEAL Board of Directors (December 1992), two parallel approaches to PE testing will be utilized for the first 2-3 rounds of the PE program, to determine which provides the best measure of accurate and reproducible laboratory performance. These two approaches are:

The LC50/IC50 Approach: Each laboratory is required to determine and report the LC50/IC50 of an unknown dry chemical.

The Classical Bioassay Approach: In addition to the LC50/IC50 determined for the dry form of an unknown chemical (LC50/IC50 Approach), each laboratory is required to determine and report the concentration of the same unknown toxicant in a solution of unknown concentration.

This is a straightforward application of the classic parallel line biological assay technique that has been historically used in pharmaceutical toxicological studies.

Both approaches allow the laboratories to use their own normal dilution/control water supply. Sets of data utilizing both approaches will be collected, evaluated and summarized. Following this initial period of parallel testing and evaluation, the CAEAL QA Committee will determine the approach to be used for subsequent testing.

Laboratories must conduct two PE rounds per year for each test for which accreditation is sought. For

each acute test, CAEAL will send the laboratory one sample of a dry chemical and four corresponding chemical samples in a liquid form, each of unknown concentration. For each sublethal/chronic test, the number of dry chemical/unknown liquid sample(s) has yet to be defined. Instructions for preparation of the toxicants, conducting the tests, turn-around time and data submission date will accompany the samples. The CAEAL Toxicologist will process the data submitted and code the laboratory identification to assure confidentiality. Acceptable deviation will be determined based on statistical analysis of submitted data which will include calculation of consensus mean and standard deviation. A summary of the statistical analyses used will be included in the report of PE results to the laboratory.

6.2.1 Approaches to Conducting Performance Evaluations

As previously discussed, both the LC50/IC50 and Classical Bioassay Approaches will be evaluated in parallel for the first 2-3 PE Rounds of the Program. In each PE round, the chemicals will be one of the recommended reference toxicants, or others, but will be completely unknown to the laboratory.

First, using the LC50/IC50 Approach, the laboratory will establish a standard curve of concentration vs. endpoint (survival, reproduction, growth, or light reduction) to determine the LC50/IC50 of the dry form of the sample sent. The laboratory is required to report the LC50/IC50 of the PE dry chemical sample to CAEAL. Then, for the Classic Bioassay Approach, the laboratory will determine the LC50/IC50 of each liquid sample sent by CAEAL and estimate the concentration of the four unknown solutions based on the LC50/IC50 of the corresponding dry chemical sample (as determined in the LC50/IC50 Approach). The laboratory is required to report the estimated concentration of the chemical in each of the four unknown solutions to CAEAL. While laboratories are required to conduct PE tests twice yearly and submit results from both approaches, the PE test results will not enter into the accreditation rating until the best approach is determined. Scores will be determined for each laboratory, however, in order to help in the determination of the best PE Approach. The scoring process is described in Section 6.4.

An example of the PE approach for one toxicant is given in Tables 3 and 4 and Figure 1:

> The laboratory first conducts a toxicity test using the dry form of the unknown chemical to establish the LC50. The value is reported to CAEAL (Table 3). The laboratory then conducts a toxicity test on the four liquid samples of the corresponding unknown PE sample and estimates the concentrations of the unknown solutions based on the Standard Toxicity Curve or the LC50 (Figure 1 and Table 4). The estimated concentrations in the unknown solutions are reported to CAEAL (Table 4).

TABLE 3 Example Of LC50s For One Dry PE Sample As Determined By 5 Laboratories

LABORATORY IDENTIFICATION	*LC50 (mg/L)*
Laboratory A	5
Laboratory B	10
Laboratory C	0.5
Laboratory D	30
Laboratory E	4

FIGURE 1 Example Of Standard Toxicity Curves Of One Dry PE Sample As Determined By 5 Laboratories

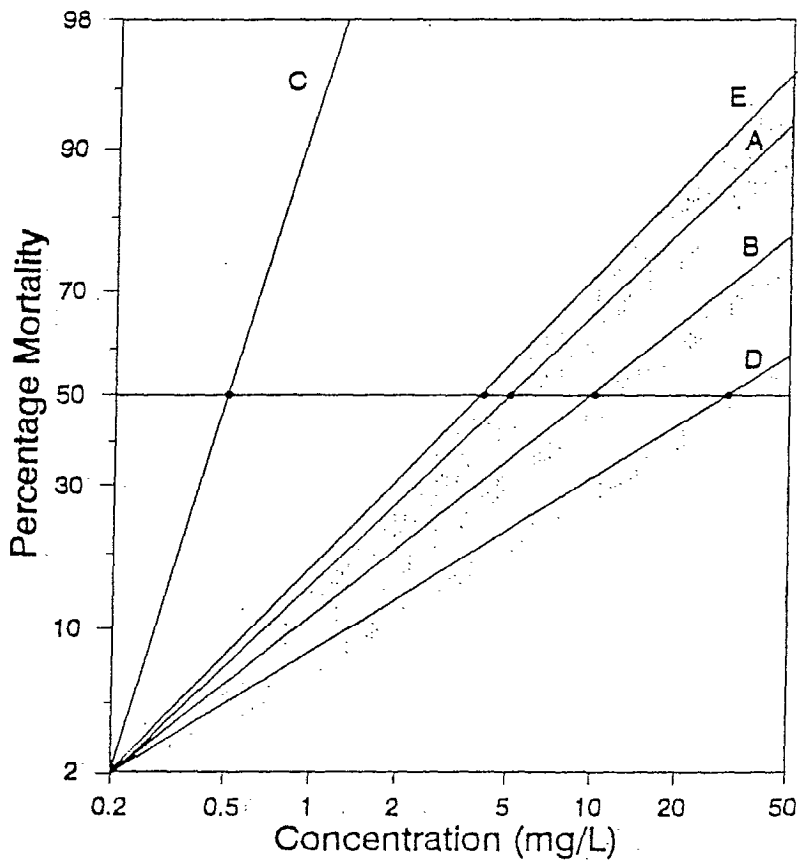


TABLE 4 Example Of PE Results For One Dry Sample And One Corresponding Liquid Sample As Determined By 5 Laboratories

Laboratory ID	*LC50 OF Dry Chemical (mg/L)*	*LC50 of Unknown Liquid Sample (% V/V)*	*Estimated Concentration of Unknown Toxicant (mg/L)*
Laboratory A	5	20	25
Laboratory B	10	20	50
Laboratory C	0.5	2.5	20
Laboratory D	30	100	30
Laboratory E	4	5	80

$$\frac{\textit{LC50 of Dry Chemical}}{\textit{LC50 of Unknown Solution}} \times 100 = \textit{Estimated Concentration of Unknown}$$

6.2.2 PE Analyses Passing Methods

The potential differences between these two approaches may be great. The criteria of whether or not a laboratory passes or fails the PE may be dependent on the selected design of the PE. To allow fairness, for the first 2-3 rounds, the laboratories will not be assessed based on PE results. However, since the PE component of the program is required for full accreditation, formal accreditation will not be provided until the best PE approach has been selected. After data have been collected and membership views have been sought the CAEAL QA Committee will recommend which method is most appropriate for future PE testing to the CAEAL Board of Directors.

6.3 Site Evaluation

The purpose of the Site Evaluation (SE) is to assess the laboratory's capabilities, determine strengths and weaknesses, and specify required and recommended corrective action.

Once every two years, member laboratories will participate in a Site Evaluation which will involve a thorough evaluation of the laboratory's facilities, organization and operational activities. Trained CAEAL staff or member assessors will conduct the inspection. Each Site Evaluation will be conducted by two

toxicologists formally trained by the National Association of Testing Authorities (NATA). NATA is internationally recognized and provides an intense four-day training course. The course deals extensively with both ISO 9001/9002 and ISO Guide 25 and trains participants in how to conduct assessments against these standards. Some assessors will receive additional training for conducting GLP Site Evaluations. The average length of the inspection will be one and one half days for those laboratories seeking accreditation for environmental testing samples only, and an additional half a day for those laboratories seeking additional accreditation GLP. Preceding the Site Evaluation, the laboratory will be sent a Pre-Audit Form which will be used by the laboratories to prepare for the Site Evaluation and may be sent to the assessors to ensure that they have the most recent information describing the laboratory's capabilities. The assessment will begin with an opening meeting with laboratory management during which the Site Evaluation process will be described, and documentation of any changes which may have occurred since submission of the Application and Pre-Audit Forms will be provided to the Lead Assessor.

A tour of the laboratory will follow the opening meeting. The Evaluators will observe facilities, equipment, procedures and log books in place and document the presence or absence of any given element of the program. The Rating Guide will be used to officially document compliance or non-compliance with the program and method requirements.

Following the inspection, the Evaluation Team will review laboratory records. The Laboratory Assessors to complete the Rating Guide and prepare a Summary of Findings Assessment Report which will list all findings, and specify required and recommended actions.

A full debriefing with key laboratory staff will be held at the conclusion of the inspection. A copy of the Rating Guide and the Summary of Findings Assessment Report will be left with the Laboratory Director.

6.4 Rating Process

The laboratory must receive a satisfactory evaluation on both the PE and Site Evaluation and must submit in-house Warning Charts annually in order to receive CAEAL Accreditation.

6.4.1 Performance Evaluation

Performance evaluations will need to be conducted for each test where accreditation is sought. As previously discussed, each acute PE test will consist of 1 sample of a dry chemical and 4 samples of the toxicant prepared in solution and two approaches to testing will be used. Data from each approach will be analyzed by the CAEAL Program Director to establish acceptable deviation criteria for each LC50/IC50 and each estimated concentration of the unknown solutions. Rating points will be assigned as demonstrated in Table 5. Acceptable PE scores will equal or exceed 70. As previously discussed, laboratories will not be assessed based on PE results for the first 2-3 rounds of the program, however, the data will be used to determine the best approach for PE testing. Once a single PE approach is chosen, full accreditation will be available and laboratories will be assessed based on PE tests well as the Site Evaluation and the submission of Warning Charts. If a laboratory fails a PE test, the laboratory must request a new PE sample within a month. Failure of two consecutive PE tests will compromise accreditation for that test.

TABLE 5 Performance Evaluation Rating Process

DEVIATION / AD VALUE	*POINTS ASSIGNED*
< 0.5	5
≥0.5 - ≤ 1.0	4
>1.0 - ≤ 1.5	2
>1.5	0

where:

Deviation = Reference Value - Reported Value

Reference Value = the actual concentration of the PE solution, either nominal or measured

AD Value (Acceptable Deviation Value) = concentration limits defining the acceptable deviation of a reported value from the reference value; based on inter-laboratory consensus (95% confidence limits) or other appropriate criteria.

The example presented in Table 6 illustrates the calculation of each laboratory's Deviation from the reference value (assume the reference value is 30 mg/L for the purpose of this example) for the five example laboratories presented in Tables 3 and 4.

TABLE 6 Example Of Deviations Calculated For One Liquid Sample As Determined By 5 Laboratories

Lab ID	Reference Value (mg/L)	Reported Value (By Classical Approach) (mg/L)	Deviation (mg/L)
A	30	25	5
B	30	50	20
C	30	20	10
D	30	30	0
E	30	80	50

It is unclear at this time how the AD value will be determined. Performance Evaluation scores, for a particular approach, will then be calculated as follows:

$$PE\ Score = \frac{Total\ Points}{No.\ of\ Samples} \times \frac{100}{5}$$

6.4.2 Site Evaluation

Legitimate variability in PE analyses can result in toxicological test data due to the genetic background of organisms and/or dilution water quality chemistry. An example of non-legitimate variability would be the laboratory's inability to properly conduct a specific test which may not be exposed until a complete site inspection is conducted. For these reasons, emphasis in the toxicological laboratory accreditation program is placed on the Site Evaluation (SE).

At the conclusion of the Site Evaluation, based on observations, an overall evaluation of laboratory activities will be made. An acceptable Site Evaluation will not necessarily mean the absence of deficiencies, but will acknowledge a basic level of assurance that a system is in place to ensure data integrity and production of acceptable quality. As outlined in Section 6.3, all corrective actions noted will fall into one of two categories: (1) required actions or (2) recommended actions. If the observed deficiencies are deemed to compromise the integrity of the tests, then required actions are specified. Required actions must be carried out by the laboratory within a specified period of time and the laboratory must provide written confirmation to **CAEAL** that the actions have been carried out. Implementation periods, whenever possible, will not exceed 30 days. **CAEAL** may recommend reassessment after a specified period of time. If the consequences of the deficiencies are not so serious as to compromise the integrity of the tests then recommended actions are specified. Although, neither specific implementation dates nor follow-up written confirmation apply to the recommended actions, the laboratory should act upon such recommendations as expediously as possible.

7.0 Records Management

Upon receipt of a completed Membership Application Form, **CAEAL** will establish a file for each laboratory. Information pertaining to laboratory accreditation received by **CAEAL** will be placed in the laboratory file. This information will include the Membership Application Form Document Submission, PE scores, SE reports, Pre-Audit Forms and annual Warning Charts for each test where accreditation is indicated. Decisions on accreditation will be made yearly. The laboratory will be notified in writing once the evaluation process is complete.

8.0 Accreditation Process

Details of the Accreditation Process are provided in Table 7. Upon successful completion of the application process (which includes: completion of application form; submission of specified documents; and enrolment in the PE Testing Program) laboratories will receive recognition in the form of a Publicity Statement, as follows:

> "CAEAL recognizes that (Laboratory Name) has enroled in the initial phase of the Toxicological Testing Laboratory Accreditation Program, and has successfully submitted the documentation which describes their capability in performing (Name of Test)toxicity test(s), describes associated QA/QC practices and has enroled in the performance evaluation testing program."

Full accreditation will not be available for the first 12 months of the program while the most appropriate PE approach is being determined.

Laboratories meeting the criteria will receive notification of official accreditation (or renewal) on an annual basis. Depending upon the type of recognition sought, various pre-requisites are required for granting recognition.

TABLE 7 Laboratory Accreditation Process

PRE-REQUISITE FOR FORMS OF ACCREDITATION	PUBLICITY STATEMENT	CERTIFICATE OF LABORATORY ACCREDITATION	CERTIFICATE OF GLP ACCREDITATION	CERTIFICATE OF LABORATORY COMPETENCY
Satisfactory completion of application process (including submission of required documents and enrolment in PE program)	✓	✓	✓	✓
Obtain Acceptable SE Report		✓	✓	✓
Obtain PE Score of 70 or above		✓	✓	
Warning Charts for review by CAEAL toxicologist		✓	✓	✓
Awarded if appropriate PE substance is unavailable				✓
Additional documentation as specified in GLP			✓	

Recognition (Publicity Statement, Laboratory Competency, or Certificate of Laboratory Accreditation for specific tests) may be suspended, subsequent to its having been granted if a laboratory; (i) fails to comply with the terms and conditions of accreditation, (ii) fails to carry out all non-test specific required actions within the time period specified, (iii) fails to carry out test specific required actions within the time period specified, (iv) fails to successfully analyze two successive sets of PE samples for a specific test, in the case of accreditation.

A laboratory that is found not to be in compliance with items (i), (ii), or (iii) above will be notified in writing and will be requested to take appropriate corrective actions. If the laboratory does not initiate appropriate corrective action and so advises CAEAL of these efforts, in writing, within 30 days of it being notified, CAEAL will give written notice that accreditation for the tests in question is suspended. If appropriate action is not taken within a further 30 days, accreditation will be withdrawn.

A laboratory that fails to achieve an acceptable PE score or SE Report for a specified test will be notified in writing. If failure to achieve an acceptable rating occurs again, the laboratory will receive written notice from CAEAL that accreditation for the area in question is suspended. If the laboratory fails to achieve an acceptable PE score on the third successive sample set of PE samples, or does not initiate corrective action for SE citations, accreditation will be withdrawn.

Within 30 days of receiving a suspension notice, for whatever reason, the laboratory has the right to appeal its case to CAEAL in writing. The subsequent decision of CAEAL, based on evidence available for review of the appeal by the CAEAL Board of Directors, will be final.

If a laboratory wishes, for whatever reason, to voluntarily relinquish accreditation for one or more tests the Laboratory Manager must provide written notice to CAEAL of its desire to do so.

Accreditation will be updated on an annual basis from the date initial accreditation was granted. Laboratories will be formally notified in writing on their anniversary date of those tests for which they have valid accreditation, via a Letter of Registration.

9.0 *Confidentiality*

All CAEAL Board members, CAEAL officials, members of the Advisory Panel and mambers of Assessment Teams are considered to be assessors and will be required to sign a confidentiality agreement that has the following elements.

1. Agreement to disclose to the CAEAL Board all involvement in personal or professional activities that would put the assessor in a position of a real or apparent conflict of interest with the performance of his/her duties as an assessor.

2. Agreement that CAEAL disclose to the applicant laboratory the assessors' involvement in such activity, that in CAEAL's opinion, represents a real or apparent conflict of interest.

3. Agreement that if a finding of real or apparent conflict of interest is made that the assessor will absent himself/herself from deliberations, of either the CAEAL Board of Directors or the Advisory Panel, which relate to the application or evaluation of the applicant laboratory.

4. Agreement to respect and safeguard the confidentiality of all information attained on an applicant laboratory including documents provided by CAEAL and any information personally observed or obtained.

5. Agreement to return to CAEAL all documents relating to the application or evaluation of an applicant laboratory.

6. Agreement to recognize that the identity of the applicant laboratory is confidential until such time as formal recognition has been granted by CAEAL.

As a further safeguard to confidentially, the CAEAL Project Manager will assign a confidential code to each applicant laboratory before performance evaluation commences. This code will be known only to the applicant laboratory and the Advisory Panel, and all communication of performance evaluation data to the CAEAL Board of Directors will be by laboratory code. Note that Advisory Panel members are normally drawn from government laboratories.

Curriculum Vitae for all assessors are kept on file by CAEAL and are available upon request.

10.0 Terms and Conditions of Accreditation

The terms and conditions of accreditation include the following:

1. Maintain CAEAL institutional membership.

2. Pay the appropriate fees as detailed in the Fee Schedule.

3. Inform CAEAL, within 30 days, of any changes in personnel, facilities, equipment or procedures, as documented in the completed Rating Guide, that may affect the ability to produce quality data for those tests for which accreditation has been granted.

4. Conform to the CAEAL Code of Ethics.

5. Conform to the publicity guidelines specified by CAEAL.

6. Analyze Performance Evaluation samples, at required intervals, as required, at the facility specified in the Application, and provide the results to CAEAL, in the manner specified, within 30 days of receipt.

7. Participate in the Site Evaluation program, at required intervals, and carry out any required corrective actions, within the times specified.

8. Undertake to make available to CAEAL, subject to mutual agreement, qualified staff that can be utilized as members of an assessment team.

9. Recognize the right of CAEAL to suspend or withdraw accreditation for any breach in the terms and conditions of accreditation or for any failure to meet the performance criteria established by CAEAL.

11.0 Publicity Guidelines

A significant benefit of CAEAL accreditation is that a laboratory may publicize its competency based on a nationally recognized accreditation program. Such activities are encouraged, and include: (i) public display of the certificate of laboratory accreditation or competence granted by CAEAL and (ii) using an approved statement on company letterhead, advertisements and test reports.

The approved Documentation Registration Publicity Statement is as follows:

> *"CAEAL recognizes that (Laboratory Name) has enroled in the initial phase of the Toxicological Testing Laboratory Accreditation Program, and has successfully submitted the documentation which describes their capability in performing (Name of Test) toxicity test(s), describes associated QA/QC practices and has enroled in the performance evaluation testing program."*

The approved accreditation statement is as follows:

> *"Accredited by the Canadian Association for Environmental Analytical Laboratories for specific tests registered with the Association".*

In all cases the name of the laboratory must appear in the immediate vicinity of the approved statement. A laboratory that is part of a larger institution may use the statement on the institutional letterhead providing that it is itself identified by name immediately preceding or following the statement.

Should accreditation be relinquished voluntarily or suspended or withdrawn by CAEAL, the laboratory shall immediately cease issuing all reference to its former accredited status.

CAEAL, for its part, will publicize laboratory accreditation by publishing a directory of accredited laboratories. The directory will include a list of the tests for which laboratories are accredited and will be available upon request.

12.0 Code of Ethics

General Principles

1. Each member shall be guided by the highest standards of ethics, personal honor, engineering or scientific integrity and professional conduct. The word "member" as used throughout this Code shall include all classes of membership.

2. Honesty, integrity, loyalty, fairness, impartiality, candour, fidelity to trust, and inviolability of confidence are incumbent upon the professional conduct of every member.

Conduct of Members in Relation to the Public

3. A member shall avoid and discourage sensational, exaggerated and unwarranted statements with regard to professional matters and shall not participate in an unsound or illegitimate undertaking.

4. A member shall not knowingly permit the publication of his/her articles or reports for an unsound or illegitimate undertaking.

5 A member shall not give a professional opinion, make a report, or give legal testimony without being as thoroughly informed as might reasonably be expected considering the purpose for which the opinion, report or testimony is desired, and the degree of completeness of the information upon which its is based should be made clear.

6. A member may publish dignified business, professional or announcement cards but shall not advertise his/her work or accomplishments in a self-laudatory or unduly conspicuous manner.

7. A member shall not knowingly issue a false statement or false information even if directed to do so by employer or client.

Conduct of Members in Relation to Employer or Client

8. A member shall protect the interest of his/her employer or client so far as it is consistent with the public welfare and his/her professional obligations and ethics.

9. A member who finds that his/her obligations to his/her employer or client conflicts with his/her professional obligations or ethics should have such objectionable conditions corrected or resign.

10. A member shall disclose to his/her prospective employer or client the existence of any interest which he/she holds, either directly or indirectly, having pertinent bearing on such employment.

11. A member shall not use, directly or indirectly, any employer's or client's confidential information in any way which is competitive, adverse or detrimental to the interests of the employer or client.

12. A member retained by one client shall not accept, without the client's consent, an engagement by another where there is likely to e a conflict of interest.

13. A member who had made an investigation for any employer or client shall not seek to profit economically from information gained, unless permission to do so is granted, or until it is clear that there can no longer be a conflict of interest with the original employer or client.

14. A member shall not divulge information provided to him/her in confidence, except when required to do so by law.

15. A member shall engage, or advise his/her employer or client to engage, and co-operate with, other experts and specialists whenever the employer's and client's interest would be best served by such service.

16. A member may shall not accept a concealed fee for referring a client or employer to a specialist or for recommending professional services other than his/her own.

17. Elected members may not seek to profit through employment with the Association however, fair and reasonable expenses shall be payable at the discretion of the Board.

18. A member shall not falsely or maliciously attempt to injure the reputation of business of another member.

19. A member shall freely give credit for work done by others to whom the credit is due and shall refrain from plagiarism in oral and written communication, and shall not knowingly accept credit rightfully due to another.

20. A member shall endeavor to co-operate with others and will encourage the ethical dissemination of useful knowledge.

21. A member of the Association shall endeavor to ensure that applicants for membership follow these standards and are otherwise qualified.

22. It shall be the duty and responsibility of every member not only to uphold these standards of ethics in precept and by example, but also, where necessary, to encourage by counsel and advice to other members their adherence to such standards.

13.0 References

Canadian Test Methods and Guidelines:

Environment Canada, 1990: Acute Lethality Tests using *Daphnia* spp. Report EPS 1/RM/11 July 1990

Environment Canada, 1990: Acute Lethality Test Using Rainbow Trout. Report EPS 1/RM/9 July 1990

Environment Canada, 1992: Growth Inhibition Test Using the Freshwater Alga *Selenastrum capricornutum* Report EPS 1/RM/25 January, 1992

Environment Canada, 1992: Test of Reproduction and Survival Using the Caldoceran *Ceriodaphnia dubia*. Report EPS 1/RM/21 February 1992

Environment Canada, 1992: Test of Larval Growth and Survival Using Fathead Minnows. Report EPS 1/RM/22 February 1992

Environment Canada, 1992: Toxicity Test Using Luminescent Bacteria (*Photobacterium phosphoreum*). Report EPS 1/RM/24

Environment Canada, 1990: Reference Methods for Determining Acute Lethality of Effluents to *Daphnia magna*. Reference Method EPS 1/RM/14, July 1990

Environment Canada, 1990: Reference Method for Determining Acute Lethality of Effluents to Rainbow Trout Report EPS 1/RM/13, July 1990

Environment Canada, 1990: Guidance Document on Control of Toxicity Test Precision Using Reference Toxicants. Report EPS 1/RM/12, August 1990.

Annex IIIb

North Carolina Biological Laboratory Certification/Criteria Procedures Document, North Carolina Department of the Environment, Health and Natural Resources, Division of Environmental Management, Water Quality Section, 1994

North Carolina

Biological Laboratory Certification/Criteria Procedures Document

North Carolina Department of Environment,
Health, and Natural Resources
Division of Environmental Management
Water Quality Section

This document has been approved for release _____

A. Preston Howard, Jr., Director

_____

Date

NORTH CAROLINA DIVISION OF ENVIRONMENTAL MANAGEMENT
BIOLOGICAL LABORATORY CERTIFICATION/CRITERIA PROCEDURES DOCUMENT

These procedures are part of the State of North Carolina's response to requirements set forth by the National Pollutant Discharge Elimination System (NPDES). This document supports the Department of Environment, Health, and Natural Resources' Administrative Code Section 15A NCAC 2H.1100. Specific laboratory facility and equipment requirements, quality assurance requirements, standard test methods/procedures, standard toxicity test reporting forms, and standard scientific reporting units pertaining to Biological Laboratory certification are described here. Procedures presented here and in subsequent versions are approved by the director before being released to the public.

METHODS AND PROCEDURES

The following documents describing NPDES test methods and procedures are recognized as standard and shall be used to measure the reporting units listed below:

(1) "Short Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms," Second Edition, EPA/600/4-89/001 or subsequent versions.

(2) "Methods for Measuring the Acute Toxicity of Effluents to Freshwater and Marine Organisms," Fourth Edition, EPA/600/4-90/027 or subsequent versions.

(3) "North Carolina *Ceriodaphnia* Chronic Effluent Bioassay Procedure," December 1985, Revised September, 1989 or subsequent versions.

(4) "North Carolina Pass/Fail Methodology for Determining Acute Toxicity in a Single Effluent Concentration," Revised September 1994 or subsequent versions.

(5) "North Carolina Phase II Chronic Whole Effluent Toxicity Test Procedure," Revised September 1994, or subsequent versions.

(6) Any other toxicological monitoring methods approved by the Director under 15 NCAC 2B.0211 or any additional methods approved and published by the Environmental Protection Agency.

(7) "Standard Operating Procedures - Biological Monitoring - Environmental Sciences Branch - Ecosystems Analysis Unit - Biological Assessment Group," February 1990 or subsequent versions.

(8) Any biological field survey analyses which either quantify or enumerate resident aquatic populations and used to evaluate attainment of Water Quality Standards as defined in 15 NCAC 2B.0211 or 15 NCAC 2B.0212.

LABORATORY FACILITY AND EQUIPMENT REQUIREMENTS

Lab facilities and equipment considered as minimum laboratory resources are as follows:

(1) Aquatic Toxicology Laboratory

(a) A minimum of 200 square feet of laboratory space.

(b) A minimum of 20 linear feet of laboratory bench space.

(c) A drained sink with hot and cold running water.

(d) Adequate control of culture environment (i.e. lighting, cooling and heating) to maintain appropriate organism requirements.

(e) A refrigerator of adequate size which will maintain sample temperatures between 0°C and 4°C.

(f) Current copies of the procedures documents written by EPA and North Carolina's Water Quality Section (see Methods section for references).

(g) Glassware, chemicals, supplies and equipment to perform any procedures included in requested certification.

(h) Instrumentation capable of measuring dissolved oxygen, pH and temperature directly from test vessels of any procedure included in certification application. Equivalent surrogate vessels may be utilized for physical measurements if injury to test organisms may result.

(i) Instrumentation or analytical capabilities to perform measurements of total residual chlorine to a level at least as low as 0.1 mg/l and total hardness to a level at least as low as 1 mg/l.

(j) A dissecting microscope with a minimum magnification of 3.5x and a compound microscope with a minimum magnification of 100x for those laboratories maintaining either of the categories of Acute Toxicity Testing/Invertebrate or Chronic Toxicity Testing/Invertebrate.

(k) A balance capable of accurately weighing fish larvae to 0.00001 g and Class "S" reference weights for those laboratories maintaining certification for the category Chronic Toxicity Testing/Vertebrate.

(l) Viable reproducing laboratory cultures of any test organisms included in the certification application. Use of test organisms for regulatory purposes not maintained as a viable laboratory culture may be accepted on a case by case basis upon receipt of written permission from the State Laboratory.

(m) Appropriate dilution water for use in whole effluent toxicity testing with chemical characteristics such that the pH is between 6.5 and 8.5 S.U. and total hardness as calcium carbonate is between 30 and 50 ppm. Should receiving waters have characteristics outside of stated ranges then alternate pH and hardness ranges may be

accepted upon demonstration that the alternate ranges are better suited to the testing objectives and quality assurance standards have been met.

(n) Appropriate Chain-of-Custody documentation blank forms, blank forms for laboratory records, and seals.

(2) Aquatic Population Survey and Analysis Laboratory

(a) A minimum of 150 square feet of laboratory space.

(b) A minimum of 8 linear feet of laboratory bench space.

(c) Binocular dissecting microscopes and compound microscopes suitable for survey type.

(d) Vials, preservatives, and space to maintain representative sample collections for at least one year after collection.

(e) Current taxonomic guides and references specified by the Division.

(f) Appropriate chain of custody documentation, laboratory records and seals are to be available.

(g) Sampling equipment to support collection of appropriate biological organisms.

(h) Settling tubes and an inverted microscope with a minimum magnification of 300x for those laboratories maintaining certification for the parameter Algae.

QUALITY ASSURANCE REQUIREMENTS

Emphasis is placed on good laboratory practices and proper documentation. Additional quality assurance requirements to those found in the previously cited documents are as follows:

(1) All instruments used in or associated with toxicity testing are to be calibrated daily or with each use and recorded in a designated notebook (i.e., automatic sampling equipment, pH meter, D.O. meter, conductivity meter, etc.)

(2) A minimum of five valid reference toxicant tests must be performed and entered on a control chart for each organism and test type for which a lab is certified. A maximum of 30 datapoints are to be entered on the control chart.

(3) A reference toxicant test should be performed every two weeks for each organism used in acute whole effluent toxicity testing, or alternatively, acute reference toxicant tests may performed such that NC NPDES acute tests are performed within one week of an acute reference toxicant test for the organism in question. In the case of the latter, to maintain acute certification for an organism, acute reference toxicant tests must be performed, at minimum, on a quarterly frequency.

(4) A reference toxicant test should be performed once per month for each organism used in chronic whole effluent toxicity testing, or alternatively, tests may be performed such that NC NPDES chronic tests are performed within two weeks of a chronic reference toxicant test for the organism in question. In the case of the latter, to maintain chronic certification

for an organism, chronic reference toxicant tests must be performed, at minimum, on a quarterly frequency.

(5) Acceptable alternative culture media utilized to culture the algae *Selenastrum capricornutum* for use as *Ceriodaphnia* food are;

(a) The MBL medium as described by Handbook of Phycological Methods: Culture Methods and Growth Measurements. 1973. J.Stein, ed. University Press, Cambridge, Mass.

(b) Additional nutrients may be used in the preparation of algae medium described in Section 12, subsection 8.2.5 of *EPA/600/4-89/001*. Specifically, the volume of nutrient stock solutions found in Table 1 on page 115 of that document may be adjusted so that solutions 1, 2, and 5 are added at a rate of 2 ml/l and solutions 3 and 4 are added at a rate of 6 ml/l.

(6) A representative of each test organism cultured shall be taxonomically identified to the species level at a minimum frequency of once per quarter. The specimen shall be preserved and held for a minimum of one year.

(7) If closed incubators are utilized for toxicity testing and/or test organism culturing purposes, culturing and testing activities may not be contained within the same incubator.

(8) Effluent samples are to be used within 72 hours of collection. The beginning of this period is defined as the time of the collection of a grab sample or the time of collection of the last subsample of a composite sample, to the time that the organisms are introduced to the test solution or the last renewal of the test solution.

PROCEDURE MODIFICATIONS

Modifications from test protocols from the cited EPA documents follow. These modifications are in addition to those specified in individual procedures documents. References to the EPA manuals are given to provide context to the modification being made to the EPA method.

(1) Acute toxicity tests using *Ceriodaphnia dubia*, *Daphnia pulex*, and *Pimephales promelas* will be conducted at 25±1°C. *(Ref. EPA/600/4-90/027. pp. 56-61, Tables 11-13.)*

(2) Organisms used in acute toxicity tests will have food made available for a minimum of two hours prior to initiation of testing. For cladoceran species, this feeding will be a minimum of 0.05 ml of YCT and 0.05 ml of a solution of the algae *Selenastrum capricornutum* (with a cell concentration of 1.71×10^7 cells/ml) per 15 milliliters of culture solution. *(Ref. EPA/600/4-90/027. pp. 56-61, Tables 11-13.)*

(3) Fathead minnows used in acute toxicity tests will be 1 to 14 days in age, and 72 hour range in age. *(Ref. EPA/600/4-90/027. pp. 60-61, Table 13.)*

(4) For each sample used in a toxicity test, pH, specific conductance, and total residual chlorine will be measured and recorded from an undiluted aliquot before use in the test. Dissolved oxygen and pH will be measured in the control and the highest toxicant concentration tested at the beginning of the test, prior to and following each renewal, and at the termination of the test. *(Ref. EPA/600/4-90/027. p. 71, 10.2.1.4.)*

(5) *Ceriodaphnia dubia* used in toxicity tests will be obtained from individual cultures, from third or subsequent broods of adults not being more than 14 days in age, containing eight or more neonates, with a average adult mortality not exceeding 20% per culture board. *(Ref. EPA/600/4-90/027. p. 138, 3.7.6.)*

(6) Chronic *Ceriodaphnia dubia* analyses will have an additional test acceptability criterion of complete third brood neonate production by 80% of the surviving control organisms. *(Ref. EPA/600/4-89/001. p. 124, Table 3.)*

(7) *Ceriodaphnia dubia* neonate reproduction totals from chronic tests shall include only organisms produced in the first through third broods.

FORMS

The forms for reporting whole effluent toxicity test results (see attachments) are as follows:

(1) AT-1 form, entitled Effluent Toxicity Report Form - Chronic Pass/Fail and Acute LC50, is used for reporting chronic pass/fail toxicity test results or acute LC_{50}s.

(2) AT-2 form, entitled Effluent Aquatic Toxicity Report Form - Acute Pass/Fail, is used for reporting acute pass/fail toxicity test results.

(3) AT-3 form, entitled Effluent Aquatic Toxicity Report Form/Phase II Chronic *Ceriodaphnia*, is used for reporting Phase II chronic toxicity test results or chronic pass/fail results.

REPORTING UNITS

The list of reporting units considered as standard are defined as:

(1) LC_{50} - The toxicant concentration killing 50% of exposed organisms at a specific time of observation.

(2) NOEC - (No Observed Effect Concentration) The highest or single concentration of toxicant to which organisms are exposed in a life cycle or partial life-cycle test, which causes no statistically significant adverse effect on the observed parameters (usually hatchability, survival, growth, and/or reproduction).

(3) LOEC- (Lowest Observed Effect Concentration) The lowest concentration of toxicant to which organisms are exposed in a life cycle or partial life cycle test, which causes a

statistically significant adverse effect on the observed parameters (usually hatchability, survival, growth, and/or reproduction).

(4) Chronic Value (ChV) - A numeric value representing the geometric mean of the numeric values of concentrations analyzed as the No Observed Effect Concentration (NOEC) and the Lowest Observed Effect (LOEC) by chronic toxicity testing. The chronic value is an estimate of the toxicant concentration that will be the actual no effect concentration based on the chronic effect tested. ChV = Antilog [Log10 LOEC + Log10 NOEC]/2.

(5) Biological Water Quality Rating - A rating, ranging from Excellent to Poor, which gives an indication of water quality based on the composition of the biological community, using standardized techniques as specified by the Division of Environmental Management.

(6) Total Taxa Richness - The total number of different taxa collected, taken to the lowest practical taxonomic level.

(7) EPT Taxa Richness - The total number of different taxa collected belonging to the orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Tricoptera (caddisflies), taken to the lowest practical taxonomic level.

(8) Diversity - The number and abundance of taxa in a specified location summarized using a mathematical formula to allow comparisons of community structure.

(9) ET(x) The relative toxicity of a toxicant measured in terms of the time it takes to elicit a given response from a given percentage (x) of the exposed test organisms.

(10) TLM - Median tolerance limit - The toxicant concentration at which 50% of test organisms survive for a specified exposure time. The term has been superseded by median lethal concentration (LC_{50}).

(11) LC(x), EC(x) - Lethal concentration (LC) or effective concentration (EC). A point estimate of the toxicant concentration that would adversely affect a given percent(x) of the test organisms.

(12) Maximum Acceptable Concentration (MATC) - Concentration to be determined within the interval bounded by the LOEC and NOEC which is used as the concentration of toxicant which has no detrimental impacts on the test population.

(13) Toxic Unit Acute - The reciprocal of the toxicant dilution that causes the acute effect by the end of the acute exposure period, for example: $1/LC_{50}$.

(14) Toxic Unit Chronic - The reciprocal of the toxicant dilution that causes no unacceptable effect on the test organisms by the end of the chronic exposure period, for example: 1/ChV.

Annex IIIc

A Register of Approved Laboratories undertaking Toxicity Testing, July 1997

A REGISTER OF APPROVED LABORATORIES UNDERTAKING TOXICITY TESTING

JULY 1997

WRc plc

R&D Note

CONTENTS Page

1. INTRODUCTION

1.1 Regulation of effluent discharges by Direct Toxicity Assessment

The EnvironmentAgency is committed to the development of effective techniques for pollution control which help industry better target investment for environmental improvements. Procedures for deriving and monitoring toxicity-based licence conditions have been developed through an extensive collaborative research programme sponsored by the Environment Agency and the Scotland and Northern Ireland Forum for Environmental Research (SNIFFER).

The Agency believes that toxicity-based techniques can be used effectively to control toxic discharges and play a key role in bringing about environmental improvement. Much of the research effort has been aimed at developing an approach which can be applied consistently across the UK using protocols which are based on sound science. The need for high quality data has been a key consideration at all times and this is reflected in the proposed test guidelines and in a series of measures aimed at promoting the quality of toxicity test data. It is this latter aspect, specifically the establishment of a Register of Approved Laboratories (RAL) to assure and control the quality of toxicity test data, which forms the subject of this document. This register will be piloted in the planned 'Demonstration Study' to investigate the application of Direct Toxicity Assessment techniques for control of effluent discharges at selected locations in the UK.

1.2 Consultation on the Register of Approved Laboratories

Initialproposals for a RAL were set out in a document and discussed at a workshop following the SETAC Conference on Direct Toxicity Assessment (DTA) at the University of Luton in July 1996. At that workshop, the overall objectives of the scheme were acknowledged but some of the specific proposals were challenged. This document builds on the earlier proposals. Specifically, it:

- Explains in more detail the need for measures to assure and promote the quality of toxicity test data;
- Revises some specific proposals, particularly those concerned with reference toxicant testing;
- Revises and provides more detail about the criteria for acceptable/unacceptable performance and the involvement of the Environment Agency in this process;
- Takes account of recent developments in the implementation of DTA as a regulatory tool, especially the establishment of a 'Demonstration Study' to evaluate the proposed measures prior to wider implementation;

- Proposes the establishment of a RAL to operate on a pilot basis as part of the 'Demonstration Study'.

1.3 Aims of this document

This document is intended to inform test laboratories of the Environment Agency's plans for a Register of Approved Laboratories and is made available for your information. There will also be an opportunity for any further explanation that may be required through the DTA Methods Workshop which has been convened by the Steering Group overseeing the planning and conduct of the 'Demonstration Study'.

2. WHY IS THE QUALITY OF TOXICITY TEST DATA AN ISSUE?

2.1 The need for reliable data

Under the proposals for the regulatory control of certain wastewater discharges by DTA, important commercial and environmental decisions will be made on the basis of ecotoxicity tests. For example, test data may determine whether or not a particular industry needs to take remedial action to reduce the toxicity of an effluent discharge. The commercial implications of such actions can be considerable, as can the environmental consequences of failing to act when it is required.

It is significant to note that, in the Environment Agency's recent consultation process, concerns about the quality and reliability of test data were amongst the issues most frequently raised by consultees (Environment Agency, 1997). Furthermore, a study tour by Agency staff to investigate US experiences with the regulatory control of effluent discharges by Whole Effluent Testing concluded that the quality of test data was one of the critical factors affecting the success of such regulatory schemes. Clearly, we must be sure that the ecotoxicity tests used for regulatory decision-making are generating valid data of the highest quality; this is a concern shared by the regulators, dischargers and the public.

2.2 Variability and bias compromise reliability

2.2.1 Some terms defined

Like any form of measurement, estimates of ecotoxicity are subject to variability. As we explain below, to be useful, ecotoxicity tests should generate data which are both sufficiently ***accurate*** and ***precise***. It is important to recognise the difference between these concepts and that they apply just as much to ecotoxicity testing as they do to a process manufacturing widgets or chemical analyses for pesticides in drinking water. These terms are defined in Table 1. In reality, we rarely actually measure accuracy and precision directly but rather, the consequences of these properties which are manifest in performance characteristics such as ***repeatability, reproducibility, within-test variability, within-laboratory variability*** or ***between-laboratory variability***. When we set out to improve, say, the repeatability or reproducibility of ecotoxicity tests, it is important that we address the root causes of these expressions of variability i.e. the precision and accuracy of toxicity measurement.

Accuracy and precision can vary independently so that, for example, it is possible to have a very precise but biased test which could lead to a false conclusion. Similarly, we may estimate the toxicity of a substance with great accuracy, but not very precisely, i.e. the average of a large number of measurements is close to the 'true' population mean but the standard deviation is large. The existence of ***bias*** (systematic error) will have a direct effect on the accuracy of measurements although it will not necessarily have any effect on their precision.

Table 1 Definitions and consequences of accuracy and precision

	PROPERTY		COMMENTS
PROPERTY OF MEASUREMENT	**Accuracy**	**Precision**	
DEFINITION	The degree of agreement between a measured value and the true value. Accuracy is a lack of **bias**	The amount of agreement between repeated measurements made under specified conditions.	Precision and accuracy can vary independently so a measurement can be accurate without being precise, and vice-versa
HOW CAN WE TELL WHEN THE PROPERTY HAS BEEN ACHIEVED?	The mean of a set of results is close to the true value	The variability around the mean of a set of results is small	When a laboratory achieves consistent measurements of toxicity using the same toxicant in repeat tests, the results are said to be precise and the method is **repeatable.** When , in addition, different laboratories achieve this state of affairs, the method can be said to be **reproducible.** Accuracy is only demonstrated when measurements are close to the expected toxicity.
REGULATORY IMPLICATIONS OF PROPERTY BEING ACHIEVED	Sources of bias are under control	Sources of error leading to variability are under control	Regulatory decisions based on measurements of toxicity are less liable to be influenced by measurement error when laboratories (a)achieve repeatable results, (b) these are reproducible across laboratories and (c) they are free of excessive bias

2.2.2 The implications of variability

(a) Deriving 'action levels' for toxicity

Bias between laboratories becomes important when effluent toxicity is being investigated because it can affect whether or not the criteria for acceptable toxicity are derived equitably between locations. This can be illustrated by an example in which the environmental impacts of truly identical discharges (with respect to toxicity, dilution in the receiving water etc.) are investigated. If the toxicity evaluation of the effluents is conducted by different laboratories which typically differ in the sensitivity of the test methods used (because of systematic error perhaps), it is quite possible that the observed toxicities of the two discharges will differ. As a result, one discharger could be required to embark on a programme of effluent toxicity reduction whilst the other is not.

(b) Monitoring effluent toxicity

If we were to carry out repeated tests on identical samples, a range of toxicity determinations will result. This reflects the random error present in any form of measurement. However, the variability involved in the measurement of effluent toxicity during monitoring means that some effluent samples will be wrongly classified, especially if variability is substantial. The sample may exceed the 'action level' when, in fact, its true toxicity is actually acceptable (Type I error or 'false positive'). Alternatively, the sample might not exceed the 'action level' but its true toxicity has been underestimated and toxicity actually exceeds the limit of acceptance (Type II error or 'false negative').

Clearly, both Type I and Type II errors should be reduced as far as practicable, the first because of misdirected expenditure, and the second because of environmental impacts which may go undetected. Thus both dischargers and regulators have a keen interest in ensuring that such errors are kept within bounds. It is for these reasons that measures are proposed for:

a) Assuring the integrity of the test data

b) Ensuring that tests are performed consistently

c) Minimising bias and variability within laboratories

Each of these is covered in detail in Sections 3, 4 and 5, respectively.

3. QUALITY ASSURANCE

3.1 Introduction

Robust regulatory decisions about the assessment and monitoring of effluent toxicity depend on tests being conducted with a high level of integrity. In other schemes where submission of test data to a regulatory body is required, e.g. for safety testing of new products or monitoring of drinking water quality, assurance of fully valid and auditable data is normally provided through compliance with the requirements of a Quality System such as Good Laboratory Practice (GLP) or accreditation by the United Kingdom Accreditation Service (UKAS) under NAMAS, which is the UK equivalent of EN45001 and ISO Guide 25.

Generally speaking, biological test laboratories have followed the GLP route although we are aware that some laboratories carrying out toxicity testing are also seeking accreditation under NAMAS. It is clear that, in order to provide adequate assurance of the integrity of test data, regulatory effluent testing should also comply with a Quality System.

3.2 Requirements to join the RAL

It is not the intention in this document to describe the requirements of GLP compliance or accreditation under NAMAS, or to discuss the differences between them. The Environment Agency takes the view that the Quality Assurance infrastructure required by both schemes is adequate to meet its objective of assuring the integrity of toxicity test data. The Agency does not believe it is cost-effective or desirable to establish a separate Quality System specifically for toxicity testing of effluents. Proof of compliance/accreditation with either Quality System is acceptable for joining the RAL.

Discussions with both UKAS and the UK GLP Monitoring Authority have been held and both accept the Environment Agency's position. Although GLP compliance is predominantly required for safety testing of products intended for registration, compliance with GLP for effluent toxicity testing is nevertheless considered acceptable to the UK GLP Monitoring Authority.

3.3 Inspection for compliance/accreditation

The UK GLP Monitoring Authority and UKAS will be provided with controlled copies of the standardised methodology (currently the 'Ecotoxicology Methods Guidelines') and guidance notes on particularly important aspects such as the procedures for external Quality Control. This information would form the basis on which the laboratory inspection will be conducted. The inspections/surveillance visits would check that the necessary infrastructure is in place for the conduct of testing and that procedures are consistent with those described in thestandard methodology. The results of any inspections would not be communicated to the Environment Agency but a laboratory would be expected to inform the Agency (specifically the National

Centre for Ecotoxicology and Hazardous Substances: NCEHS) of the outcome of inspections/surveillance visits.

One of the requirements for accreditation under NAMAS is that the test laboratory should participate in an external Quality Control scheme. The scheme described in Section 5 would meet that particular requirement.

3.4 Conduct and reporting of effluent toxicity tests

Effluent toxicity tests which are intended for regulatory decision-making (i.e. testing to characterise the toxicity of a discharge or monitoring data) should conform to the requirements of either NAMAS or GLP, in terms of the planning, authorisation, conduct, reporting and archiving of the studies. The use of pro-forma reports is encouraged and examples are included in the 'Ecotoxicology Methods Guidelines'. This pro-forma may be used alone or appended to the laboratory's regular test report.

It is recognised that some of the Principles of Good Laboratory Practice cannot reasonably be met in the case of effluent toxicity tests. For example, it is inappropriate to undertake chemical analysis to characterise effluent samples. Indeed, one of the main reasons for adopting the DTA approach to wastewater control is because of the chemical complexity and variability of many effluents. Nor is it necessary to confirm test exposure concentrations or to take steps to control the stability of the effluent during exposure other than any stipulated in the 'Ecotoxicology Methods Guidelines'. Archiving of effluent samples used as test substances will not be required, nor is it necessary to reconcile the usage of an effluent sample at the end of a study. It would not be necessary to declare such unavoidable non-compliances with the Principles of Good Laboratory Practice in the test report.

4. CONSISTENCY IN METHODOLOGY

4.1 Introduction

The need for consistency in the way toxicity data are generated has led to the development of standardised methodology (currently the 'Ecotoxicology Methods Guidelines') which provides detailed guidance on all aspects of testing, including those which influence the quality and consistency of test data. Therefore, the RAL attaches great importance to the conduct of tests according to Standard Operating Procedures (SOPs) which are based on this methodology.

4.2 The standardised methodology (currently the 'Ecotoxicology Methods Guidelines')

The 'Ecotoxicology Methods Guidelines' provide guidance on the following:

- Collection, storage and preparation of samples for testing
- Disposal of samples
- Sample custody records
- Cleaning and storage of apparatus
- Quality Control requirements
- Sources and culturing of test organisms
- Test methods, including data analysis and test validity criteria
- Reporting formats

In most cases, the methods described are derived from standard international guidelines e.g. OECD, ISO guidelines. The guidance given is sufficiently detailed to design Standard Operating Procedures based on these guidelines. In many cases, acceptable SOPs may already exist with the possible exception of procedures to meet the Quality Control requirements of the RAL.

Controlled copies of the guidelines have been circulated for comment and will be updated accordingly. The Environment Agency has yet to decide on the final 'vehicle(s)' for documenting this standardised methodology. This documentation will be made available to any laboratory wishing to join the RAL. Its contents will be updated at intervals to reflect, for example, the development of new test methods, revisions to existing test methods and to Quality Control criteria. Issue of the standardised methodology and subsequent revisions will be controlled by the Environment Agency's National Centre for Ecotoxicology and Hazardous Substances (NCEHS).

5. QUALITY CONTROL

5.1 Aims

The following proposals have been developed to constrain variability and bias in the results obtained from ecotoxicity tests, and thereby help promote a 'level playing field' and to encourage consistency in the conduct of tests. In addition, information about the extent of measurement error will enable dischargers and the Environment Agency to take this into account in decisions about the acceptability (or otherwise) of the toxicity of particular effluent discharges. We are therefore concerned with both the precision and accuracy of ecotoxicity test measurements.

Quality Control criteria for precision and accuracy have been derived, expressed as ranges within which we would expect acceptable results for specific chemicals to fall. The next step is to monitor performance to ensure that laboratories are generating toxicity test data which fall within those limits. The questions being asked are:

- Is the range of toxicity measurements sufficiently narrow (are they sufficiently precise)?
- Do individual measurements of toxicity fall within acceptable bounds (is there adequate control over bias)?

The use of Quality Control measures in chemical analysis has been established for some time but suitable, technically-based procedures that can be applied to ecotoxicity testing are currently lacking. We are not suggesting that test data generated currently are poor but rather we are seeking to ensure that regulatory decisions are soundly based and to reassure dischargers and the public of this fact. Currently, laboratories are rarely able to demonstrate that test data are being generated with sufficiently small bias and control over variability because a suitable mechanism is not in place.

5.2 Operation of the Quality Control Scheme

5.2.1 Overview

Unlike Quality Control procedures for analytical chemistry, it is not possible to run an internal standard in ecotoxicity tests or to calibrate an ecotoxicity test. The only practical approach is to carry out parallel tests with a reference toxicant and to use the information generated to indicate the degree of control over variability and bias. But how do we determine what is acceptable? External Quality Control criteria for precision and bias have been derived from data generated during an interlaboratory ring-test carried out in 1995 (Whitehouse *et al*, 1996). Assessment against these Quality Control criteria is judged from EC_{50} measurements obtained from the results of 'reference toxicant' tests as summarised in Table 2 and as illustrated in Figure 1.

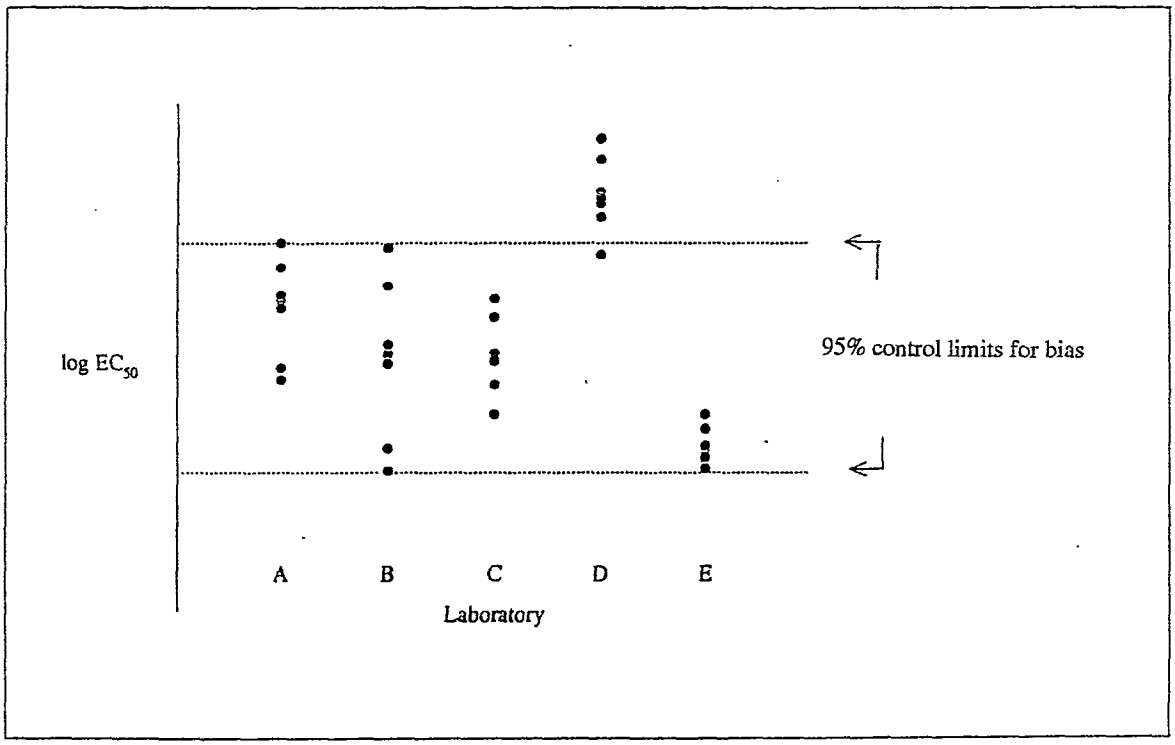


Figure 1 Assessing the precision and accuracy of EC50 measurements derived from reference toxicant tests (hypothetical data)

Explanation: The points represent individual EC50 measurements. Data from laboratories A and C indicate good precision and control over bias. Laboratory E also achieved good precision and there is a suggestion of bias, but not sufficient to exceed the control limits. Although data from laboartory B would be acceptable in terms of performance against the criteria for bias, precision is poor. Conversely, laboratory D has generated data which is precise but shows strong evidence of bias.

Table 2 **Summary of procedures for assessing the precision and accuracy of ecotoxicity measurements**

	PROPERTY		COMMENTS
	Accuracy	**Precision**	
MEANS OF DETERMINING PROPERTY	Comparison of individual EC/LC50 measurements with expected EC/LC50 to provide an estimate of bias	Comparison of range of EC/LC50 measurements with acceptable range of measurements	Performance against accuracy and precision criteria can be judged from repeated tests with a specified reference toxicant.
CRITERIA FOR ASSESSING PROPERTY	Expected EC/LC50 is based on 'consensus' mean determined from interlaboratory ring-test with specified toxicant. Acceptable measurements fall within 95% confidence limits around 'consensus' mean	'Critical' variance for EC/LC50 measurements from repeated tests with a specified reference toxicant is determined from interlaboratory ring-test.	Performance against targets for precision is judged from several, consecutive measurements whilstaccuracy is assured only if the results also demonstrate adequate control over bias. Both criteria must be met.

Assessing precision

As noted earlier in Table 1, precision is manifest in the repeatability of measurements achieved by a laboratory and so, to monitor precision, it is obviously necessary to carry out repeat tests with the reference toxicant. The number of these tests should be sufficient to permit a valid assessment of repeatability (i.e. the variance is reasonably stable) but should be as low as possible to keep costs to a minimum. The recommendations described below have been made with these considerations in mind.

Assessing accuracy

The degree of accuracy achieved is indicated by the bias exhibited by measurements from the 'true' value for that substance/test method. Therefore, some means of estimating the 'true' toxicity of a substance is necessary. This is relatively simple to achieve in chemical analysis because the response of an analytical instrument can be related to a known concentration of the substance of interest. However, we never actually know the 'true' toxicity of a substance because there is no external reference point, i.e. it is defined by measurement. Nevertheless, the 'true' toxicity of a substance can be approximated from the 'consensus' mean obtained during ring-testing. The accuracy with which tests are performed can be judged by comparing individual EC_{50} measurements with the criteria for bias.

Choice of reference toxicant

The variability exhibited by a test can depend on the substance whose toxicity is being determined. Therefore, different Quality Control criteria result from the use of different

substances (Whitehouse *et al*, 1996). Because of this, test laboratories joining the RAL are obliged to use the same substance in the reference toxicant tests as the one from which the criteria have been developed. The Agency accepts that this may result in a change in procedure for some laboratories who already carry out reference toxicant tests to monitor repeatability but consistency across laboratories is essential if the objectives of the scheme are to be met. At present, suitable data are available for two reference toxicants; 3,4-dichloroaniline and zinc sulphate, but zinc sulphate is favoured because of the lower analytical costs, ease of preparing aqueous solutions and greater sensitivity to deviations in important aspects of test conduct such as pH and hardness of the dilution water. Although the speciation and bioavailability of zinc can become complex at extreme pHs, its bioavailability is not affected within the range of pH values usually stipulated in toxicity tests.

5.2.2 Quality Control criteria

For assessments of precision, the variance of EC_{50} values from repeated tests within a laboratory with a reference toxicant (S^2) is compared with the 'underlying' variance ($\sigma^2_{repeatability}$), based on results obtained from a programme of ring-testing. Table 3 shows how $\sigma^2_{repeatability}$ for three test methods has been compared with S^2 estimated from between 7 and 10 repeat tests using the same method and toxicant. The criteria for bias are expressed as control limits around a 'consensus mean' and examples are shown in Table 4. For more information about how these Quality Control criteria are derived, the reader is referred to Whitehouse *et al* (1996).

Table 3 Assessment of precision for EC_{50} values obtained when zinc sulphate is used as the reference toxicant (Repeatability is acceptable if: $S^2/\sigma^2_{repeatability}$ (n-1) < $\chi^2_{(n-1,\alpha)}$)

Method	df (n-1)	S^2	$\sigma^2_{repeatability}$	test statistic ($S^2/\sigma^2_{repeatability}$. (n-1))	critical $\chi^2_{0.05, n-1}$	p value*
Microtox™	9	0.055	0.007	66.50	16.92	0.000
Daphnia, 48h acute	6	0.089	0.033	8.93	12.59	0.177
Pacific oyster, 24h embryo development	9	0.041	0.060	11.11	16.92	0.268

*p values < 0.05 indicate a significant difference in variances with a probability of >95%
*p values < 0.01 indicate a significant difference in variances with a probability of >99%

Explanation: in this example, the calculated ratios of $S^2/\sigma^2_{repeatability}$ for the Daphnia and oyster tests were less than the critical χ^2 value and so these test results display acceptable precision. The Microtox™ results, by comparison, indicate unacceptable precision and failed the criteria for precision with a high probability.

Table 4 Quality Control criteria for accuracy (expressed as EC_{50} limits for zinc when zinc sulphate is used as the reference toxicant)

Test method	*'Consensus' mean EC_{50}*	*95% control limits**	
	(mgl^{-1} zinc)	*upper*	*lower*
Microtox™	0.3	0.62	0.13
Daphnia, 48h acute	1.03	3.51	0.20
Pacific oyster, 24h embryo development	0.12	0.55	0.01

*derived as 'consensus' mean of $\log x \pm 1.96 \times \sigma_{reproducibility}$ of $\log x$

Updating Quality Control criteria

In some cases, the current Quality Control criteria provide a rather wide range of acceptable EC_{50} values and there is some concern that they may not, therefore, be an effective driver of improvement in precision or accuracy. This is a consequence of a relatively small dataset and the relatively high variation in the data provided by the laboratories involved in the initial ring-testing. It is likely that, as more reference toxicant data are generated, recalculation of the criteria will result in more stringent limits. Any revisions to criteria would be issued as a controlled document to laboratories on the RAL.

It may also be possible to revisit the ring-test data to identify laboratories which appear to achieve particularly good precision (as judged by low within-test and between-test variability) and to calculate a second set of criteria based on their data, reflecting 'best practice'. These criteria would not be used for judging the acceptability of performance in reference toxicant tests but could be useful in providing an indication of the precision and accuracy that is possible. In time, we would expect the Quality Control criteria derived from reference toxicant data will narrow, perhaps eventually approaching the limits representing 'best practice'.

5.2.3 Conduct and reporting of 'reference toxicant' tests

Laboratories would be required to conduct tests to generate reference toxicant data from each test method for which they wish to be registered. These will provide the data on which precision and accuracy are assessed. The reference toxicant will be zinc sulphate, supplied as a stock solution by the NCEHS from which test solutions can be prepared.

Reference toxicant tests must be documented in full using a pro-forma and, in addition, failed tests and tests which fail the Quality Control criteria must also be documented. The reference

toxicant tests should also be recorded in the appropriate test control documentation system (e.g. Master Schedule or sample handling register) which will be available for inspection.

To join the RAL, a laboratory will need to generate data from five separate tests using a specified reference toxicant which all meet the Quality Control criteria. These test data are submitted to the NCEHS who would then register that laboratory for the tests for which acceptable data have been submitted. Once the laboratory is registered and is carrying out regulatory testing, reference toxicant tests should be performed on each occasion effluent tests are performed for regulatory purposes. If several effluent samples are tested simultaneously, only one reference toxicant test need be performed. Reference toxicant tests should be carried out on at least five occasions in a year, using different batches of test organisms, dilution waters and reagents. The number and choice of concentrations used in these reference toxicant tests is left to the discretion of the test laboratory but the test should be designed in such a way that an EC/LC_{50} with 95% confidence limits may be calculated. Because the Quality Control criteria are based on EC/LC_{50}s, it is not suficient to carry out reference toxicant tests using a single exposure concentration (a 'positive control').

Further details of test methodology to generate the necessary Quality Control data are provided by the standardised methodology (currently the 'Ecotoxicology Methods Guidelines') but the testing requirements are summarised here:

1. The reference toxicant would be specified (currently zinc sulphate heptahydrate);

2. A range of test concentrations would be required but the number and choice is left to the discretion of the test laboratory;

3. The number of test organisms at each exposure concentration would be specified;

4. Exposure concentrations of reference toxicant need not be confirmed by chemical analysis, although this would be encouraged in the event of an extreme result. EC50 estimates would therefore normally be based on nominal concentrations.

5. Calculations of the specified end-point should be performed using an approved method (as detailed in the'Ecotoxicology Methods Guidelines').

5.2.4 Responsibility for assessing laboratory performance

Laboratory performance against the Quality Control criteria will be monitored by the laboratory's own QA function, according to whichever Quality System is in place. There is no need, therefore, to routinely submit all results of reference toxicant tests to the Environment Agency; only when there appears to be a problem of complying with the Quality Control criteria is it necessary to inform the NCEHS. A DTA Methods Working Group, set up with representation from the regulators, industry and contract laboratories, will oversee the operation of the Quality Control Scheme. The group will assist with:

- Identifying the need for action in the event of unacceptable performance

- Updating and publicising changes to Quality Control criteria

- Monitoring and recommending improvements to the operation of the Quality Control scheme

- Collating information on factors which may affect accuracy and repeatability of test data and making such information available to laboratories joining the RAL

- Producing a short annual report on the operation of the external Quality Control Scheme.

The Methods Working Group will work to ensure that the Quality Control scheme operates in a fair and even-handed fashion and to identify ways in which it can be improved. It will not be involved in assessing the performance of laboratories against the Quality Control criteria; that role rests with the laboratory and, where there are problems, with the NCEHS. However, the NCEHS may ask the Working Group to advise on what action should be taken in the event of serious and persistent poor performance, as described below.

5.2.5 Judging performance against Quality Control criteria

As with any other regulatory submission, significant deviations from test validity criteria (e.g. temperature, water quality, control responses) may invalidate a test. This would not in itself lead to suspension from regulatory testing or removal from the RAL unless it was clear that the laboratory was consistently unable to meet basic test validity criteria.

Laboratories can monitor accuracy continuously so that they may respond to any indications of unacceptable variability or bias quickly. Assessments of precision are carried out after every five reference toxicant testsbut bias may be assessed after each reference toxicant test. Laboratories would monitor their own performance against the Quality Control criteria and to consult with the NCEHS in the event of any concerns. Steps taken by a laboratory to resolve Quality Control difficulties at an early stage are encouraged. The NCEHS will exercise a degree of judgement about appropriate courses of action in the event of poor performance, especially during the piloting of the scheme in the 'Demonstration Programme', where the practical consequences of the scheme can be assessed and refined as necessary.

Excessive variability in repeat tests or deviations outside the range of expected EC_{50} for the reference toxicant may indicate inadequate control over precision and bias. However, the assessment criteria are all probability-based (e.g. 95% control limits) and so some measurements will fall outside the Quality Control criteria by chance alone (for example, 1 in 20 of acceptable results can be expected to fall outside the 95%-ile criteria). This could lead to false positives and unecessary remedial action by test laboratories. Therefore, in the performance assessment procedure (Figure 2) we have sought to strike a balance between the desire to promote quality data whilst not requiring laboratories to take action without good cause. Basically, a tiered approach is proposed in which more demanding remedial action is expected for exceedances about which we can be more confident. Assessments of reference toxicant data for precision and accuracy are broadly similar but there are some important differences, and these are detailed below.

Precision

Assessments of precision necessarily involve a number of EC_{50} estimates. Judgements of precision must be retrospective because it is necessary to first accumulate a number of measurements. It is inappropriate to monitor precision on a 'rolling' basis (i.e. after every new reference toxicant test) because the estimate of S^2 will not be independent of the previous one. Practically, even a large increase in random error could be masked by a previously good performance against the precision criteria. 'Warning' and 'action' limits can be introduced, based on the use of different significance factors in the calculations of the control limits for critical S^2 value (see Table 3 for examples). If we wished to reduce the chances of falsely declaring a laboratory's performance to be unnacceptable, then more relaxed failure criteria might be applied (e.g. p=0.01 instead of p=0.05) although the chances of Type II errors (failing to act when it is warranted) would increase accordingly. The chances of exceeding the 95%-ile criteria (p=0.05) by chance alone are, by definition, 1 in 20, compared to a probability of 1 in 100 for the 99%-ile criteria (p=0.01). Thus, exceeding the repeatability criteria at p=0.05 would serve as a warning of possible excessive error and the comparison at p=0.01 is the threshold at which the process is judged to be inadequately controlled and remedial action is required. This tiered approach is consistent with the 'warning' and 'action' limits associated with the ± 2 and ± 3 x standard deviations thresholds conventionally applied to Shewhart control charts.

Accuracy

Assessments of bias will be based on: (a) the magnitude of individual deviations from the acceptable range and (b) the frequency with which exceedances occur.

(a) As for the assessments of precision, we can reduce the chances of falsely declaring a laboratory's performance to be unnacceptable by using more relaxed failure criteria (e.g. p=0.01 instead of p=0.05). In other words, remedial action is required only when a measured value lies a long way outside the acceptable range. Whilst this approach permits judgements on single EC_{50} measurements from reference toxicant tests, it has the disadvantage that it could penalise a laboratory whose fundamental control over accuracy is very good but which had generated an isolated, abnormal EC_{50} value. For this reason, approach (b) has some merits.

(b) Judgements are based on the frequency with which measured EC_{50} values exceed the criteria for accuracy. For example, a single observation exceeding the 95%-ile criteria could occur by chance with a probability of 1 in 20 but two consecutive results lying outside would occur by chance only on 1 in 400 occasions, and would therefore be regarded as a cause for remedial action. When both EC_{50} estimates lay on the same side of the 'consensus' mean, this would suggest unacceptable bias. Consecutive values on opposite sides of the mean would be evidence for excessive random error, although this should also be reflected by unacceptable precision.

These options are not mutually exclusive and will be applied in concert. They place varying demands on the number of measured EC_{50} values required but address rather different aspects of performance. In summary, the NCEHS should be informed in the event of a single

exceedance of the 99%-ile criteria or two consecutive exceedances of the 95%-ile criteria, as indicated in Figure 2.

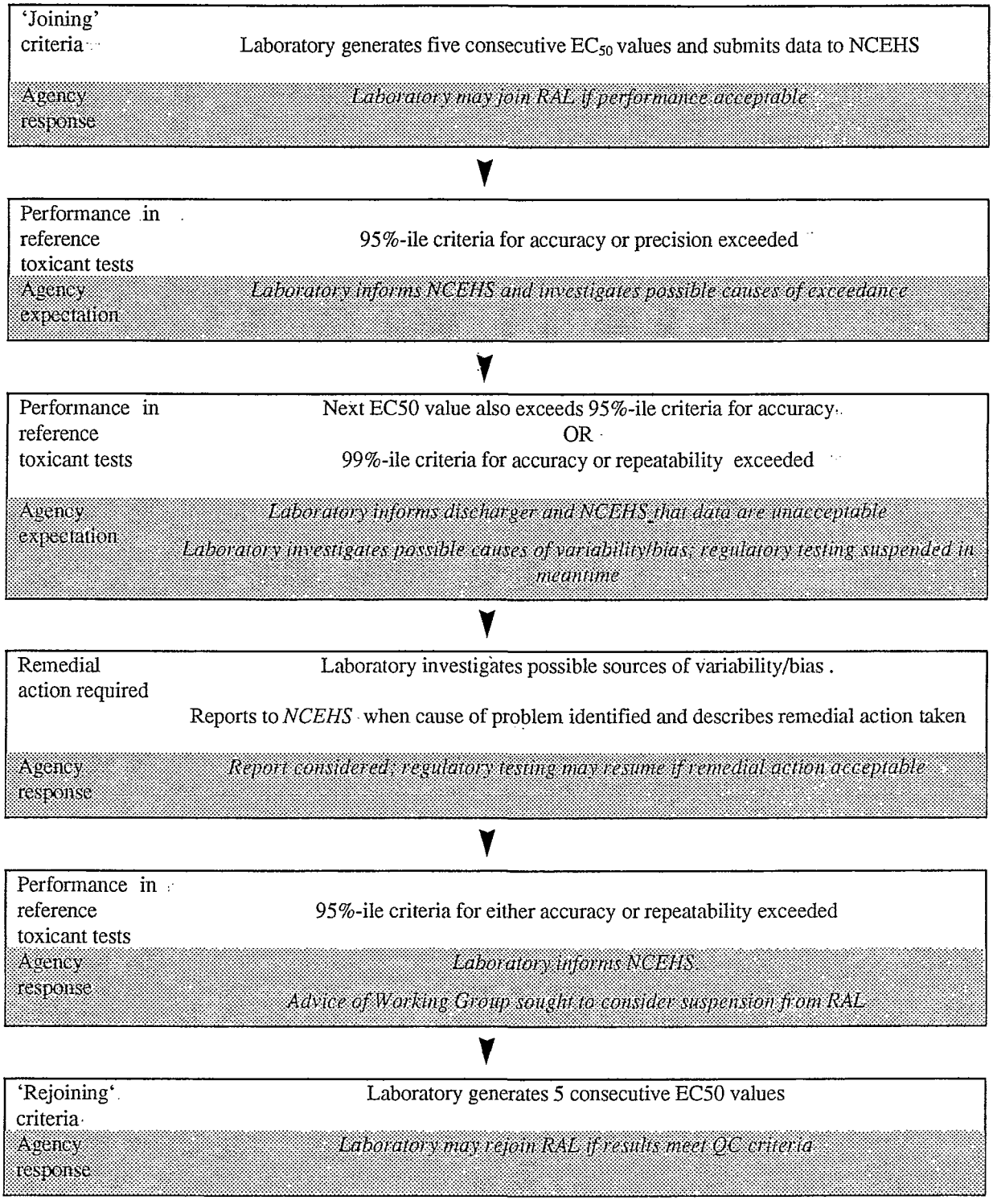


Figure 2 Proposed assessment of laboratory performance against external Quality Control criteria

6. ORGANISATIONAL ASPECTS

Several organisations and groups are involved in the RAL and in the generation of data intended for regulatory decision-making. These include the testing laboratories, dischargers, the Environment Agency (specifically, the NCEHS), the Methods Working Group responsible for overseeing the operation of the Quality Control scheme, and Inspectorates responsible for monitoring compliance with GLP or NAMAS (UK GLP Monitoring Authority and UKAS, respectively).

It is important to be aware that registration as an 'approved laboratory' is a quite separate activity from the submission of effluent toxicity test data for regulatory consideration. Indeed, a laboratory could join the RAL and never carry out any effluent toxicity tests. Dischargers undertaking toxicity testing of their effluent discharges will be provided with an up-to-date copy of the RAL by the NCEHS. The results of effluent toxicity tests are transferred to the discharger who would make them available to the Environment Agency, if required. The GLP/NAMAS Inspectorates are not directly involved in issues concerned with the regulation of effluent discharges by DTA.

7. PROCEDURE FOR REGISTRATION

7.1 GLP/NAMAS accreditation

Laboratories wishing to join the RAL should notify the NCEHS directly, copying this to either the GLP Monitoring Authority or UKAS. This applies to laboratories seeking to join the UK GLP Compliance Programme or gaining accreditation under NAMAS for the first time and also to laboratories which are already compliant/accredited. On receipt of this notification, the NCEHS will provide a copy of the standardised methodology and the laboratory should become proficient in the methods for which it is seeking registration, demonstrated by generating acceptable reference toxicant data. At the same time, it may take measures to set up the necessary Quality System infrastructure. For laboratories which are not currently GLP-compliant or NAMAS-accredited, this could be a major undertaking but for other laboratories, only modest changes to sample handling procedures, equipment, ensuring familiarity by QA staff, training and SOPs should be required.

An inspection by either the GLP Monitoring Authority or UKAS will follow. It is recognised that some laboratories may already be monitored under the UK GLP Monitoring Scheme and that some may already be accredited under NAMAS for certain test procedures. In these situations a limited inspection, confined to the procedures required for effluent toxicity testing, is possible. Alternatively, inspection of the 'RAL-specific' elements will be included in the next scheduled inspection. This is still to be resolved with the relevant Inspectorates.

7.2 Submission of Quality Control data

Following a successful GLP or NAMAS inspection, the laboratory is asked to submit a copy of its certificate of compliance/accreditation to the NCEHS along with the Quality Control data (Section 5.2). Because the Quality Control requirements are specific to each test method, it follows that registration will be applied on a test-by-test basis, i.e. a laboratory may be approved for some methods but not others. If acceptable performance is demonstrated, then the laboratory is added to the RAL with immediate effect and a letter to that effect, listing the test methods for which the laboratory is approved would be sent to the laboratory by the NCEHS. This information would also be recorded in the RAL.

8. SUMMARY

8.1 Aims of the RAL

Ecotoxicity tests used for regulatory decision-making should generate valid data of the highest quality. This is a concern shared by regulators, dischargers and the public and the Environment Agency is committed to ensuring that the quality of data will not be an issue in discussions about regulation of effluent discharges by DTA.

The procedures described are consistent with other regulatory schemes in placing high priority on the validity and auditability of test data. Measures to ensure consistency in the methods used for testing have also been introduced. In addition, the RAL places emphasis on measures to constrain variability and bias, in recognition of their effects on regulatory decisions based on effluent toxicity.

8.2 Under what circumstances is registration required?

Dischargers are required to ensure that toxicity test data which are to be submitted to the Environment Agency are generated wholly within approved laboratories, whether they are independent contractors, regulatory laboratories or within the dischargers' own organisation. Other data are not acceptable in regulatory decisions about the toxicity of effluents or monitoring against 'action levels'. Studies on effluent toxicity which are not intended for regulatory decision-making, e.g. 'Toxicity Reduction Evaluations' or 'Toxicity Identification Evaluations' need not be performed by an approved laboratory and are not subject to the requirements set out in this document. The RAL is concerned only with ecotoxicity testing of samples; procedures involving flow measurement, chemical analysis of individually licenced chemical constituents, and any other non-toxicological requirements of licence monitoring are not considered here. The RAL will be maintained by the Environment Agency and a list of members made available to dischargers who are required to investigate toxicity of their effluent discharges or to monitor toxicity against an agreed 'action level'.

8.3 Requirements to register as an approved laboratory

Essentially, there are three distinct sets of requirements which, if met, allow a test laboratory to register as an approved laboratory.

8.3.1 Quality System

Formal compliance with GLP or accreditation under NAMAS is essential to provide the Environment Agency and dischargers with the necessary assurance of high quality and auditable data. Laboratories wishing to register would therefore be subject to inspection/monitoring by the relevant Inspectorate (GLP Monitoring Authority or UKAS). Those inspections would take into account specific methodological requirements for effluent

testing (such as sample handling requirements and Quality Control procedures) described in the 'Ecotoxicology Methods Guidelines'.

8.3.2 Methodology

Laboratories must perform tests according to an approved methodology. This has resulted in the documentation of standard methods for ecotoxicity testing of effluent samples intended for regulatory purposes, currently described in an 'Ecotoxicology Methods Guidelines'. Laboratory SOPs should be consistent with the requirements of these guidelines.

8.3.3 External Quality Control

A key aspect of the RAL is the establishment of an external Quality Control scheme. This specifies certain performance criteria for joining, and remaining within, the RAL, and also criteria for exclusion from the RAL in the event of unacceptable control of variability or bias. Administration of this scheme will be by the National Centre for Ecotoxicology and Hazardous Substances of the Environment Agency but this will be overseen by the DTA Methods Working Group. Registration places certain responsibilities on test laboratories to conduct regular reference toxicity tests and to take appropriate action in the event of unacceptable performance against the Quality Control criteria.

8.4 Pilot phase of the RAL

It is intended that the RAL should run on a pilot basis as part of the planned 'Demonstration Study' to investigate wider aspects of DTA for regulatory control of effluent discharges. It is the Environment Agency's intention that all regulatory data collected in this 'Demonstration Study' should be from approved laboratories only. This pilot phase will be an opportunity to implement the proposals described in this document and it is likely that some aspects will change in the light of the practical experience gained. During this pilot phase, there will be no charge levied for joining the RAL.

9. REFERENCES

DoH (1989) Good Laboratory Practice. The United Kingdom Compliance Programme. Department of Health, London.

Environment Agency (1997) The application of toxicity-based criteria for the regulatory control of wastewater discharges. Response Compendium, Document SO/3/97/1000/B/AYAM

NAMAS Executive (1991) NAMAS Accreditation Standard. General Criteria of Competence for Calibration and Testing Laboratories. M10. National Physical Laboratory, Teddington. EPA Acute Manual

NAMAS Executive (1993) Supplement to NAMAS Accreditation Standard. Measurement and Calibration Systems. M10 Supplement. National Physical Laboratory, Teddington.

Whitehouse P., VanDijk P.A.H., Delaney P., Roddie B.D., Redshaw C.J. and Turner C. The precision of aquatic toxicity tests: its implications for the control of effluents by Direct Toxicity Assessment. pp 44-53 in: 'Toxic Impacts of Wastes on the Aquatic Environment' eds. J.F. Tapp, J.R. Wharfe and S.M. Hunt, RSC, 1996.

Annex IIId

Options for Trialling Quality Control procedures for Ecotoxicity Tests, WRc 1998

OPTIONS FOR TRIALLING QUALITY CONTROL PROCEDURES FOR ECOTOXICITY TESTS

Paul Whitehouse, WRc Medmenham

Introduction

It is now widely recognised that, in order to achieve sound regulatory decision-making in a variety of fields, the data on which those decisions are based should be reliable and free of excessive bias and variability. All methods of measurement are liable to these shortcomings but a number of Quality Control (QC) schemes have been developed to promote the performance of chemical analysis and more recently, for microbiological testing. With the advent of ecotoxicity testing as a regulatory tool for wastewater control, concerns have also been raised about the consequences of variability in the results from these tests.

It is quite possible that false conclusions about the toxicity of a discharge could result because tests carried out by a laboratory are particularly sensitive or insensitive (bias). The former could lead to unnecessarily requiring toxicity reduction (false positive), possibly entailing significant costs to industry. The latter can result in not asking for toxicity reduction when it is actually warranted (false negative) with the result that environmental damage follows. Similarly, variability between laboratories, and even within laboratories, can lead to ambiguous or false conclusions about whether or not a discharge is achieving the toxicity criteria which are required. With a national scheme, it is all the more important that DTA controls are implemented equitably between locations and so variability between laboratories assumes a particularly high importance. Unfortunately, there are a number of examples where laboratory bias has resulted in false conclusions, particularly in the area of product hazard and risk assessment. Decisions about toxicity assessment and the need for toxicity reduction are equally liable to this problem.

Measures to standardize methodology go a long way to removing the most extreme cases of bias and variability but accurate and precise estimates of ecotoxicity do not automatically follow. As with chemical analysis and microbiological analysis, there is a strong case for defining performance criteria ('targets') for accuracy and precision, and monitoring performance against those criteria. This has been recognised in Canada and certain US states where laboratory certification is a prominent feature of the regulatory infrastructure for wastewater control by Whole Effluent Toxicity. Respondents to the DTA consultation document in 1996 also highlighted concerns about test variability and its implications for wastewater regulation.

Detailed proposals for an external QC scheme have been developed by WRc on behalf of the Environment Agency ('A Register of Approved Laboratories involved in Ecotoxicity Testing') and debated with test laboratories who would be directly affected. Whilst the need for measures to constrain bias and variability is agreed, the perceived implications to

laboratories (cost and regulation) associated with proposed external QC measures remain a cause for concern. A demonstration of the benefits offered by different options and the associated costs is therefore required. A suitable approach would be one that:

- provides reliable monitoring of the performance of effluent toxicity testing in laboratories
- raises the standard of testing
- incurs costs which are in proportion to the benefits

Scope of this note

The need for standardization in methodology to avoid gross differences in conclusions about effluent toxicity between locations is now accepted by all those involved in the implementation of DTA. To this end, detailed 'Methods Guidelines' have been produced. The need to assure the integrity of data is also generally accepted and to achieve this objective, laboratories are expected to become accredited under NAMAS or to be compliant with GLP. The need for and approaches to achieve standardization and data integrity are taken as read and will be considered no further.

In contrast, the approaches to QC remain contentious. At a meeting of the DTA Methods Working Group in September, the Agency proposals were discussed alongside possible refinements of the proposals and other, rather distinct approaches to QC. It was felt that the forthcoming DTA Demonstration Study provided an ideal opportunity to carry out a proper and open evaluation of the possible options, leading to a recommendation for the QC of regulatory ecotoxicity testing in the future. This note:

- sets out the options that were identified at the September DTA Working Group meeting
- identifies the strengths and limitations of each option
- identifies the costs (incurred by the Steering Group in financing the organisation and analysis of different options, and by test labs in carrying out additional testing) of trialling these options the Demonstration Programme

Description of Options

1. Concentration-response tests with reference toxicants

This is the approach currently proposed by the Agency. Essentially, it requires laboratories to carry out tests with a reference toxicant. The results are used to monitor for bias and variability, using statistically-derived criteria ('targets') for acceptable accuracy of tests and for the degree of agreement in toxicity between repeated tests within a laboratory (precision). There would be an upper and lower limit on the allowable deviation from the target for accuracy but only a lower limit for precision

(there is no limit on good precision!). The underlying assumption of the approach is that responses to reference toxicants are an acceptable surrogate for effluents. To date, criteria have been developed for Microtox™, acute *Daphnia*, *Acartia* and OEL tests but not fish, algal or *Tisbe* tests.

If trialled during the Demonstration Programme, laboratories would be asked to carry out concentration-response tests using around five concentrations of a reference toxicant (probably zinc sulphate) on several occasions. They would then estimate EC50 values and make those data available to an external co-ordinator. Laboratory performance would be judged against the predefined criteria for accuracy and precision*. The criteria are substance-specific which means that criteria for one substance cannot be used to judge the accuracy and precision of tests performed with another. This means that reference toxicant testing must be performed using the same substance and this will involve additional work for some laboratories.

There are two possible variants of this approach, differing in the frequency of reference toxicant testing.

(a) Reference toxicant testing with every effluent toxicity test

This approach permits an assessment of test performance on every occasion that a regulatory effluent test is carried out. The implication is that, if testing fails the criteria for accuracy, the corresponding effluent toxicity data may be rejected because it is also likely to be subject to unacceptable bias. This approach may be regarded as rather 'heavy handed' but provides the regulator and discharger with a degree of confidence about individual measurements of effluent toxicity that would not be possible any other way. It is recognised that this option may entail considerably more reference toxicant testing at some sites than others.

A key advantage of trialling this option is that it also permits an assessment of other options. By sub-sampling the dataset generated, several of the other possible options (options 1(b), 2(a) and 2(b)) can be simulated because they are effectively sub-sets of option 1(a).

(b) Reference toxicant testing 'uncoupled' from effluent toxicity testing

This is equivalent to the modified scheme tabled in July 1997. Laboratories carry out a fixed number of reference toxicant tests (say, 5 during the period of the Demonstration Programme). Consequently, many effluent tests would not have an associated reference toxicant test. This approach may be more palatable to the laboratories (because it entails a smaller workload than option 1(a)) but, because reference toxicant data are*

*These criteria are calculated from ring-test data, explained in detail in the Agency's proposals for a 'Register of Approved Laboratories for Laboratories undertaking Toxicity Testing'.

effectively 'uncoupled' from most effluent tests, it removes the regulator's ability to reject the data from particular effluent toxicity tests.

The main value of incorporating this option is that it would allow us to determine whether or not the information generated was as useful for monitoring test performance as more 'demanding' options (particularly option 1(a)).

2. *'Abbreviated' tests with reference toxicants*

The approach here is the same as that in option 1 except that laboratory's workload is reduced by eliminating some of the test concentrations that must be tested. Basically, only two concentrations of the reference toxicant are tested and, if they bracket the 50% response concentration, then it is reasonable to assume that the EC50 lies between these concentrations. The two concentrations used could be the upper and lower thresholds for accuracy described above.

On the face of it, this approach is appealing because the costs of reference toxicant testing become very low compared to the costs of an effluent toxicity test (see Table 2). However, simulations using typical data highlight some statistical problems, giving rise to an appreciable amount of ambiguity i.e. it may not always be possible to tell whether or not a particular reference toxicant test has given rise to acceptable data or not. The value of this approach would only really become apparent if we can obtain realistic data and assess whether or not these ambiguities really undermine our ability to confidently assess the performance of tests. The Demonstration Programme provides such an opportunity, perhaps by opting for an approach which calls for full concentration-response testing and then sub-sampling the resulting set of reference toxicant data to 'create' datasets which comprise just two concentrations of the reference toxicant. This approach would entail no extra costs to participating laboratories but would involve a significant amount of analysis by the co-ordinator.

This 'abbreviated' approach might be the only realistically acceptable approach to carrying out reference toxicant tests with the more expensive algal tests and fish tests (where there are additional concerns about animal welfare). As with option 1, two sub-options can be identified which differ in the frequency of testing that is required.

(a) Reference toxicant testing with every effluent toxicity test

Reference toxicant testing may be 'coupled' to regulatory effluent tests. As with option 1(a), this approach provides the regulator and discharger with a degree of confidence about individual measurements of effluent toxicity that would not be possible any other way.

(b) Reference toxicant testing 'uncoupled' from effluent toxicity testing

Alternatively, testing may be carried out on a limited number of occasions, thereby 'uncoupling' the reference toxicant tests from effluent testing. Whilst this requires considerably less testing it removes the regulator's ability to reject the data from particular effluent toxicity tests. The main value of incorporating this option is that it would allow us to determine whether or not the information generated was as useful for monitoring test performance as the more costly approach described in option 2a.

3. *Testing of 'split' samples of effluent*

Not unreasonably, the emphasis on the use of single 'well-behaved' single substances as reference toxicants has been questioned when the aim is to predict performance in tests with effluents. Much of the reason for relying on single substances is practical: effluent samples are often unstable and their toxicity can change - sometimes drastically - over short periods. In interlaboratory comparisons, this makes it difficult to ensure that laboratories carry out tests on the same sample and impossible to assess precision from repeat tests because we cannot distinguish changes in test sensitivity from changes in the composition and toxicity of the sample being tested. Because there are no criteria for accuracy there is also the problem of determining what is and what is not acceptable in terms of the resulting toxicity.

In the few examples where toxicity of 'split' effluent samples has been compared, variability tends to be less than that obtained in corresponding tests with single substances. This may lead to us to conclude that assessments of variability based on single substances (such as those described in options 1 and 2) provide an over-pessimistic view of test variability. However, this tendency probably has more to do with the fact that effluent toxicity is usually expressed as a <u>percentage</u> whilst single substance toxicity is expressed as a <u>concentration</u>. The former has an upper ceiling of 100 whilst there is no upper ceiling on the concentration of a substance. The result is that variability, e.g. between laboratories, for single substances appears to be greater than it does for effluents.

There remains an understandable desire to incorporate some element of effluent testing as part of the QC measures being evaluated in the Demonstration Programme. Despite the shortcomings highlighted above, 'split' effluent samples might be very useful in investigating the incidence of false conclusions about compliance with a real action level produced during the Demonstration Programme. If a discharge can be identified which appears to be marginal with respect to compliance, then it would be useful to 'split' a sample of that effluent on one or two occasions and ask all laboratories in the Demonstration Programme to test it. The amount of disagreement about compliance/non-compliance judged from those tests would be important information about the practical consequences of variability. Such an exercise would also provide an opportunity to assess how well reference toxicant tests were at predicting variability between laboratories in this situation. However, the approach is not without potential problems: it is possible that interlaboratory variability may be exacurbated if a highly

unstable effluent or one containing a high proportion of volatiles was used. Careful selection of suitable effluents would be necessary to avoid such samples. In the event that a suitable effluent could not be identified, a blend of single substances whose identity and composition is known only to the co-ordinator could be used instead.

Testing of split effluent samples would only really be useful as a supplement to a trial of single substance reference toxicant testing rather than as a replacement. This approach also provides an element of 'blind' testing and so addresses any concerns about bias or fraud that might be raised externally.

4. *Predicting concentrations of unknown samples from 'calibration' curves*

The Canadian Toxicological Testing Laboratory Accreditation Program has trialled an approach to external QC (administered by the Canadian Association for Environmental Analytical Laboratories: CAEAL) which uses reference toxicants in a way that overcomes some of the uncertainties involved in defining the accuracy of ecotoxicity test data.

Basically, samples of substances (whose identity is unknown to the laboratory) are circulated and EC50 values estimated from a concentration-response test. At the same time, solutions containing different concentrations of the same substance are supplied. Using the concentration-response tests as a calibration curve, the laboratory estimates the concentration of toxicant in the solution samples. These estimates are then compared with the concentration which is known to be present in these solutions. The advantage of this approach is that the performance critrion for accuracy is no longer expressed in terms of toxicity - which is always subject to some uncertainty - but in terms of a concentration, which is known with certainty. There still remains the problem of judging how much deviation from the true concentration is acceptable. Furthermore, this approach does not provide any way of monitoring the precision of tests and so relies on internal QC (see option 5) for that aspect.

As the basis of an external QC scheme, the CAEAL approach has limitations but the ability to use a known chemical concentration as a performance criterion might be useful as a supplementary measure. It could easily be incorporated into the Demonstration Programme, perhaps using solutions of zinc sulphate on just one occasion, to determine whether or not it adds value to assessments of performance using other approaches.

5. *Reliance on internal QC alone*

The simplest approach to performance assessment is to rely on a laboratory's own in-house assessments of repeatability using a reference toxicant chosen by that laboratory. There are simple procedures borrowed from analytical chemistry ('Control charts')

which employ previous performance to define internal action and warning limits within which 'normal' results are expected to lie. In toxicity testing, this type of internal QC has three major limitations:

- because the performance criteria (action and warning limits) are defined on the basis of previous experince with that substance, they can actually reinforce bias
- internal QC penalises laboratories with historically good control over variability because narrower (and therefore more demanding) performance criteria result
- it addresses only precision. It does not provide any means for assessing the accuracy of test results or the degree of variability between laboratories

It is hard to see how a scheme based on internal QC can provide the required assurance over performance of ecotoxicity tests. However, such measures should never be discouraged and, in combination with an approach which addresses accuracy and variability between laboratories (e.g. options 1-4), internal QC can provide some useful information.

Recommendations

The Demonstration Programme provides an opportunity to evaluate a number of different approaches to QC of ecotoxicity tests. Based on the results obtained, the preferred option for a QC scheme to be used in routine testing, would be identified. The favoured approach would be one which delivers adequate QC information without incurring an unecessary burden on participating laboratories or the regulator.

From the descriptions above, a number of 'core' options can be identified which have the potential to provide information on both the accuracy and precision of ecotoxicity test data. There are also a number of 'supplementary' options which, on their own, would fall short of a robust QC scheme but which can add valuable information to the 'core' approach.

Option 1(a) is recommended as the 'core' option. This involves reference toxicant tests by each participating laboratory on each occasion an effluent is tested, using a full range of test concentrations. The resulting dataset would then be sub-sampled to simulate lower frequencies of testing and also to simulate testing at only two concentrations. In this way, options 1(b), 2(a) and 2(b) are automatically covered and it should be possible to identify an approach (testing frequency and number of test concentrations) which delivers adequate QC data at lowest cost to the participating laboratories identified.

Further advantages of option 1 are that this approach serves to build up a useful database of reference toxicity data. Currently, some of the QC criteria which have been calculated suffer from being derived from a modest amount of ring-test data; recalculating the criteria using a larger dataset will make them more robust. A secondary benefit is that it should be

possible to produce 'supplementary' criteria which describe the performance (acuracy and precision) achieved by the best performing laboratories, thereby encouraging further performance improvements.

In addition to this programme of reference toxicant testing, it is proposed that option 3 ('split' sample of an effluent on two occasions) is also carried out. This should provide a valuable assessment of the effects of variability and bias under realistic conditions. Option 3 also permits a degree of 'blind testing' of samples, a feature which is only available through options 3 and 4.

The project would concentrate on interpreting data generated from tests with Microtox™, *Daphnia* and OEL because these are the tests for which QC criteria are currently available. The Demonstration Programme also provides an opportunity to gather reference toxicant data from other test methods e.g. algal, fish and *Tisbe/Acartia* toxicity tests, but that exercise should be considered separately from the QC issues covered here.

Costs

Some tasks are common to each option (e.g. seminar, collation of data, reporting) and so only need to be accounted for once. In other words, the costs of trialling several options (as proposed here) are appreciably less than the sum of the costs shown in Table 2. The expected cost to the Steering Group from an external contractor to manage a trial of all the different QC options described in this document is indicated below. This includes only the contractor's costs and not those incurred by the test laboratories, which would be additional; it is difficult to estimate the costs incurred by the test laboratories because they will depend on the design of the Phase II Demonstration studies. The costs further assume that six laboratories would be involved in the trial.

Analysis and reporting of Option 1(a):	£11000
Additional analysis for Option 1(b):	£2600
Additional analysis for Option 2 (a):	£4000
Additional analysis for Option 2 (b):	£2000
Additional analysis for Option 3:	£3300
Total:	£22900

These estimates make provision for the following:

- seminar for all laboratories involved in the Demonstration Programme
- supply, co-ordinate and follow-up testing of reference toxicants and effluents
- chemical analysis of stock solutions
- collate all resulting test data
- analyse and interpret test data
- report results and recommendations to the Steering Group

Other options

The extremes - in terms of cost and the value of the information generated - are represented by the following:

- the recommended approach above (options 1a and 3, and reanalysis of data to evaluate options 1b, 2a and 2b as well). The associated costs of this study are £22900.

- trialling only option 5 (reliance on internal QC). The costs of this study are £4000.

Several other options lying between these extremes are possible. Obviously they sacrifice some information compared to the recommended approach but would incur smaller costs. Two specific examples are listed below along with their associated costs:

- Option 1(a), trialled along with option 3 and the data re-analysed to investigate the value of adopting option 1(b) - this approach should enable the optimum number of repeat tests to be identified. However, it provides no information on the effects of reducing the number of test concentrations in each reference toxicity test. As a consequence, the possibility of savings by test laboratories are lost. The associated costs would be £18300.

- Option 1(b), trialled along with option 3 and the data re-analysed to investigate the value of adopting option 2(b) - this approach allows us to assess any benefits resulting from reducing the number of test concentrations in reference toxicant tests. However, it provides no information on the value (and costs) of different testing frequencies or 'coupling' reference toxicity tests to effluent tests. Consequently, the ability to confidently identify the minimum number of reference toxicant tests which is needed for effective QC is eroded. The associated costs would be £13500.

Table 1 **Summary of QC options which may be trialled during Demonstration Programme**

	Option						
Consideration	**1a**	**1b**	**2**	**2a**	**3**	**4**	**5**
Are objective performance criteria available?	Y	Y	Y	Y	N	Y	Y for precision (but flawed)
Are objective criteria for monitoring performance against criteria available?	Y	Y	Y but assessment may be ambiguous	Y but assessment may be ambiguous	N	N	Y for precision (but flawed)
Do the data generated make it possible to suggest improvement measures?	possibly	possibly	N	N	N	possibly	N
Is it possible to refine QC criteria to make them more robust?	Y	Y	possibly	possibly	N	N	N/A
Promotes improvement in control over bias and variability?	Y	Y	possibly	possibly	possibly	possibly	maintains status quo
Permits 'blind testing' of samples?	N	N	N	N	Y	Y	N

Table 2 Estimated costs of trialling different QC options during Demonstration Programme (£)

	Option						
	1a	**1b**	**2a**	**2b**	**3**	**4**	**5**
Assumed frequency of testing	testing on 20 occasions	testing on 5 occasions	testing on 20 occasions	testing on 5 occasions	exercise is repeated twice	exercise is performed once	labs carry out at least 5 internal tests
*Cost to lab	M: 1000 D: 4200 O: 3400	M: 500 D: 1300 O: 1000	M: 1000 D: 3000 O: 2500	M: 400 D: 1000 O: 800	M: 200 D: 900 O: 500	M: 250 D: 400 O: 400	M: 400 D: 1500 O: 1200
$Cost of trialling each option	£11000	£8200	£9800	£7800	£4200	£5800	£4000

*M=Microtox™, D= *Daphnia* acute, O=OEL

*It is assumed that tests would always be carried out in conjunction with effluent tests, so only variable costs need be considered (e..g. test organisms, assessment, reporting).

*A total cost to the laboratory for participating in the trail is subject to the following uncertainties: (a) which tests are used (b) the complexity of the study site (no. of discharges) and (c) amount of testing specified in Planning Phase

$ These are costs incurred by the Steering Group in financing the organisation, co-ordination, analysis and reporting of different trail options. The costs are for trialling the options independendently (i.e. not at the same time time). The differences between options reflect differences in the complexity of analysis and the amounts of data analysed.

$ Trialling several options simultaneously gives rise to significant savings, as discussed in text ('Recommendations')

Annex IIIe

Options for Trialling Quality Control Procedures for Ecotoxicity Tests - Outstanding Issues, WRc 1998

OPTIONS FOR QUALITY CONTROL FOR REGULATORY ECOTOXICOLOGICAL TESTING

OUTSTANDING ISSUES:

Choice of reference toxicants

The choice of toxicant can influence one's view about the performance characteristics of ecotoxicity tests. Experience shows that organic toxicants tend, on the whole, to be 'better behaved' (i.e they give rise to smaller CVs for repeatability and reproducibility) than many inorganics. Even within inorganics (where most data have been generated) there can be considerable variation in calculated CVs although whether this is a function of the toxiacnt or the design of the study is debatable. Many studies have calculated CVs on the basis of just 2 or 3 measurements; that is just not enough because the variances will not have stabilised. There is much less information available for complex substances (because effluents, for example, will themselves be variable and unstable) but US studies with drilling muds and effluents suggest that more complex materials give rise to similar or less variability in toxicity estimates. The 1995 interlaboratory ring-test in the UK which used zinc sulphate and 3,4-DCA as toxicants showed greater variation with the former.

There is a temptation to advocate a substance as a reference toxicant because 'it gives repeatable results'. However, I would argue that this is actually a good reason for <u>not</u> choosing such a substance: a 'well behaved' toxicant is less likely to provide useful Quality Control information than one which is sensitive to variations in water quality parameters, experimental practice and variations in genotype. Such factors give rise to the variations in 'output' variability (i.e. variation in measured toxicity) that we are seeking to determine with the reference toxicant. It follows that a suitable reference toxicant is one which responds to variations in at least some of these factors. The reason why metals tend to give larger CVs for repeatability and reproducibility than organics is probably because their bioavailability can be influenced by water quality parameters such as pH, hardness and the presence of chelating agents. The additional 'output' variation we are seeing with metals may reflect variations in the composition of test media and I would argue that a reference toxicant should be capable of revealing such variations if they are important to the standardisation of the test. In many cases, they are (e.g. the method requires a certain stanadrd of water quality) and will be addressed in the test validity criteria.

Of course, we need to ensure that any reference toxicant is not so sensitive to water quality parameters that it simply precipitates out on contact with the test medium. The speciation of zinc is fairly simple within the pH range specified in test guidelines we are concerned with, so any effects on bioavailability and hence toxicity should only become evident if the water quality parameters, e.g. pH, have not been properly controlled.

In selecting a reference toxicant, we need to put these considerations together with practical ones such as ease of handling (water solubility and stability), safety and ease of analysis. There is nothing special about the use of zinc sulphate and 3,4-DCA with respect to these characteristics; several other substances would do just as well.

However, there is one very important factor that should be added:

If we are planning to use toxicity data generated using reference toxicants as part of an external QC scheme then it is essential that we have some external 'target' against which to measure accuracy and precision. Without these external reference points, any judgement of performance could only be based on internal QC. Any attempt at external QC would be meaningless. It would be like trying to keep score in a football match without knowing where the goal is!

This is where zinc sulphate and 3,4-DCA score (excuse the weak pun) and other toxicants do not. It is only for these substances that we have accumulated enough data to derive these targets. If we accept this as an important criterion in choosing a suitable reference toxicant then most of the other issues are pretty academic. As far as choosing between zinc sulphate and 3,4-DCA is concerned, the decison is largely a practical one. Zinc is much cheaper to analyse for than 3,4-DCA and so this would help keep down the costs of the trial (even if laboratories do not conduct any chemical analyses themselves) but if there were strong preferences for 3,4-DCA, this could be used.

Alternative substances would be entirely acceptable - even mixtures - if it was possible to derive 'targets' for accuracy and precision for them. This would require at least some interlaboratory ring-testing and the ring-test would need to be designed in such a way as to allow us to partition sources of variability. It is possible that procedures based on Bayesian statistics might be used which allow us to make more effective use of existing data on particular substances (including previous ring-tests) and cut down the amount of new ring-testing required as a result.

Quality Assurance and Quality Control

It is generally accepted that regulatory effluent testing should be undertaken within a Quality-assured (QA) environment i.e. in a laboratory which is either inspected under the UK GLP Monitoring Programme or accredited under NAMAS. This requirement is not to be confused with QC requirements: the QA provided by these Quality Systems is primarily intended to maintain the integrity of regulatory effluent test data and in this respect it is entirely consistent with the requirement for QA for, say, new chemicals or offshore chemicals notification. QC fulfils a different need because it is concerned with the performance of testing. Whilst accreditation under NAMAS requires some degree of QC (although it is not specified exactly how this should be achieved), GLP compliance makes no reference at all to QC. QC is intended to complement the benefits resulting from QA. The options suggested in the previous note are intended to identify a cost-effective approach to QC.

We are proposing to trial one or more different QC options in parallel with Phase 2 of the DTA Demonstration Programme. The question is whether the QC (i.e. reference toxicant and/or split effluent) testing carried out during that programme should be subject to QA in addition to effluent toxicity data. The arguments for and against QA of QC data are as follows:

For	Against
1. Assures integrity of QC data (no cheating!)	1. Increases costs of QC testing (see below)
2. Encourages good data management so data more likely to be auditable, if required	2. Can delay reporting of QC data
	3. Inappropriate for trial programme
	4. QA personnel may not give high priority to non-regulatory data

Costs of QC testing

The costs for carrying out QC testing in the previous note made a number of assumptions:

- All reference toxicant/split effluent tests would be carried out at the same time as effluent tests. This means that the only costs involved are variable ones associated with initiation, monitoring of tests and data analysis, and would be carried out when these tasks are being performed anyway

- The QC tests would not be subject to QA (see above)

If the more intensive option (option 1a) was to be pursued, it seems likely that a proportion of 'dedicated' QC tests may need to be carried out i.e. not coupled with effluent toxicity tests. In this case, both variable and fixed costs would be incurred. If it was felt that QA was necessary, this would further increase costs. In the following Table, costs for carrying out QC tests according to option 1a are estimated when (a) all testing is performed alongside effluent tests, (b) when only half the tests are carried out alongside effluent tests and (c) if all QC tests are carried out independently of effluent tests:

Scenario	Average cost per test (£)		
	Microtox	*Daphnia*	OEL
(a) all tests carried out in parallel with an effluent test	50 (without QA) 120 (with QA)	250 (without QA) 320 (with QA)	200 (without QA) 270 (with QA)
(b) 50% of tests carried out in parallel with an effluent test (the rest are 'dedicated' QC tests)	100 (without QA) 190 (with QA)	420 (without QA) 520 (with QA)	450 (without QA) 550 (with QA)
(c) all 'dedicated' QC tests)	150 (without QA) 250 (with QA)	600 (without QA) 750 (with QA)	700 (without QA) 850 (with QA)

PW 22/12/97