Analysis of 1995 Survey Data. Phase 2 Post-Survey Appraisal

Unit III: Post-Survey Appraisal

R&D Technical Report E102

M T Furse, R T Clarke, J Davy-Bowker, K E Vowles

Research Contractor
Institute of Freshwater Ecology

Publishing Organisation
Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, Bristol BS12 4UD
Tel: 01454 624400 Fax: 01454 624409
Website: www.environment-agency.gov.uk

ISBN 1 85705 659 0

Dissemination Status
Internal: Released to Regions
External: Released to Public Domain

Statement of Use
This document presents the findings of a questionnaire of practitioners' views on the procedures used to undertake the 1995 biological General Quality Assessment (GQA). It also includes analyses of some of the results of external quality audits of macro-invertebrate sample processing, where these are of relevance to the 2000 GQA Survey. It will be used by the Agency to help determine the procedures to be used during the 2000 GQA Survey.

Key Words
General Quality Assessment; questionnaire, methodologies, survey design, new procedures, reporting, audit, biases (systematic errors)

Research Contractor
This document was produced under R&D Project E1-036 by:
Institute of Freshwater Ecology, River Laboratory, East Stoke, Wareham, Dorset BH20 6BB
Tel: 01929 462314 Fax: 01929 462180

Environment Agency's Project Manager
The Environment Agency's Project Manager for R&D Project E1-036 was:
Dr R.A Dines, Southern Region

Further copies of this report are available from:
Environment Agency R&D Dissemination Centre
WRc, Frankland Road, Swindon, Wilts. SN5 8YF
Tel: 01793 865000 Fax: 01793 514562 E-mail: publications@wrcplc.co.uk

ACKNOWLEDGEMENTS

This project was funded collaboratively by the Environment Agency and the Natural Environment Research Council.

The authors are grateful for the advice and support provided by the Environment Agency Project Manager, Bob Dines and Research and Development Management Support Officer, Pamela Mardon

The study would not have been possible without the co-operation of the members of staff of the Environment Agency, Scottish Environmental Protection Agency and the Department of the Environment (Northern Ireland) who offered their views by answering the questionnaire.

CONTENTS

		Page
ACKNOWLEDGEMENTS		**i**
LIST OF TABLES		**v**
LIST OF FIGURES		**vii**
EXECUTIVE SUMMARY		**ix**
1	**OBJECTIVES**	**1**
1.1	Context	1
1.2	Specific Objectives	2
2	**THE HISTORY OF NATIONAL RIVER QUALITY SURVEYS**	**3**
2.1	Context	3
2.2	The 1958 Survey	3
2.3	The 1970 Survey	3
2.4	The 1972 Survey	4
2.5	The 1973 Survey	4
2.6	The 1976 Survey	4
2.7	The Biological Monitoring Working Party	5
2.8	The 1980 Survey	6
2.9	The Development of RIVPACS	6
2.10	The 1985 Survey	7
2.11	Development of a System for Classifying the Biological Condition of Sites	7
2.12	The 1990 Survey	7
2.13	External Quality Audit	8
2.14	Internal Analytical Control	9
2.15	The 1995 Survey	9
2.16	The Development of RIVPACS III+	10
2.17	The 2000 GQA Survey	12
3	**METHODS**	**13**
3.1	General Approach	13
3.2	Development of the Questionnaire	13
3.3	Circulation of the Questionnaire	14
3.4	Presentation and Interpretation of the Responses	14
4	**RESULTS OF THE QUESTIONNAIRE**	**15**
4.1	The level of Response	15
4.2	The Replies to Each Question	16

5 EVALUATION OF QUESTIONNAIRE REPLIES 79
5.1 Number of Samples 79
5.2 Sampling Methods 80
5.3 Sample Sorting 83
5.4 Identification and Quantification 86
5.5 Internal AQC and External Audit 91
5.6 Banding (Grading) of Biological Condition 94
5.7 Other Forms of Data Collection and Interpretation 100
5.8 Environmental Data 104
5.9 Equipment 110
5.10 Reporting 110
5.11 Survey Design 112
5.12 Summary of Recommendations 115

6 EFFECTS OF SAMPLE ANALYTICAL ERRORS ON THE DETECTION OF CHANGE IN BIOLOGICAL CONDITION 125
6.1 Introduction 125
6.2 Methods of Assessment 127
6.3 Detectable Differences in Relation to Analytical Bias 127
6.4 Summary 133

7 FACTORS ASSOCIATED WITH POOR PERFORMANCE IN SAMPLE PROCESSING 135
7.1 Availability of Audited Samples in 1995 135
7.2 Sample Processing Biases in Relation to Site Quality 135
7.3 Sample Processing Biases in Relation to Environmental Characteristics of Sites 136
7.4 Summary 140

8 REFERENCES 141

LIST OF TABLES

Page

Table 2.1 The EQI grade ranges (the 5M system) for three separate forms of BMWP Index, BMWP score (EQI_{BMWP}), number of scoring taxa (EQI_{Taxa}) and ASPT (EQI_{ASPT}), as used in the 1990 River Quality Survey. 8

Table 2.2 The EQI grade ranges for two separate forms of BMWP index; number of scoring taxa (EQI_{Taxa}) and ASPT (EQI_{ASPT}), as used in the 1995 GQA. 10

Table 3.1 The named biologist in each Agency Region and country to whom the questionnaire was circulated 14

Table 4.1 The response to the questionnaire 15

Table 5.1 Minimum and maximum times, in minutes, taken to sort different forms of sample in different Regions/Area 86

Table 5.2 A list of the taxa (and life stages) cited by Environment Agency biologists as being the most difficult to identify. 87

Table 5.3 All BMWP families which were missed on at least 25% of occasions on which they were recorded as present in audit samples by IFE in either or both of 1990 and 1995 88

Table 5.4 A suggested new system of abundance categories to be used during the 2000 GQA Survey 90

Table 5.5 A list of alternative procedures, other than RIVPACS, for the interpretation of macro-invertebrate data collected during the 2000 GQA Survey 101

Table 5.6 Suggested improvements or developments of RIVPACS to aid interpretation of national surveys. 101

Table 5.7 Suggested taxonomic groups which should be included in biomonitoring surveys with notes on their relevance to stress detection. 103

Table 5.8 The degree of difficulty of obtaining time invariant data from maps. 106

Table 5.9 The degree of difficulty of obtaining time variant data in the field. 108

Table 5.10 Number of times each season is recommended for sampling in replies received to Questions S2 and S4 of the supplementary questionnaire 115

Page

Table 6.1 Estimates of average net under-estimation of the number of taxa (termed the bias) in single season samples taken from each region in the 1990 RQS and 1995 GQA surveys. Average "losses" per audited sample in 1995 are also given. 126

Table 6.2 Critical sizes of the changes in EQI values between 1990 and 1995 for the matched sites which were detected as statistically significant ($p<0.05$), either (a) uncorrected for bias (equivalent to no bias) or (b) after correcting for bias using the best available estimates of biases for each Region. 128

Table 6.3 Minimum sizes of real differences (D_m) in EQI_{TAXA} and EQI_{ASPT} values between two samples which have either at least a 50% or at least a 90% chance (i.e. power) of being detected as statistically significant ($p<0.05$) for various fixed levels of sample bias B in both samples. %sites denotes the percentage of matched 1990 RQS and 1995 GQA sites with bias-corrected differences > D_m. 131

Table 6.4 Percentage of the (n=3018) matched 1990 RQS and 1995 GQA sites Detected by RIVPACS III+ as having statistically significant ($p<0.05$) changes in values of either EQI_{TAXA} or EQI_{ASPT} for various assumed fixed levels of sample bias B in both samples. 132

Table 7.1 Sample processing biases amongst GQA sites in 1995 in relation to site quality as measured by EQI_{TAXA}. 136

Table 7.2 Average sample processing biases in 1995 in each NRA/Environment Agency Region based on the total of 393 audited samples matched to 1995 GQA sites. 136

Table 7.3 Average sample processing biases (number of samples in brackets) for 1995 audit samples from sites in England and Wales classified into six categories of each RIVPACS environmental variable 137

Table 7.4 Average sample processing biases (number of samples in brackets) for 1995 audit samples from sites in England and Wales in relation to site altitude (m) and slope (m km^{-1}). 138

Table 7.5 Average sample processing biases (number of samples in brackets) for 1995 audit samples from sites in England and Wales in relation to (a) Region and sorting method (whether done on the live or preserved sample), (b) Region, site altitude (m) and sorting method. 139

LIST OF FIGURES

Page

Figure 6.1 Percentage of matched sites with a statistically significant ($p<0.05$) difference in EQI (i.e. O/E) for (a) number of taxa and (b) ASPT between the 1990 RQS and 1995 GQA surveys in relation to the size of the estimated difference. This is given for the EQI values uncorrected for bias (and hence, for this exercise, assuming no bias exists) and corrected for the estimated biases, as given in Table 6.1 129

Figure 6.2 Distribution of changes in EQI for ASPT after correcting for biases (using The best estimates for each Region in 1995) in relation to the 'face' (i.e. uncorrected) value of EQI for ASPT. 130

EXECUTIVE SUMMARY

The reporting of Phase 2 is divided into three units :

Unit I: Taxon distribution studies : R&D Technical Report E103 (Davy-Bowker *et al.* 2000)
Unit II: Changes in biological condition : R&D Technical Report E101 (Clarke *et al.* 1999)
Unit III: Post-survey appraisal : R&D Technical Report E102 (Furse *et al.* 1999) - this report

Unit I contains:

- a description of the incorporation of the 1990 RQS and 1995 GQA survey biological and environmental data into IFE's Quinquennial Survey Database (QSD). This includes procedures to establish matching pairs of sampling locations for use in analyses of change between surveys
- distribution studies of each BMWP taxon providing information on their geographic distribution, their environmental ranges (in terms of the RIVPACS environmental variables) and their tolerance or susceptibility to particular sources of environmental stress thought to be operating at individual sites

Unit II contains summaries of the:

- patterns of distribution of biological condition in 1995, especially in relation to RIVPACS environmental variables
- changes in biological condition between matched sites in 1990 and 1995, incorporating measures of the statistical significance of change in biological grade
- changes in biological condition in relation to site environmental characteristics and ITE landscape type
- data obtained from Environment Agency Regions on the known or suspected sources of environmental stress operating on each of the GQA sites
- relationships between biological condition or change in condition and the type and severity of any environmental stress or pollution

Unit III (this report) contains summaries of the :

- history of national River Quality Surveys in the Environment Agency area (England and Wales)
- responses to the 1995 GQA post-survey questionnaire to Agency staff developed within this project
- results and conclusions from an investigation using the bias specification options in RIVPACS III+ to assess the effect of alternative analytical quality targets for macro-invertebrate samples
- analysis of the 1995 quality audit to determine which factors, if any, can be associated with poor levels of performance
- recommendations for future surveys

The major recommendations of the review of replies to the questionnaire are as follows:

Number of samples

There should be no reduction in the coverage of sites in the 2000 GQA Survey, in comparison with 1995, unless there are resource limitations that cannot be overcome.

Where feasible, Regions should adjust their coverage of particular site types upward to rectify deficiencies they identified in the 1995 GQA. This appears to apply to headwaters in particular.

Sampling Methods

Where site characteristics are suitable, the pond-net sampling procedures adopted in 1995 should be retained for the 2000 GQA Survey. This will provide a reliable basis for the application of the RIVPACS III+ procedures for detecting temporal change.

When appropriate, standardised deep-water sampling procedures are available, sampling with the standard FBA-style pond-net, with handle length of approximately 1m, should be confined to sites that are wadeable for at least 25% of their total width.

The Environment Agency should introduce standard procedures for the sampling of deep water sites, i.e. those that are not suitable for standard pond-net sampling.

All biologists involved in field sampling of deep-water sites should receive appropriate training prior to undertaking sampling.

Sample Sorting

Whilst it remains unclear whether bankside sample processing is as efficient and comprehensive as laboratory processing, this issue is considered to be so important that a standardised laboratory-based approach should continue to be prescribed for use in the 2000 GQA Survey. This action will ensure that observed BMWP index values and those predicted by RIVPACS III+ are based on the same sorting procedure.

Identification and Quantification

It is essential that a system of allocation of abundance classes to each BMWP family is adopted by all Regions for the 2000 GQA Survey.

This system should be standardised capable of being standardised between Regions. It should also be compatible with the $\log_{10}$ system recommended for the 1995 GQA Survey.

Whatever system of categorisation is adopted by each Region for the 2000 GQA Survey, the data must be presented to the Environment Agency's National Database in the standard categories adopted for the survey.

If abundance-based indices are to be used to report on the results of the 2000 GQA Survey, then abundance checks should be incorporated in quality control procedures.

Internal AQC and External Audit

The use of internal AQC should be continued for the 2000 GQA Survey.

Grading of Biological Condition

Although there is perceived to be scope to improve the grading system used in the 1995 survey, it is recommended that it is retained for the 2000 GQA Survey in order to maintain compatibility with the 1995 survey. Continual changes in the evaluation procedures can create the impression that the message of the surveys is being obfuscated by the shifting methodologies.

Other Forms of Data Collection and Interpretation

The use of RIVPACS and EQIs alone to examine the biological data collected during GQAs fails to optimise the cost-effectiveness of the survey. The following recommendations, if adopted, will help this to be evaluated.

The Agency should apply the abundance-based indices being developed for RIVPACS and the Artificial Intelligence procedures being developed at the University of Staffordshire.

The Agency should also conduct trials of the application of the LIFE (Lotic Invertebrate index for Flow Evaluation) and CCI (Community Conservation Index) systems, the use of the Trophic Diatom Index (TDI) and should try and should try and synchronise River Habit Survey reaches and 2000 GQA Survey biological sampling sites.

Environmental Data

Alkalinity is an important RIVPACS predictor variable and the reliability of EQI evaluations are reduced by its unavailability. Its regular collection should be retained for all GQA sites until long-term average values can be substituted.

All time invariant values used for RIVPACS predictions in the 2000 GQA Survey should be re-measured independently by two people.

Logical checks of environmental data for individual rivers should be made to ensure that rivers flow downhill and the discharge of any site should be no less than the discharge of the next site upstream of it.

Reporting

The widespread and varied application of GQA data for regional purposes is to be encouraged and where possible extended.

Trial species level identification of a sub-set of the 2000 GQA Survey data should be undertaken and the new indices, LIFE and CCI should be applied and evaluated.

The results of the 2000 GQA Survey should be prominently published in the public domain.

Survey Design

Quinquennial GQAs should be replaced by annual surveys in the form of a rolling programme.

The 2000 GQA Survey should be based on two seasons' macro-invertebrate sampling per site.

2000 GQA Survey sampling should be undertaken in spring and autumn in order to provide a standard basis for inter-survey comparisons of distributional changes.

The major conclusions of the analysis of the effects of sample analytical errors on the detection of change in biological condition are as follows:

The effect of having biases arising from sample analytical errors (and allowing for them) is to make it more difficult to estimate the biological condition of a site in terms of EQI value and hence grade. Consequently, it also becomes more difficult to detect and estimate the size of a change over time or the difference between two sites.

Analyses suggest that there would undesirable loss of statistical power to detect change in biological condition if the tolerable sample analytical bias was allowed to increase from 1.75 to a level of over 3 taxa per sample.

The gain in power to detect change obtained by reducing the sample analytical bias from 1.75 to 1.0 taxa is not great.

We recommend that the Environment Agency continue to aim to achieve and maintain a gross level of missed taxa of no more than 2.0 per sample, which has been equivalent to a sample analytical bias of 1.75 taxa.

The major conclusions of the analysis of the factors associated with poor levels of performance in sample processing are as follows:

The size of sample processing errors does not appear to depend on the quality of the site (i.e. its value for EQI_{TAXA}); except for very poor quality sites with GQA grades e/f, which have very few taxa. This agrees with Furse *et al.*(1995) who found no relationship between sample biases and the taxonomic richness of the sample.

Sample processing errors were examined in relation to the environmental characteristics of the sites. Those few samples from sites at high altitudes (i.e. >200m) and/or with steep slopes (i.e. >25m/km) had, on average, larger processing errors. This was partly associated with the higher sample errors in North West Region.

No other environmental characteristics, including substratum sediment type, appeared to be associated with higher than average sample processing errors.

Whether a sample was sorted live or after being preserved did not seem to consistently influence the size of sample processing errors across all Regions in general. However no single Region processed adequate numbers of **both** live and preserved samples for meaningful comparisons of sorting performances to be made.

It is recommended that the Agency improves their consistency in the use completely standard names and codes for each site and river used on the sample audit forms and in biological, environmental and chemical databases.

1 OBJECTIVES

1.1 Context

Phase 2 of this Research and Development project aims to provide a comprehensive appraisal of the information content and performance of the 1995 General Quality Assessment (GQA) survey and of the changes between the 1990 River Quality Survey (RQS) and 1995 GQA survey. The implications of its results are to be taken into consideration in formulating the procedures to be used in the 2000 GQA Survey.

This phase aims to increase understanding of the spatial and temporal relationships between taxonomic distribution, biological condition, environmental characteristics, Landscape type and the sources of environmental stress and pollution thought to be operating on each site.

The reporting of Phase 2 is divided into three units :

Unit I: Taxon distribution studies : R&D Technical Report E103 (Davy-Bowker *et al.* 2000)
Unit II: Changes in biological condition : R&D Technical Report E101 (Clarke *et al.* 1999)
Unit III: Post-survey appraisal : R&D Technical Report E102 (Furse *et al.* 1999) - this report

Unit I contains:

- a description of the incorporation of the 1990 RQS and 1995 GQA survey biological and environmental data into IFE's Quinquennial Survey Database (QSD). This includes procedures to establish matching pairs of sampling locations for use in analyses of change between surveys

- distribution studies of each BMWP taxon providing information on their geographic distribution, their environmental ranges (in terms of the RIVPACS environmental variables) and their tolerance/susceptibility to particular sources of environmental stress thought to be operating at individual sites

Unit II (this report) contains summaries of the :

- patterns of distribution of biological condition in 1995, especially in relation to RIVPACS environmental variables

- changes in biological condition between matched sites in 1990 and 1995, incorporating measures of the statistical significance of change in biological grade

- changes in biological condition in relation to site environmental characteristics and ITE landscape type

- data obtained from Environment Agency Regions on the known or suspected sources of environmental stress operating on each of the GQA sites

- relationships between biological condition or change in condition and the type and severity of any environmental stress or pollution

Unit III contains summaries of the :

- history of national River Quality Surveys in the Environment Agency area (England and Wales)
- responses to the 1995 GQA post-survey questionnaire to Agency staff developed within this project
- results and conclusions from an investigation using the bias specification options in RIVPACS III+ to assess the effect of alternative analytical quality targets for macro-invertebrate samples
- analysis of the 1995 quality audit to determine which factors, if any, can be associated with poor levels of performance
- recommendations for future surveys

1.2 Specific Objectives

The specific objectives addressed in this report were to:

- review the effectiveness of the biological component of the survey in meeting its objective of assessing the biological condition of the watercourses in the Environment Agency Regions
- make recommendations that maximise the application of the biological data collected during the survey for other Agency operational purposes
- consider the implications of the preceding analyses for the refinement of the methodology for future surveys

2 THE HISTORY OF NATIONAL RIVER QUALITY SURVEYS

2.1 Context

A brief history of the development of national River Quality Surveys of England and Wales, the Environment Agency area, is presented here. Attention is focused on the biological component of the survey and the methodologies associated with the collection, interpretation and presentation of the data. Where known, brief details are provided of post-survey appraisals of the biological component of the survey.

The purpose of this chapter is to provide a context for the procedures adopted in the 1995 General Quality Assessment (GQA) and for the post-survey appraisal of these procedures.

2.2 The 1958 Survey

The first national survey of the quality of British rivers was conducted in 1958 by the Ministry of Housing and Local Government. The entire survey, the results of which were not published, was based on chemical data and was reported to be "less exact and less thorough" than the 1970 survey (Department of the Environment & The Welsh Office 1971). The survey involved no biological monitoring.

2.3 The 1970 Survey

The 1970 survey, entitled the "River Pollution Survey" included both chemical and biological monitoring. In order to facilitate biological monitoring and appraisal, a first attempt was made to produce a standard national biological classification scheme for evaluating the quality of rivers (Department of the Environment & The Welsh Office 1971). The scheme was a simple one in which four quality classes were recognised. Allocation to classes largely depended on the relative frequency of the freshwater shrimp ("Amphipoda") and of three orders of insects: Ephemeroptera, Plecoptera and Trichoptera. The nature of the fishery was also taken into consideration but no quantitative rules or guidelines were provided.

Allocation to quality class was inevitably a subjective process depending on how the practitioner distinguished between phrases such as *"an appreciable proportion of Plecoptera and/or Ephemeroptera, Trichoptera and Amphipoda"*, which partly defined Class A, and *"Plecoptera and Ephemeroptera populations may be restricted. Trichoptera and Amphipoda usually present in reasonable numbers"*, which partly defined Class B. Furthermore, it appears that no attempt was made to provide a standard data collection protocol and that existing data was often used in lieu of new sampling programmes. Thus, in volume 1 of the report of the survey (Department of the Environment & The Welsh Office 1971) it was stated that:

> *"...biological information about non-tidal rivers was supplied where available, mainly in the first instance to enable some studies to be made of the co-relation between chemical and biological classifications."*

The 1970 survey report contained no immediate post-survey appraisal of the effectiveness of the procedures used to collect, analyse and report the biological results of the survey. However, the report of the survey included an acknowledgement that:

> *" there must be some reservation over the adequacy of the* [biological] *classes as now defined for application to all types of rivers. Further consideration is being given to the classification and it is possible that there may be a revision of the definition of the biological classes in the future."*

2.4 The 1972 Survey

The 1972 survey involved an update of the findings of that conducted in 1970. The survey was exclusively chemical and the reason given for excluding biological monitoring was that the biological classification used in the 1970 survey did not apply equally to all rivers (Department of the Environment & The Welsh Office 1972).

2.5 The 1973 Survey

The only new sampling undertaken in the 1973 survey was chemical. However, as part of the reporting exercise, authorities were asked to re-appraise the results they submitted in 1970, this time using a revised biological classification

Details of the process leading to the revised biological classification are given in the report of the 1973 survey (Department of the Environment & The Welsh Office 1975). The impetus for change arose from the comparison of the chemical and biological classifications in place in 1970. Indeed, this comparison was seen to be as important an objective of the 1970 survey as making a biological assessment of national water quality.

When the biological and chemical classifications were compared, it was shown that there was considerable regional variation in the degree of concordance of the two methods of appraisal. Inter-regional differences were attributed to the influence of varying current velocities on the biological response to pollution (Department of the Environment & The Welsh Office 1975).

The new biological classification, used for the first time in 1973, resulted from a meeting of river authority biologists held at the Department of the Environment. The new classification contained only minor revisions from the 1970 version (Department of the Environment & The Welsh Office 1975), except that it was requested that River Authorities were asked to treat fish as secondary in importance to macro-invertebrates when applying the system. No new data were requested for the 1973 survey and there were no new published instructions on the protocols to be used to collect and sort the macro-invertebrate samples.

2.6 The 1975 Survey

The residual short-comings of the biological classification scheme used in the 1973 survey were such that the use of biological data was once again omitted from the major 1975 survey (Department of the Environment & The Welsh Office 1978).

Reasons for the exclusion of biology from the 1975 survey included the perception that more needed to be first known about the ecological requirements of plants and animals in relation to flow, substratum type and chemical water quality. Nonetheless, it was recognised that plants and animals are valuable indicators of new, intermittent or unsuspected pollution and that a suitable procedure was required in order to reinstate biological monitoring in future surveys. The Department of the Environment & The Welsh Office (1978) stated:

> "*There is no doubt that* in situ *assessments of water quality in biological terms are extremely valuable"*

2.7 The Biological Monitoring Working Party

In order to devise that "more satisfactory" biological classification a working party, the Biological Monitoring Working Party (BMWP), was commissioned in 1976. Its terms of reference were to:-

- recommend a biological classification of river water quality for use in the River Pollution Survey
- consider ways and means of implementing that classification
- consider relationships, if any, between chemical and biological classifications

Significantly, in making its final report, the working party felt unable to recommend a system of biological classification of *"river quality"* (Biological Monitoring Working Party 1978). Instead it recommended a procedure, named the BMWP Score system, for assessing the *"biological condition"* of a river. This system underwent a series of revisions (Chesters 1980, National Water Council 1981) before being used in the assessment of data collected during the renamed "River Quality Survey" of 1980. Details of the evolution, format and application of the BMWP score system were belatedly introduced into the public domain by Hawkes (1997).

Hawkes (1997) outlined the working party's conclusions concerning the sampling procedures to be adopted when applying the BMWP score system. In summary, they recognised the advantages of adopting a standardised method but found themselves unable to lay down definitive procedures because of the wide range of river types in which those procedures would have to be applied. In Hawke's (1997) view, this failure to standardise procedures was largely responsible for the deviations in the scores achieved in the early application of the score system.

The working party also considered that chemical and biological data provided different but complementary measures of the condition of a site (Biological Monitoring Working Party 1978). They felt that the biological assessment was of greatest value when it failed to match that interpreted from chemical analyses and stated that:-

> *"it does not serve any purpose to attempt to correlate the results of the chemical and biological assessments. If correlations were established there would be no justification to carry out both forms of assessment."*

2.8 The 1980 Survey

With the advent of the BMWP score system, biological monitoring was re-introduced, into the 1980 survey, alongside chemical monitoring,. The results of the survey were published in text and map form. On the maps, the biological condition of sites was represented by BMWP scores depicted alongside each individual site (National Water Council 1981).

In the report of the survey it was stated that, in general, the higher the BMWP score the better the biological condition of the site (National Water Council 1981). However, it was also acknowledged that:

> *"interpretation of the biological scores is .. a matter for professional experts as the diversity of families present at a site depends not only on the degree of any pollution but also the nature of that pollution and, more particularly, on what would be present without any pollution."*

The authors of the report recognised that there were substantial differences between the fauna of lowland and upland streams. They drew attention to work underway at the Freshwater Biological Association which was aimed at developing a better understanding of the differences in the natural macro-invertebrate community composition in rivers of different environmental types. It was this work which was to lead on to the development of the system which came to be known as RIVPACS.

On the basis of contemporary understanding, it was recommended that the results of the 1980 biological survey primarily represented a basis for future comparisons.

2.9 The Development of RIVPACS

The aim of the research programme begun at the FBA and continued by its successor organisation, the Institute of Freshwater Ecology (IFE), was to quantify the links between the environmental characteristics of sites and the macro-invertebrate assemblages that occur at them when they are unstressed by physical or chemical perturbations. The output of the research programme has been a series of increasingly sophisticated versions of a software package known as the River InVertebrate Classification and Prediction System (RIVPACS) (Wright *et al*.1993).

Essentially RIVPACS is a system of prediction by analogy. Through the use of multivariate statistical procedures, the system provides a prediction of the fauna which should be captured at a site, using standard sampling methods, if that site is not significantly stressed (Wright *et al*.1993). On this basis, each site can be provided with a specific biological target against which its observed (ie sampled) fauna can be judged. The degree of compliance between the expected (ie RIVPACS-predicted) fauna and that observed has been quantified in the form of the Ecological Quality Index or EQI (Sweeting *et al*. 1992).

The EQI of a site is the ratio of its observed BMWP index value and that predicted by RIVPACS. It can take three forms depending upon whether the function used is the BMWP score, number of scoring taxa or ASPT. In each case the EQI of a site is unity if the observed index value exactly meets expectations but zero if no taxa are present. Most sites lie within this value range but EQIs of some sites have a sufficiently diverse and high-scoring fauna that their EQI values can exceed one.

Unlike previous indices, the EQI provided the opportunity to make direct and meaningful comparisons between the fauna of sites of entirely different character or geographic location. This is because the EQI is a measure of the extent to which each individual site meets its biological potential and this is a common factor by which all sites may be judged.

2.10 The 1985 Survey

With the exception of estuaries, which had their own specific systems of evaluation, biological monitoring was once again dropped from the 1985 River Quality Survey.

The freshwater survey was thus based on chemical monitoring alone. No formal reason for this was presented in the report of the survey (Department of the Environment and the Welsh Office 1986) but it is probable that the non-availability of an operational system, such as RIVPACS, that would allow direct spatial comparisons between sites, was an important contributory cause.

2.11 Development of a System for Classifying the Biological Condition of Sites.

Version I of RIVPACS was made available for operational use in 1988, with RIVPACS II released to the NRA in 1990. It was version II, therefore, that was current at the time of the 1990 survey. This version contained information on the macro-invertebrate assemblages and environmental characteristics of 438 sites throughout Great Britain.

The EQI values generated by RIVPACS II could also be divided into a series of numerical bands (or grades, as more commonly now used) for classifying the biological condition of sites. The principles and suggested practice of grading were outlined to the NRA in a series of IFE reports (Wright *et al.* 1991, Clarke *et al*. 1992, Clarke *et al*. 1994). In each of these reports the highest quality grade width was set at the level attained or exceeded by a set percentage of the sites in RIVPACS II, which in turn were perceived to be sites with the best achievable biological condition for their environmental type. Different grade widths were suggested for different functions of the BMWP system and different seasonal combinations of samples.

For the purposes of the 1990 River Quality Survey, a 95%ile attainment rate was suggested to set the lower limit of the highest ASPT band, whereas a 90%ile was suggested for number of scoring taxa and ASPT.

2.12 The 1990 Survey

The availability of a tested and fully operational version of RIVPACS provided an important spur for the re-inclusion of biology, alongside chemical monitoring, in the 1990 River Quality Survey. A total of 23,083 biological samples were collected, for survey purposes, from 8796 sites throughout the United Kingdom.

The procedures for undertaking the biological survey were the subject of detailed advanced discussions and a formal set of instructions were issued, as described in the final report of the survey (National Rivers Authority 1991). These covered when to sample (separate samples in each of spring, summer and autumn), the choice of sites, the objectives of macro-invertebrate sampling, the sampling methods and equipment, sample sorting procedures, levels of identification, assignment of abundance categories and a biological grading system based on EQI values derived from RIVPACS II (Table 2.1).

Table 2.1 The EQI grade ranges (the 5M system) for three separate forms of BMWP Index; BMWP score (EQI_{BMWP}), number of scoring taxa (EQI_{Taxa}) and ASPT (EQI_{ASPT}), as used in the 1990 River Quality Survey. Grade "a" represents the best biological condition and grade d represents the worst.

Grade	EQI_{BMWP}	EQI_{Taxa}	EQI_{ASPT}
a	≥ 0.75	≥ 0.79	≥ 0.89
b	≥ 0.50	≥ 0.58	≥ 0.77
c	≥ 0.25	≥ 0.37	≥ 0.66
d	< 0.25	< 0.37	< 0.66

The overall grade of a site was based on calculating the three separate indices for site taxon lists based on combining the results of each individual season's samples. The overall biological grade of the site was then determined by the EQI_{ASPT} grade, where this was ranked lowest, or, where this was not the case, the middle ranking EQI.

2.13 External Quality Audit

In an attempt to understand the errors associated with the sorting and identification of macro-invertebrate samples collected during the 1990 River Quality Survey and to try and impose a uniformity of standard, the NRA let a contract to the IFE to re-sort and identify a sub-set of samples from each Agency Region.

In order to achieve this, specimens removed from each sample were retained in a single labelled vial per sample. The residual sorted material from each sample, together with any unpicked specimens of macro-invertebrates were reconstituted and re-fixed or re-preserved in a labelled container or containers. The sealed vial of identified specimens was placed in the appropriate container of re-constituted sample from its site and season of origin. All but a very small number of samples were then sent to the IFE River Laboratory for long-term storage and for external Quality Audit.

A sub-set of sixty samples per NRA Region were randomly selected, for Quality Audit, by the IFE. The sample taxon lists obtained for each audit sample by the NRA were requested by IFE from the appropriate Area or Regional Laboratory. Samples were then re-sorted by the IFE, who also checked the identification of the animals stored in the sealed vial. The taxon list compiled by IFE was compared to the list obtained by the NRA and differences between the lists noted and returned to the NRA for their consideration and action. Most changes involved taxa found by IFE but not by the NRA. These were termed gains. Taxa claimed by the Agency but not found by IFE were termed losses. The net gain for a site was the gross number of taxa gained minus the gross number of taxa lost.

Further details of the audit procedures are given in a series of separate reports produced for each NRA Region by IFE (e.g. Gunn *et al.* 1991), in an internal Agency procedures manual, BT003 (Environment Agency 1996a and later versions), and in a forthcoming publication by Dines & Murray-Bligh (in press).

Following the introduction of the audit there was a gradual reduction in the mean number of net gains per sample in most Regions between spring and autumn of 1990 (e.g. Gunn *et al.* 1991).

The external Quality Audit provided by IFE was extended on a similar basis until 1994.

2.14 Internal Analytical Control

For the 1995 river quality survey and beyond, the NRA (subsequently Environment Agency) introduced their own internal Analytical Quality Control (AQC) procedure for evaluating the performance of their staff at sorting and identifying macro-invertebrate samples. The AQC procedures adopted were devised and developed for the NRA by the Water Research Centre (WRc) (Kinley & Ellis 1991, van Dijk 1994), and subsequently modified by the Environment Agency (1996a).

Full details of the AQC procedures were set out in the WRc reports and by the Environment Agency Manual BT003 (Environment Agency 1996a and later versions). They are outlined more briefly in Environment Agency Manual BT001 (1997a and later versions) and Dines & Murray-Bligh (in press).

As a general principle, the AQC standards were set such that an average of no more than two BMWP taxa should be missed per sample (Dines & Murray-Bligh in press).

As a consequence of the introduction of the internal AQC procedures, subsequent IFE Quality Audits were undertaken on both "primary" samples (i.e those not also subject to internal AQC) and also samples that had received both primary processing and an internal AQC check (e.g. Gunn *et al*. 1996a, 1996b)

2.15 The 1995 Survey

The 1995 survey, termed the 1995 General Quality Assessment (GQA), again included both biological and chemical sampling.

Although the instructions given for biological monitoring in 1990 had been much more precise than ever previously, the NRA felt that the extent of compliance with instructions had been variable between Regions and Areas within Regions. Therefore the instructions for biological sampling in 1995 were even more prescriptive than those given previously. These instructions were developed within the NRA, following outside consultation with IFE and others, and were documented in two internal reports (Environment Agency 1996b (Manual BT002), 1997a (Manual BT001)). BT002 was a specifically concerned with the 1995 GQA whereas BT001 concentrated on the collection of macro-invertebrate samples in connection with RIVPACS.. Although the final drafts of each document post-dated the survey, both were sufficiently well developed by the beginning of 1995 to be circulated to each Agency Region as the definitive procedures to be adopted during the 1995 GQA.

The procedures adopted for the 1995 GQA included a revised system for grading the biological condition of sites. The number of EQI index types used was reduced to two (EQI_{Taxa} and EQI_{ASPT}), the number of separate grades was increased to six (a-f – Table 2.2), to match the chemical classification, and the number of sampling seasons was reduced to two (spring and autumn).

Table 2.2 The EQI grade ranges for two separate forms of BMWP index; number of scoring taxa (EQI_{Taxa}) and ASPT (EQI_{ASPT}), as used in the 1995 GQA.

Grade	Description	Lower grade limits	
		EQI ASPT	EQI number of taxa
a	Very good	1.00	0.85
b	Good	0.90	0.70
c	Fairly good	0.77	0.55
d	Fair	0.65	0.45
e	Poor	0.50	0.30
f	Bad	0.00	0.00

The overall site grade, based on the two seasons combined taxon lists, was now the lower of the separate grades derived from EQI_{Taxa} and EQI_{ASPT}

The results of the 1995 GQA Survey were published in very little detail in a brief summary report (Environment Agency 1997b). Full details of the 1995 GQA biological survey were confined to an unpublished internal report (Environment Agency 1996c).

2.16 The Development of RIVPACS III+

The objectives of the quinquennial River Quality Surveys are to provide an overview and summary of the condition of British watercourses and to provide an indication of the temporal trends in change of quality (Environment Agency 1997b) in order to best formulate river management strategies.

The introduction of RIVPACS III and the use of EQIs provided a much more rigorous method of grading sites and for making comparison between sites on different types of river. Nevertheless, in 1990, biological grading still lagged behind the chemical grading system in terms of its ability to be used to estimate errors in grade (class) allocation and the detection of temporal change. This situation was largely resolved by further work undertaken for the NRA (and subsequently Environment Agency by the IFE (Furse *et al*. 1995, Clarke *et al*. 1997, Clarke in press). The research programme was directed at identifying the variation associated with the collection of macro-invertebrate samples for use with RIVPACS, the errors and systematic biases associated with the sorting and identification of samples and the errors associated with measuring of environmental variables used to make RIVPACS predictions. Collectively, variation, biasses and errors were termed uncertainties (Clarke in press). Uncertainties associated with collecting biological and environmental data were studied by replicate sampling/measurements. Uncertainties associated with sorting and identification were derived from the results of IFE's internal audits of NRA/Environment Agency samples.

Furse *et al*. (1995) found that uncertainties in biological and environmental data were independent of the type or biological condition of the site being sampled. Thus, they found that, for operational purposes, the square roots of the number of taxa captured and BMWP score of a site should be assumed to have constant sampling variances dependant upon the number of season's data being used to represent the site. They also recommended that sampling variance of the observed ASPT should be best estimated by a series of constants that are also dependent on the number of season's data being used.

Similarly, it was recommended that the errors in expected values of BMWP indices due to variation in environmental measurements are expressed by constant standard deviations, irrespective of the number of samples used to derive the index but dependant on whether the index used is number of taxa, BMWP score or ASPT.

Finally, they considered the errors introduced by the failure of NRA/Agency staff to remove specimens of all families present in a sample or by incorrectly identifying those specimens removed (Furse *et al*. 1995). By analysis of the results of external audits conducted by IFE (e.g. Gunn *et al.* 1991) they demonstrated that sorting and identification errors led to fewer taxa being correctly identified as present in a sample than was truly the case. They termed this systematic error or net underestimate of the number of taxa present as the "bias" and this term is used throughout this report and in RIVPACS III+. They recommended that the bias associated with sorting a single sample should be set at the most recent performance rate for the organisation or laboratory concerned, as judged by external Quality Audit (Furse *et al*. 1995). Where combined seasons RIVPACS predictions and EQI calculations were being made, Furse *et al.* (1995) found that about 50% of taxa missed in a single sample were not subsequently found in a second sample from that site in another season of the same year. Furthermore, 37% of taxa missed in a single sample were not found in either of the samples taken in the two other sampling seasons. Based on these data, procedures were provided for adjusting for bias in each of number of taxa, BMWP score and ASPT for samples.

The relationships developed by Furse *et al*. (1995) were incorporated into the version of RIVPACS available at the time of the 1995 GQA. This was RIVPACS III and it contained two linked modules. Module one was for Great Britain and contained 614 sites in 35 groups, whilst module 2 was for Northern Ireland and contained 70 sites in seven groups. The new version of RIVPACS, incorporating the uncertainty procedures, was termed RIVPACS III+ and the first operational version was released in 1997 (Clarke *et al*. 1997). The new features it contained comprised facilities to:

- enter numerical ranges for one or more sets of grades of biological condition

- enter estimates of the net-underestimation (bias) of the number of BMWP taxa present in each sample, as regulated by the analytical quality control and audit procedures used in specific sampling programmes

- calculate confidence limits for EQI index values for samples

- test for the statistical significance of the difference between EQI index values of pairs of samples, where each of the pair may be based on one single collection or combined faunal lists of any seasonal combination of single collections made in spring, summer and autumn

- assign samples probabilistically to defined grades of biological condition, where each sample may be based on one single collection or combined faunal lists of any seasonal combination of single collections made in spring, summer and autumn

- allow for bias in the calculation of confidence limits, application of statistical tests and probalistic assignment of samples to grades of biological condition

The pre-publication release of the results of the IFE study (Clarke *et al.* 1997) allowed the Environment Agency to undertake their own preliminary statistical analyses of changes in the biological condition of sites between 1990 and 1995. Results were documented in an unpublished internal Agency report (Warn, 1996).

The application RIVPACS III+ to the results of the 1995 survey has underpinned much of the contents of two sister reports to the current document (Clarke *et al.* 1999, Davy-Bowker *et al.* 2000).

2.17 The 2000 GQA Survey

An extensive biological and chemical GQA is planned for 2000. The biological survey will benefit from the advances made in the standardisation of procedures, quality control and audit schemes and the development of RIVPACS III+. However, the precise procedures implemented during the 2000 GQA Survey will benefit from a critical review of the procedures used in 1995. This document contributes to that review.

3 METHODS

3.1 General Approach

A questionnaire, sent to selected Environment Agency biologists in each Region, was the principal method of acquiring the information and opinions upon which the major section (Chapters 2 and 3) of this report is based.

In addition, separate investigations of two specific issues, of relevance to the objectives of the appraisal of the effectiveness of the survey, were undertaken by IFE.

These were to:

- investigate the use of the error module in RIVPACS III+ for setting the most cost-effective analytical quality targets.
- analyze the results of the external audits of 1995 GQA macro-invertebrate samples in order to determine which factors, if any, can be associated with poor levels of performance.

The methods adopted for the production and circulation of the questionnaire are described in the following sections but the methods adopted for the two specific investigations are described in their single relevant chapters, (Chapters 4 and 5 respectively).

3.2 Development of the Questionnaire

The questionnaire was designed to elicit the opinions of Environment Agency biologists on all aspects of the design, implementation and reporting of the 1995 survey including field and laboratory procedures and equipment, sample processing and identification, quality control measures and data analysis, presentation and use. In addition to providing information on the 1995 procedures, the questionnaire was so designed as to allow recipients to make suggestions on the ways in which they could be improved for future surveys

Each draft of the questionnaire was in large font format, to reduce the apparent density of questionning and, wherever possible simple tick boxes were provided with multiple choice response options.

The first draft of the questionnaire was produced by IFE and circulated to the Project Leader (Dr R A Dines) and Project Board member, Dr J A D Murray-Bligh in January 1999. This document was designed to be as comprehensive as possible in order to ensure that all aspects of the survey procedures were fully covered. This draft was 43 pages long and contained 131 separate questions.

Circulation of the document was followed by an IFE/Environment Agency meeting on 1st March, 1999. Attending that meeting were Mike Furse, Ralph Clarke and John Davy-Bowker, representing IFE, and Bob Dines and John Murray-Bligh, representing the Agency. At that meeting, the Environment Agency prioritised the key issues that needed to be dealt with and recommended elimination of a number of other questions in order to shorten the length of the questionnaire and increase the chances of a positive response. It was also decided to divide the questionnaire into two sections, main and supplementary and to encourage recipients of the questionnaire to regard a response to the main part as obligatory and the second part as optional but valuable.

An initial revision of the questionnaire was produced by IFE and circulated to Bob Dines and John Murray-Bligh for their final comments. Their comments on the revised draft were acted upon and a final draft for circulation was produced and approved. The main part of the revised questionnaire comprised 54 questions on 24 pages. This part was divided into ten sections: number of sites, sampling procedures, sample processing, sample identification, quality control, quality grading, other approaches, environmental data, equipment, reporting and general. The supplementary part of the questionnaire comprised 13 questions on seven pages. This part was divided into three sections: number of samples, sample processing and quality grading.

3.3 Circulation of the questionnaire

At the 1st March IFE/Agency meeting it was decided that a single copy of the two-part questionnire would be circulated to the Regional Biologist, or nominated substitute, in each Agency Region (Table 1.3) and that biologist would be asked to complete and return the documents after consultation with relevant colleagues in each of the Regions' Area Laboratories. It was also agreed that complimentary copies be sent to the Scottish Environmental Protection Agency (SEPA) and DoE (Northern Ireland), with a covering letter to state that responses were welcome but should not be considered obligatory.

Table 3.1 The named biologist in each Agency Region or country to whom the questionnaire was circulated. RPBs = River Purification Boards

Region			Biologist
1999 (Environment Agency)	1990 RQS (NRA)	1995 GQA (NRA)	
Anglian	Anglian	Anglian	Sarah Chadd
North East (Two Regions in 1990)	Northumbrian Yorkshire	Northumbria & Yorkshire	Brian Hemsley -Flint
North West	North West		Elaine Fisher
Midlands	Severn-Trent	Severn-Trent	Shelley Howard
Southern	Southern	Southern	Bob Dines
South West (Two Regions in 1990)	South West Wessex	South Western	George Green
Thames	Thames	Thames	Paul Logan
Welsh	Welsh		Graham Rutt
SEPA (Scotland)	RPBs	RPBs	David Lowson
DoE (Northern Ireland)	DoE (NI)	DoE(NI)	Peter Hale

Prior to circulation of the questionnaire, each recipient was circulated with a memo from Bob Dines (Agency Project Leader), informing them of the imminent arrival of the questionnaire and encouraging them to complete it after consultation with senior colleagues involved in the 1995 survey. The memo was circulated on 12th March, 1999. The main questionnaire, and an accompanying letter, were circulated to all selected recipients on 1st April, 1999. Replies were requested by 3rd May but this date was subsequently extended to 31st May at the request of some Regions.

3.4 Presentation And Interpretation Of The Responses

The replies to each question on the main and supplementary questionnaire are presented on a question-by-question basis in Chapter 4. Interpretation of the responses and principal recommendations are presented in Chapter 5. In the latter chapter, the opinions of the respondents are supplemented by the opinions and interpretations of the authors of this report.

4 RESULTS OF THE QUESTIONNAIRE

4.1 The Level of Response

Each of the eight circulated Regions responded in detail to the main questionnaire (Table 4.1), as did SEPA and DoE (Northern Ireland). Most Regions also provided answers to some, or all of the questions on the supplementary questionnaire.

The level of consultation between the Senior Biologists and their colleagues was not always stated clearly on the returned forms but seemed to vary between Regions.

Table 4.1 The response to the questionnaire. M = Main questionnaire and S = Supplementary questionnaire. Regions as per Environment Agency 1999.

Region	Respondant	Colleagues stated as having been consulted	Sections completed
Anglian	Sarah Chadd	Richard Chadd Julia Stansfield	M
North East	Brian Hemsley-Flint (Ridings)	Elaine Axford (Dales) Jim Heslop (Northumbria)	M & S
North West	Elaine Fisher	Ed Mycock Ray Prigg & the Biology teams	M
Midlands	Shelley Howard	Gary Fretwell Phil Harding Alan George Ayleen Clements (each completed all or part of the form independently before collation by Shelley Howard)	M & S
Southern	Bob Dines	Phil Smith Shirley Medgett Shelagh Wilson	M & S
South West	George Green	None	M & S
Thames	Paul Logan	John Murray-Bligh Judy England Julie Jeffery Claire Gladdy	M & S
Welsh	Graham Rutt	None	M & S
SEPA (Clyde)	Maureen Cook	None	M & S
SEPA (Solway)	David Rendall	None	M & S
DoE (NI)	Peter Hale	Imelda O'Neill	M & S

In assessing the replies to the questionnaire it needs to be borne in mind that the answers given for the South West Region largely apply to the current Agency South Wessex Area and may differ from the views of the staff of the other Area Laboratories. Furthermore the replies from SEPA are based solely on the views of biologists from the former Clyde and Solway River Purification Boards (RPBs) and may not represent the views of the other former RPB's.

4.2 The Replies to Each Question

The text provided in this section is taken, verbatim, from the written and electronic replies received from Environment Agency staff.

4.2.1 The main questionnaire

Question 1: For each of the following categories of watercourse type, please indicate your opinion on the adequacy of the number of sites sampled in order to get a reliable representation of the biological condition of rivers in your Region. Tick only one category per watercourse type.

All watercourses as a whole

Anglian	Approximately right number
North East	Slightly fewer than necessary
Midlands	Approximately right number
Southern	Slightly fewer than necessary
Thames	Slightly fewer than necessary
Welsh	Slightly fewer than necessary
SEPA (Clyde)	Many fewer than necessary
DoE (NI)	Many fewer than necessary

Clean large, deep rivers

Anglian	Approximately right number
North East	Approximately right number
Midlands	Approximately right number
Southern	Approximately right number
South West	Approximately right number
Thames	Slightly fewer than necessary
Welsh	Approximately right number
SEPA (Clyde)	Approximately right number
DoE (NI)	Slightly fewer than necessary

Polluted large, deep rivers

Anglian	Approximately right number
North East	Approximately right number
Midlands	Approximately right number
Southern	Approximately right number
Welsh	Approximately right number
SEPA (Clyde)	Approximately right number
DoE (NI)	Slightly more than necessary

Question 1 (continued): The adequacy of the number of sites sampled

Clean middle reaches

Anglian	Approximately right number
North East	Approximately right number
Midlands	Approximately right number
Southern	Approximately right number
South West	Approximately right number
Thames	Approximately right number
Welsh	Approximately right number
SEPA (Clyde)	Slightly fewer than necessary
DoE (NI)	Approximately right number

Polluted middle reaches

Anglian	Approximately right number
North East	Approximately right number
Midlands	Approximately right number
Southern	Approximately right number
Thames	Approximately right number
Welsh	Approximately right number
SEPA (Clyde)	Slightly fewer than necessary
DoE (NI)	Approximately right number

Clean upland headwaters

Anglian	Many fewer than necessary
North East	Many fewer than necessary
Midlands	Approximately right number
Welsh	Approximately right number
SEPA (Clyde)	Many fewer than necessary
DoE (NI)	Many fewer than necessary

Polluted upland headwaters

Anglian	Slightly fewer than necessary
North East	Slightly fewer than necessary
Midlands	Approximately right number
Welsh	Many fewer than necessary
SEPA (Clyde)	Approximately right number
DoE (NI)	Many fewer than necessary

Question 1 (continued): The adequacy of the number of sites sampled

Clean lowland headwaters

Anglian	Many fewer than necessary
North East	Slightly fewer than necessary
Midlands	Slightly fewer than necessary
Southern	Many fewer than necessary
South West	Many fewer than necessary
Thames	Many fewer than necessary
Welsh	Approximately right number
SEPA (Clyde)	Approximately right number
DoE (NI)	Many fewer than necessary

Polluted lowland headwaters

Anglian	Slightly fewer than necessary
North East	Approximately right number
Midlands	Approximately right number
Southern	Many fewer than necessary
South West	Many fewer than necessary
Thames	Slightly fewer than necessary
Welsh	Many fewer than necessary
SEPA (Clyde)	Approximately right number
DoE (NI)	Many fewer than necessary

Acidified sites

Anglian	Approximately right number
North East	Approximately right number
Midlands	Slightly fewer than necessary
Southern	Approximately right number
Welsh	Slightly fewer than necessary
SEPA (Clyde)	Slightly fewer than necessary
DoE (NI)	Slightly more than necessary

Agriculturally enriched sites

Anglian	Approximately right number
North East	Slightly fewer than necessary
Midlands	Approximately right number
Southern	Many fewer than necessary
South West	Approximately right number
Thames	Approximately right number
Welsh	Slightly fewer than necessary
SEPA (Clyde)	Many fewer than necessary
DoE (NI)	Approximately right number

Question 1 (continued): The adequacy of the number of sites sampled

Urban watercourses

Anglian	Approximately right number
North East	Approximately right number
Midlands	Slightly more than necessary
Southern	Approximately right number
South West	Approximately right number
Thames	Slightly fewer than necessary
Welsh	Approximately right number
SEPA (Clyde)	Approximately right number
DoE (NI)	Approximately right number

Canals

Anglian	Approximately right number
North East	Slightly fewer than necessary
Midlands	Slightly more than necessary
Southern	Approximately right number
Welsh	Approximately right number
DoE (NI)	We are excluding them

Drains and ditches

Anglian	Approximately right number
North East	Slightly fewer than necessary
Midlands	Approximately right number
Southern	Many fewer than necessary
Welsh	Approximately right number
DoE (NI)	We are excluding them

Other

Anglian	Brackish water: Slightly fewer than necessary
Welsh	Sites potentially affected by sheep dip: Many fewer than necessary

Question 2: If you feel that there are any differences between your component Areas/laboratories, in the adequacy of the number of sites sampled, which are significant and need recording, please give them:

Anglian	The responses did differ, probably reflecting the distribution of watercourse types across the Region. For example Northern Area felt that slightly fewer than necessary drains and ditches were sampled whilst Eastern Area felt that in their Area the number was approximately right. Apart from headwaters it is felt that the spread of sampling is generally about right, covering all types of 'river'.
North East	Dales have too few headwater and acid streams Northumbria has no canals hence no sites
North West	This will be looked at in the stretch review in preparation for GQA2000.
Midlands	Upper Severn - fewer sites than necessary overall, fewer agriculturally enriched sites There was a high proportion of clean watercourses not adequately covered. Canals were optional and most Areas ignored most of the stretches. One Area sampled out of season and results not reported. This will affect perceived need. Upper Trent, which has the most canal stretches, thought the number of stretches in the GQA network were about right but as these were not sampled according to the full protocol considered the actual survey inadequately covered the canals. Lower Severn comment: no sampling done on Severn estuary (left to South West and Welsh Regions)
Southern	Hampshire is largely chalk and has relatively few headwaters. Kent and Sussex have most of the drains and ditches.
South West	Not qualified to respond
Thames	No comment
SEPA (Clyde)	Coverage of sites in former Solway RPB Area inadequate, this has now been remedied.
DoE (NI)	We monitor approximately 2,500 km we require 2X this length. The number of sites is therefore inadequate for purpose, which is an issue that we are attempting to resolve.

Question 3: Please provide any additional comments you may wish to make on the number and type of sites sampled.

Anglian	Samples from upper/middle stretches of chalk streams (Lincolnshire Wolds), headwaters in general and extensive system of fen drains somewhat lacking in 1995 survey. Sampling of clean canalised stretches possibly excessive (lower river stretches and actual canals). Some definitions of headwater, middle, large and deep would help standardisation.
North East	This will hopefully be addressed in the review of biology monitoring to be carried out this year.
Southern	The main deficiencies are in headwaters – this explains the poor cover for acidified sites and agricultural influences. Sussex estimates an additional 50+ sites to give adequate coverage. We have few sites on winterbournes – better sampling methods may be needed (Phil Smith suggested this – Shelagh Wilson questions whether these should have a place in a basic surveillance programme as they would need specialised techniques and sample timing. I (Bob Dines) am inclined to her view.).
South West	None.
Thames	No comment.
Welsh	In early 1994 NRA Welsh Region, as was, carried out a GQA sampling site review in order to rationalise the number of sites sampled. The review was principally geared to water quality sampling points and LAPWING analysis was used to look for redundancy in sites. Sites were eliminated where a site downstream could adequately describe quality within a stretch further up provided that the biological quality was similar (defined, from memory, as being the same GQA grade). This review resulted in a considerable number of sites being cut out of the survey especially in what were thought to be clean headwaters. Recent experience shows, however, that these headwaters can be seriously affected by sheep dip pesticides such that quality assessment in headwaters may have been artificially high in Welsh Region. There is one case, for example, of 3 upland headwater stretches being defined sampling within in the 4th stretch down from the top of the catchment. In autumn 1998 the three topmost stretches were practically afaunal due to sheep dip pollution – who knows what the position was up there in 1995!
DoE (NI)	If the numbers sampled is inadequate then so are the types. We hope to extend RIVPACS on this basis next year when we are in a better position to know what we are dealing with.

Question 4: For the purposes of the 1995 GQA, a manual (BT 001; Version 1.0) was produced that offered recommendations on the manner in which samples should be collected during the survey The recommended methods for collecting macro-invertebrate samples exclusively using a pond-net are laid out in sections 3.7.1 to 3.7.3f (pp 3.25 - 3.32) and 3.7.6 (p 3.41). Did you fully follow these procedures for collecting samples in 1995 GQA?

Anglian	Yes
North East	Yes
North West	Yes
Midlands	Yes
Southern	Yes
South West	Yes
Thames	Yes
Welsh	Yes
SEPA (both)	Yes
DoE (NI)	Yes

Question 5: If you answered no to Q4, what modifications did you make to the recommended method?

Midlands	1 min search may have been variable

Question 6: If you answered yes to Q4, in what way, if any, did the method you adopted for sampling by pond-net and search in the 1995 GQA differ from the procedures you generally adopted prior to that survey? Please include the main differences, if any between the approaches adopted for the 1990 RQS and 1995 GQA.

Anglian	No difference in Northern Area. In Eastern Area, samples taken prior to 1995 were generally not timed. Sampling in 1990 involved first sampling the riffle area, then the margins and then doing a hand search. It did not, therefore, sample habitats in proportion to their occurrence in the same way as in 1995.
North East	Very little difference if any.
North West	The NW had previously carried out bankside assessments to give rapid feedback on WQ issues. This operational work highlighted where detailed site assessments were needed and detailed biological investigations using a range of techniques (metals analysis, Surber sampling, macrophytes etc) were then carried out. The bankside assessments were carried out 3 times a year at the same sites throughout the catchments (many more sites than GQA network) and also allowed time for investigative work where necessary.

Question 6 (continued): Differences in sampling by pond-net and search in the 1995 GQA and prior to that survey

North West (cont.)	In 1990 a lack of communication (somewhere along the line from national to biologist) meant that a mixture of 'methodology' was incorporated into the survey. Biologists had been used to field sorting up to the point of Class 1a and then not sorting anymore. Sufficient information had been collected for normal purposes. When the sample was then sent for a formal audit it was therefore not surprising that errors of 14 missed taxa were recorded – the biologist had used methods for a different end point of data use. This led to an unhealthy (and unfair) reputation for NW biology. In 1995 GQA methodology was carried out as in BT001.
Midlands	1990 1 min search was optional. 1990 recommendation for emptying net every minute.
Southern	More diligent with 1 min search & with zigzag diagonal sampling.
South West	One minute search (20 secs surface dwellers: 40 secs attached animals). Not undertaken prior to 1995.
Thames	No comment.
Welsh	Differences were very minor. Previously, to reduce fatigue and speed sampling, we had used two samplers per site taking 1.5 minutes each + 30 seconds stone search each. We adopted the manual method completely in 1995.
SEPA (Solway)	Before 1995 we collected 3 minute kick samples only. The additional 1 minute search was new in 1995.
DoE (NI)	We have UKAS accreditation for the method and our version is slightly more systemised

Question 7: In the 1995 GQA sampling manual (BT001) alternative methods are recommended for sampling deep water sites. General procedures were recommended for dredge sampling in section 3.7.1 (pp 3.25 - 3.28) and more specific procedures in section 3.7.4 (pp 3.32 - 3.36). Recommended airlift procedures are given in 3.7.1 (pp 3.25 - 3.28) and 3.7.5 (pp 3.36 - 3.40). Further recommendations on additional sweep sampling and emptying nets are given for both methods in sections 3.7.3g (p 3.32) and 3.7.6 (p 3.41) respectively. Which of the following methods did you use to sample sites that were too deep to sample by pond-net except from the bank? Please list any that apply.

Anglian	Dredge/ bankside netting
North East	Dredge/ air-lift
North West	Bankside netting only
Midlands	Dredge/ bankside netting
Southern	Dredge/ bankside netting
South West	Dredge/ bankside netting
Thames	Bankside netting only
Welsh	Dredge
SEPA (both)	None
DoE (NI)	Prefer bankside netting only but sometimes use dredge

Question 8: If you used a dredge and/or an airlift, did you fully follow the recommended procedures for collecting samples in 1995 GQA?

Anglian	No
North East	Yes
Southern	Yes
South West	No
Welsh	Yes
DoE (NI)	Yes

Question 9: If you answered no to Question 8, what modifications did you make to the recommended methods?

Anglian	Dredge sampling (Eastern) Trawling parallel to the bank rarely done. Sieving to sub-divide sample not done - samples divided approximately from the dredge if necessary. Discarded material not checked except for large unionids.
Midlands	Lower Severn used smaller dredge than BT001, trawled behind boat in R Severn for Health & Safety reasons.
South West	Depending on the nature of the substrate, occasionally took fewer or more than recommended 3-5 dredges.

Question 10: If you answered yes to Question 8, in what way, if any, did the methods you adopted for deep water sampling in the 1995 GQA differ from the procedures you generally adopted prior to that survey? Please include the main differences, if any between the approaches adopted for the 1990 RQS and 1995 GQA.

Anglian	A larger mesh was use prior to 1995 - 1mm instead of 0.5mm.
North East	No difference
Midlands	Dredge not used on 1990, only bankside netting.
Southern	Change from dredge & marginal kicking to dredge & sweep only. The new technique was not liked and the dredge was replaced by net sampling at many sites.
Welsh	We only used the dredge method at two sites both of which we had never previously sampled. In the event both sites proved to be brackish and were abandoned.
DoE (NI)	None that I regard as being significant. Slightly more formalised due to UKAS

Question 11: If you feel that there are any differences between your component Areas/laboratories, in the compliance with sampling procedures set out in manual BT0001, which are significant and need recording, please give them.

Anglian	None
North West	None for GQA.
Midlands	Upper Severn and Lower Trent used net only. For Lower Severn see reply to Question 10.
Southern	Kent Area has used a dredge in some situations (small channels) where the other Areas would have used kick/sweep. This comment was made by Phil Smith who has recently taken over West Sussex from Shelagh Wilson and has noticed from records that a dredge has been used in small channels. Shelagh Wilson disputes this (though the records are presumably correct) and says they use it very infrequently – I (Bob Dines) suspect that their use of the dredge has declined recently as they have become more disillusioned with it. Kent sometimes use a 1 kg dredge (but more throws). They find some of their staff simply cannot use the standard dredge.
South West	Not qualified to respond.
Thames	No comment.
SEPA (both)	No difference.
DoE (NI)	We only have the one laboratory and management structure

Question 12: Please provide any additional comments you may wish to make on the recommended methods to be used to sample macro-invertebrates in national surveys.

Anglian	See comments under Question 20
Midlands	Upper Severn comment: Three minutes probably over-sampled small streams.
DoE (NI)	None other than a long pond net has proved much more effective than a dredge in most circumstances.

Question 13: The 1995 GQA manual (BT 001) laid down methods of sorting and identifying macro-invertebrate samples (Section 3.9, pp 3.42 – 3.47 and Section 3.10, pp. 3.47 – 3.60). Laboratory sorting of samples was requested in preference to bankside sorting and live identification. Did you ever undertake live bankside sorting and identification of samples during the 1995 survey?

Sorting

Anglian	Never
North East	Never
North West	Never
Midlands	Never
Southern	Never
South West	Never
Thames	Sometimes
Welsh	Never
SEPA (Clyde)	Never
SEPA (Solway)	Always
DoE (NI)	Never

Identification

Anglian	Never
North East	Never
North West	Never
Midlands	Never
Southern	Never
South West	Never
Thames	Sometimes
Welsh	Never
SEPA (Clyde)	Never
SEPA (Solway)	Always
DoE (NI)	Occasionally, to ensure that the sample was representative, but always given a complete sort in the laboratory.

Question 14: If you answered Aalways≅ or Asometimes≅ for either part of Question 13 (on bankside sorting and identification) please describe the circumstances where you used bankside sorting and what the advantages of this approach were.

North West

Although it was **not used for GQA**, bankside sorting/id has a range of advantages:

a) movement can help in the identification

b) can tell which species were dead already – important in pollution incidences

c) quick and cheap assessment for water quality – can phone the water quality section with results immediately if a problem.

d) for operational purposes it is a very efficient and effective method – can do far more sites throughout the catchment in one day.

e) Strongly felt by some ecologists (and should not be underestimated)– the beasties are returned alive.

f) Motivation of ecologists – it is easy to see where the results are going and what action is taken. If samples for lab sort are taken, it can reduce interest in other field factors. Also if the ecologists are sorting samples in lab which were taken weeks previously, and the data is only presented on maps in an end of year report, it can remove the interest in the sample.

Thames

SORTING: Rapid assessment – no need for preservative. Preferred method of one surveyor.

IDENTIFICATION: Live animals move – and in a characteristic way. Preferred method of one surveyor.

SEPA (Solway)

SORTING: When carrying out a survey we generally always then field sort it at the bankside. However, this is information for our Pollution Prevention Officers, and is not used for reporting. It is only a quick 'look and see' analysis.

IDENTIFICATION: The field scores were calculated on all sites. These samples were then subsequently analysed back in the lab to family level, the results of which were used in the 1995 survey.

Question 15: After sampling, how did you transport the samples to the laboratory?

Anglian	Live
North East	Live
North West	Preserved in alcohol
Midlands	Live
Southern	Fixed in formaldehyde
South West	Live
Thames	Preserved in alcohol/ Live
Welsh	Fixed in formaldehyde
SEPA (Clyde)	Live
SEPA (Solway)	Fixed in formaldehyde / Live
DoE (NI)	Live

Question 16: If you answered "live" to Q15, please give details of the **exact** procedure used.

Anglian	Sample transferred from pond net to airtight lidded 3 or 5 litre bucket with as little water as possible and labelled. Refrigerated immediately upon return to laboratory, on same day. May be carried in cold box, including cold packs, in hot weather. They were sorted live within four days of collection.
North East	Samples are drained of water in net, damp sample placed in marked polythene bags, or pots in the case of Northumbria. These were then returned to the lab for preservation. In Ridings we now use cool boxes for storage of samples in vehicles.
Midlands	Sample returned to lab in non refrigerated car boot, kept in fridge, sorted within 48 hours (if not, preserved). Samples in container with water (wet, not damp).
South West	Samples returned to lab in cool-box to be preserved in alcohol.
Thames	Drained majority of water off. Kept in sampling pots.
SEPA (Clyde)	Sample placed in pot on site. Formaldehyde added on return to the lab. This is always within 8 hours and never overnight.
SEPA (Solway)	Only a small amount of samples are fixed in the field (due to distance from the lab). The vast majority of samples are collected in one day and brought back to the lab. These samples are then preserved immediately and sorted. The preserved samples are stored in ventilated cabinets and fixed with formaldehyde.
DoE (NI)	Samples back to the lab live, sieved and sorted ASAP and always within 36 hours, fridged during hot periods.

Question 17: Do your laboratories have adequate facilities for the safe handling of formaldehyde? If appropriate, explain how this differs between Areas.

Anglian	Yes. Differences between Areas are portable, bench-clamped extractor (operating with GAC filter) owned by Northern Area laboratory.
North East	Yes.
North West	Yes, but one out of the eight does not. The south Area ecology does not in the new lab.
Midlands	Yes. Lower Severn have no sink inside the fume cupboard. Use IMS not formaldehyde.
Southern	Yes. No significant differences – all have good facilities.
South West	Yes. Not qualified to respond for differences between Areas.
Thames	Yes.
Welsh	Yes.
SEPA (both)	Yes.
DoE (NI)	Yes. No differences between Areas.

Question 18: Do any of your laboratories have adequate facilities for the safe handling of formaldehyde but do not use them for the applying formaldehyde as a preservative for GQA macro-invertebrate samples? If so please give reasons.

Anglian	Yes. All samples are stored and sorted live wherever possible. This avoids the environmental and health and safety risks associated with the unnecessary use of formaldehyde. COSHH regulations state that, if a viable alternative exists (which it clearly does) then COSHH regulated substances should not be used. In-house controlled experiments (as a MSc project) have demonstrated that there is no statistically significant difference in recovery of invertebrates (and therefore, data obtained) from a sample sorted live compared with one sorted preserved in Formalin. When necessary, e.g. for AQC samples or if samples cannot be sorted within a few days they are preserved using the available safety equipment.
North East	No.
North West	Yes. It is used as a fixative for storage once the sample is back in the laboratory. Formaldehyde is not used in the field because the ecologists use their own cars. Many ecologists have small children and did not want them to be exposed to formaldehyde when travelling in the car. It was felt that however carefully it is treated, there is always the chance of small amounts being spilt and spoiling the car.
Midlands	Yes. For Lower Severn the reply to Question 17. Also member of staff has allergy to formaldehyde. Upper Severn use Propane for safety reasons.
Southern	No.
South West	No. Health & Safety reasons. Limit use to marine samples in South Wessex Area.
Thames	No.
Welsh	No.
SEPA (Solway)	No. Our lab/storage facilities are fully equipped for handling formaldehyde. We also have procedures for the safe handling and storage of formaldehyde, which are always followed.
DoE (NI)	No.

Question 19: If you feel that there are any differences between your component Areas/laboratories, in the way in which you process samples, which are significant and need recording, please give them.

Welsh	In South East Area a stack of sieves of different sizes was always used to separate animals out into different size fractions; this technique was occasionally used in Northern Area. One sieve only was used in South West Area for all samples. This difference in methods seems to give rise to no clear differences in AQC results.
DoE (NI)	None that I can think of other than we did not photograph all sites in 1995

Question 20: Please provide any additional comments you may wish to make on the recommended methods to be used to process macro-invertebrate samples for national surveys.

Anglian	Staff concerns <u>will not allow</u> the regular use of Formalin as a preservative during the 2000 survey. It is a general opinion in Anglian Region that the use or non-use of Formalin is not, in any way, a 'standardisation' issue, but an operational (i.e. logistic) one. In-house work shows that its use makes no difference to the specimen recovery process. Therefore, if other Regions wish to use it, this should be justified on the basis of operational need (e.g. spate flows in Welsh mountain streams reducing the 'window of opportunity' for sampling). If the argument is put forward that "standardisation" equals "everybody does exactly the same, regardless of the effect on the data obtained", then we had better design "standard" biologists, with the same sized feet, musculature, eyesight, hair colour, etc. In fact, samplers of different build (& 'enthusiasm') are more likely to have an effect on the data obtained than the use or non-use of a preservative!
North West	The ecologists are strongly against the use of formaldehyde.
South West	Split sample into different fractions using sieve, and use plenty of white trays.

Question 21: The 1995 GQA manual required that all the aquatic macro-invertebrates in the sample be identified, including the pupae caddis and dipterans (Section 3.10.3, p3.55 – 3.56). What level of identification did you achieve for your samples from the 1995 GQA? Please give answers for each of four taxonomic levels below.

BMWP families

Anglian	Always
North East	Always
North West	Always
Midlands	Always
Southern	Always
South West	Always
Thames	Always
Welsh	Always
SEPA (both)	Always
DoE (NI)	Always

All families

Anglian	Sometimes
North East	Sometimes
Midlands	Sometimes
South West	Sometimes
Thames	Sometimes
Welsh	Always
SEPA (both)	Sometimes
DoE (NI)	Sometimes

RIVPACS species level (some taxa only)

Anglian	Sometimes
North East	Sometimes
North West	Sometimes
Midlands	Sometimes
Southern	Sometimes
South West	Sometimes
Thames	Never
Welsh	Never
SEPA (both)	Never
DoE (NI)	Sometimes

RIVPACS species level (all taxa)

Anglian	Sometimes
North East	Never
Midlands	Never
South West	Never
Thames	Never
Welsh	Never
SEPA (both)	Never
DoE (NI)	Never

Question 22: Which families do you have the greatest difficulty in identifying? Include difficult life stages of groups you otherwise find easy to identify (e.g. Goeridae pupae, immature Leuctridae/Capniidae etc).

Anglian	Northern Area: Caddis pupae (available key (Hickin, 1967) unreliable). Species level identification virtually impossible. Eastern Area: All pupae - never done. Tipulidae
North East	Some dipteran larvae and pupae. Some caddis pupae. Small cased caddis
Midlands	Dipteran larvae. Caddis pupae. Very small specimens e.g. Limnephilidae, Lepidostomatidae, Leptoceridae and Beraeidae
Southern	Early instar stoneflies & caddis. Diptera. Caddis pupae...Beetle larvae, particularly those that look like Hydrophilidae
South West	Capniidae, Beraeidae, Lepidostomatidae, Psychomiidae, Platycnemidae, Coenagrionidae, Hydrophilidae, Dryopidae
Thames	Non-scoring dipterans
Welsh	We don't consider identification itself to be a very great issue, far more relevant is animals being completely overlooked during the sorting process. However, most common problems in identification related to small caddis (especially Odontoceridae), Planorbiidae/terrestrial snails, *Asellus*/wood-lice, beetle larvae and adults (especially hydrophilids), Zygoptera, larval Diptera.
SEPA (Clyde)	Small cased caddis may occasionally cause problems.
SEPA (Solway)	Family level identification was usually okay – small caddisfly occasionally caused problems, but no particular family.
DoE (NI)	None I think but in lab white planarians can give a problem and some of the smaller cased caddis. What is a tipulid? I think this is now sorted. We address errors by training as part of the UKAS regime.

Question 23: If you feel that there are any differences between your component Areas/laboratories, in the level of identification achieved, which are significant and need recording, please give them.

Anglian	Routine species-level identification of long standing in Northern Area (more than 20 years). Only recently established (on a routine basis) for other two Areas (last 2-3 years).
South West	Not qualified to respond
Thames	Waltham always identify "other taxon" families. Fobney did not.

Question 24: Please provide any additional comments you may wish to make on the identification of taxa. Amongst the issues you may wish to comment on are the value ranges of abundance classes and difficulties in the estimation of the correct classes for each family.

Identification

Anglian	Northern Area operates under the principle that the highest practical level of identification must be undertaken as a matter of policy. There is a continual willingness to extend this range as far as possible (for example, simuliid larvae are now identified to species/species complex level, where possible).

Assignment of abundance classes

Anglian	The abundance classes are fine as they stand. Quality Control of the process should be considered, however, as a matter of priority, if only to emphasise the importance of accurate assessment for use in processes such as the LIFE score. Such a scheme **must** be practical; it would not be acceptable to count every individual and set rigidly defined rules as to what is correct and what is not.
North East	These are subjective and susceptible to operator differences. This is most problematical around class boundaries. I believe it is better to provide a count of abundance and then allocate an abundance scale.
North West	The log 10 abundance scale is not felt to be useful for interpretation and although it is reported for GQA, the ecologists use other schemes as well: 1 2-5 5-20 This can be translated into the log 10 abundance. 20-100 100-500 500+
Midlands	Abundance class A must be divided at least into 1-2 and 3-9.
Southern	Subdivision of the 1-10 class is valuable. We use 1 and 2-10 but 1-3 and 4-10 may be an option. A split 11-50, 51-100 may also be useful (we do not use this).
South West	Require guidance.
Thames	Sort out classes nationally.
Welsh	This is quite a difficult process if individuals aren't counted but provides a useful rough guide to abundance.
DoE (NI)	We do this routinely and it should be a component of RIVPACS for the future.

Question 25: Analytical Quality Control (AQC) and Quality Audit (Audit QA) procedures were required for the 1995 GQA. AQC is an internal procedure in which experienced analysts check the sorting and identification performance of colleagues for a pre-set proportion of samples processed. The audit is an external procedure in which the performance of Agency staff at sorting and identification is assessed by experts from another organisation, based upon a pre-determined number of all samples processed. Samples may be subject to external audit before or after internal AQC checks. Do you think that internal AQC was of value in helping to control the performance of sample sorting and identification. Please give reasons for your answer.

Sorting

Anglian	Yes
North East	Yes
North West	Yes
Midlands	Yes
Southern	Yes
South West	Yes
Thames	Yes/No
Welsh	Yes
SEPA (both)	Yes
DoE (NI)	Yes

Identification

Anglian	Yes
North East	Yes
North West	Yes
Midlands	Yes
Southern	Yes
South West	Yes
Thames	Yes/No
Welsh	Yes
SEPA (both)	Yes
DoE (NI)	Yes

Reasons

Anglian	Provides a rapid, effective & immediate management of training and quality of work/data.
North East	Enables on going assessment of laboratory quality and identifies possible sorting problems early. Demonstrated to work.
North West	Raised awareness of process and methodology.
Southern	The rapid feedback of AQC is a big advantage over the external audit. It is more immediate for the lab staff and provides a constant nudge to maintain quality. Staff find the feedback valuable.

Question 25 (continued): Reasons for opinions on the value of internal AQC .

South West	Demonstrates personal pride and professionalism in consistently producing high quality information.
Thames	Some find it helpful – an other thinks it's an unnecessary check on the work of experienced biologists.
Welsh	We have always been firm believers in AQC in Wales and were one of the first Regions to initiate a scheme. The AQC scheme in 1995 (especially the internal) was of immense value in maintaining quality in the Region. It's one drawback was that there was no constraint to pull an AQC for each batch of 10 samples sorted such that by chance several weeks could go by without an AQC being drawn. The new stratified random approach, which ensures one AQC per batch of 10 is much better.
SEPA (Solway)	AQC procedures are vital to ensure less experienced staff are performing effectively.
DoE (NI)	It spots problems before they become ingrained – in this respect we have been lucky in maintaining a static field team.

Question 26: Please describe the process you currently use for selecting samples for internal AQC

Anglian	Northern Area: Samples retained in batches of ten and arranged in random order (1 to 10). Random number generator used to choose sample to be 'tested' against random selection of marbles from opaque bag (9 blue marbles and 1 black). Marble removed from bag & placed to one side, new random number generated, new marble selected & removed and so on, until black marble selected. Randomly selected sample corresponding to random selection of black marble is sample to be put through AQC process. Other 9 samples discarded. Eastern Area: Similar process to above but a marble is drawn each time a sample is completed. Marbles are not replaced until all 10 have been drawn.
North East	Dales: Black ball in bag with 9 other different coloured balls. Sorted samples all logged in sequence in book and balls used on every 10 samples logged. Ridings: Black marble with 9 other coloured marbles in old sock, marble selected as soon as sample sorted, enables AQC to follow straight after initial analysis. Northumbria: Random selection of 1 in 5 samples from a batch of samples.
North West	North Area – 10 marbles, one black. Marble picked out after each sample and if it's the black one the sample is AQCed. Central Area – after 10 samples have been sorted they are shelved and the team leader picks one out at random.
Midlands	Samples entered into lab log in batches of ten in as analysed. After primary sort all labels removed and replaced by AQC batch and number. Bag of 10 numbered balls is used to blind pick (by non biologist) sample. AQC analyst does not know origin of sample.
Southern	Nine white and one black pot lids in a bucket – consecutive lids are picked out and placed with each of the batch of ten samples until the black one appears. That sample is selected.
South West	Once processed the samples are placed in a crate with spaces numbered one to ten. Once the crate has ten samples then one is selected at random by picking a bottle top out of a bucket with ten tops numbered one to ten, i.e. using a lucky dip system. The selected sample is given an anonymity code based on random numbers.

Question 26 (continued): Process used for selecting samples for internal AQC

Thames	One in ten samples – picking ball from bag – nine blue and one green – without replacement. This process carried out by AQC controller.
Welsh	We store processed samples in a fridge until a batch of 10 has been completed. Then we use a bag of balls containing 9 plain and one marked ball to chose the sample for AQC. Balls are drawn from the bag until the marked ball is pulled e.g. if the sixth ball drawn is marked ball, we carry out AQC examination on the sixth sample processed in the batch. The other 9 samples are then discarded to reduce our usage of formaldehyde.
SEPA (Clyde)	Random selection from each analyst. No analyst can be sure whether a particular sample will be selected or not.
SEPA (Solway)	A minimum of 5% of all samples are subject to internal AQC. Analysts record samples analysed and one sample is chosen at random every 20 samples.
DoE (NI)	Black and pink balls out of an envelope.

Question 27: Please describe the process you currently use for selecting samples for external audit

Anglian	Process undertaken centrally (Regional Biologist), once sample has been internally Quality Controlled. There are a known number of GQA samples and therefore it is possible to calculate how many AQC samples each Area will undertake. A random selection of 10 samples is made using marbles and a table is drawn up prior to each sampling season which identifies which AQC samples are to be sent for audit. Once an Area has completed an AQC check they 'phone up to find out if it is to be audited.
North East	Dales: These chosen from the samples identified for internal AQC. Samples AQCed as a result of new sorters are not included for external audit. If external AQC samples do not require to have undergone internal AQC then these are chosen on a sorter basis (1in 10 per sorter). Ridings: One in two AQC samples selected using marbles in sock technique. Northumbria: All AQC sample number written down and a third party picks the required number of samples from the list.
North West	For GQA years (when there are sufficient samples), each internal AQC is put on a shelf. When five have been collected, one is randomly chosen by the team leader. In non GQA years, all the internal AQC samples may be sent for external audit.
Midlands	Similar process at Region. List divided into blocks dependent on the expected number of samples for the year for that laboratory. Regional Biologist picks ball from bag and arranges for these samples to be sent to IFE.
Southern	Three of every four samples (based on average throughput) using a similar procedure to AQC samples.
South West	This depends on the number of samples to be processed during the year. In 1998/99 we needed to select 1 in 10 primary samples to meet our quota of 20 for external audit. Hence no further selection was required. In previous years we have needed to select 1 in 20 primary samples. The AQC samples were placed in pairs and were selected by a toss of the coin.
Thames	Fifteen samples per lab per season are randomly selected by AQC controller – unidentified AQC pots selected from total number of pots at end of season.

Question 27 (continued): Process currently used for selecting samples for external audit

Welsh	Currently we have a very low number of samples being processed by lab sorting in the average year ca 100-150. Thus every internal AQC sample drawn goes for external audit (see answer to Question 26) and these are supplemented by one primary sample per batch of 10 drawn in a similar fashion to the internal AQCs.
SEPA (Clyde)	Random selection from each analyst.
SEPA (Solway)	We do not have a set number of samples selected for external audit. A contract is arranged every year. A typical number of samples would be 4-6 per analyst.
DoE (NI)	Lab manager at random and I do hers

Question 28: What type of action do you take when samples fail to pass the national AQC target in your Region?

Anglian	Internal management process. Staff always informed of performance immediately upon availability of AQC data. The matter is discussed with the relevant member of staff to see if there are any extenuating circumstances (sample or sorter). If not, discuss the need for greater care, changes of technique etc. Failure to achieve target becomes significant only if repeated by same sorter or if 'extreme' (4 or 5 omissions, careless identification, etc.). Personal pride of sorter is usually adequate spur for improved performance; most biologists are professional enough not to need telling when performance is unacceptable.
North East	Dales: Inform relevant sorter, request more care and attention on future samples. Ridings : Rarely occurs, sorter is informed and reasons identified. Northumbria: Re-pick the batch containing the failed sample.
North West	We have never been in the situation where batches of samples have to be resorted. Where a poor result is obtained with a sample, the reason for the poor result is investigated (time, method or poor identification) and rectified.
Midlands	According to national guidelines, review training, time / care taken over samples (Can usually relate to cause e.g. workload, personal problems such as family health, job interviews/insecurities etc).
Southern	Range of actions as per AQC manual. From nothing to discussion of problem taxa etc.
South West	Never happens.
Thames	Action only taken to keep pots once in defer status – alarm status never reached.
Welsh	I'm not sure what this question means exactly as the target is an overall 2.0 gains per sample and does not apply to individual samples. However, our AQC analysts provide feedback to the primary sorters on every AQC sample once it has been re-processed. This serves to iron out any sorting problems before they get out of hand. We never require samples to be reprocessed. Thus, although every lab has been in the 'Defer' state from time to time we have never reached the 'Alarm' situation where the average No gains recorded by internal AQC departs from the national target (2.0 gains per sample) to a degree that is statistically significant.

Question 28 (continued): Action taken when samples fail the national AQC target

SEPA (Clyde)	This involves: checking the next few samples analysed. If targets are still not met then retaining may be necessary.
SEPA (Solway)	We have clearly defined procedures, which are followed. If a misidentification is suspected an AQC investigation form is completed and involved re-examining specimens. If failure is in the sorting procedure then a proportion of samples are resorted. If the failure still exists the analyst could be retrained.
DoE (NI)	All results are discussed with the individual officers as UKAS requirement. We haven't been too bad that retraining has been required but it is an option and we are strongly into IIP.

Question 29: What type of action do you take when samples fail to pass the external audit target in your Region?

Anglian	Similar to the reply to Question 28. Particular issues (e.g. typographical errors) may be raised.
North East	Not been an issue here
North West	Again, the reason for a poor result is investigated, and appropriate methods put in place to rectify the problem (training, time allowance etc).
Midlands	After adjustment of control figures, similar to reply to Question 28.
Southern	N/A!!!
South West	Never happens
Thames	Discussion of possible reasons for failure between line manager, analyst and Regional Biologist.
Welsh	I assume that this means departure from the average 2.0 gains per sample. In this case we would carry out a thorough review of our procedures based upon a careful assessment of the sources of error as identified by the results sheets.
SEPA (both)	As given in Question 28.
DoE (NI)	All results are discussed with the individual officers as UKAS requirement. We haven't been too bad that retraining has been required but it is an option and we are strongly into IIP

Question 30: If you feel that there are any differences between your component Areas/laboratories, in your AQC and audit procedures, which are significant and need recording, please give them.

Anglian	The Northern Area approach to selecting AQC samples is better as it avoids the risk of staff knowing that an AQC sample has already been selected from a batch of 10 samples. Specimen placed in vial (i.e. removed from sample), but not recorded on data sheet, is not subsequently recorded as an omission error by the AQC analyst, in one laboratory.
North East	Northumbria operate a very strict one in five AQC batch system as they tend to sort in batches, periodically. Ridings have a more or less continuous throughput of samples and operate a continuous AQC process. Problems are generally sorted by co-operation and asking for help before the sample process has finished resulting in good results overall. All new staff involved in sorting have all samples re-picked until they meet adequate sorting standards.
Midlands	Time between primary sort and AQC sort varies according to Area.
South West	Not qualified to respond.
Thames	No comment.
Welsh	By error in 1995, South East Area used an AQC scheme based on the individual rather than the lab and this was not recognised at a Regional level until quite recently. Thus the current South East Area practice is for balls to be drawn at the point where each individual has processed 10 samples in order to ensure that each individual has their samples checked on a regular basis. This is known to contravene the national guidance but is unlikely to greatly affect the overall results; all Areas will use the nationally accepted practice in 2000.
SEPA (Clyde)	There are differences in methods of selecting samples.
DoE (NI)	Not applicable

Question 31: Please provide any additional comments you may wish to make on the use of AQC and audit procedures for the sorting and identification of macro-invertebrate samples, with special reference to national GQA surveys.

Anglian	Proposed process for species level audit is unnecessarily unwieldy, prescriptive and impractical. This does not, of course, currently apply to the national survey. Current family level AQC/audit process works very effectively. External (internal?) AQC of abundance categories should be considered (see reply to Question 24).
North West	The audits/AQCs are taken seriously by the labs and individuals can get quite despondent if they have bad results. It is questionable, therefore, if it is necessary to have individuals results published in the annual reports when the lab overall results should suffice.
Midlands	Comment from one Area that AQC takes a lot more time and resources, limiting operational work. It has all the disadvantages and no advantages such as adjustment of the data.
Southern	Valuable for individuals As well as for overall performance.
DoE (NI)	Very much in favour of the fact that it is a necessary tool to optimise quality

Question 32: The macro-invertebrate data collected during the 1995 GQA was used to determine the BMWP score, number of scoring taxa and ASPT (Average Score Per Taxon) for each sample. Individual season's taxon lists for spring and autumn were also combined to form a site taxon list for the year. BMWP index values were computed for the combined season's lists for each site. RIVPACS was then used to produce optimal (= expected) faunal lists and BMWP index values for each sample or combination of samples. The ratio of observed to expected BMWP index values (often referred to as EQI or Ecological Quality Index) was used to band sites into grades of biological condition. Separate grades were determined for each season=s and for paired seasons= faunal lists for each site based on each of ASPT and number of BMWP taxa. An overall site grade for was determined by taking the lower of the grades determined separately for ASPT and number of taxa for the two seasons combined list.

The EQI band ranges used for assessing the combined seasons ASPT and number of taxa grades for the 1995 GQA were as follows:

Grade	Description	Lower grade limits	
		EQI ASPT	EQI number of taxa
a	Very good	1.00	0.85
b	Good	0.90	0.70
c	Fairly good	0.77	0.55
d	Fair	0.65	0.45
e	Poor	0.50	0.30
f	Bad	0.00	0.00

Please give any comments you wish on this banding system and how it has worked in interpreting the 1995 GQA data. Please record any differences between your component Areas/laboratories that are significant and need recording.

Anglian

Generally works well, but must be recognised as a tool for management of water quality, not the 'universal answer', independent of local interpretation.

Category f ('Bad') is rather unscientific; wouldn't 'Very Poor' be better?

Less to do with the GQA scheme but, it is felt that RIVPACS under-predicts for certain types of river. For example, sites in lower reaches (slow flowing, deep rivers) are often under-predicted. RIVPACS does not predict well the rich pond-like fauna that you find in these rivers. This does not present a large problem in that the rivers are classified as GQA a and it is not believed that there are any significant quality issues anyway.

North East

Northumbria found that the production of GQA system did not achieve any significant differences in identification of poor and bad sites. Sites with apparently similar good quality are often arbitrarily ascribed a or b classes.

Since I was heavily involved in the establishment of the GQA system I feel it adequately defines the quality of rivers in Ridings Area and hope that it works for the Region as a whole, and have seen little evidence that it does not.

Question 32 (continued): Comments on the GQA banding system.

North West	Any band system will not fit every situation and it was generally felt that this system was OK.
Midlands	Grades are unrealistic for small urban headwaters. Samples with Gammarus, Baetis, Hydropsyche and molluscs do not get better than Band e. RIVPACS grades confusing due to the varying use of single season or combined season. Combined season not liked.
Southern	In Southern we only really use the top 3 bands (mostly!) as most of our rivers are quite clean. This reduces the value of the classification for operational purposes. BUT, it is supposed to be a national surveillance classification, not an operational tool.
South West	Generally satisfied with banding system.
Thames	Chosen number of bands to compare with chemistry – next time biology needs primacy.
Welsh	The banding system was arrived out through an iterative process involving all Regions and in Wales we feel that is was very successful. Our main concern is that some upland stretches received an inflated rating because there were cases where relatively impoverished spring and autumn samples could combine to yield a healthy taxon list - this was especially true of acidified streams some of which received very optimistic Band b ratings when the spring fauna was really pretty poor. The same situation could apply to other intermittent pollutants such as sheep dip and wastes from dairy and beef farming. This potential problem was highlighted at the time when the grading system was being developed. An alternative approach would be to produce separate spring and autumn EQIs for each site and take the average. However, this would not be using RIVPACS at its best and would overemphasise the influence of the odd poorly processed sample. Basically the scheme chosen will be somewhat optimistic about quality highlighting the best that the system can achieve, the other possible method would reflect rather more upon the worst case. You pays your money… etc.
SEPA (Clyde)	Generally I am quite happy with this system, but I feel that: a) an ASPT EQI of 1.00 for very good is too exacting b) very high EQIs (e.g. >1.25) should cause the bad to revert to good. These very high EQIs may be a reflection of some enrichment of the water course.
DoE (NI)	We found it too generous. Bands E and F might as well not have been there. Even rivers that we acknowledge as being badly polluted were included in Class D. This cannot be good as a management tool. We also should be able to flag seasonal problems.

Question 33: The evaluation of the biological condition of river stretches in the 1995 GQA depended exclusively on the use of RIVPACS to interpret macro-invertebrate data. What other methods would you like to see applied to the interpretation of the type of macro-invertebrate assemblage data collected during the 1995 GQA?

Region	***System of evaluation***	***Type of stress/site for which the proposed system is of relevance***
Anglian	LIFE (Lotic Invertebrate index for Flow Evaluation) CCI (Community Conservation Index) Diagnostic methods	Low flows/drought Conservation value (applicable to flowing water and stillwater sites. Data analysis which allows distinction between different types of stress e.g. habitat, toxic pollution
North East	Make use of abundance data and community structure	ID of initial stress in clean waters.
North West	Artificial intelligence Intelligent ecologists!	 All
Midlands	More interpretation of actual fauna e.g. Absence Asellus/Gammarus Absence molluscs/leeches Low diversity/high ASPT LIFE when developed	 Pesticides/sheep dip Ammonia Acidification Low flows
Southern	LIFE Detailed analysis of basic data	Flow problems Operational needs
South West	LIFE index	Low flows
Thames	Amended BMWP Artificial Intelligence LIFE Multivariate analyses with new environmental variables, including flow related variables, RHS, Geology and Temperature model (See John Murray-Bligh) Family distributions Conservation scores RHS	 Diagnostic analyses

Question 33 (continued): Other methods for the interpretation of macro-invertebrate data collected during the 1995 GQA?

Region	***System of evaluation***	***Type of stress/site for which the proposed system is of relevance***
Welsh	Could use acid indicator system developed by Rutt *et al*. (1990) as for 1990 data set.	Acidification
	Use Bill Walley's artificial intelligence approaches.	Potentially full range of stresses
DoE (NI)	Numerical abundance	All types of stress but especially pesticides
	Seasonal fluctuations	Flagging periodic problems e.g. sheep dips and apple spraying
	RIVPACS for smaller rivers	We are expanding our current network and will need to address this problem
	RIVPACS for drains and canals	Objective still water evaluation

Question 34: The evaluation of the biological condition of river stretches in the 1995 GQA depended exclusively on the use of macro-invertebrate data. What other taxonomic groups and methods of approach, **if any**, would you like to see applied to the interpretation of the biological condition of the river stretch?

Region	***System of evaluation***	***Type of stress/site for which the proposed system is of relevance***
Anglian	Diatoms (TDI) Plants (MTR) Phytoplankton/Chlorophyll Habitat evaluation	Nutrient rich areas (eutrophication issues). As above Only applicable in deep rivers: same issue as above. A quantitative system for evaluating habitat quality would be useful
North East	Diatoms and macrophytes	Eutrophication studies.
North West	Diatoms	Eutrophication
Midlands	Macrophytes Diatoms Fish	Eutrophication Eutrophication/general WQ
Southern	MTR/TDI (and others?) Better methods for headwaters (e.g. reduced sampling time) Conservation value (e.g. species Id?)	Eutrophication
South West	MTR TDI	Nutrient enrichment and eutrophication.
Thames	Diatoms, fish, macrophytes (?), phytoplankton, hydro-morphology and RHS/ habitat	
Welsh	Diatom Quality Index approach	Eutrophication
SEPA(Clyde	Trophic Diatom Index	Eutrophication
DoE (NI)	The compartments defined in the Water Framework Directive Macrophytes and algae Fish Geomorphology	 Primarily but not exclusively eutrophication Naturalness The degree of physical influence We are only one laboratory

Question 35: If you feel that there are any differences between your component Areas/laboratories, in the importance attached to alternative approaches to assessing biological condition, which are significant and need recording, please give them.

Southern	I (Shelagh Wilson) worry a bit about using the same procedure for GQA sampling as for WQ investigations. The 1min search while adding to the biological knowledge of the site, does tend to push the BMWP up and pull the ASPT down, as the margins tend to be the back waters where the pondy animals hang out, and they also tend to be less sensitive to organic input due to their pond loving nature. We probably do not concentrate on the full technique when undertaking investigatory work.

Question 36: Please provide any additional comments you may wish to make on the use of alternative approaches to the assessment of the biological condition of river stretches.

Anglian	Alternative approaches will only become truly viable if a complete change in the national mind-set occurs, so that an annual sampling programme, with sites sampled at least twice yearly and analysed to species level, becomes the norm. A quinquennial 'snapshot', at an inadequate taxonomic level, is not sufficient to address all relevant issues e.g. flow, conservation value or even water quality except superficially.
North West	It is important to relate the quality of data needed to the methodology i.e. rapid assessment for rapid turnover of results and GQA quality for national surveillance. The Artificial Intelligence system appeared promising because it 'learnt' as more data added and therefore could reassess the baseline.
SEPA (Clyde)	A system to assess the effects of acidification is important in South West Scotland. We have such a system, however it is based on species level identification and as such it is not appropriate for GQA surveys. Family level assessment of acidification can often give an approximate indication of conditions but in my experience can occasionally give misleading results.
DoE (NI)	We use macrophytes for routine assessment to explain chemical anomalies

Question 37: RIVPACS requires annual mean alkalinity values in order to make the best available faunal predictions. If alkalinity values are not available, either of hardness, Calcium or conductivity may be used as a surrogate. Which determinant was most commonly used for predictions in your Region?

Anglian	Alkalinity
North East	Primarily alkalinity, otherwise conductivity
North West	Alkalinity
Midlands	Alkalinity
Southern	Alkalinity
South West	Alkalinity
Thames	Alkalinity
Welsh	Alkalinity
SEPA (both)	Alkalinity
DoE (NI)	Alkalinity

Question 38: The manual recommends (p2.21) that an absolute minimum of three evenly spaced alkalinity or surrogate values are used to calculate the annual mean value but recommends that a minimum of twelve monthly values are obtained. **Approximately** what proportion of the annual mean alkalinity values you obtained for the 1995 GQA survey were based on >9 values.

Anglian	>75%
North East	>75%
North West	>75%
Midlands	>75%
Southern	>75%
South West	>75%
Thames	>75%
Welsh	>75%
SEPA (both)	<25%
DoE (NI)	>75%

Question 39: On what year(s) were most of your annual mean alkalinity values based?

Anglian	1995
North East	1993-95
North West	1995
Midlands	1995
Southern	1995
South West	1995
Thames	1990
Welsh	1995
SEPA (both)	1995
DoE (NI)	1995

Question 40: For sites that were common to the 1995 GQA and 1990 RQS, approximately what proportion of the time invariant values (NGR, altitude, slope, distance from source and discharge category) were derived from the following categories:

Newly calculated in 1995

Anglian	100%
North East	100%
North West	100%
Southern	100%
South West	Sorry – can't remember
Welsh	100% for all but discharge at sites previously samples in 1990. Thus figure for discharge would be ca 40%
DoE (NI)	100%

Based entirely on 1990 values

Thames	100%
SEPA (both)	100%

No Regions used averaged data from 1990 and 1995.

Question 41: For the majority of sites, which of the following determinants were measured by more than one independent person as a means of quality control on the accuracy of the data acquisition? Please mark all that apply.

Region	***NGR***	***Altitude***	***Slope***	***Distance from source***	***Discharge category***
Anglian	**X**	**X**	**X**	**X**	
North East	**X**	**X**	**X**	**X**	**X**
North West	**X**	**X**	**X**	**X**	
Midlands	**X**	**X**	**X**	**X**	**X**
Southern	**X**	**X**	**X**	**X**	
South West					
Thames	**X**	**X**	**X**	**X**	**X**
Welsh	**X**	**X**	**X**	**X**	
SEPA (both)					
DoE (NI)	**X**	**X**	**X**	**X**	**X**

Question 42: Please indicate how difficult you find it to record each of the following RIVPACS variables and what the difficulties were, if any?

NGR

Anglian	Low difficulty.
North East	Low difficulty. Main difficulty is grid letters often getting confused at the boundaries.
North West	Low difficulty at one office
Midlands	Low difficulty
Southern	Low difficulty
South West	Low difficulty
Thames	Low difficulty. We should use GPS or 1:25,000 maps.
Welsh	Low difficulty
SEPA (Clyde)	Low difficulty.
DoE (NI)	Low difficulty

Altitude

Anglian	Low difficulty
North East	Moderate difficulty. Often difficult to ascribe altitude as no adequate marks on 1:50,000 maps
North West	High difficulty at one office. Urban areas are the main problem
Midlands	Low difficulty
Southern	Low difficulty
South West	Low difficulty
Thames	Moderate difficulty. From Landranger maps, distances between contour lines in lowland areas can be large, requiring estimation. 1:25,000 maps.
Welsh	Moderate difficulty. Main difficulty is identifying where contours cross rivers especially urbanised areas or where gradients are severe and contours closely packed.
SEPA (Clyde)	Low difficulty.
SEPA (Solway)	High difficulty. Little experience of calculating altitude so don't really know how accurate you have been. Time consuming.
DoE (NI)	It was not always obvious but will improve with GIS. Low difficulty.

Question 42 (continued): Difficulty of measurement

Slope

Anglian	Moderate difficulty. Main difficulty: Ability to find relevant contours and accurate measurement of distance for calculation.
North East	High difficulty. Difficult to locate contours that cross the river, 1:50,000 map too small for this detail.
North West	High difficulty at one office. Urban areas are the main problem.
Midlands	High difficulty. Tricky in urban, very low and very high gradients.
Southern	Moderate difficulty.
South West	High difficulty. Accurate measurement of distance between contours.
Thames	Low to moderate difficulty. No contours on the urban London map.
Welsh	Moderate difficulty. Main difficulty as for altitude.
SEPA (Clyde)	Moderate difficulty. Occasional problem where contour lines on map are not clear.
SEPA (Solway)	As above. Contour lines not always easy to see on map.
DoE (NI)	Main difficulty was not always accurate GIS.

Discharge

Anglian	Low difficulty
North East	High difficulty. Distinguishing the different widths on the definitive maps. Northumbria used Water Resources to extrapolate discharge from flow data.
North West	Low difficulty.
Midlands	Low difficulty.
Southern	N/A – determined by Water Resources staff.
South West	High difficulty. Difficulty in obtaining reliable data.
Thames	Low difficulty. Had microlowflows readings been used?
Welsh	Moderate difficulty. Data obtained from hydrology – bit of our black box process from biologists point of view.
SEPA (Clyde)	Low difficulty. Data supplied by hydrology section.
SEPA (Solway)	Discharge category given by hydrologists. Low difficulty.
DoE (NI)	IOH provide values.

Question 42 (continued): Difficulty of measurement

Distance from source

Anglian	High difficulty. Main difficulties: Identifying relevant source and accurate measurement along watercourse (even using a map wheel).
North East	Moderate difficulty. Main difficulty is determination of longest tributary from the site. Measuring wheel not brilliantly accurate.
North West	Moderate difficulty at one office. Fiddly.
Midlands	Low difficulty. Source often not shown in urban areas, sometimes inaccurate on 1:50,000 maps.
Southern	Low difficulty.
South West	High difficulty. Problem is identifying/deciding source of river. We measured from the furthest upstream point on any headwater within the catchment.
Thames	Low difficulty.
Welsh	Low difficulty. Main difficulty is identifying source correctly.
SEPA (Clyde)	Low difficulty
SEPA (Solway)	Moderate difficulty. Used a digitiser to calculate distance. Quite time consuming.
DoE (NI)	Moderate difficulty. Main difficulty is that it can be difficult in some longer/cross-border rivers

Alkalinity

Anglian	Low difficulty
North East	Low difficulty
North West	Low difficulty at one office
Midlands	Low difficulty. Post 1995 may be difficult to obtain this for new sites, 1995 data to be used on 2000.
Southern	N/A – Water Quality data.
South West	Low difficulty.
Thames	Low difficulty.
Welsh	Low difficulty. No real problem. However issue of which of consistency and appropriateness of the various analytical methods for alkalinity needs to be explored. Did all Regions use the same method in 1995? Was it the same method used for the RIVPACS data analysis?
SEPA (both)	Low difficulty.
DoE (NI)	Low difficulty. Main difficulties are low alkalinities below 10 are a pain.

Question 42 (continued): Difficulty of measurement

Width

Anglian	Low difficulty. Main difficulty : deep rivers, with no bridge available, must be estimated or measurements taken using a rangefinder.
North East	Moderate difficulty. Northumbria do not use rangefinders and find difficulties, With rangefinders task is eased.
North West	Low difficulty.
Midlands	Low difficulty.
Southern	Low difficulty.
South West	Low difficulty.
Thames	Low difficulty.
Welsh	Moderate difficulty. Main difficulty is choosing point at which to measure width. Was happier with earlier suggestions of measuring at top middle and bottom of sampling area. Also problems with wide rivers which can't be crossed on foot.
SEPA (Clyde)	Low difficulty.
SEPA (Solway)	Low difficulty. Always measured at site during survey.
DoE (NI)	Low difficulty. Main difficulty except in bigger rivers – not convinced by rangefinders

Mean depth

Anglian	Low difficulty. Main difficulty: deep, unwadeable stretches need to be estimated. Also where channel depth is highly variable.
North East	Moderate difficulty. In deep stretches or if cannot get across full width of river.
North West	Low difficulty.
Midlands	Low difficulty. Problem for the three field variables (width, depth and substratum) having to obtain third season data when only two samples required. Extra resources needed
Southern	Low difficulty.
South West	Low difficulty.
Thames	Low difficulty.
Welsh	Moderate difficulty. Main difficulties as for width.
SEPA (Clyde)	Low difficulty.
SEPA (Solway)	Low difficulty. Always measured at site during survey.
DoE (NI)	Low difficulty. Main difficulty very deep rivers.

Question 42 (continued): Difficulty of measurement

Substratum cover

Anglian	High difficulty. Main difficulty : ultimately, this is always an estimate. No practical means of measuring this exists. Also difficult when thin layer of superficial sediment and where bed is covered with algae or deep, turbid water.
North East	Moderate difficulty. Very subjective measure.
North West	High difficulty. Extremely variable between people. Easy to do but errors can be great. Poor reproducibility.
Midlands	Moderate difficulty. Medium /deep rivers. Can be subjective and variable depending on individual.
Southern	Moderate difficulty.
South West	High difficulty as it is very subjective.
Thames	Low to moderate difficulty. Difficult to see through turbid water etc.
Welsh	Moderate difficulty. Main difficulty is coming up with an overall average for the whole sampling area. Inevitably pretty subjective especially in Wales around the boulders/cobbles and pebbles/gravel boundary. Helped by reaching consensus between two individual assessors.
SEPA (Clyde)	Moderate difficulty. Occasional problem of disagreement between samplers. This was overcome by discussion.
SEPA (Solway)	Low difficulty. Average percentage substratum cover calculated easily.
DoE (NI)	Moderate difficulty. Main difficulty is that I am not convinced of the accuracy but then this applies to the original survey on which RIVPACS was based. We double man to reduce errors.

Question 43: For the majority of sites, which of the following determinants were measured by more than one independent person in each season as a means of quality control on the accuracy of the data acquisition? Tick all that apply.

Region	***Depth***	***Width***	***Substratum***
Anglian			
North East			
North West			
Midlands	None – all sites sampled by lone workers		
Southern	**X**	**X**	**X**
Thames			**Some**
Welsh			**X**
SEPA (Clyde)			**X**
SEPA (Solway)	**X**	**X**	**X**
	Different samplers can collect the samples in each season, so you could have three completely different results from the previous study.		
DoE (NI)			**X**

Question 44: What other environmental variables would you like to see recorded during GQAs for use for predictive or interpretative purposes?

Anglian	Plants; Land use
North West	Record of known incidents
Midlands	Natural/modified bed/channel
Southern	Macrophyte cover and flow
South West	None
Thames	Plant cover. Geology (rather than alkalinity?). Other habitats (e.g. tree roots). Flow variables. RHS. Investigation of low alkalinities. Use of long term averages.
Welsh	None. There's enough to do already!
DoE (NI)	Geomorphology for beginners.

Question 45: If you feel that there are any differences between your component Areas/laboratories, in the difficulties associated with recording environmental data for RIVPACS, which are significant and need recording, please give them.

North West	More urban in south Area, therefore more difficulty with map work.
Thames	I suspect that data checking was more rigorous in Waltham Cross. At Fobney more reliance was put on data already in the Regional database.

Question 46: Please use the following text box to provide any additional comments you may wish to make on the difficulties associated with recording environmental data for RIVPACS

North East	The training day prior to the 1995 survey was very helpful in addressing some of these problems.
North West	The necessity for 3rd season environmental measurements at new sites is impractical – may as well take a sample!

Question 47: The 1995 GQA manual (BT 001, Chapter 4) included extensive recommendations concerning the specifications of the equipment to be used during the survey. Please make any comments you wish on the equipment specifications provided in the manual, including difficulties encountered in using any of the equipment and recommendations for better alternatives. Please note any significant differences between Areas.

North East	Some concern over the weight and safety implications in using the dredge were expressed.
Midlands	BT001 dredge too heavy. Duncan naturalist dredge used wt 2kg, aperture 10x30 cms 1mm mesh.
Southern	The dredge is TOO HEAVY! Use of a Roamer (or similar system) for NGRs should be mandatory.
South West	Not qualified to respond.
DoE (NI)	Deep water sampling needs to be resolved we need better standardisation that we can all live with.

Question 48: In addition to any use that was made of the biological and environmental data for national reporting or evaluation of the 1995 GQA, were any reports on the survey produced at your Area or regional level?

Anglian	Yes
North East	Yes
North West	No
Midlands	Yes
Southern	Yes
South West	No
Thames	Yes
Welsh	Yes
SEPA (both)	Yes
DoE (NI)	Yes

Question 49: If you answered yes to Question 48, please list the reports. Please indicate with an asterisk which, if any, of these reports included comparisons between the results of the 1990 RQS and the 1995 GQA.

Anglian	1. Monthly data summaries and failure reports for staff in water quality, planning, fisheries, water resources, conservation, etc.
	2. Regional water quality report, produced as collaborative effort by regional/Area staff.
	3. One-off, site/stretch/area specific reports, to address specific issues or concerns (multi-functional use).
North East	General report covering all sites but no comparison made with 1990 data in Dales
	Northumbria and Ridings did not produce any separate report.
North West	LEAPs
Midlands	Regional WQ annual report, tables have chemistry and biology together and compare with objective, previous survey and previous quinquennial survey.
	LEAPS
Southern	Catchment reports at Area level
	Regional 90-95 Biological Quality Report
	All included temporal comparisons.
Thames	Biological map and report (includes a comparison of the 1990 RQS and 1995 GQA results).
	GQA and BMWP banding.

Question 49 (continued): Reports, including comparisons between the results of the 1990 RQS and the 1995 GQA.

Welsh	Simple reports of preliminary results were produced for every catchment in Welsh Region to indicate progress achieved with catchment management plans and to highlight quality problems. All include comparison with 1990 data at a fairly simplistic level.
SEPA (Clyde)	The following report made use of the provisional spring data for 1995: Solway review, 1954-1996. Ed. F.H. Begg, A. McNeill and D.A. Rendall. Pub. Solway River Purification Board.
SEPA (Solway)	An internal report was produced for circulation within this Area only. The report was a summary of the results from our Area. The results from 1990 survey were included for comparison.
DoE (NI)	A site by site summary and overview of river quality running to 666 pages with maps. Includes macrophyte and chemical classes.

Question 50: Please give a brief description of the type of reports produced in your Region which made use of the 1995 GQA biological and environmental data for purposes other than recording the proportions of river length in different biological grades.

Anglian	See 1 & 3 in Question 49. As well as reporting the proportion of river in each class the Regional Report produced a map of classification, net upgrade of quality by km, ranked `hit-lists` of significant upgrades and downgrades by stretch together with explanatory text and mismatches between chemical and biological quality to aid interpretation of water quality. Recommendations for use of biological data for assessing and interpreting water quality.
North East	Data used in answering information requests from within and external to the Environment Agency. Annual quality reports include past data including the 1995 survey results.
Midlands	Routine survey reports for operational / surveillance purposes are produced for EP and FER staff and others. These will compare all historical data. Quinquennial survey will be used but less useful as it is combined season. Most use is made of single season assessments including those in the quinquennial year.
Southern	Catchment & Regional Reports (see reply to Question 49) give much more detail of quality and changes. Data extensively used in operational situations eg impact assessments, pollution incidents etc.
Thames	LEAPs. Pollution reports. Conservation. Planning proposals. Abstraction licence. Discharge consents. Regional State of the Environment Report. Public enquiries. Student projects (education).
Welsh	See reply to Question 49.
DoE (NI)	We produced a glossy summary report.

Question 51: Please list, very briefly, any other uses you would like to see made of the 1995 GQA data at regional or national level.

Anglian	1.Flow indexing (contribution to setting flow objectives-need annual data-sets for this). 2.Identifying stretches/sites of conservation importance (indexed, if necessary). Need species-level data for this.
North East	Use to produce taxa distribution maps, and assessment of environmental requirements of taxa.
Southern	A proper, published report of the biological survey.
South West	Included in LEAP reports.
Thames	Biological Quality Objectives Definition of "Good Ecological Status". A national State of the Environment Report.
DoE (NI)	Given the cost this is a matter that should be considered by BTG. I would dearly love up-to-date distribution maps.

Question 52: Do you consider that in the 1995 GQA your Region had a consistent approach to all aspects of the GQA (which could affect the data) across its component Areas?

Anglian	Yes
North East	No
North West	Yes
Midlands	Yes (No for Upper Severn)
Southern	Yes
South West	Yes
Thames	Yes (but see reply to Question 53)
Welsh	Yes
SEPA (both)	Yes
DoE (NI)	Yes

Question 53: If you answered No to Question 52, please list the major variations that occurred in the following text box.

North East	Links to Chemical sites and the procedures associated with the review of sites for 1995 were tackled in different ways in each Area resulting in different approaches to similar problems and a lack of continuity within the Region.
Midlands	Upper Severn 90/95 comparison - loss of high numbers of long term sites and therefore continuity. Improved AQC meant apparent not real improvement in quality.
Thames	More likely to be differences between samplers. The six "old retainers" would have known their sites whilst the twelve "new" samplers (of which we had many) would be more error prone and less efficient.

Question 54: Please use the following text box to record any comments you wish to make about the 1995 GQA which are relevant to the design and implementation of future surveys and which have not been covered by the preceding question.

North East	The 1995 survey was driven by requirements of Water Quality. Future surveys need to be chosen to be biologically relevant. The new review should address this but may not because of the constraints being imposed at the start of the exercise.
North West	Need to continue the development of systems for canals/still waters and lowland ditches. One in five years causes confusion when chemistry reports 1999 and biology still reports 1995. . How do we cope with atypical years? Present cycle of 5 yearly surveys misses AMP cycle.
Southern	The Agency must decide ASAP on the requirement for surveys in non-quinquennial years.
Welsh	Just to re-inforce the points. As described above 1995 GQA may have overestimated quality in headwaters for two reasons:- 1. There were no sampling sites in a number of classified headwater stretches that are now known to suffer severely from sheep dip pollution. 2. In some cases acidification impacts were not clearly flagged by the survey because relatively poor spring and autumn taxon lists could combine to produce quite a respectable N-Taxa EQI and push sites up into Band b when c was more reasonable or c when d was perhaps the most appropriate.

4.2.2 The supplementary questionnaire

Question S1: The 1990 River Quality Survey involved the collection of macro-invertebrate samples from each of three seasons, spring, summer and autumn, partly on the advice of the Institute of Freshwater Ecology team responsible for the development of RIVPACS. In order to enhance the level of coverage of sites in 1995 within the available budget and to maintain the other operational duties of the Agency biologists, the number of visits to each site was reduced to two. Single samples were taken in each of spring and autumn. It was claimed that this would not result in unacceptable reduction in the reliability of evaluations of environmental quality derived from RIVPACS. Under the system of collecting a single sample per visit, what do you consider the **optimal number** of sampling visits to each site to provide a reliable estimate of the biological condition of a site over the year of sampling as a whole?

North East	Two
Midlands	Three
Southern	Three
South West	Two
Thames	Two (Paul Logan) or three (John Murray-Bligh)
Welsh	Two
SEPA (Clyde)	Two
SEPA (Solway)	Two
DoE (NI)	Three

Question S2: Under your optimal sampling programme, when do you consider sampling should take place?

North East	Spring and Autumn. NB New sites will require a third visit to establish environmental variables.
Midlands	Spring, summer and autumn
Southern	RIVPACS spring summer autumn
South West	Spring and autumn
Thames	Spring and autumn (Paul Logan) Spring, summer and autumn? (John Murray-Bligh)
Welsh	Spring and autumn are OK but as mentioned in the main questionnaire poor samples in spring and autumn can combine to give a reasonable banding – is this a rare example of 'two wrongs making a right'?!
SEPA (Clyde)	Spring and summer and/or autumn.
SEPA (Solway)	Spring (essential) and autumn or summer. Spring and another season is achievable given present resources.
DoE (NI)	Spring, summer and autumn in order to give effective comparisons with RIVPACS – do we need winter RIVPACS?

Question S3: Under the system of collecting a single sample per visit, what do you consider the **minimum number** of sampling visits to each site to provide an acceptable estimate of the biological condition of a site over the year of sampling as a whole.

North East	One
Midlands	Two
Southern	Two
South West	Two
Thames	One or two
Welsh	Two
SEPA (Clyde)	Two
SEPA (Solway)	Two
DoE (NI)	Three

Question S4: Under your minimum sampling programme, when do you consider sampling should take place?

North East	Autumn.
Midlands	Spring and autumn. Note need to extend seasons i.e. Feb 1st – May 31st and Sept 1st – Dec 31st.
Southern	Any 2 seasons with at least 3 months between samples.
South West	Spring and autumn
Thames	Autumn (Paul Logan) Depends on the variability of the sampling site and when impacts are most likely. Need to have some long term data to make the decision. (John Murray-Bligh).
Welsh	See response to Question S2
SEPA Clyde)	See response to Question S2
SEPA (Solway)	See response to Question S2
DoE (NI)	See response to Question S2

Question S5: If you have had experience of sampling programmes that have involved both three and two seasons single sample collections (e.g. the 1990 RQS and 1995 GQA), to what extent do you think the reliability of the assessments made from two single season samples was poorer than three seasons?

North East	No apparent difference
Midlands	Slightly poorer
Southern	No apparent difference
South West	Slightly poorer
Thames	No apparent difference or slightly poorer.
Welsh	No apparent difference.
SEPA (both)	No apparent difference.
DoE (NI)	Much poorer.

Question S6: Do you think that replicate sampling would improve the quality of assessments of the biological condition of sites?

North East	No
Midlands	No
Southern	Yes
South West	Don't know
Thames	Yes and no. Rather increase the number of replicate sites in a reach.
Welsh	No
SEPA (both)	No
DoE (NI)	No

Question S7: If you answered yes to Question S6, what is your optimal replicate sampling regime, in terms of number of seasons and numbers of replicates per season?

Thames	Five samples per season for two seasons.

Question S8: If you answered yes to Question S6 and completed Question S7, what seasons/months would you recommend for sampling?

Southern	Replicate sampling would be based on statistics but is completely out of the question regarding resources.
Thames	Requires R&D to sort this out.

Question S9: If you feel that there are any differences between your component Areas/laboratories, in the number of samples that should be collected and the timing of sampling, which are significant and need recording, please give them.

Midlands	This Region suggests that the RIVPACS grading of combining seasons is not the optimal system. We accept that the system may give a better assessment at clean water sites where natural seasonal differences in the fauna may be important. However in polluted or intermittently polluted watercourses the combining of samples may result in a masking of a seasonal input e.g. sheep dip particularly where presence / absence data is used. We would recommend use of an average EQI . Greater confidence in the assessment could be made by increasing sampling at variable sites or using a different time period, perhaps the same as the WQ period.

Question S10: Please provide any additional comments you may wish to make on. the number of samples that should be collected and the timing of sampling

DoE (NI)	Three season gives not just classification, it is a monitoring system on which problems can be detected and resolved. I even have a problem with which two seasons would be preferable. We have different problems in different seasons.

Question S11: Whilst accepting that this is a "how long is a length of string?" question, please estimate the **APPROXIMATE AVERAGE LENGTH OF TIME**, in minutes, you took to sort the following type of sample in the 1995 GQA. Your answers should take account of the range of samples from those with few individual taxa to those with numerous and/or diverse taxa.

Mainly gravel or coarser substratum with little detritus or macrophyte material

North East	Dales: 90 Ridings: 60 Northumbria: 60.
Midlands	75
Southern	180
South West	No relevant data available!
Thames	60-90
Welsh	120
SEPA (Clyde)	I do not feel that it is reasonable to answer this type of question with actually measuring. The supposed answer can be used in all sorts of unexpected and unpleasant ways!
DoE (NI)	45

Mainly gravel or coarser substratum with copious detritus and/or macrophyte material

North East	Dales: 360 Ridings: 240 Northumbria: 120.
Midlands	120
Southern	400
Thames	60-120
Welsh	220
DoE (NI)	90

Mainly sand with little detritus or macrophyte material

North East	Dales: 120 Ridings: 60 Northumbria: 90.
Midlands	75
Thames	60-90
Welsh	Very rare – 200?
DoE (NI)	45

Mainly sand with copious detritus and/or macrophyte material

North East	Dales: 360 Ridings: 240 Northumbria: 150.
Midlands	120
Thames	60-120
Welsh	Very rare – 220?
DoE (NI)	90

Question S11 (continued): Duration of sample sorting.

Mainly silt with little detritus or macrophyte material

North East	Dales: 90 Ridings: 120 Northumbria: 150.
Midlands	60
Southern	250
Thames	60-120
Welsh	Very rare – 220?
DoE (NI)	120

Mainly silt with copious detritus and/or macrophyte material

North East	Dales: 360 Ridings: 300 Northumbria: 240.
Midlands	120
Southern	600
Thames	60-180
Welsh	Very rare – 240?
DoE (NI)	120

Question S12: If you feel that there are any differences between your component Areas/laboratories, in the length of time you take to process various types of sample, which are significant and need recording, please give them.

Southern	Figures are regional means (guestimates!). A difference in site types & quality across the Region mean that Sussex and especially Hampshire samples take about 25% longer than Kent.
SEPA (Solway)	I think it is impossible to even estimate the average time for sorting various sample types. The range is anything from 20 minutes to one day (7hours) but we do not record this information and would worry if we were ever asked to. Every sample/sampler is different.

Question S13: What is your opinion of the following text descriptions of each grade of the 1995 RQS system? Please suggest any alternative wording you think appropriate.

GRADE a - VERY GOOD*: The biology is similar to (or better than) that expected for an average and unpolluted river of this size, type and location. There is a high diversity of Families, usually with several species in each. It is rare to find a dominance of any one Family.*

Southern	We have to say OK to all of these as I largely wrote them!
South West	OK
Thames	Take out reference to "high diversity". What if predicted has low number of families (e.g. mountain streams). "Better than" may include some polluted so remove. High status?
Welsh	OK
SEPA (Clyde)	OK
SEPA (Solway)	Fine

GRADE b - GOOD*: The biology shows minor differences from Grade a and falls a little short of that expected for an unpolluted river of this size, type and location. There may be a small reduction in the number of Families that are sensitive to pollution, and a moderate increase in the number of individuals in the Families that tolerate pollution (like worms and midges). This may indicate the first signs of organic pollution.*

North East	Dales comment: Suggests very little difference form Grade a but if losing any sensitive taxa and increasing the number of individuals in the pollution tolerant taxa then this suggests a more obvious effect.
South West	OK.
Thames	Remove organic or add to all others. Don't refer to other grades! Suggested revised wording: ***GOOD**: The biology falls a little short of that expected for an unpolluted river of this size, type and location. There may be a small loss of number of Families that are sensitive to pollution, and a moderate increase in the number of individuals in the Families that tolerate pollution (like worms and midges). This may indicate the first signs of pollution.*
Welsh	OK
SEPA (Clyde)	Too much emphasis on organic pollution to the exclusion of toxic pollution.
SEPA (Solway)	If this is the definition of good compared to very good then it only distinguishes differences due to organic pollution. What if a site was classified as good but you suspected it was due to other types of pollution (such as sheepdip) this definition would be wrong. Although the results indicate "good" – sensitive groups may have reduced but tolerant groups may not have increased.

Question S13 (continued): Text descriptors.

GRADE c - FAIRLY GOOD*: The biology is worse than that expected for an unpolluted river of this size, type and location. Many of the sensitive Families are absent or the number of individuals is reduced, and in many cases there is a marked rise in the number of individuals in the Families that tolerate pollution.*

North East	Dales comment: Biology is <u>obviously</u> worse than expected
South West	OK
Thames	Should "many" of the sensitive families be absent in "Fairly Good"? This might be the politically acceptable group! Moderate status.
Welsh	OK
SEPA (Clyde)	OK
SEPA (Solway)	Fine

GRADE d - FAIR*: The biology shows big differences from that expected for an unpolluted river of this size, type and location. Sensitive Families are scarce and contain only small numbers of individuals. There may be a range of those Families that tolerate pollution and some of these may have high numbers of individuals.*

North East	Dales comment: "Big differences" from that expected may be should be renamed **POOR**
South West	OK
Welsh	OK
SEPA (Clyde)	OK
SEPA (Solway)	I personally think fair describes quality that isn't too bad. So it would depend on what 'big differences' actually means.

GRADE e - POOR*: The biology is restricted to animals that tolerate pollution with some families dominant in terms of the numbers of individuals. Sensitive families will be rare or absent.*

North East	Dales comment: Restricted faunas- should be **VERY POOR**
South West	OK
Welsh	OK
SEPA (Clyde)	OK
SEPA (Solway)	Fine

Question S13 (continued): Text descriptors.

GRADE f - BAD*: The biology is limited to a small number of very tolerant families, often only worms, midge larvae, leeches and the water hoglouse. These may be present in very high numbers. Even these may be missing if the pollution is toxic. In the very worst case there may be no life present in the river.*

North East	Dales comment: How about **AWFUL**. Description too similar to grade e.
Midlands	Should read "no macroinvertebrate life" as there will be bacteria etc.
South West	OK
Thames	New concept of "toxic" added only for last group. What about mild toxicity earlier – or ignore and look for a new diagnostic score. Only stoneflies if acid pollution.
Welsh	OK
SEPA (Clyde)	OK
SEPA (Solway)	I think this should be something like 'grossly polluted' rather than just bad.
DoE (NI)	I think this is a BTG issue for the future.

5 EVALUATION OF QUESTIONNAIRE REPLIES

The references to individual Environment Regions and Areas in the following sections are based on the situation pertaining in 1999, unless otherwise stated.

5.1 Number of Samples

None of the respondents considered that, overall, too many sites were sampled and only two (25%) of the eight Regions who replied to this question felt that approximately the right number of sites had been sampled. These two Regions were Anglian and Midlands. No opinions were provided by North West and South West Environment Agency Regions.

Of the remaining respondents, the four Environment Agency Regions who replied each felt that slightly too few sites had been sampled in 1995. In contrast the Scottish and Northern Ireland respondents both felt that many fewer than necessary sites had been sampled in 1995. In particular, Northern Ireland felt that they needed to double the length of watercourses they monitor. The Scottish shortfall was stated to be in the Solway Area. Both there, and in Northern Ireland, steps are in hand to increase site coverage in 2000.

When the number of sites sampled was considered on a "site-type" basis there was a more varied response. In some cases there was a general consensus that site coverage was adequate. In particular, most Regions were satisfied with the number of large deep rivers that were sampled, irrespective of whether they were clean or polluted. In one case (Northern Ireland), it was even felt that too many polluted, large, deep rivers were sampled. Similarly, almost all respondents were satisfied with the number of clean and polluted middle reaches which were sampled, whilst Midlands and Thames Regions were the only two that wanted to increase the coverage of urban streams.

On the other hand there was a strong consensus that fewer than necessary headwater streams were sampled, particularly polluted watercourses. Every respondent felt that at least one category of headwater was under-represented in the 1995 survey, even though the number of headwaters in that survey was almost double the number sampled in 1990. Of the 30 separate responses by Region and headwater type, 15 (50%) were judged to have many fewer sites than necessary sampled in 1995, 6 (20%) slightly fewer and only 9 (30%) approximately right. The greatest perceived shortfall of headwater sites was in Anglian, North East, Southern, South West and Thames Regions and in Northern Ireland. Most satisfied with the previous intensity of headwater sampling were Midlands, Welsh and SEPA Regions, although Welsh Region drew attention to their need to sample many more polluted, upland headwaters.

Welsh Region also noted a preference to sample slightly more acidified sites, a view they shared with two other Regions Midlands and SEPA, where such sites are relatively common. In contrast, Regions such as Anglian, Southern and Northern Ireland, where such sites are relatively rare, felt no need to increase coverage. Welsh was also concerned to provide better coverage of headwaters that might be susceptible to sheep dip pollution. A preferred increase in the number of agriculturally enriched sites sampled was shared by North East, Welsh and especially Southern and SEPA Regions.

There was little pressure to increase the coverage of ditches and canals with the exception of North East Region, except that Southern Region felt the need to sample many more drains and ditches than 1995.

There were no significant inter-Area differences recorded by the responding Regions. Where differences were noted, then these were generally differences in the balance of the site types sampled, which reflected the particular geographic character of individual Regions, rather than gross differences in the needs to increase or decrease sampling of particular stream types in one Area rather than another.

Conclusion: In replying to the questionnaire the respondents make no case for reducing the number of sites sampled overall or for any particular site type. Most Regions would prefer an increase in coverage of one or more site types and only Northern Ireland suggest any category of site where some compensatory reductions in coverage could be made.

Recommendations: There should be no reduction in the coverage of sites in the 2000 GQA Survey, in comparison with 1995, unless there are resource limitations that cannot be overcome.

Where feasible, Regions should adjust their coverage of particular site types upward to rectify deficiencies they identified in the 1995 GQA. This appears to apply to headwaters in particular. This viewpoint re-iterates the findings of Furse & Symes. (1997) who demonstrated that, pro rata to their relative total stream length, headwaters were considerably under-sampled in comparison with larger watercourses. Furthermore, Furse *et al.* (1993) showed that headwater streams, i.e. those within 2.5km of their source, accounted for approximately 20% of the total species richness of individual river catchments.

Where resources are inadequate to meet the coverage recommended by Regional Biologists, consideration should be given to a rolling programme of GQA monitoring, spread over the full five years currently separating each GQA. Recommendations of this type were previously made by Furse (1995), in the context of biomonitoring of headwaters.

Although not specifically mentioned by respondents, staged sampling of deep water sites and canals would allow time for effective monitoring and evaluation techniques to be developed and tested.

5.2 Sampling Methods

5.2.1 Pond-net sampling

All Regions, including Scotland and Northern Ireland, adhered to the recommendations on pond-net sampling set out in BT001 (Environment Agency 1997a).

Adoption of the recommended techniques led to some changes in the procedures adopted by several Regions for their 1990 survey. In particular, the implementation of the one minute search was uniformly adopted in 1995, whereas many Regions did not undertake any form of search at some or all of their sites in 1990.

The biggest difference between surveys occurred in North West Region, where rapid biological appraisal of sites had long been used as an effective substitute for chemical monitoring. Their approach involved stopping searching samples for new taxa once the top biological grade had been achieved. Whilst this technique provided a means of quality grading individual sites, its incompatibility with the RIVPACS approach reduced the reliability of the application of the techniques for the statistical analysis of change that are available in RIVPACS III+. The adoption, by North West Region, of the recommended techniques in the 1995 survey will provide a better basis for analysis of temporal change between then and 2000, assuming that similar techniques are used in the latter year.

In answers to other sections of the questionnaire, both Southern and Midland Regions queried whether a different approach to the method and/or timing of pond-net samples in headwaters needed to be considered.

Conclusion: The methods of sampling used to collect macro-invertebrate samples in the 1995 survey were accepted and implemented by all Regions, despite the fact that some elements of the recommended procedures differed from those some Regions adopted in 1990.

It is possible that, although the general recommendations of BT001 (Environment Agency 1977) were followed by each Region, there may have been minor, inter-regional or inter-Area differences in the precise manner in which pond-net sampling was conducted. No such changes are evident from the responses to the current questionnaire.

Recommendations: The pond-net sampling procedures adopted in 1995 should be retained for the 2000 GQA Survey. This will provide a reliable basis for the application of the procedures for detecting temporal change incorporated in RIVPACS III+.

Where differences in the implementation of the recommended procedures are known to the Environment Agency, and where these may possibly have a significant effect on the results obtained, then the recommended procedures in the revised BT001 should be made more even more prescriptive, if that is possible, in order to eliminate these differences.

Special instructions should be included in the revised version of BT001 to be adopted for the 2000 GQA Survey, giving the modified sampling techniques that may need to be adopted in order to collect a three minute pond-net sample from narrow and shallow headwater streams. These include more active use of the hands to disturb the substratum and the creation, by foot, of shallow pools of sufficient depth to allow the net-bag to be positioned under water and downstream of the disturbed area.

Headwater sites incorporated in RIVPACS III+ were each sampled for three minutes, plus one minute's search. Despite the difficulties involved, this duration of sampling should be retained for the 2000 GQA Survey, in order that Ecological Quality Index (EQI) values obtained for these sites are based on a common level of sampling contributing to both observed and expected BMWP Index values. This will also best facilitate reliable temporal comparisons.

When appropriate, standardised deep-water sampling procedures are available, sampling with the standard FBA-style pond-net, with handle length of approximately 1m, should be confined to sites that are wadeable for at least 25% of their total width.

5.2.2 Deep-water sampling

There was considerable variation in the deep water sampling techniques used in the 1995 survey.

Four Regions, Anglian, Midlands, Southern and South West, used combined dredge and bankside netting. However, neither Anglian nor South West entirely followed the procedures recommended in BT001. Northern Ireland occasionally combined dredge sampling with bankside netting but preferred to use only the latter procedure. The method of bankside netting recommended to accompany deep-water techniques involved sweeping of the surface and vegetation only and not the benthos. This method was not liked by Southern who often reverted to net sampling only.

Two Regions, North West and Thames used bankside netting only, whilst Welsh used solely dredge sampling (albeit at two brackish water sites only). North East Region used either dredge or airlift samplers according to circumstances. According to their replies, and contrary to the recommendations shown in Figure 2.8 of BT001, neither Welsh nor North East Region combined their deep water sampling with additional bankside netting.

Variations from the recommended dredging procedures included not dredging parallel to the bank, using a dredge which differed from the recommended dimensions, deviating from the recommended number of hauls (both lower and higher numbers) and departing from the preferred method of sub-sampling the material collected. Regions also varied in their use of dredging from the bank or a boat, although both techniques are allowed for in BT001.

There were few differences between the procedures used by most Regions in 1995 and those used prior to that date. Thus, for example Anglian reduced their mesh size in 1995, Midlands introduced dredging for the first time and Southern replaced additional netting of the benthos with bankside sweeps only.

Conclusion: Deep water sampling was not standardised during the 1995 GQA and most Regions used the same procedures, or slightly modified versions of the same procedures, that they had operated prior to the 1995 GQA. Similar conclusions were drawn by Wright *et al*. (1999).

Until recently, no attempt was made to compare the efficiency of the different sampling procedures at obtaining a diverse and representative fauna. This issue is now in hand through research being carried out for the Environment Agency by IFE (National R&D Project E1-007). The current research is being carried out on six rivers, two each in Yorkshire, Lincolnshire and Somerset. Sampling equipment being compared include dredges, air-lifts and long- and standard-handled pond-nets. Sampling is being undertaken from both the bank and from a boat.

At two sites replicate sampling of either dredge or air-lift samples is being undertaken by both IFE and Environment Agency personnel in order to determine the sort of variation that is associated with each form of sampling. This will provide an initial example of the sort of information on sampling variation that was obtained by Furse *et al*. (1995) for standard pond-net sampling and subsequently incorporated in RIVPACS III+ (Clarke *et al*. 1997). A more detailed investigation would be required in order to provide a reliable indication of the sort of variation associated with deep water sampling procedures.

The results of the IFE study are not due to be reported by March 2000 and will therefore not be available during the planning stage of GQA 2000 Survey.

Recommendations: The Environment Agency should introduce standard procedures for the sampling of deep water sites, i.e. those that are not suitable for standard pond-net sampling as defined in 5.2.1 above.

The Agency should be advised by the findings of the National R&D Project (E1-007) in selecting the most appropriate, standard, deep-water sampling procedures for use in national GQAs.

All biologists involved in field sampling of deep-water sites should receive appropriate training prior to undertaking sampling.

The Environment Agency should consider staging sampling for the 2000 GQA Survey, with deep-water sampling taking place in 2001. This will allow more detailed consideration of the results of the National R&D Project (E1-007) and time for adequate training and practice in the application of standard deep-water sampling techniques.

5.3 Sample Sorting

5.3.1 Location of sample processing

In manuals BT001 and BT002, laboratory sorting of samples was stated to be mandatory and bankside sorting was banned. Within the Environment Agency this request was stated to have been adhered to by all Environment Agency Regions and biologists, with the exception of a single individual.

Whilst bankside sorting and identification was not used by North West Region for the 1995 survey, they present a range of compelling arguments in favour of this approach where rapid assessment of a wide number of sites is required for operational purposes. Both North West and Thames Regions argue the advantages of bankside processing where rapid assessment is required and both Regions feel the characteristic movements of different families of invertebrates are aids to quick identification.

A dual approach was adopted in the former Solway RPB Region of Scotland, with preliminary field sorting and identification, for operational purposes, being followed by a more detailed laboratory examination for GQA purposes.

Conclusion: Whilst questionnaire replies indicated that all but one biologist participating in the 1995 GQA sorted and identified samples in the laboratory, not all Regions were convinced that this represented the most cost-effective procedure. The advantages and disadvantages of bankside sorting and identification have long been a contentious issue within the water industry with protagonists and detractors for both approaches.

Whether bankside processing provides as comprehensive and accurately identified taxon list as laboratory processing remains unproven. In a recent and very detailed Australian study of this issue Humphrey *et al*. (in press) illustrated the potential of bankside processing to result in higher error rates, particularly as the result of poor recovery of small and cryptic taxa. Many of the missed taxa were common in samples. They concluded that these errors could lead to increased frequency of misclassification and inaccurate model predictions and support this conclusion by simulations based on actual errors recorded by QA/QC audits.

Recommendations: Whilst it remains unclear whether bankside sample processing is as efficient and comprehensive as laboratory processing, this issue is considered to be so important that a standardised, laboratory-based approach should continue to be prescribed for use in the 2000 GQA Survey. This action will ensure that observed BMWP index values and those predicted by RIVPACS III+ are based on the same sorting procedure.

5.3.2 Transport, fixation and preservation of samples

Six of the eleven respondents stated that all samples were transported live from the sampling site to the laboratory. In two other Regions/Areas live transport was commonly, though not exclusively used. Only Southern and Welsh Regions consistently transported samples fixed in formaldehyde. North West Region (always) and Thames Region (sometimes) used alcohol to preserve and transport samples.

Where samples were transported live, two factors which were common to several Regions' approaches were to drain the sample of as much water as possible prior to transportation and to store the sample in a cool box, often containing ice packs. In-house research conducted on behalf of Anglian Region showed that in their laboratories there were no statistical differences between the recovery of macro-invertebrates from live samples or those preserved in Formalin.

Some Regions fixed or preserved their samples on return to the laboratory but others sorted the samples live. In the three Regions where live sorting was practised a maximum period, after return to the laboratory, was set during which time the sample had to be sorted. This time varied between 36 hours (Northern Ireland), 48 hours (Midlands) and four days (Anglian). In each Region samples were stored in a refrigerator prior to processing.

Most Environment Agency laboratories had suitable laboratory facilities for the handling and use of formaldehyde. The only two noted exceptions were single Area laboratories in each of the North West and Midlands Regions. However, two Regions which have suitable facilities, Anglian and South West appear not to use them for Health and Safety reasons, although the reply from South West is somewhat ambiguous in its possible interpretation.

Conclusion: The responses to the questions relating to formaldehyde indicate a growing reluctance to use it for any stage of the collecting, processing and retaining of macro-invertebrate samples. As Anglian Region correctly point out, it may not be necessary to standardise all elements of the processes of transporting, storing and sorting samples and this allows for regional differences in attitude to the use of formaldehyde. Nevertheless, Health and Safety requirements are likely to become increasingly stringent and certain recommendations can be made which will limit the necessity to use Formalin.

Recommendations: It is recommended that, wherever local Health and Safety protocols allow, samples should be fixed in Formalin for transportation to the laboratory. In all usages of Formalin for sample fixation and storage neutral-buffered solutions are preferred. Where this is not permitted by local Health and Safety protocols, then samples should be transported live to the laboratory, on the day of collection, for either live sorting or for fixing in formalin.

During live transport, samples should be drained of as much water as possible and carried in a cool box containing ice packs. These measures will reduce predation, and de-oxygenation.

Live samples should always be stored in a refrigerator.

Live samples should always be fully sorted within two working days of their day of collection. If this cannot be achieved then the samples should either be fixed in Formalin or preserved in alcohol.

Sorted samples, once re-constituted, may be either fixed in formalin or preserved in alcohol. The former is preferred. It must be ensured that the quantities of Formalin and, **especially** alcohol added to re-constituted samples should be of sufficient strength to exclude the possibility of the sample decaying, especially if they contain large quantities of organic material

All samples which are re-constituted for internal AQC or external audit should be fixed in formalin or preserved in alcohol, according to the Health and Safety procedures operating in the respective Area, immediately after sorting is completed. Removed animals should always be preserved in alcohol.

Alcohol is a preservative and not a fixative. Use of alcohol without prior fixation may lead to soft-bodied animals breaking up and becoming un-identifiable. Re-constituted samples which have not been fixed previously with formalin, and which are subject to AQC or audit, should be re-analysed within two weeks. All samples for external audit should be dispatched to the auditors as soon as possible and within two weeks of the date on which the last sample was analysed for AQC (Environment Agency 1996a).

5.3.3 Sorting time

The sorting time for particular sample types varied considerably from Region to Region (Table 5.1).

Table 5.1 Minimum and maximum times, in minutes, taken to sort different forms of sample in different Regions/Areas

Sample type	Minimum sorting time	Maximum sorting time
Mainly gravel or coarser substratum with little detritus or macrophyte material	**45**	**180**
Mainly gravel or coarser substratum with copious detritus and/or macrophyte material	**60**	**400**
Mainly sand with little detritus or macrophyte material	**45**	**120**
Mainly sand with copious detritus and/or macrophyte material	**60**	**360**
Mainly silt with little detritus or macrophyte material	**60**	**250**
Mainly silt with copious detritus and/or macrophyte material	**60**	**600**

Conclusions: It is impractical to standardise sample sorting times and this should never be attempted. The estimates provided are approximate "guesstimates" for loosely defined categories of samples. Methodologies and sample types will vary considerably between Regions, even within the six defined categories.

Notwithstanding this, in their reply to the questionnaire, all Regions claimed to be meeting their internal AQC targets for gross gains. Similarly, the two latest primary audit reports (Gunn *et al.* 1998, 1999) show that only 18% of regions per year exceeded a mean gross gain of 2 taxa (maximum 2.3) and 18% of Area laboratories (maximum 3.1). Yet sample processing time varies, per sample type (Table 5.1), by factors that range from 2.67 to 10.0 and absolute time differences, which range from 75 minutes to 540 minutes (nine hours). Within this variation there may be some opportunity for the quicker Regions to discuss with the slower Regions the techniques they use to achieve AQC standards in so much shorter time. This may include differences in their field sampling techniques, types of river being sampled and the experience of the people undertaking sampling, all which may result in different size samples being retained for processing.

Recommendations: As a consequence of the wide variation in sampling times, the Environment Agency should promote the exchange of ideas on the best methods of meeting the AQC standard, for the mean number of missed taxa per sample, in the most cost-effective manner.

5.4 Identification and Quantification

5.4.1 Identification

The only level of identification common to all Regions for the 1995 GQA was BMWP family level.

All Regions attempted to identify some of the non-BMWP families but only Welsh Region attempted the identification of all families.

Similarly, most Regions attempted species level identification of some taxonomic groups in some samples. Only Thames and Welsh Regions and SEPA never attempted any species level identifications for GQA samples. Despite this, it is known that both Thames and Welsh Regions have staff with the expertise to undertake species identifications when specifically required for the purposes of the sampling programme.

Anglian Region was the only one to attempt the same comprehensive level of identification used by RIVPACS.

The types of taxa which present the most difficulties vary both within and between Regions as does the level of precision at which each Region lists their difficulties. An attempt to summarise the problem groups/taxa illustrates this variability (Table 5.2).

Table 5.2 A list of the taxa (and life stages) cited by Environment Agency biologists as being the most difficult to identify.

Taxon (and life stage)	Number of Regions/Areas listing difficulties
Tricladida	2
Terrestrial Molluscs	1
Asellidae v terrestrial Isopoda	1
Hydrophilidae (larvae and/or adults)	3
Dryopidae	1
Capniidae	1
Early instar Plecoptera	1
Platycnemidae	1
Coenagrionidae	1
Zygoptera	1
Psychomyiidae	1
Small cased Trichoptera	7
Trichoptera pupae	4
Tipulidae	2
Diptera larvae (+ Diptera in general)	5
Diptera pupae	1
Pupae in general	1

The main difficulties appear to be associated with adult and larval Hydrophilidae, the larval stages of Trichoptera and Diptera and pupae of all groups, although several other taxa get one or two mentions.

These replies may be compared with analyses of the proportions of occurrences of each taxon which were missed from approximately 200 audit samples selected at random for each of the 1990 RQS (Furse *et al*. 1995) and 1995 (GQA). The taxa which the NRA as a whole seemed to have the most difficulty picking out and identifying, i.e. those not recorded in over 25% of the samples in which they were present in each year were Dendrocoelidae, Hydrometridae, Hydrophilidae, Scirtidae, Psychomyiidae, Hydroptilidae, Lepidostomatidae and Beraeidae (Table 5.3).

Table 5.3 All BMWP families which were missed on at least 25% of occasions on which they were recorded as present in audit samples by IFE in either or both of 1990 and 1995.

BMWP scoring taxa	Proportion of samples in which the taxon was missed by the NRA primary sample processing in 1990.	Proportion of samples in which the taxon was missed by the NRA primary sample processing in 1995.
Mesoveliidae	100.0	0.0
Dryopidae	100.0	0.0
Dendrocoelidae	53.0	27.8
Hydrometridae	50.0	33.3
Scirtidae	50.0	30.8
Hydroptilidae	50.0	32.4
Beraeidae	50.0	66.7
Valvatidae	43.0	18.9
Hydrophilidae	42.0	29.4
Psychomyiidae	40.0	26.0
Goeridae	40.0	17.0
Viviparidae	33.0	0.0
Naucoridae	33.0	0.0
Brachycentridae	33.0	0.0
Molannidae	33.0	12.5
Lepidostomatidae	32.0	8.5
Physidae	29.0	7.8
Planorbidae	29.0	12.0
Leptophlebiidae	25.0	6.8
Corophiidae	20.0	75.0
Perlidae	17.0	33.3
Siphlonuridae	0.0	100.0
Libellulidae	0.0	50.0
Hygrobiidae	0.0	100.0

The number of taxa missed on 25% of their occurrences decreased from 19 in 1990 to twelve in 1995 (Table 5.3).

The taxa most commonly incorrectly assumed present in audit samples from 1990 and 1992 (Furse *et al*. 1995) were Dendrocoelidae, Mesoveliidae, Gerridae and Lepidostomatidae (in both 1990 and 1992), Valvatidae (1990 only), and Physidae (1992 only). However, the number of cases of each was usually no more than one per Region per year. Comparable analyses have not yet been undertaken with the 1995 GQA external audit samples.

Conclusions: All Regions appear to be competent in the level of identification required to determine BMWP and EQI index values. Most problems are encountered with juvenile caddis, non-BMWP Diptera larvae and the pupae of most taxonomic groups. Results of the IFE audits of 1992 suggest that mis-identifications were rare and the impact upon BMWP and EQI index values much less significant than the failure to remove all taxa present from samples. However, the development of species-based indices, such as LIFE, places an increasing requirement upon the Environment Agency to train their biologists to identify most major taxonomic groups to species level, or to generic or species group level for most families of Diptera.

Recommendations: Each Environment Agency Region should have a policy in place to train biologists to the level of competence in macro-invertebrate identification to meet the needs of the 2000 GQA Survey. Wherever possible, such training should be provided in house but, where necessary, specialist external training courses should be arranged.

The need to achieve more precise levels of identification, including during national surveys, will require some biologists within each Area laboratory to be competent at species level identification and appropriate training should be planned, where necessary.

5.4.2 Quantification

Most Regions accepted the value of the current abundance category system; 1 = 1-9 individuals, 2 = 10-99 individuals, 3 = 100-999 individuals, 4 = 1,000-9,999 individuals and 5 = >10,000 individuals per sample. However many Regions wish to sub-divide the categories because they believe this to provide more meaningful information. Several Regions adopted a more detailed categorisation during the 1995 GQA. The greatest need for sub-division of categories was stated to be in the range 1-9 individuals per sample.

Although not explicitly stated, most Regions appeared to estimate abundance classes rather than apply direct counts of individuals per family.

Conclusions: The development of the LIFE index and the new procedures being developed for incorporating abundance categories in RIVPACS (National R&D Project E1-007) and Artificial Intelligence applications (Walley & Martin 1997; Walley *et al*. 1998) emphasise the importance of obtaining estimates of the number of individuals of each taxon in each sample. In most applications related to GQAs this will be numbers of individuals of each BMWP family.

Experience with the current R&D programmes has shown national interpretation of GQA data crucially depends on a common level of data from all sources. This includes a common level of quantification. Failure to comply with this recommendation, as was the case with North West and, to a lesser extent, Midlands Regions in 1995 can present considerable difficulties (Walley & Martin 1998; Davy-Bowker *et al*. 2000).

In establishing a system of categorisation it must be remembered that pond-net sampling is not a quantitative procedure and abundance categories are intended to give an indication of the relative abundance of taxa and not absolute numbers. Any system of categorisation should avoid a giving a false impression of the precision with which abundances of individual taxa are estimated.

The introduction of a detailed level of quantification also presents practitioners with a more difficult task of allocating the "correct" abundance class to each taxon. North East Region draw attention to the problems they incur in selecting which category taxa should be assigned to when their numbers appear to be near a class boundary. Anglian Region feels that a practical form of AQC on abundance estimates would be helpful.

Recommendations: It is essential that a system of allocation of abundance classes to each BMWP family is adopted by all Regions for the 2000 GQA Survey.

The system adopted by each Region should be standardised, or be capable of standardisation, between Regions.

The number of categories should be sufficient to distinguish significant differences in abundance but not so many as to give a false impression of the accuracy of the quantification or to present unacceptable difficulties in allocating taxa to categories.

The coding of abundance categories should reflect and be easily convertible to the five categories recommended for the 1995 GQA and given above. This is best achieved by retaining the original category numbers as prefixes to the more detailed new categories. A system of categorisation is given for consideration (Table 5.4).

Table 5.4 A suggested new system of abundance categories to be used during the 2000 GQA Survey.

Number of individuals per taxon per sample	New abundance category	Log_{10} based abundance category (as per 1995 GQA)
1	11	1
2-4	12	1
5-9	15	1
10-49	20	2
50-99	25	2
100-999	30	3
1000-9999	40	4
>10000	50	5

Whatever system of categorisation is adopted by each Region for the 2000 GQA Survey, the data must be presented to the National Database in the standard categories adopted for the survey (e.g. the suggestion above).

Where present, and where more precise identification is not subsequently required, then a minimum of ten representatives of each BMWP family should be **counted** for each sample. This will facilitate the correct allocation of taxa to any sub-division of abundance category 1 (*sensu* the 1995 GQA) and also the correct allocation of taxa on the borderline of categories 1 and 2.

Numbers of individual taxa in excess of ten, and allocation of these taxa to abundance categories should be estimated as recommended in manual BT001 (Environment Agency 1997a).

Where the abundance category of a taxon with more than nine individuals present is in doubt then the taxon should always be assigned to the lower abundance category of the two possible categories in question.

AQC analysts and auditors should also assign each taxon to an abundance category when re-processing samples (see also 5.5.5). They should use the same procedures for assigning taxa to categories as those adopted by the primary analyst.

Frequent disparities between abundance categories assigned by the primary sorter and the AQC analyst should be investigated, particularly if these differences are skewed in a constant direction. Total counts may be necessary to resolve differences.

If abundance-based indices are to be used to report on the results of the 2000GQA Survey, then abundance checks should be incorporated in quality control procedures.

Recommendations on the quantification of individual species would be more complex than those given above for BMWP families and are outside the scope of Question 24 of the questionnaire.

5.5 Internal AQC and External Audit

5.5.1 Value of the internal AQC

With few exceptions, there was a near-universal approval for the use of internal AQC procedures, as laid out in the Agency's draft procedures for quality assurance manual, BT003 (Environment Agency 1996a), to check on the efficiency of sample sorting and identification.

The most commonly cited benefit of the system was that it provided rapid feedback on the quality of individuals' performances and allowed immediate remedial measures to be put into effect. In general staff found the process positive and encouraging.

The exceptions were an Area Laboratory which questioned the cost benefit of the procedure, especially as results were not modified following identification of errors, and a single staff member who thought it unnecessary to check the quality of experienced biologists' performance.

Conclusion: The value of the AQC system is accepted by a large majority of biologists and the exercise is seen to be valuable.

Recommendations: The use of internal AQC should be continued for the 2000 GQA Survey.

5.5.2 Method of selection of samples for internal AQC and external audit

Internal Analytical Quality Control: All Regions have well defined systems of sample selection which are in line with the recommendations made in BT003. With the single exception of the Northumbria Area of North East Region, these are based on the selection of one in ten samples for AQC. The process used by most Regions is an objective one, selecting blind from either ten coloured balls, one of which is coloured differently from the others, or from ten sample pot lids. However, a more subjective approach to sample selection is inferred from the answers provided by the Central Area of North West Region, Northumbria Area and SEPA. Even so, all three of these Areas/Regions claim random selection without describing the processes involved.

Whilst all Regions appear to operate similar selection procedures, there is one major dichotomy in the practices adopted by different laboratories. In some cases the random selection procedure is operated immediately after the sample has had its primary processing completed. In other cases batches of ten samples are assembled or logged and once all ten are available one of these is selected at random. The former approach provides rapid feedback but may lead to prolonged periods without a sample being selected. The latter approach, which is strongly advocated by Welsh Region, and which appears to be more in line with BT003, provides a regular supply of AQCed samples but often introduces a delay between primary sample processing and the AQC.

Conclusion: Each Region either describes a random AQC selection procedure or claims to operate a random procedure but does not describe that procedure. Differences exist between the precise procedures used but all approaches appear to fulfil the requirements of the Agency's AQC scheme, as laid out in procedures manual BT003. The authors of the current report have a slight preference for the practice, advocated by Welsh Region and applied by them and most other laboratories, in which sample selection for audit only occurs after ten samples are available. This system, which is the procedure currently advocated nationally by the Agency, guarantees a regular rate of samples being AQCed without introducing major delays in the feedback process.

Recommendations: The AQC sample selection procedures operated by each Region/Area appear to fulfil the requirements of the quality control system and no need is envisaged to prescribe greater standardisation of approach.

External audit: The range of practices used to select samples for external audit is considerably more variable than those used to make selections for internal AQC. Each Region is aware of the total number of samples that have been contracted out for audit and takes suitable steps to ensure that that number are selected. Normally only samples which have been subject to internal AQC are sent for audit. In this way both the primary and AQC processing can be checked.

In one case (Anglian), the AQC samples to be sent for external audit are selected at the beginning of the year. In another Region (Thames) the selection process is not undertaken until the end of the year, though most Regions supply samples to external auditors on several occasions throughout the year. In all other cases samples are selected after AQC has been completed. This normally entails assembling small sets of samples and then applying some process of random selection to select the appropriate proportion of samples for sending for external audit. North East Dales Area confirm, in their response, that new staff's' samples are not sent for audit until they have been adequately trained. This is in line with the recommendations of BT003.

Conclusions: Each Region undertakes fixed procedures to ensure that the contracted number of samples is selected at random for sending out for external audit. However, not all of these procedures meet the recommendations set out in BT003. In particular it is recommended that "audit samples should be distributed evenly so that changes in quality throughout the year are taken into account" (Environment Agency 1999b).

Recommendations: Selection of samples for external audit is being operated inconsistently between Regions and a greater level of conformity and closer adherence to the recommended procedures and guidance set out in BT003 are advocated.

5.5.3 Action taken when internal AQC and external audit targets are not met

AQC procedures allow the on-going level of performance in the processing of macro-invertebrate samples to be in three states (van Dijk 1994). "Accept" is the state when performance for a sample is better than or equal to the reference (target) value and no action is required. "Defer" is the state when performance for a sample is worse than the reference (target) value and performances must be closely monitored until the cumulative performance (Cusum Record) becomes acceptable again or until the cumulative performance has not sufficiently improved over a defined period (Decision Interval). "Alarm" is a state in which the average level of performance is deemed to be unsatisfactory. In this state corrective action must be taken to restore an acceptable level of performance.

Question 28 of the questionnaire was ambiguous in that it failed to distinguish between actions taken by each Region when the alarm state is reached and actions taken when a particular sample is poorly sorted in relation to the reference value, which was set at 2 for the 1995 GQA and subsequently. Nonetheless the replies received provide a good indication of the remedial actions taken in response to poor levels of performance.

It is inferred from the replies that the alarm state is rarely, if ever, reached in any Region and that the Defer state is relatively rare. However, most Regions have established procedures for reporting back to individuals on their level of performance and, if necessary, for improving the standard of individuals' work. Guidance on action to be taken is given by van Dijk (1994) and in BT003, where it is stated that the recommendations on this topic are "advisory rather than prescriptive" (Environment Agency 1996a). In general the response to the defer (or alarm) states involve a process of review of time and care taken over samples and the level of training received. This includes direct discussion between line managers and individual staff members on specific problems with individual samples or taxa, levels of workload, further training needs and, in some cases, personal problems which may impact on performance.

In two cases, North East Northumbria Area and SEPA (Solway), a more extreme course of action, of re-sorting a batch of samples, is taken after an individual sample fails to meet the standard required. In contrast, according to the reply received, only in the South West Region is no remedial action ever required.

Most Regions adopt similar remedial actions if the results of the external audit show that the Region's average, or single sample, level of performance fails to meet the required standard (assumed to be the reference value of an average of no more than two missed taxa per sample). In some cases, e.g. Thames and Welsh, replies indicate that the examination of the reasons for failure are more rigorous than for poor internal AQC results. Welsh Region's response involves a thorough review of procedures. Two Regions, North East and South West report that external audit results never indicate an unacceptable level of performance.

Conclusions: All Regions, except South West, report formalised procedures for remediating levels of performance which fail to reach the reference value of two missed taxa per sample. These procedures generally follow recommended national guidelines (Environment Agency 1996b), although there is an acceptable degree of variation between Regions that does not appear to need reducing by further prescriptive measures of standardisation.

Recommendations: Remediation of levels of performance which are shown, by internal AQC or external audit, to fall short of the reference value is conducted effectively within each Region and no further action is required to prescribe more standardised procedures. *(This recommendation necessarily excludes South West Region where no procedures are reported and are apparently never necessary)*.

5.5.4 Variation between laboratories in AQC and audit procedures

Conclusions: Few significant differences between laboratories were reported. The practice adopted by the Ridings Area of North East Region, of re-picking all new staffs' samples until a satisfactory level of performance is assured is recommended in BT003. It is also practised by IFE but the extent to which it has been adopted by other Environment Agency Regions is unclear from the responses to the questionnaire.

Recommendations: All Regional and Area laboratories should adhere to the three-phase process, set out in BT003, of training inexperienced sample processors and integrating them in the full AQC scheme

5.5.5 General comments on internal AQC and external audit procedures

Recommendations: The Environment Agency and IFE Project Leaders for the external audit contracts should consult Anglian Region on procedures for species level audits and audits of the assignment of abundance classes (see also section 5.4.2).

The Environment Agency and IFE Project Leaders for the external audit contracts should consider the request of North West Region that audit reports should not contain information on the performance of individual biologists.

5.6 Grading of Biological Condition

5.6.1 The grading system

The reaction to the grading system and the RIVPACS III techniques upon which it was based were variable. Some Regions (North West, Southern, South West and Welsh) or Areas (Ridings Area of North East Region) expressed more or less unqualified satisfaction. Anglian Region and Northumbria Area of North East Region pointed out certain parts of the system which they thought operated less well than others. Some Regions had more fundamental criticisms of the system (Thames, SEPA and DoE (NI)) or of the predictive accuracy of RIVPACS III in certain river types (Anglian and Midlands). Although happy with the grading system as a whole, Welsh Region shared concerns with Midlands Region and DoE Northern Ireland over whether combined season grading was the best procedure for reporting on the biological condition of sites for a single year.

A common concern between North East Region and SEPA was the separation of Grade a and Grade b sites. North East felt that sites of similar good quality are ascribed to either grades a or b on an apparently arbitrary basis, whilst SEPA's view was that requiring an EQI_{ASPT} of ≥ 1.00 was too demanding. SEPA also thought that in some instances, where very high EQIs occurred, this might indicate a slight loss of biological condition due to mild organic enrichment.

Others have found problems at the other end of the quality scale, although their perceptions have sometimes been contradictory. Thus, DoE (Northern Ireland) thought qualifications for grades e and f were too severe and that their worst sites never fell to these levels. On the other hand, Midlands thought that their urban headwaters did not exceed these levels on occasions where the Agency biologists felt the nature of their macro-invertebrate assemblages warranted a higher grade.

Two Regions, Midlands and Welsh, were concerned that using combined seasons' taxon lists to assign sites to quality classes was insensitive to changes in biological condition throughout the year. It was their view that the combined season system led to sites being assigned to the best biological conditions at the site in any one single season.

At a more fundamental level, Thames Region appeared to be requesting another review of the grading system after the 2000 GQA Survey. Also, both Anglian and Southern Region stressed the limitations in the applicability of the grading system for purposes other than national surveillance classification. Thus, for example, the Water Framework Directive envisages only five grades of "ecological quality" (Commission of the European Communities 1998) in which the highest grade of macro-invertebrate assemblage is defined as:

> *"Species composition, abundance and share of sensitive species in comparison to tolerant species correspond totally or* ***nearly totally*** [our emphasis] *to the type-specific conditions"*

The reported type of watercourse for which RIVPACS appeared to under-predict was slow-flowing, deep rivers of East Anglia with a pond-like fauna.

Conclusions: No requests were made to modify the current grading system for the 2000 GQA Survey but Thames Region appear to be reflecting a variety of concerns from individual Regions when they request a review of the grading system for surveys after 2000. This need for such a review is likely to be re-enforced by the requirements of the Water Framework Directive (Commission of the European Communities 1998).

The current authors share the concern of some Regions about the discrimination between Grade a and Grade b sites. The top 1990 grade was set as the 95 percentile (ASPT) and 90% (Number of Taxa) range for the RIVPACS II sites. This reflected the fact that, even amongst the RIVPACS reference sites, there was a variation in condition centred on mean EQI values of almost exactly one. However, setting the lower grade "a" EQI limit at 1.00 means that to be "very good" a site must exceed the mean value established for the best available reference sites in Britain. This appears to be at variance with the definition of the highest grade of "ecological quality" set out in the Water Framework Directive (Commission of the European Communities 1998), where the benthic macro-invertebrate assemblage may nearly totally meet the type-specific (i.e. reference) condition. This approach was not adopted for EQI_{Taxa} in the 1995 grading system, where paradoxically, it is recognised that the best condition sites have an EQI value centred on unity and not equal to or in excess of it.

The contradictory views of Midlands Region and DoE (NI) on the value of grades e and f may reflect each Region's perception of what represents relatively good and poor communities in their Regions. Thus, it may well be that few sites in Northern Ireland are in such poor condition as the Environment Agency sites in grades e and f. However, the concerns of Midlands Region over the grading of headwater sites are shared by the authors who feel that RIVPACS III+ still does not have an adequate representation of headwater sites, with their naturally lower species richness (Furse & Symes 1997).

The best means of representing the biological condition of sites over the calendar year of a GQA, using EQI values, has been a long-running issue. The options were addressed by Clarke *et al*. (1994) in a report to the National Rivers Authority prior to the 1995 GQA. This was based on the 1990, four-grade banding system. The principal conclusions are repeated on the following pages:

Advantages of three seasons' combined taxon lists

- The use of three seasons' data ensures that a high proportion of the taxa present at a site over a twelve month period are vulnerable to capture by the sampling programme

- Errors, inefficiencies and chance factors associated with collecting single samples are partially overcome by combining samples [see also Clarke *et al*. (1997)]

- Errors, inefficiencies and chance factors associated with sorting single samples are partially overcome by combining samples [see also Clarke *et al*. (1997)]

- A greater number of more precise quality grades can be set using combined seasons' faunal lists

Disadvantages of three seasons' combined taxon lists

- Combined seasons' faunal lists can simulate an acceptable quality level even though real quality is poor in one or more seasons

- Multiple season sampling is expensive in terms of both finances and staffing resources

- Samples may not be available for each of the three RIVPACS seasons

Advantages of two seasons' combined taxon lists

- Grade widths remain comparatively narrow despite the reduced sampling effort

Disadvantages of two seasons' combined taxon lists

- Combined seasons' faunal lists can simulate an acceptable quality level even though real quality is poor in one or more seasons

- Samples may not be available for two of the three RIVPACS seasons

Advantages of a single season taxon list

- The basis of the procedure is especially clear and meaningful

- A single season taxon list comprises an actual assemblage captured at a defined time and can therefore more readily be associated with prevailing quality conditions

- Samples may not be available for two of the three RIVPACS seasons

- Single season sampling is cheap and requires less staffing resources

Disadvantages of a single season taxon list

- The use of just one season's data means that only a sub-set of the taxa present at a site over a twelve month period are vulnerable to capture by the sampling programme

- Errors, inefficiencies and chance factors associated with collecting [and sorting] single samples are high [see also Clarke *et al*. (1997)]

- Only wide, imprecise quality grades can be set from single season's faunal lists if the same mis-grading rate as combined seasons lists is to be maintained

Advantages of the use of individual season's values from multi-season sampling

- The minimum grade is derived from a genuine assemblage captured at a defined time and can, therefore, more readily be associated with prevailing quality conditions at that time of the year

- The minimum value suggests the poorest quality attained by the site over an extended period

Disadvantages of the use of individual season's values from multi-season sampling

- Errors, inefficiencies and chance factors associated with collecting [and sorting] single samples lead to an increase in the width and decrease in the number of [meaningful grades of biological condition] in comparison with [combined season evaluations]

- Minimum EQI [values have] an intrinsic bias towards under-estimating quality

Advantages of the use of average EQI values from more than one season's sampling

- The use of separate individual season's taxon lists means that each component EQI contributing to the overall average is based on a real faunal assemblage at a specific time

- The application provides a better picture of the quality conditions pertaining at the site throughout the year

Disadvantages of the use of average EQI values from more than one season's sampling

- Use of average EQIs obscures the range and pattern of change in [the biological conditions] of [a] site [which are known because of the need to calculate individual season's values for the purposes of averaging]

- Use of single season EQIs to compute the site average incorporates all the problems of high error rate and sampling variability associated with this approach

General comments

- The best way to represent the condition of a site throughout the year, when using BMWP indices directly or indirectly, is to present each single season's grade or EQIs in chronological order

Where the need for simplicity of presentation dictates that a single annual grade per site is required then the balance of advantages and disadvantages appear to be tilted marginally in favour of using EQIs based on combined seasons' faunal lists rather than averaging EQI values from individual seasons. The reasons for this are:

- Combined seasons' EQIs allow narrower grade widths for the same rate of misclassification

- Combined seasons' faunal lists reduce the effects of variation and bias in sample collection and processing

- The conceptual basis of averaging values is dubious

- Interpretation of the presence, absence and abundance of taxa, and changes in these, has the potential for being much more informative than the use of EQI values

Readers are directed to the original publication (Clarke *et al*. 1994) for a more detailed presentation of the arguments and logic that underpin these conclusions

Anglian Region's view that RIVPACS III under-predicts BMWP index values in some deep-water sites is likely to be resolved by modifications to RIVPACS which are expected to result from the current R&D programme on deep water sampling

Recommendations: Although there is perceived to be scope to improve the grading system used in the 1995 survey, it is recommended that it is retained for the 2000 GQA Survey to maintain compatibility with the 1995 survey. Continual changes in the evaluation procedures can create the impression that the message of the surveys is being obfuscated by the shifting methodologies.

Notwithstanding the previous recommendation, following the 2000 GQA Survey, and prior to its use in future surveys, the effectiveness of the current grading system should be thoroughly reviewed, with particular reference to the Commission of the European Communities Water Framework Directive.

The results of the 2000 GQA Survey for England and Wales should be presented in a single national report supported by eight separate Regional Reports.

In addition to lists of taxa present and their abundance category, the following information on each site should be held in the national database for the survey:

- their overall grade of biological condition based on the lower of the individual EQI_{ASPT} and EQI_{Taxa} grades for combined seasons taxon lists

- the probability that the site's overall biological grade has changed grade as one of two alternate categories (1) no = <50% probability, (2) yes = $\exists$50% probability.

- separate single season grades of biological condition based on EQI_{ASPT} and EQI_{Taxa} independently.

- a statistical evaluation of the significance of the change in EQI values between the two sampling seasons and a summary of the significance of the change at $p>0.05$ level, expressed in three categories; (1) improved, (2) no change, (3) deteriorated.

The principal means of recording the biological condition of each site in the national report should be the overall site grade based on the minimum of the individual EQI_{ASPT} and EQI_{Taxa} grades for combined seasons taxon lists.

The national report should present information on both the proportion of sites in each overall grade of biological condition and also on the proportion of sites which have changed their overall grade between 1995 and 2000 with a greater than 50% probability, as determined using RIVPACS III+. Both sets of information should be included in the national report at both national and regional levels of interpretation.

The eight separate Regional Environment Agency Reports should include information and interpretation of each of the four types of output statistic listed above.

5.6.2 Grade (Band) descriptors

The supplementary questionnaire (Question S13) provided an opportunity for respondents to comment on the short name of each grade of biological condition and the supporting text that describes the type of macroinvertebrate assemblage associated with it.

Setting of names and descriptions is a contentious issue and one where it is known that achieving consensus was difficult (Bob Dines personal communication). Under these circumstances, the number of comments received was fewer than the authors of this report had expected.

Alternative grade names to those used in 1995 were:

Grade a: Very good	None suggested
Grade b: Good	None suggested
Grade c: Fairly good	Moderate
Grade d: Fair	Poor. A name indicating a condition which is worse than "fair"
Grade e: Poor	Very poor
Grade f: Bad	Awful. Grossly polluted

A thrice-stated objection to the text description of grade b was the inclusion of the term "organic pollution" but failure to mention any other potential sources of stress. Thames Region offered the following, full alternative text which excludes this term:

> *The biology falls a little short of that expected for an unpolluted river of this size, type and location. There may be a small loss of number of Families that are sensitive to pollution and a moderate increase in the number of individuals in the Families that tolerate pollution (like worms and midges). This may indicate the first signs of pollution.*

In connection with Grade c, Dales Area recommended the inclusion of the word obviously, to provide the following alternative definition:

> *The biology is obviously worse than that expected for an unpolluted river of this size, type and location. Many of the sensitive Families are absent or the number of individuals is reduced, and in many cases there is a marked rise in the number of individuals in the Families that tolerate pollution.*

In connection with Grade f, Midlands suggest the addition of the word macro-invertebrate, to provide the following alternative definition:

> *The biology is limited to a small number of very tolerant families, often only worms, midge larvae, leeches and the water hoglouse. These may be present in very high numbers. Even these may be missing if the pollution is toxic. In the very worst case there may be no macro-invertebrate life present in the river.*

Conclusion: Although a small number of suggestions were made, there seemed to be a general acceptance of the existing grade descriptions. The current authors feel that the alterations proposed to Grade b, by Thames Region and Grade f, by Midlands Region, improve these definitions. However, the abbreviated text names are a politically contentious issue that the Agency itself must decide upon.

Recommendations: The revisions to the text descriptors of Grade b and Grade f suggested by Thames and Midlands Regions respectively, should be accepted.

The Agency should consider the alternative abbreviated names suggested by the respondents.

5.7 Other Forms of Data Collection and Interpretation

5.7.1 Macro-invertebrates

Several other procedures were suggested for the interpretation of the macro-invertebrate data collected during the 2000 GQA Survey. The most frequently cited were use of the LIFE index (Lotic Invertebrate index for Flow Evaluation) (Extence *et al*. in press), the Artificial Intelligence systems being developed at the University of Staffordshire and detailed analyses of the raw data (Table 5.5)

Other approaches mentioned twice included the Community Conservation Index (CCI), indices of acidification (e.g. Rutt *et al.* 1990) and the use of abundance indices. The CCI has currently been developed for Anglian Region but has the potential for development as a series of alternative Regional models. The use of abundance data is being developed further by both the IFE (R&D Project E1-007) and the University of Staffordshire (E1-056).

There were also suggestions for the further development of the functionality of RIVPACS (Table 5.6).

Table 5.5 A list of alternative procedures, other than RIVPACS, for the interpretation of macro-invertebrate data collected during the 2000 GQA Survey.

System of evaluation	Type of site/stress for which the proposed system is of relevance	Number of Regions that suggested the procedure
LIFE (Lotic Invertebrate index for Flow Evaluation)	Sites which may be suffering from stress due to low flows/drought	5
Artificial Intelligence or other diagnostic procedures	Sites where it is wished to know the type of stress(es) causing an impact	4
Detailed analysis of the basic data including spatial and temporal taxon distribution studies.	General – to meet operational needs.	3
CCI (Community Conservation Index) or other conservation index	Sites requiring an assessment of their conservation value	2
Indices of acidification	Acidified sites	2
Abundance indices	Sites where it is wished to detect the early onset of a stress. Pesticide impacted sites.	2
Revised BMWP score systems	General	1

Table 5.6 Suggested improvements or developments of RIVPACS to aid interpretation of national surveys.

Suggested modification	Purpose
Introduction of new environmental variables	To simplify the collection of environmental data by extended use of map and GIS-derived variables and to improve the predictive accuracy of the system.
Development of a headwaters version of RIVPACS	To improve the evaluation of the biological condition of small, near-source streams.
Development of a canals and drains version of RIVPACS	To improve the evaluation of the biological condition of canals and drains.

Conclusions: The current practice of using the biological data collected during GQAs to provide a single index of the biological condition of sites provides continuity in meeting the basic national reporting needs of government. However, by concentrating on a single index value, it fails to optimise the information content of the data collected. This point is strongly made by Anglian Region that advocates annual sampling with identification to species level.

All Environment Agency Regions, plus DoE (NI), in their responses to the questionnaire, each proposed additional techniques for interpreting the data, including further improvements to the functionality of RIVPACS III.

Most popular of the suggestions are those which include a diagnostic element in their methodologies. In particular the biologists are keen to develop the use of the LIFE index (Extence *et al.* in press) for interpreting the impact of low flows, the Artificial Intelligence systems being developed by the University of Staffordshire, which have the potential for diagnosing a wide range of stresses, and index systems developed for the assessment of the extent of acidification of streams and rivers.

There is also a continuing requirement for improved interpretation of abundance data. Options for meeting this need include new versions of the BMWP score system which take account of both taxon abundances and river types, diagnostic Artificial Intelligence systems and the development of more refined abundance indices within RIVPACS III. Other suggestions for the improvement of RIVPACS are largely directed towards development of additional modules for specific watercourse types (headwaters, drains and canals). With the exception of a headwaters module for RIVPACS, all the suggestions listed in this paragraph are being addressed by national Environment Agency R&D projects placed at the University of Staffordshire and at IFE. The latter includes an investigation of the use of other environmental variables for predictive purposes.

The IFE are also undertaking a national R&D project, on behalf of the Environment Agency, for the development of an improved system for the biological detection of acidification. The acidification status of a river is an important parameter for use in the classification of its "ecological status (Commission of the European Communities 1998).

The request for a more fundamental examination of the data collected during quinquennial surveys, with particular reference to taxon distributions, has been met for the 1995 GQA by the reports to the Environment Agency by Clarke *et al*. (1999), Davy-Bowker *et al*. (2000), Walley & Martin (1997, 1998) and Walley *et al.* (1998).

Recommendations: The use of RIVPACS and EQIs alone to examine the biological data collected during GQAs fails to optimise the cost-effectiveness of the survey. The following recommendations, if adopted, will help this to be evaluated.

The results of the 2000 GQA Survey should also be the subject of trials of the two indices LIFE and CCI, the algorithm for detecting acidification being developed by IFE, the abundance-based indices being developed for RIVPACS and the diagnostic artificial intelligence procedures being developed by the University of Staffordshire. Use of CCI will require additional development work in order to establish appropriate regional models.

The types of analysis of the distribution of taxa and changes in that distribution undertaken by IFE and the University of Staffordshire should be repeated using the 2000 GQA Survey data.

Where this cannot be achieved internally by the Agency then it should form the basis for R&D research programmes.

Some or all of the applications of the CCI and LIFE indices will require species level identification. A target of 10% of the samples collected during the 2000 GQA Survey should be identified to species level for this purpose and these should represent a good geographical spread of samples and good coverage of all river sizes and types.

Species level identification may require that the primary sorting of samples from selected sites should involve a more detailed sorting process in which all, or a known proportion of taxa are removed from samples from the selected sites.

Where species level identification cannot be achieved in house then this should be contracted out to organisations with proven and reliable skills at species level identification.

An audit system for species level identification is necessary (see also Section 5.5.5).

Full identification of these samples may not be achievable during 2000 and may have to be deferred to a later year. Any deferment should not be a cause for delay in the publication of the primary survey report.

Existing species level data on nearly 2000 sites held by IFE may be useful to the pursuit of these recommendations.

5.7.2 Other taxonomic groups

Several taxonomic groups or evaluation procedures, other than those using macro-invertebrates, were suggested as appropriate for reporting on river quality. The most frequently cited taxonomic groups were diatoms (9) and macrophytes (7) and the most cited procedures, using these two groups respectively were Trophic Diatom Index (TDI) (Kelly 1998) and Mean Trophic Ranking (MTR) (Holmes *et al.* 1999) (Table 5.7).

Table 5.7 Suggested taxonomic groups which should be included in biomonitoring surveys with notes on their relevance to stress detection.

Taxonomic group	Number of Regions suggesting it	Relevant application
Diatoms	9	Detection of organic enrichment/ eutrophication using the Trophic Diatom Index
Macrophytes	7	Detection of eutrophication using the Mean Trophic Ranking index
Fish	3	Naturalness – given by one Region only
Phytoplankton	2	None given

Other systems of evaluation that were each mentioned once but without any application or relevance to a particular stress being suggested were: habitat evaluation, RHS, hydromorphology and geomorphology.

Conclusions: There was widespread interest in the use of diatoms and macrophytes in biomonitoring surveys but little collective enthusiasm for any other non-macro-invertebrate procedures. The aquatic flora, including macrophytes, phytobenthosv and phytoplankton are each important biological parameters for se in the classification of the "ecological status" of rivers (Commission of the European Communities 1998).

The relative information content, operational value and cost benefit of collecting macro-invertebrate, diatom and macrophyte data are unknown. A better understanding of the relative merits of these groups and the redundancy of information involved in a multimetric approach to biomonitoring, using all three groups (and possibly RHS) are equally poorly understood.

The 2000 GQA Survey could provide a cost effective way of obtaining some comparative data on all three approaches in order to assess the relative and collective benefits that sampling each could provide.

Recommendations: Single diatom samples should be collected during the spring macroinvertebrate sampling of 10% of the sites in the 2000 GQA Survey. MTR samples should be taken at the same sites in summer. These should also be the same sites for which species level macro-invertebrate identification is undertaken (see Section 5.7.1).

Staff collecting TDI and MTR data should have received full training in these techniques prior to sampling.

Full identification and/or interpretation of these samples is unlikely to be achievable during 2000 and would probably have to be deferred to a later year. Identification of diatom samples may need to be contracted out. Any deferment should not be a cause for delay in the publication of the primary survey report.

If RHS surveys are planned for 2000, then it would be beneficial if these could include the test sites used to compare different biomonitoring procedures.

An analysis should be undertaken of the relative merits of the variety of biomonitoring techniques available to the Agency.

5.8 Environmental Data

5.8.1 Time variant chemical data (alkalinity)

Alkalinity is one of the most important predictors in the suite of variables used in RIVPACS. Surrogate variables may be used but only after their conversion to alkalinity using imperfect regression equations.

All Regions had access to, and applied alkalinity values in assessing the biological condition of sites using RIVPACS. Only North East Region was occasionally constrained to use a surrogate variable. This Region was the only one to use data spread over more than one year, 1993-95, in order to obtain an adequate number of individual measurements. All other Regions exclusively used 1995 data, except Thames who used 1990 data.

All Environment Agency Regions plus DoE (Northern Ireland) were able to obtain more than nine individual alkalinity values for >75% of their sites. However, SEPA could only obtain this number of individual values for <25% of their sites.

All Regions reported that they had little difficulty in obtaining the appropriate data although concerns were raised about consistency of analytical methods and the representation of values below 10 mg l^{-1}.

Of greater concern to the application of RIVPACS for data analysis was the possibility that alkalinity determinations may not be made in Midlands Region in 2000. Personal communication between the authors and Agency staff suggests that this may be a more widespread problem than in Midlands alone.

Conclusion: The acquisition of alkalinity data for the 1995 survey was generally satisfactory but, for the purposes of future GQAs it is desirable that a long term run of data are averaged (see below for details). When this approach is taken for all time variant variables it will be possible to obtain long term fixed values for each site of their expected BMWP indices.

In an earlier report to the National Rivers Authority, Clarke *et al*. (1994), p75, recommended:

- *All appropriate environmental data collected between 1990 and 1995 should be used to obtain medium-term averages for predicting the 1995 fauna*

- *An NRA objective should be to obtain fixed long-term averages by the time of the 2000 River Quality Survey* [sic]

- *Data from very atypical years should be eliminated from the averaging process.*

- *Identification of these anomalies should be at the discretion of experienced local staff [but some guidelines were also provided in the report].*

- *The minimum number of individual estimates required is that which reduces the %SE of the estimated long-term average of that variable to a level where the estimated value of any resultant EQI differs from the "true" value by no more than 10% of the equivalent quality band width.* [It is probable that at least five years' data will be required for this purpose]

Recommendations: Alkalinity is an important RIVPACS predictor variable and the reliability of EQI evaluations are reduced by its unavailability. Its regular collection should be retained for all GQA sites until long-term average values can be substituted.

5.8.2 Time invariant physical variables.

Six Regions newly measured the values of the time invariant variables for the 1995 Survey. Thames Region and SEPA used the values derived in 1990, whilst Midlands and South West were unable to reply. In the case of the former South Wessex NRA Area it is known that 96% of sites were new to the 1995 survey and their variable values were most probably newly measured, unless they had been calculated earlier for non-GQA purposes.

No Regions used averaged data from 1990 and 1995. However, with the exception of South West and SEPA all remaining Regions used two independent assessors to measure the values of NGR, altitude, slope and distance from source. This was in line with the recommendation of Clarke *et al*. (1994). Four Regions also claimed to double record discharge categories.

In the view of the respondents, the easiest of the time variant variables to measure was the National Grid Reference (NGR) and the most difficult was slope (Table 5.8).

Table 5.8 The degree of difficulty of obtaining time invariant data from maps.
L = Low M = Moderate H = High

Environmental variable	Region										
	Anglian	North East	North West	Midlands	Southern	South West	Thames	Welsh	SEPA (Clyde)	SEPA (Solway)	DoE (NI)
NGR	L	L	L	L	L	L	L	L	L	L	L
Altitude	L	M	H	L	L	L	M	M	L	H	L
Slope	M	H	H	H	M	H	LM	M	M	H	M?
Distance from source	H	M	M	L	L	H	L	L	L	M	M
Discharge	L	H	L	L	L	H	L	M	L	L	L

National Grid Reference: The only listed difficulty was in the correct attribution of grid letters near boundaries and several examples of this were noted when correcting the 1990 RQS and 1995 GQA data-sets for further analysis (Clarke *et al*. 1999; Davy-Bowker *et al*. 2000). In a separate analysis of a similar data-set Walley & Martin (1998) recorded 14 out-of-Region grid reference errors. Thames questioned whether more detailed map scales would increase accuracy and Southern recommended that the use of a "Roamer", or similar system, should be mandatory for deriving NGRs. The use of GPSs represents an alternative approach.

Altitude: The main difficulties were experienced in urban areas, where contour lines were often not shown, in lowlands with widely-spaced contours and in extreme uplands with very closely spaced contours. Not surprisingly, the difficulties encountered were closely related to the predominant type of terrain in the Region. Logical checks of altitude against slope revealed many occasions where rivers apparently flowed uphill. Walley & Martin (1998) used autoassociative neural network techniques for similar purposes.

Slope: This variable presented the greatest difficulty in measurement. The chief difficulties were in identifying and tracing the relevant contour lines, particularly where they crossed the river. Problems were most extreme in uplands and urban areas but Midlands Region also reported difficulties in areas with very low gradients.

Distance from source: Many Regions reported moderate to high difficulties in the measurement of this variable. The principal difficulties were in identifying the relevant source, lack of representation of river courses and sources in urban areas and the inaccuracies involved in the use of a map wheel to obtain the required values. SEPA used a digitiser to obtain distances.

Discharge: Many Regions reported that acquisition of values of this variable was easy because it was supplied by their Hydrology Sections, although Welsh Region viewed this as a black-box process. Where discharge categories were obtained from the river quality maps produced with the report of the 1985 RQS (Department of the Environment and the Welsh Office 1986), e.g. in some Areas of North East Region, difficulties were experienced in reading the correct discharge category from the width of the watercourse. This was despite the fact that a cut-out key was supplied with each map. South West reported the greatest difficulty in obtaining reliable values, although it was not clear whether this was due to their unavailability from their Hydrology Section or because the river quality maps were unavailable, or both.

Conclusions: The derivation of the time invariant data is a time-consuming and difficult process. Nevertheless, it is important that the high levels of accuracy and logical consistency are attained in order to set accurate biological targets for each site.

The double measuring of most time invariant values should have eliminated most errors in the time invariant environmental data-sets from 1990 and 1995. However, when IFE checked the values of time variant data used for the 1995 survey, for the purposes of other reports (Clarke *et al*. 1999; Davy-Bowker *et al*. 2000), a high proportion of obvious errors were detected (e.g. rivers flowing uphill or getting nearer to source as they moved downstream, etc).

Manual measurements of cartographic variables will again be necessary for the 2000 GQA Survy and will provide an opportunity to eliminate persistent errors from previous surveys. However, developments in GIS within the Agency, particularly at their National Data Centre at Twerton, will eventually result in the electronic generation of accurate information on all cartographic variables for each site. This may include any additional variables which may be held cartographically, such as geology and soil type, and which may be shown to be useful through the RIVPACS development R&D programme (E1-007) currently being undertaken by IFE.

Recommendations: All time invariant values used for RIVPACS predictions in the 2000 GQA Survey should be re-measured independently by two people. Any pair of new values whose standard error of their mean lies outside the tolerable range given in BT001 or Clarke et al. 1994 should be re-measured by both individuals until an acceptable level of agreement is reached. Values should then be cross-checked against 1995 values and all disparities >10% should be investigated and corrected.

Logical checks of environmental data for individual rivers should be made to ensure that rivers flow downhill and the discharge of any site should be no less than the discharge of the next site upstream of it. The values used in connection with the 2000 GQA Survey should be the average of the corrected values measured for the 1995 and 2000 GQA Surveys. For new sites, the values obtained in 2000, after double measurement and subsequent validation and correction, should be used.

Southern Regions recommendation, that the use of a "Roamer", or similar system, should be mandatory for deriving NGRs, is endorsed here.

The Environment Agency should continue to develop GIS procedures for the accurate derivation of values of cartographic variables.

5.8.3 Time variant physical variables.

Three time variant environmental variables were measured in the field during the 1995 RQS. Two, width and depth were generally regarded as easy to measure but greater difficulties were associated with the derivation of substratum cover data. (Table 5.9).

Table 5.9 The degree of difficulty of obtaining time variant data in the field.
L = Low M = Moderate H = High

Environmental variable	Region										
	Anglian	North East	North West	Midlands	Southern	South West	Thames	Welsh	SEPA (Clyde)	SEPA (Solway)	DoE (NI)
Width	L	M	L	L	L	L	L	M	L	L	L
Depth	L	M	L	L	L	L	L	M	L	L	L
Substratum	H	M	H	M	M	H	LM	M	M	L	M

Width: The principal listed difficulty was measurement of wide rivers. Some Regions used rangefinders but not all were happy with the accuracy of this approach. One Region, Welsh, favoured taking the average of the width at the top, middle and bottom of the sampling area.

Depth: The only real difficulty expressed was that of depth estimation in deep, unwadeable rivers.

Substratum: Only one Area, SEPA (Solway) considered that collecting this form of data presented low difficulty. Most other Regions drew attention to the subjective nature of the data collection process. Turbid water also presented a problem.

Five Regions; Southern, Thames, Welsh, SEPA and DoE (NI) attempt to reduce the errors in the estimation of substratum cover by using two independent recorders who reach a consensus/average for each sampling visit. This is in line with recommendations made by Clarke *et al*. (1994) who advocate either double recording for each sampling visit or the use of different people to obtain the estimate in the three different recording seasons.

General: Apart from substratum, only Southern Region and the Solway Region of SEPA used multi-person measurements of width and depth to arrive at their final estimates.

A concern for two Regions was the resource implications of collecting three separate seasons' environmental data when only two of these seasons were used to collect macro-invertebrate samples.

One Region extolled the value of the field sampling training days provided by the IFE and the NRA prior to the 1995 survey and a general feeling that these were valuable has since been reported (Bob Dines personal communication).

Conclusions: Only substratum cover caused frequent difficulties in its estimation. The procedure is not time consuming but is subjective in its nature. This subjectivity can be reduced, but not entirely eliminated, by using several people to independently estimate cover in each category.

The difficulties of obtaining reliable cover data and the resource problems associated with three seasons site visits will be eliminated, or greatly reduced, once long-term average values are available for each site. The benefit of long-term values is endorsed by Thames Region in their reply.

Recommendations: All sites in the 2000GQA Survey should be visited in spring, summer and autumn to collect time variant environmental data. All time variant variables recorded during site visits should be measured independently by two people on each visit and an average or consensus value recorded. Where this is not possible, for example because of single-manning, then different staff members should be deployed to obtain values in different field visits.

Where adequate data are available (see Section 5.8.1) then long-term average data should be used in lieu of new field recording of time variant site data.

Field sampling training days, similar to those provided by the IFE and the NRA prior to the 1995 survey, should be run again prior to the 2000 GQA Survey. It is unlikely that any standard protocols for deep water sampling will be agreed prior to spring 2000 and deep water sampling and prior training should be delayed until these issues are resolved.

5.8.4 Other variables

There were no consistent requests for additional variables and one firm request that no additional variables were to be measured because of the resource implications of such a strategy. Only macrophyte information (Anglian, Southern and Thames) and flow characteristics (Southern and Thames) were listed by more than a single Region.

However, several other factors were suggested by a single Region only. These were land use, geomorphology, extent of channel modification, geology and habitat types. Many of these variables are components of RHS and the application of this form of survey was advocated by Thames Region.

Conclusions: The current National R&D project (E1-007) being conducted at IFE includes an investigation of the possibility of using GIS derived variables, including several listed above.

Recommendations: No further variables should be measured during the 2000 GQA Survey, except where these are required for any of the alternative bioassessment techniques discussed in Sections 5.7.1 and 5.7.2

The Environment Agency should be guided by the results of National R&D project E1-007 as to the accessibility and predictive power of additional, GIS-derived environmental variables other than those currently used in RIVPACS III.

5.9 Equipment

Recipients of the questionnaire were given the opportunity to comment on any facet of the equipment recommended in manual BT001 (Environment Agency 1977) for use in the 1995 GQA.

Almost all of the comments received were concerned with deep water sampling and in particular the weight (5kg) of the Medium Naturalist Dredge recommended for use in BT001. Three Regions (North East, Midlands and Southern) considered this to be too heavy and Midlands suggested the use of a 2kg Duncan Naturalist's Dredge. Similar concerns were raised in response to an earlier questionnaire circulated to Environment Agency biologists by IFE (Wright *et al*. 1999).

Conclusion: The results of the earlier questionnaire showed that at least four different dredge weights were being used for deep water sampling (Wright *et al*. 1999). Weights ranged from 2kg to 7kg and included the 5kg dredge recommended in BT001. The latter was most commonly used of the various devices.

The IFE is currently undertaking an Agency-funded comparative trial of different deep-water sampling procedures (National R&D Project E1-007). The sampling devices being compared include the 5kg Medium Naturalist's Dredge recommended in BT001 but do not include any other weight or design of dredge.

Recommendations: If the results of R&D Project E1-007 indicate that the use of a dredge should be included in a standardised deep-water sampling protocol then the Agency should consider further trials on the most appropriate specifications of the dredge to be used. These trials should cover a wide range of substratum types.

5.10 Reporting

5.10.1 Current usage

The replies indicate that, in addition to national reporting, extensive further use was made of the biological data collected during the 1990 RQS and 1995 GQA. Prominent amongst these were Regional Biological Quality Reports (Anglian, North East, Midlands, Southern, Thames, SEPA and DoE (NI), some of which were stated to include quality maps (Anglian, Thames). Other Regions produce detailed catchment reports instead of (Welsh) or as well as (Southern) Regional reports whilst North West, Midlands and Thames stated that they use the data for producing LEAPs (Local Environment Area Plans).

In many of these cases (most of North East, Midlands, Southern, Thames, Welsh, SEPA) the 1995 GQA data is compared with 1990 RQS results and often with other years too. Midlands and Anglian Regions also explicitly state that comparisons are also made with chemical data from the quinquennial surveys and, in Anglian, use is made of mismatches to aid interpretation of water quality.

Whilst the preceding paragraphs provide a summary of the replies received it is possible that other Regions produce output and comparisons of the types listed but have not given full details in their answers. For example, it is now known that Southern Region also produced a Regional Biological Quality Report, based on the 1995 survey, in which interpretative comparisons were made with the results of the chemical survey (Bob Dines personal communication).

Thames Region lists several of the many other uses to which the data are put, which include: pollution reports, conservation planning proposals, abstraction licensing, discharge consents, public enquiries, educational projects and Regional State of the Environment Reports. To this list, Southern and Anglian add a variety of operational uses and North East draw attention to their use in answering enquiries from external sources to the Agency.

Conclusion: In addition to national reporting, the RQS and GQA data are a valuable source of information, which already has a wide portfolio of existing uses.

Recommendation: The widespread and varied application of GQA data for regional purposes is to be encouraged and where possible extended (see Sections 5.7.1 and 5.7.2)

5.10.2 Potential future use

North East Region and DoE (NI) both called for published distribution maps and an assessment of the environmental requirements of taxa. In England and Wales these requirements have been addressed by Walley & Martin (1997) and by Davy-Bowker *et al*. (2000) but this has yet to be attempted for Scotland and Ireland. These analyses have been at BMWP level but Anglian Region have produced two indices, LIFE and CCI which utilise information on the distribution of individual species. LIFE also allows analyses at family level but the differing flow requirements of the component species of several families means that the use of family level data introduces a higher level of uncertainty than using species level. Anglian feel strongly that identification at species level is the way forward if the resources spent on bioassessments and the information content of the results are to be fully capitalised on.

Thames Region regards the data as appropriate to the processes of defining the term “Good Ecological Status” (as defined in the European Framework Directive (Commission of the European Union 1998)) and the setting of Biological Quality Objectives.

More fundamentally, Southern Region emphasise the need for a proper, public domain report setting out the basic results of the 2000 GQA Survey. In comparison with previous surveys the report of the 1995 survey (Environment Agency 1997) was extremely brief and lacking in detail.

Conclusion: The request for published distribution maps and environmental ranges of BMWP families, based on previous RQS and GQA data, has been met in R&D reports by Walley & Martin (1997), Walley *et al.* (1998) and Davy-Bowker *et al*. (2000). However, the comparison of change data, between 1990 and 1995, was partially compromised by apparently poor identification and retrieval of some families from samples in the former survey. Improved levels of identification in 1995, which should be maintained in 2000, will mean that the comparisons of family distributions between 1995 and 2000 will be even more reliable than the 1990 to 1995 comparison.

The lack of any consistent species level identification of samples collected during national surveys limits any opportunity to apply CCI to past data and restricts the application of LIFE to family level information only.

Recommendations: The family distribution and comparison of change studies conducted on the 1990 RQS and 1995 GQA should be continued in order to incorporate the 2000 GQA Survey data.

Trial species level identification of a sub-set of the 2000 GQA Survey data should be undertaken and the new indices, LIFE and CCI should be applied and evaluated in accordance with the recommendations in Sections 5.4.1 and 5.7.1.

The previous two recommendation should be implemented as Phase 3 of the current R&D Research project.

The results of the 2000 GQA Survey should be prominently published in the public domain.

5.11 Survey Design

5.11.1 General comments

Few replies were received to Question 54, which offered an opportunity for respondents to make suggestions for the improvement in the design and implementation of future surveys. However, North West and Southern Regions shared a common concern about the adequacy of only undertaking GQAs every five years. Amongst the disadvantages of this system are that the frequency of reporting of chemical surveys is annual, whereas biological reporting is quinquennial. North West Region also queried the implications of the quinquennial survey coinciding with a year of atypical [climatic] conditions.

Anglian Region re-iterated their concerns about the relevance of surveys that are entirely driven by water quality issues and would like to see national surveys having greater biological relevance.

Concern was again expressed about the adequacy of coverage of canals, ditches and standing water bodies.

Conclusions: The authors of this report share North West and Southern Regions' concerns that quinquennial surveys are not the best approach to monitor changes in the biological condition of rivers. Elsewhere in this report (see Sections 5.1 and 5.2.2) we draw attention to the advantages of staged (rolling programme) surveying as an alternative to all-inclusive surveys every five years. In particular we have drawn attention to the need to stage sampling of headwaters (Furse 1995), large deep rivers (Sections 5.1 and 5.2.2) and canals (Section 5.1).

One possible approach to a rolling sampling programme, subject to resource availability, would involve an annual sampling regime divided into two component parts. One part would be headwaters, large, deep rivers, canals and ditches (and standing freshwater bodies?) and the other would be all other rivers and streams. Each component would be sampled alternately, with a two year cycle, meaning that all sites would be sampled five times over every ten year period and reporting would be annual with major review of trends of temporal changes at the end of each decade. Alternatively half of each of the two groups of sites could be sampled each year with an eight year major reporting cycle.

The authors also believe that, whilst the element of reporting changes in biological condition should continue (see Section 5.7.1), increasingly sophisticated and diverse techniques and taxonomic groups should be incorporated into future surveys in order to provide a more holistic and informative review of the state of the freshwater environment.

Recommendations: Quinquennial GQAs should be replaced by annual surveys in the form of a rolling programme (see also Section 5.1).

The Environment Agency should consider how this might best be achieved within the context of the resources available and the other elements of the work programme of biology sections. One option for consideration should be the biennial sampling of each site, as described above.

The current practice of using the biological data, collected during GQAs, in order to provide a single index of the biological condition of sites offers continuity in meeting the basic national reporting needs of government. It should be continued as part of a wider interpretative programme (see Section 5.7.1).

The use of RIVPACS and EQIs alone to examine the biological data collected during GQAs may fail to optimise the cost-effectiveness of the survey. The following recommendations, if adopted, will help this to be evaluated.

The results of the 2000 GQA Survey should also be the subject of trials of the two indices LIFE and CCI, the algorithm for detecting acidification being developed by IFE, the abundance-based indices being developed for RIVPACS and the diagnostic artificial intelligence procedures being developed by the University of Staffordshire (see Section 5.7.1).
Diatom and MTR sampling should be introduced into future surveys on a trial basis, in order to assess the relative merits of the variety of biomonitoring techniques available to the Agency (see Section 5.7.2)

If RHS surveys are planned for 2000 then it would be beneficial if these could include the test sites used to compare different biomonitoring procedures (see Section 5.7.2).

5.11.2 Number of samples per site per year

Number of samples: Most Regions felt that two samples per season represented the optimal balance between the resources available and the reliability of the data obtained for evaluating the biological condition of each site. Generally, the Regions felt that, in comparison with three seasons, two seasons' sampling provided no apparent difference in the reliability of the data obtained (4/5 Regions) or produced only slightly poorer reliability (2/3). However, there was very little support for a reduction in the number of samples to one per site. Nor was there support for replicate sampling because of the resource implications involved in such an approach.

However, there were significant exceptions to these generalisations. Three Regions, plus a single biologist in a fourth Region regarded three as the optimal number of samples to be taken per site per year. DoE (Northern Ireland) felt particularly strongly about this. They considered taking two samples only led to much poorer reliability and saw three samples as the minimum acceptable number per site per year.

Conclusions: In an earlier report to the Environment Agency concerning numbers of samples per site per year, Clarke *et al*. (1994) conclude that, in order to obtain an equivalent rate of correct classification to biological grade:

> *"paired season EQI bands [need] only [be] slightly wider than three seasons combined and [paired-season sampling] offer[s] a viable alternative to [three seasons] where financial and staffing resources are limited."*

The results of the questionnaire show that most Regions readily accept this view point. A further factor supporting this approach is the development, in RIVPACS III+ of procedures which allow statistical comparisons of temporal (and spatial) change based on differing numbers of samples being taken in the two component sets of samples being compared (Clarke *et al.* 1997).

Where individual Regions wish to collect more than two separate seasons' data, and this may be an attractive option if three site visits are required in order to collect environmental data, then they may do so. However, this should not be allowed to impact upon the density of site coverage that they achieve as their contribution to the national survey. DoE (NI) would clearly find this course of action attractive.

Recommendations: The 2000 GQA Survey should be based on two seasons' macro-invertebrate sampling per site.

Individual Regions may choose to sample more frequently, as long as this does not impact on the density of coverage of sites incorporated in the national survey.

DoE (NI) and SEPA report separately and may elect to sample three times a year for their own national requirements without jeopardising the opportunity for unified reporting across the United Kingdom.

Choice of sampling season: Spring and Autumn were the most favoured sampling seasons in replies received to the questionnaire (Table 5.10). Summer was much less favoured.

Table 5.10 Number of times each season wass recommended for sampling in replies received to Questions S2 and S4 of the supplementary questionnaire

Season	Number of recommendations
Spring	12
Summer	4
Autumn	11
Summer or autumn	3
Spring, summer or autumn	1

Conclusion: The selection of spring and autumn as the most favoured months for sampling is in line with the 1995 GQA. As such, it provides much more reliable national estimates of the rates of gain or loss, or changes in abundance, of individual taxa (see Davy-Bowker *et al*. 2000). In a statistical analysis, Clarke *et al*. (1994) also found that paired-season grading best replicated three seasons' when they were taken in spring and autumn rather than any other pair of seasons.

Recommendation: GQA sampling should be undertaken in spring and autumn in order to provide a standard basis for inter-survey comparisons of distributional changes.

5.11.3 Canals

The development of appropriate techniques for sampling canals and for assessing their biological condition is being undertaken by Pond Action (Williams *et al.* 1998). Hence, canals have not been generally been a subject of the questionnaire upon which the current report is based.

However, the provisional procedures proposed by Williams *et al.* (1998) have been considered by the IFE, as part of R&D Project E1-007. On this subject Wright et al (1999) state:

> *Pond Action has made a useful start in the development of a classification-prediction system for canals ... but ... there is still much work to be done before an operational methodology for assessing the biological quality of canals is in place.*
>
> *It is now apparent that a fully-fledged system cannot be in place prior to the GQA Survey in 2000. However, the field protocols developed by Pond Action could be formalised and used as the basis of a sampling programme to be undertaken on canals during the GQA Survey in 2000. Such a survey would include a wide geographical range of sites encompassing both high quality and impacted sites. If there is a need to supplement the existing reference dataset in terms of additional sites or seasons, then selected samples collected during the GQA Survey could be passed on to a contractor for processing at species level. Once the full dataset was assembled, the classification and prediction exercises could commence, and on completion of an operational system, it should still be possible to make an appraisal of the full range of canal sites sampled during the GQA Survey in 2000.*

5.12 Summary of Recommendations

The recommendations made in the preceding sections are summarised here in two categories, principal recommendations are in bold and subsidiary recommendations in regular font.

Number of samples

There should be no reduction in the coverage of sites in the 2000 GQA Survey, in comparison with 1995, unless there are resource limitations that cannot be overcome.

Where feasible, Regions should adjust their coverage of particular site types upward to rectify deficiencies they identified in the 1995 GQA. This appears to apply to headwaters in particular.

Where resources are inadequate to meet the coverage recommended by Regional Biologists, consideration should be given to a rolling programme of GQA monitoring, spread over the full five years currently separating each GQA. Recommendations of this type were previously made by Furse (1995), in the context of biomonitoring of headwaters.

Although not specifically mentioned by respondents, staged sampling of deep water sites and canals would allow time for effective monitoring and evaluation techniques to be developed and tested.

Sampling Methods

Pond-net sampling

The pond-net sampling procedures adopted in 1995 should be retained for the 2000 GQA Survey. This will provide a reliable basis for the application of the procedures for detecting temporal change incorporated in RIVPACS III+.

Where differences in the implementation of the recommended procedures are known to the Environment Agency, and where these may possibly have a significant effect on the results obtained, then the recommended procedures in the revised BT001 should be made more prescriptive in order to eliminate these differences.

Special instructions should be included in the revised version of BT001 (Environment Agency 1997a), giving the specialised sampling techniques that may need to be adopted in order to collect three minute pond-net samples in headwater streams.

Headwater sites incorporated in RIVPACS III+ were each sampled for three minutes, plus one minutes search. Despite the difficulties involved, this duration of sampling should be retained for the 2000 GQA Survey, in order that Ecological Quality Index (EQI) values obtained for these sites are based on a common level of sampling contributing to both observed and expected BMWP Index values. This will also best facilitate reliable temporal comparisons.

When appropriate, standardised deep-water sampling procedures are available, sampling with the standard FBA-style pond-net, with handle length of approximately 1m, should be confined to sites that are wadeable for at least 25% of their total width.

Deep-water sampling

The Environment Agency should introduce standard procedures for the sampling of deep water sites, i.e. those that are not suitable for standard pond-net sampling.

The Agency should be advised by the findings of the National R&D Project E1-007 in selecting the most appropriate, standard, deep-water sampling procedures for use in national GQAs.

All biologists involved in field sampling of deep-water sites should receive appropriate training prior to undertaking sampling.

The Environment Agency should consider staging sampling for the 2000 GQA Survey, with deep-water sampling taking place in 2001. This will allow more detailed consideration of the results of the National R&D Project E1-007 and time for adequate training and practice in the application of standard deep-water sampling techniques.

Sample Sorting

Location of sample processing

Whilst it remains unclear whether bankside sample processing is as efficient and comprehensive as laboratory processing, this issue is considered to be so important that a standardised laboratory-based approach should continue to be prescribed for use in the 2000 GQA Survey. This action will ensure that observed BMWP index values and those predicted by RIVPACS III+ are based on the same sorting procedure.

Transport, fixation and preservation of samples

It is recommended that, wherever local Health and Safety protocols allow, samples should be fixed in Formalin for transportation to the laboratory. In all usages of Formalin for sample fixation and storage neutral-buffered solutions are preferred. Where this is not permitted by local Health and Safety protocols, then samples should be transported live to the laboratory, on the day of collection, for either live sorting or for fixing in formalin.

During live transport, samples should be drained of as much water as possible and carried in a cool box containing ice packs. These measures will reduce predation, and de-oxygenation.

Live samples should always be stored in a refrigerator.

Live samples should always be fully sorted within two working days of their day of collection. If this cannot be achieved then the samples should be fixed in Formalin. Fixed sample may subsequently be preserved in alcohol.

Sorted samples, once re-constituted, may be either fixed in formalin or preserved in alcohol. The former is preferred. It must be ensured that the quantities of Formalin and, **especially** alcohol added to re-constituted samples should be of sufficient strength to exclude the possibility of the sample decaying, especially if they contain large quantities of organic material.

All samples which are re-constituted for internal AQC or external audit should be fixed in formalin or preserved in alcohol, according to the Health and Safety procedures operating in the respective Area, immediately after sorting is completed. Removed animals should always be preserved in alcohol.

Alcohol is a preservative and not a fixative. Use of alcohol without prior fixation may lead to soft-bodied animals breaking up and becoming un-identifiable. Re-constituted samples which have not been fixed previously with formalin, and which are subject to AQC or audit, should be re-analysed within two weeks. All samples for external audit should be dispatched to the auditors as soon as possible and within two weeks of the date on which the last sample was analysed for AQC (Environment Agency 1996a).

Sorting time

As a consequence of the wide variation in sampling times, the Environment Agency should promote the exchange of ideas on the best methods of meeting the AQC standard, for the mean number of missed taxa per sample, in the most cost-effective manner.

Identification and Quantification

Identification

Each Environment Agency Region should have a policy in place to train biologists to the level of competence in macro-invertebrate identification to meet the needs of the 2000 GQA Survey. Wherever possible, such training should be provided in house but, where necessary, specialist external training course should be arranged.

The need to achieve more precise levels of identification, including during national surveys, will require some biologists within each Area laboratory to be competent at species level identification and appropriate training should be planned, where necessary.

Quantification

It is essential that a system of allocation of abundance classes to each BMWP family is adopted by all Regions for the 2000 GQA Survey.

The system adopted by each Region should be standardised, or be capable of standardisation, between Regions.

The number of categories should be sufficient to distinguish significant differences in abundance but not so many as to give a false impression of the accuracy of the quantification or to present unacceptable difficulties in allocating taxa to categories.

The coding of abundance categories should reflect and be easily convertible to the five categories recommended for the 1995 GQA. This is best achieved by retaining the original category numbers as prefixes to the more detailed new categories see Table 5.4, section 5.4.2).

Whatever system of categorisation is adopted by each Region for the 2000 GQA Survey the data must be presented to the National Database in the standard categories adopted for the survey.

Where present, and where more precise identification is not subsequently required, then a minimum of ten representatives of each BMWP family should be **counted** for each sample. This will facilitate the correct allocation of taxa to any sub-division of abundance category 1 (*sensu* the 1995 GQA) and also the correct allocation of taxa on the borderline of categories 1 and 2.

Numbers of individual taxa in excess of ten, and allocation of these taxa to abundance categories should be estimated.

Where the abundance category of a taxon with more than nine individuals present is in doubt then the taxon should always be assigned to the lower abundance category of the two possible categories in question.

AQC analysts should also assign each taxon to an abundance category when re-processing samples (see also 5.5.5). They should use the same procedures for assigning taxa to categories as that adopted by the primary analyst.

Frequent disparities between abundance categories assigned by the primary sorter and AQC analyst, should be investigated, particularly if these differences are skewed in a constant direction. Total counts may be necessary to resolve differences.

If abundance-based indices are to be used to report on the results of the 2000 GQA Survey, then abundance checks should be incorporated in quality control procedures.

Recommendations on the quantification of individual species would be more complex than those given above for BMWP families and are outside the scope of Question 24 of the questionnaire.

Internal AQC and External Audit

Value of the audit

The use of internal AQC should be continued for the 2000 GQA Survey.

Method of selection of samples for internal AQC and external audit

Selection of samples for external audit is being operated inconsistently between Regions and a greater level of conformity and closer adherence to the procedures set out in BT003 are recommended.

Action taken when internal AQC and external audit targets are not met

Remediation of levels of performance which are shown, by internal AQC or external audit, to fall short of the reference value external audit is conducted effectively within each Region and no further action is required to prescribe more standardised procedures.

Variation between laboratories in AQC and audit procedures

All Regional and Area laboratories should adhere to the three-phase process, set out in BT003, of training inexperienced sample processors and integrating them in the full AQC scheme.

General comments on internal AQC and external audit procedures

The Environment Agency and IFE Project Leaders for the external audit contracts should consult Anglian Region on procedures for species level audits and audits of the assignment of abundance classes.

The Environment Agency and IFE Project Leaders for the external audit contracts should consider the request of North West Region that audit reports should not contain information on the performance of individual biologists.

Grading of Biological Condition

The grading system

Although there is perceived to be scope to improve the grading system used in the 1995 survey, it is recommended that it is retained for the 2000 GQA Survey to maintain compatibility with the 1995 survey. Continual changes in the evaluation procedures can create the impression that the message of the surveys is being obfuscated by the shifting methodologies.

Notwithstanding the previous recommendation, following the 2000 GQA Survey, and prior to its use in future surveys, the effectiveness of the current grading system should be thoroughly reviewed, with particular reference to the Commission of the European Communities Water Framework Directive.

The results of the 2000 GQA Survey for England and Wales should be presented in a single national report supported by eight separate Regional Reports.

In addition to lists of taxa present and their abundance category, the following information on each site should be held in the national database for the survey:

- their overall grade of biological condition based on the lower of the individual EQI_{ASPT} and EQI_{Taxa} grades for combined seasons taxon lists
- the probability that the site's overall biological grade has changed grade, as one of two alternative categories (1) no = <50% probability, (2) yes = $\exists$50% probability)
- separate single season grades of biological condition based on EQI_{ASPT} and EQI_{Taxa} independently, for each of EQI_{ASPT} and EQI_{Taxa}
- a statistical evaluation of the significance of the change in EQI values between the two sampling season and a summary of the significance of the change at $p>0.05$ level, expressed in three categories; (1) improved, (2) no change, (3) deteriorated.

The principal means of recording the biological condition of each site in the national report should be the overall site grade based on the minimum of the individual EQI_{ASPT} and EQI_{Taxa} grades for combined seasons taxon lists.

The national report should present information on both the proportion of sites in each overall grade of biological condition and also on the proportion of sites which have changed their overall grade between 1995 and 2000 with a greater than 50% probability, as determined using RIVPACS III+. Both sets of information should be included in the national report at both national and regional levels of interpretation.

The eight separate Regional Environment Agency Reports should include information and interpretation of each of the four types of output statistic listed above.

Band (Grade) descriptors

The revisions to the text descriptors of Grade b and Grade f suggested by Thames and Midlands Regions respectively (pp 99-00), should be accepted.

The Agency should consider the alternative abbreviated names suggested by the respondents.

Other Forms of Data Collection and Interpretation

Macro-invertebrates

The use of RIVPACS and EQIs alone to examine the biological data collected during GQAs fails to optimise the cost-effectiveness of the survey. The following recommendations, if adopted, will help this to be evaluated.

The results of the 2000 GQA Survey should also be examined using the two indices LIFE and CCI, the algorithm for detecting acidification being developed by IFE and the diagnostic artificial intelligence procedures being developed by the University of Staffordshire.

The types of analyses of the distribution of taxa and changes in that distribution undertaken by IFE and the University of Staffordshire should be repeated using the 2000 GQA Survey data.

Where this cannot be achieved internally by the Agency then it should form the basis for R&D research programmes.

Some or all of the applications of the CCI and LIFE indices will require species level identification. A target of 10% of the samples collected during the 2000 GQA Survey should be identified to species level for this purpose and these should represent a good geographical spread of samples and good coverage of all river sizes and types.

The resource implications of species level identification may require that the primary sorting of samples from selected sites should involve a more detailed sorting process in which all, or a known proportion of taxa are removed from samples from the selected sites.

Where species level identification cannot be achieved in house then this should be contracted out to organisations with staff with proven and reliable skills at species level identification.

An audit system for species level identification is necessary (see also Section 5.5.5).

Full identification of these samples may not be achievable during 2000 and may have to be deferred to a later year. Any deferment should not be a cause for delay in the publication of the primary survey report.

Existing species level data on nearly 2000 sites held by IFE may be useful to the pursuit of these recommendations.

Other taxonomic groups

Single diatom samples should be collected during the spring macroinvertebrate sampling of 10% of the sites in the 2000 GQA Survey. MTR samples should be taken at the same sites in summer. These should also be the same sites for which species level macro-invertebrate identification is undertaken (see Section 5.7.1).

Staff collecting TDI and MTR data should have received full training in these techniques prior to sampling.

Full identification and/or interpretation of these samples is unlikely to be achievable during 2000 and would probably have to be deferred to a later year. Identification of diatom samples may need to be contracted out. Any deferment should not be a cause for delay in the publication of the primary survey report.

If RHS surveys are planned for 2000, then it would be beneficial if these could include the test sites used to compare different biomonitoring procedures.

An analysis should be undertaken of the relative merits of the variety of biomonitoring techniques available to the Agency.

Environmental Data

Time variant chemical data (alkalinity)

Alkalinity is an important RIVPACS predictor variable and the reliability of EQI evaluations are reduced by its unavailability. Its regular collection should be retained for all GQA sites until long-term average values can be substituted.

Time invariant physical variables

All time invariant values used for RIVPACS predictions in the 2000 GQA Survey should be re-measured independently by two people and any values differing by >5% should be re-measured by both individuals until an acceptable level of agreement is reached. Values should then be cross-checked against 1995 values and all disparities >10% should be investigated and corrected.

Logical checks of environmental data for individual rivers should be made to ensure that rivers flow downhill and the discharge of any site should be no less than the discharge of the next site upstream of it. The values used in connection with the 2000 GQA Survey should be the average of the corrected values measured for the 1995 and 2000 GQA Surveys. For new sites, the values obtained in 2000, after double measurement and subsequent validation and correction, should be used.

Southern Region's recommendation, that the use of a "Roamer", or similar system, should be mandatory for deriving NGRs, is endorsed here.

The Environment Agency should continue to develop GIS procedures for the accurate derivation of values of cartographic variables.

Time variant physical variables.

All sites in the 2000GQA Survey should be visited in spring, summer and autumn to collect time variant environmental data. All time variant variables recorded during site visits should be measured independently by two people on each visit and an average or consensus value recorded. Where this is not possible, for example because of single-manning, then different staff members should be deployed to obtain values in different field visits.

Where adequate data are available then long-term average data should be used in lieu of new field recording of time variant site data.

Field sampling training days, similar to those provided by the IFE and the NRA prior to the 1995 survey, should be run again prior to the 2000 GQA Survey. It is unlikely that any standard protocols for deep water sampling will be agreed prior to spring 2000 and deep water sampling and prior training should be delayed until these issues are resolved.

Other variables

No further variables should be measured during the 2000 GQA Survey, except where these are required for any of the alternative bioassessment techniques discussed in Sections 5.7.1 and 5.7.2

The Environment Agency should be guided by the results of National R&D Project E1-007 as to the accessibility and predictive power of additional, GIS-derived environmental variables other than those currently used in RIVPACS.

Equipment

If the results of R&D Project E1-007 indicate that the use of a dredge should be included in a standardised deep-water sampling protocol then the Agency should consider further trials on the most appropriate specifications of the dredge to be used. These trials should cover a wide range of substratum types.

Reporting

Current usage

The widespread and varied application of GQA data for regional purposes is to be encouraged and where possible extended.

Potential future use

The family distribution and comparison of change studies conducted on the 1990 RQS and 1995 GQA should be continued in order to incorporate the 2000 GQA Survey data.

The previous recommendation should be implemented as Phase 3 of the current R&D Research project.

Trial species level identification of a sub-set of the 2000 GQA Survey data should be undertaken and the new indices, LIFE and CCI should be applied and evaluated.

THE RESULTS OF THE 2000 GQA SURVEY SHOULD BE PROMINENTLY PUBLISHED IN THE PUBLIC DOMAIN.

Survey Design

General comments

Quinquennial GQAs should be replaced by annual surveys in the form of a rolling programme.

The Environment Agency should consider how this might best be achieved within the context of the resources available and the other elements of the work programme of biology sections. One option for consideration should be the biennial sampling of each site, as described above.

The current practice of using the biological data collected during GQAs, in order to provide a single index of the biological condition of sites offers continuity in meeting the basic national reporting needs of government. It should be continued as part of a wider interpretative programme.

The sole use of RIVPACS and EQIs to examine the biological data collected during GQAs fails to optimise the cost-effectiveness of the survey.

The results of the 2000 GQA Survey should also be examined using the two indices LIFE and CCI, the algorithm for detecting acidification being developed by IFE and the diagnostic artificial intelligence procedures being developed by the University of Staffordshire.

Diatom and MTR sampling should be introduced into future surveys on a trial basis, in order to assess the relative merits of the variety of biomonitoring techniques available to the Agency (see Section 5.7.2)

If RHS surveys are planned for 2000 then it would be beneficial if these could include the test sites used to compare different biomonitoring procedures (see Section 5.7.2).

Number of samples per site per year

The 2000 GQA Survey should be based on two seasons' macro-invertebrate sampling per site.

Individual Regions may choose to sample more frequently, as long as this does not impact on the density of coverage of sites incorporated in the national survey.

DoE (NI) and SEPA report separately and may elect to sample three times a year for their own national requirements without jeopardising the opportunity for unified reporting across the United Kingdom.

Choice of sampling season

2000 GQA Survey sampling should be undertaken in spring and autumn in order to provide a standard basis for inter-survey comparisons of distributional changes.

6 EFFECTS OF SAMPLE ANALYTICAL ERRORS ON THE DETECTION OF CHANGE IN BIOLOGICAL CONDITION

6.1 Introduction

RIVPACS III+ (Clarke *et al*. 1997) provides an assessment of the statistical significance of an observed change in the estimated biological condition of either one site at two points in time or two different sites sampled at the same or different times. The Ecological Quality Indices (EQI) of site condition are based on the ratio (O/E) of the observed (O) to RIVPACS expected (E) values of both number of BMWP taxa and ASPT (Average Score Per Taxon).

In RIVPACS III+, the errors in the estimate of the expected values are assumed to arise from errors in measuring the values of the RIVPACS environmental predictor variables for the site.

The errors or variation in the observed (and recorded) fauna arise from two sources: (i) natural sampling variation and (ii) errors in processing the macro-invertebrate samples due to sampling, sorting and identification errors. The effects of sampling variation within RIVPACS are in a sense fixed and result from a prescribed method of sampling (usually with a pond-net) for an (active) period of three minutes (see Environment Agency 1997a for detailed sampling protocols).

Having obtained a sample, the sample processing errors arise from Agency biologists not finding and recording all the taxa present in the sample and mis-identifying other taxa. Such errors usually result in the number of taxa recorded as present in the sample being less than the number actually present. This systematic net under-estimation of the number of taxa is referred to in RIVPACS III+ terminology as the 'bias'. The extent of errors in processing and in sample analyses is controlled within the Environment Agency by an internal Analytical Quality Control scheme (AQC), whereby a fraction of all samples are re-analysed by their more experienced taxonomists. In addition, since 1990, the Environment Agency has contracted IFE to monitor error rates by auditing an agreed number of their RIVPACS samples. Approximately 500 Agency samples are audited per year.

Very experienced IFE staff identified the macro-invertebrates in the high quality reference sites used to derive RIVPACS predictions and expected values. As the same very high quality of IFE staff are used to audit the Environment Agency's samples, IFE is assumed to equally correctly record all the taxa present in the audit samples (or more importantly, to the same very high standard used for the original RIVPACS reference samples).

Together, the internal AQC and external audit by IFE enable the Agency to record, at both Regional and Area level, their analytical quality and to quantify their sample analytical errors in their estimates of the observed fauna. The number of taxa found by IFE in their audit of a sample that were not recorded by the Environment Agency biologist is termed the sample "gains" in the audit reporting. The number of taxa recorded as present by the Environment Agency biologist but not found by IFE in their audit are termed the sample "losses".

The bias for a sample is the net under-estimation of the number of taxa present and is equal to the "gains" minus the "losses". These biases are also referred to as "net gains" and can be estimated from the column headed "mean net effect on no. of taxa" in the summary tables in each of the audit reports (e.g. Gunn *et al.* (1996a-i)).

Section 2.1.2 of Clarke *et al.* (1999) describes the method of deriving the best estimates ofthe average biases (net under-estimation of number of taxa) achieved by each Region in 1990 and 1995; Table 2.2 in that report is repeated here as Table 6.1.

Table 6.1 also gives the average "losses" per sample in each region in 1995, calculated from Gunn *et al.* (1996a-i). The average "losses" are around 0.25, which is equivalent to one taxon being incorrectly recorded as present in every four samples processed. This level of "losses" is the same as found by Furse *et al.* (1995) in their analysis of the 1990 audit results. Thus the "losses" are much smaller than the "gains". Therefore it is reasonable to assume, for computational tractability, that no taxa have been incorrectly recorded as present and that the average number of taxa missed was equal to net under-estimation of the number of taxa present, namely the average "gains" minus the average "losses", referred to as the bias. These methodological approximations are used in RIVPACS III+ to derive bias-corrected estimates and confidence limits for site condition which allow for sample analytical errors.

Table 6.1 Estimates of average net under-estimation of the number of taxa (termed the bias) in single season samples taken from each region in the 1990 RQS and 1995 GQA surveys. Average "losses" per audited sample in 1995 are also given.

Regions in 1990	Bias in 1990	Regions in 1995	Bias in 1995	Average "losses" per sample in 1995
Anglian	3.40	Anglian	1.98	0.27
Northumbrian	2.67	Northumbria & Yorkshire	1.45	0.07
Yorkshire	1.13			
North West	3.13	North West	2.18	0.33
Severn-Trent	3.77	Severn-Trent	1.64	0.20
Southern	1.57	Southern	1.02	0.32
South West	1.13	South Western	1.42	0.08
Wessex	3.93			
Thames	1.97	Thames	1.78	0.27
Welsh	1.95	Welsh	1.73	0.23

Over the past few years, the Environment Agency target has been to achieve and maintain a gross level of missed taxa of no more than 2.0 per sample. If "losses" are assumed to be about 0.25 taxa per sample, then "gains" of about 2.0 taxa are equivalent to a sample analytical bias of 1.75 taxa.

RIVPACS III+ estimates the joint effects of errors in estimating the RIVPACS expected (E) values, the sampling variation and the sample processing errors by generating computer simulations which incorporate random components representing each of these effects (see Section 7 of Clarke *et al*. 1997 for further details). For example, when the average bias is assumed to be 1.75, then, for each simulated observed (O) value for number of taxa, a statistically independent random number of taxa is added to the simulated observed value obtained by allowing for sampling variation. In this case, the number added is taken as a random number from a statistical Poisson distribution with a mean of 1.75; this is likely to be 0, 1, 2, 3 or 4, but could very occasionally be as much as 7 or 8.

Thus, the effect of having analytical errors and biases is to increase the range of observed differences in EQI values between two samples which could occur and hence increase the width of the confidence interval for the real difference. This then makes it harder to correctly detect changes in biological condition. Because the "losses" are relatively small, this simulation approach of correcting for the net under-estimation of number of taxa will adequately represent the true distribution of sample analytical errors resulting from both "gains" and "losses".

These bias estimates in Table 6.1 were used to correct for bias in RIVPACS III+ in all assessments of the biological condition of GQA sites in each of 1990 and 1995 and of the changes in biological condition.

The aim of this chapter is to use RIVPACS III+ to assess the effect of permitting different levels of analytical quality (i.e. bias or mean net-underestimation of the number of taxa present in samples) on the ability to detect changes in biological condition.

6.2 Methods of Assessment

It was considered important to assess the effects of bias across a wide range of site qualities and differences in qualities. Therefore, the assessment was based on the differences in biological condition between all the 3018 matched sites sampled in both the 1990 RQS and 1995 GQA surveys. The assessments were made on the difference in the EQI values for either number of taxa or ASPT.

RIVPACS III+ derives a frequency distribution of the simulated differences in EQI values for the two samples being compared (by default 500 simulations are used). This frequency distribution is used to provide a statistical test probability of getting the observed, or more extreme difference if there was no real difference in biological condition between the two samples (or, more precisely, the two sites or the same site at two points in time).

Traditionally, in statistical tests, the "null hypothesis" of no real difference is rejected if the test probability is less than 0.05. This 1 in 20 chance is an arbitrary convention; 0.10 or 0.01 may be more appropriate depending on the relative costs of not detecting real differences compared to acting on non-existing differences – referred to as statistical type II and type I errors respectively. However, 0.05 is a convenient test probability to use here to help summarise the effects of varying biases. A difference was therefore considered to be statistically significant if the test probability was less than 0.05.

6.3 Detectable Differences in Relation to the Analytical Bias

6.3.1 Size of difference detected between the 1990 and 1995 matched sites ignoring bias and corrected for actual bias

Figure 6.1 shows how the percentage of matched sites for which the observed difference was considered to be statistically significant ($p<0.05$) increases with the size of the difference. (A more detailed analysis which treated positive and negative differences separately showed that the results were similar regardless of whether the biological condition had improved or declined).

When the bias is ignored (and effectively treated as zero), differences in EQI based on both number of taxa and ASPT need to be at least 0.08 to be detected as statistically significant. To be almost certain (i.e. at least 99% of the time) of detecting a difference in EQI as statistically significant, it needs to be at least 0.23 when based on number of taxa, but only 0.11 when based on ASPT (Table 6.2). This difference is because values of EQI, when uncorrected for bias, are inherently less variable for ASPT than for number of taxa (see Section 2.2 of Clarke *et al*. 1999).

After allowing for the actual sample analytical quality and biases obtained for each Region in the 1990 RQS and 1995 GQA surveys, no differences in EQI for number of taxa were detected as significant unless they were at least 0.12. However, to be almost certain of identifying a difference in EQI_{TAXA} as significant, the difference needed to be at least 0.30, equivalent to a loss of at least 30% more of the expected number of taxa (Table 6.2).

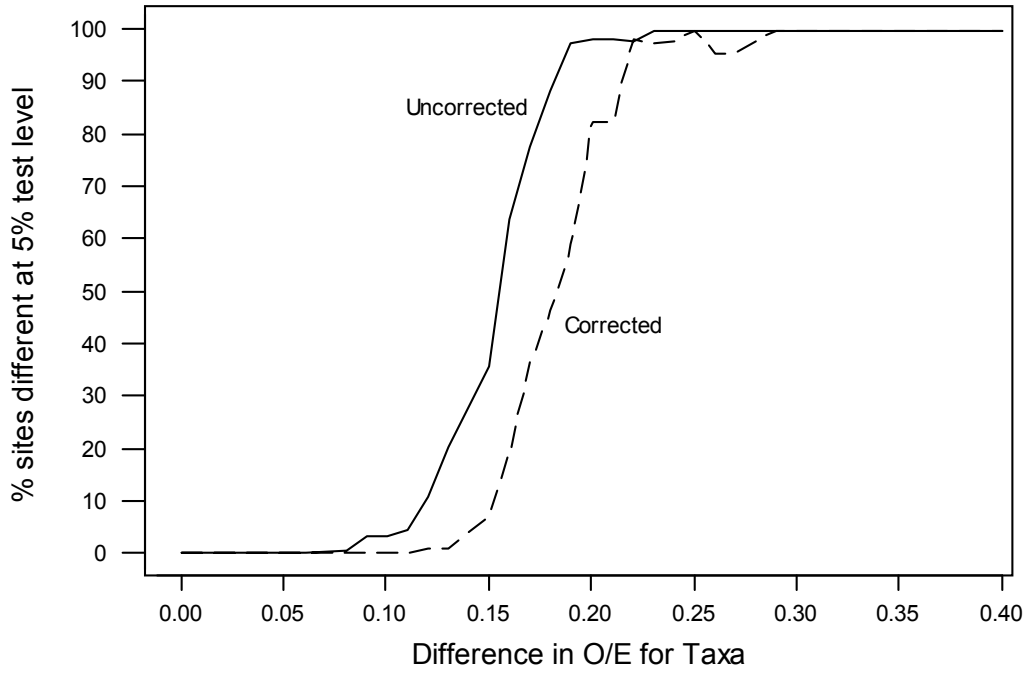
Table 6.2 Critical sizes of the changes in EQI values between 1990 and 1995 for the matched sites which were detected as statistically significant ($p<0.05$), either (a) uncorrected for bias (equivalent to no bias) or (b) after correcting for bias using the best available estimates of biases for each Region.

EQI based on	smallest difference in EQI which was detected as significant		smallest difference in EQI which was nearly always (i.e. in >99% of cases) detected as significant	
	(a) uncorrected	(b) corrected	(a) uncorrected	(b) corrected
Number of taxa	0.08	0.12	0.23	0.30
ASPT	0.08	0.08	0.11	0.20

Sample analytical errors have much less effect on EQI values based on ASPT. In particular, allowing for the biases does not tend to alter the estimate of the change in EQI much, but merely increases the uncertainty in its value. After allowing for biases, some matched sites whose EQI for ASPT was estimated to have changed by only 0.08 were still detected as statistically significant. However, to be almost certain of detecting a difference in EQI based on ASPT, the difference needed to be at least 0.20 (Table 6.2)

The effect of correcting for bias using RIVPACS III+ on the estimate of the 'true' observed (O) value of ASPT, and hence on the EQI value for ASPT, depends on the 'face' (i.e. uncorrected) value of ASPT recorded by the Environment Agency biologists (Figure 6.2). When the 'face' value of EQI is above 0.90, correcting for bias does not usually alter the EQI value by more than ±0.01. However, for sites estimated to be of poorer biological condition in terms of their EQI for ASPT, correcting for bias leads to an increase in the estimate of the 'true' EQI value. In the extreme, for those few (n=14) of the 3018 matched sites with 'face' EQI values less than 0.5, correcting for bias in 1995 added, on average, 0.07 to the estimate of the EQI value. The maximum amount added to the EQI value of any site of any biological condition was 0.09 (Figure 6.2).

a)



(b)

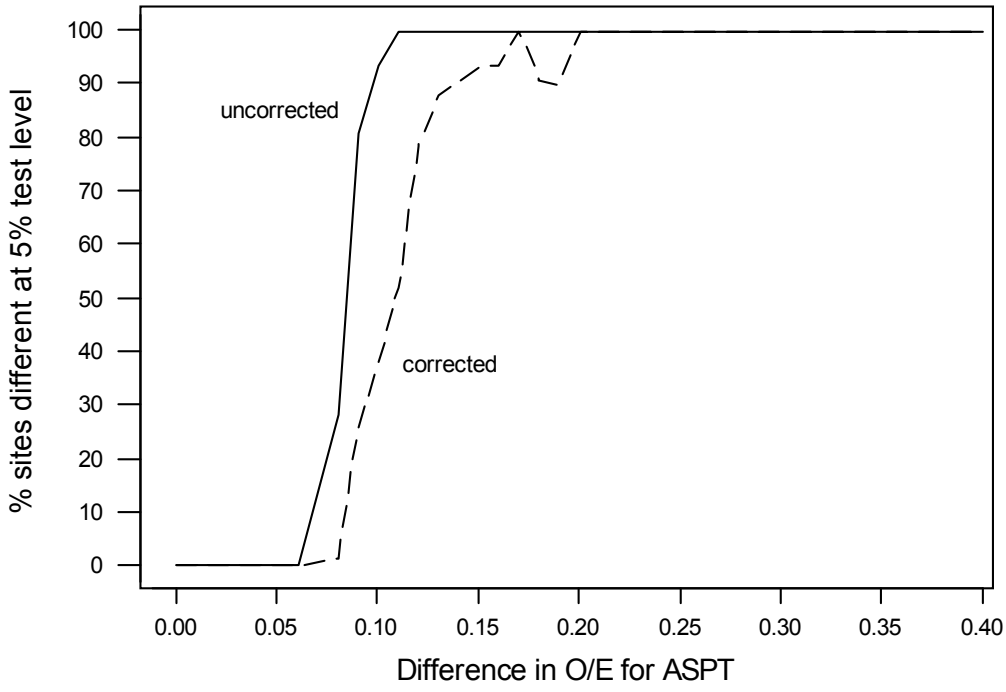


Figure 6.1 Percentage of matched sites with a statistically significant ($p<0.05$) difference in EQI (i.e. O/E) for (a) number of taxa and (b) ASPT between the 1990 RQS and 1995 GQA surveys in relation to the size of the estimated difference. This is given for the EQI values uncorrected for bias (and hence, for this exercise, assuming no bias exists) and corrected for the estimated biases, as given in Table 6.1

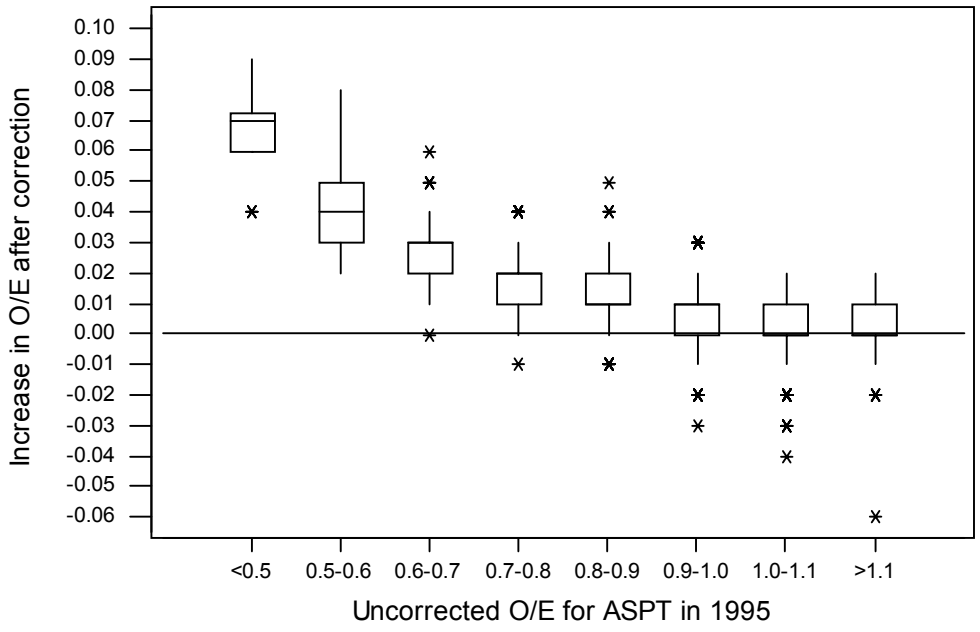


Figure 6.2 Distribution of changes in EQI for ASPT after correcting for biases (using the best estimates for each Region in 1995) in relation to the 'face' (i.e. uncorrected) value of EQI for ASPT. Sites have been grouped into classes of 0.1 of EQI (e.g. 0.45=0.4-0.5). Overall n = 6016 sites.

6.3.2 Effects of alternative analytical quality standards on the ability to detect change in site biological condition

At present the Environment Agency sets an analytical quality standard for processing RIVPACS samples with the target of not exceeding an average of two missed taxa per sample. This roughly equivalent to a sample bias (i.e.net under-estimation of number of taxa) of 1.75 (see Section 6.1).

To investigate the general effect of setting alternative targets, the 1990 and 1995 samples for the 3018 matched sites were re-assessed for changes in EQI on the assumption that the observed or, in RIVPACS III+ terminology, the 'face' fauna for each of the two years' samples for all sites were obtained from a sample analytical standard with a bias of B taxa. RIVPACS III+ procedure Compare was then re-run for all the sets with the fixed bias B set to 0.0, 0.5, 1.0, 1.5, 1.75, 2.0, 2.5, 3.0, 3.5 or 4.0 taxa. Larger sample biases are expected to make it more difficult to detect changes with statistical confidence. The actual 'face' EQI values for the matched samples in 1990 and 1995 were treated as the uncorrected EQI values in these simulations. This gave observed differences in EQI which tended to be greater than the true changes between 1990 and 1995. However this did not invalidate this exercise but rather it made it easier to assess the likelihood of detecting large differences as statistically significant.

Table 6.3 summarises the effects of each bias in terms of the difference in EQI necessary to have at least either a 50% or a 90% chance of being detected as statistically significant at the 95% statistical significance level (i.e. $p<0.05$). In statistical terms the probability of a statistical test detecting a particular size of difference as statistically significant is referred to as the power of the test. Table 6.3 gives the differences in EQI for which RIVPACS III+ has a power of either at least 50% or 90% (Linear interpolation has been used to give smoothed intermediate values).

If no taxa were missed (i.e. zero bias), then it is more likely than not (i.e. power >50%) that differences in EQI_{TAXA} of at least 0.155 and differences in EQI_{ASPT} of at least 0.084 would be detected as statistically significant. This detectable difference rises steadily as the bias increases, such that when average bias equals four taxa, the corresponding critical differences must be at least 0.202 and 0.113 respectively. This represents a 30-35% increase in the minimum size of differences which have at least a 50% chance of being detected.

Moreover, if the bias was increased from the current Agency target of 1.75 to 4.0, the minimum size of difference with at least an even chance of being detected by RIVPACS III+ as statistically significant would increase by 15% when based on EQI for number of taxa and by 11% when based on EQI for ASPT.

Table 6.3 Minimum sizes of real differences (D_m) in EQI_{TAXA} and EQI_{ASPT} values between two samples which have either at least a 50% or at least a 90% chance (i.e. power) of being detected as statistically significant ($p<0.05$) for various fixed levels of sample bias B in both samples. %sites denotes the percentage of matched 1990 RQS and 1995 GQA sites with bias-corrected differences > D_m.

bias B	based on EQI_{TAXA}				based on EQI_{ASPT}			
	50% power		90% power		50% power		90% power	
	D_m	%sites	D_m	%sites	D_m	%sites	D_m	%sites
0.0	0.155	31.5	0.181	24.9	0.084	19.5	0.097	14.1
0.5	0.162	29.5	0.190	22.7	0.090	16.9	0.110	10.2
1.0	0.170	27.5	0.195	21.4	0.095	14.9	0.125	7.6
1.5	0.175	26.6	0.197	20.8	0.101	13.0	0.127	7.3
1.75	0.176	26.3	0.200	20.2	0.102	12.7	0.130	6.9
2.0	0.178	25.7	0.206	18.8	0.104	12.1	0.135	6.2
2.5	0.184	24.0	0.224	15.7	0.106	11.5	0.143	5.3
3.0	0.189	23.4	0.230	14.8	0.108	10.9	0.148	4.6
3.5	0.195	21.4	0.234	14.2	0.111	10.0	0.150	4.3
4.0	0.202	20.0	0.243	12.9	0.113	9.6	0.153	4.0

If a more rigorous test is required to identify a change in EQI (i.e. one with power >90%), then the change needs to be larger (Table 6.3). Based on EQI_{TAXA} and with the current target bias of 1.75 taxa (maximum), the change needs to be at least 0.20 to have at least a 90% chance of being detected as statistically significant by RIVPACS III+. Roughly 20% of all 1990 and 1995 matched sites had greater changes in biological condition. If however, the bias was 4.0 taxa in each survey, then only about 13% of matched sites would have had changes which had such a high probability of being detected as statistically significant. Thus, allowing the sample bias to increase from 1.75 to 4.0 taxa would mean that over one-third (1-12.9/20.2) of all the sites "very likely" (i.e. power>90%) to have been identified as having changed in biological condition by RIVPACS III+ would no longer be "very likely" to be so identified.

It is interesting and slightly surprising that the effect of increased sample biases on the ability to be "very likely" to detect differences appears to even greater when changes are based on EQI_{ASPT} rather than when based on EQI_{TAXA}. When the bias is allowed to increase from zero to 1.75 to 4.0 taxa, the minimum change in EQI_{ASPT} that is "very likely" to be detected as statistically significant increases from 0.097 to 0.130 to 0.153 respectively; these differences were exceeded by 14.1, 6.9 and only 4.0% respectively of all matched sites. Thus, if the permitted sample bias was allowed to increase from 1.75 to 4.0 taxa, over 40% (1-4.0/6.9) of all sites currently "very likely" to be detected as having changed in biological condition would no longer be so likely. If the currently permitted sample bias of 1.75 could be totally eliminated then over twice as many sites would be "very likely" to be identified as having changed in biological condition in terms of their EQI_{ASPT}.

Table 6.3 indicates that at all levels of sample bias, a higher percentage of the 1990 RQS and 1995 GQA matched sites would be identified as having statistically significant (i.e. $p<0.05$) changes in biological condition when based on their changes in EQI_{TAXA} values than when based on their EQI_{ASPT} values. This finding is supported by the real results of running RIVPACS III+ using the actual sample biases in each year in each Environment Agency region (as given in Table 6.1). In this case only about half as many (13.6% versus 26.2%) of the matched sites were detected as having changed in biological condition when based on EQI_{ASPT} compared to when based on EQI_{TAXA}.

Another method of summarising the effect of a range of sample biases on the detection of change is to calculate the proportion of matched sites for which the RIVPACS III+ procedure "Compare" indicated a statistically significant change ($p<0.05$), for each assumed level of bias in both surveys' samples (Table 6.4).

Table 6.4 Percentage of the (n=3018) matched 1990 RQS and 1995 GQA sites detected by RIVPACS III+ as having statistically significant ($p<0.05$) changes in values of either EQI_{TAXA} or EQI_{ASPT} for various assumed fixed levels of sample bias B in both samples.

bias B	% of sites with statistically significant ($p<0.05$) change based on :	
	EQI_{TAXA}	EQI_{ASPT}
0.0	39.7	30.2
0.5	37.1	24.5
1.0	34.9	21.5
1.5	33.4	19.1
1.75	32.8	17.9
2.0	32.0	16.6
2.5	30.2	14.8
3.0	28.9	14.0
3.5	27.3	12.4
4.0	26.2	11.8

As the bias increases there is a steady fall in the percentage of sites whose differences in EQI values are detected as statistically significant, even though the bias-corrected difference in EQI_{TAXA} for a site did not change with the size of bias as the bias was assumed to be the same in both years. A reduction in sample bias for both surveys from 1.75 to 1.0 taxa would have led to only a 6% increase (34.9%/32.8%) in the proportion of sites detected ($p<0.05$) as having changed in value of EQI_{TAXA}. The effect of the reduction is greater for EQI_{ASPT} (17.9% to 21.5%), but most dramatic if sample analytical errors could be completely eliminated, with two-thirds (30.2%/17.9%) more matched sites being detected ($p<0.05$) as having changed in EQI_{ASPT} (Table 6.4).

6.4 Summary

Over the past few years, the Environment Agency target has been to achieve and maintain a gross level of missed taxa of no more than 2.0 per sample. With average audit "losses" of about 0.25 taxa per taxa, audit "gains" of about 2.0 taxa are equivalent to a sample analytical bias of 1.75 taxa.

The effect of sample analytical errors, and the systematic bias they introduce, is to make it more difficult to estimate the biological condition of a site in terms of EQI value and hence grade. Consequently, it also becomes more difficult to detect and estimate the size of a change over time, or the difference between two sites.

The likelihood of detecting an observed change in bias-corrected EQI as statistically significant, will obviously partly depend on the size of the change. However, for a given real change in EQI value at a site, increased biases make it harder to detect the change with any particular statistical confidence.

Correcting for sample biases obviously increased the RIVPACS 'observed' (O) value for number of taxa. The ASPT of the missed taxa tends to slightly higher than that of the recorded taxa, except for sites of high biological condition (Furse *et al.* 1995). Correcting for biases tends to increase the EQI_{ASPT} value (occasionally by as much as 0.09) for sites with low face EQI values, but has little or effect on the estimate EQI_{ASPT} value for sites in good biological condition with face EQI values greater than 0.9.

For the 1990 RQS and 1995 GQA matched sites, the smallest change in EQI values which was detectable as statistically significant ($p<0.05$), after correcting for the actual biases which occurred in each region in each year, was 0.12 using EQI_{TAXA} and 0.08 using EQI_{ASPT}. However, to nearly always (i.e. in >99% of cases) identify changes as being significant, the changes in EQI need to be 0.30 and 0.20 respectively. These are substantial practical changes with the former equivalent to one sample having 30% less of its expected number of taxa than the other sample..

Simulations were used to test the effect of a range of biases (0.0, 0.5, 1.0, 1.5, 1.75, 2.0, 2.5, 3.0, 3.5 or 4.0 taxa) on the ability to detect change in the matched samples. The actual 'face' EQI values for the matched samples in 1990 and 1995 were treated as the uncorrected EQI values in these simulations.

The smallest statistically detectable change in EQI increases with the size of the biases. To be "very likely" (i.e. greater than 90% probability) to detect a change in EQI_{TAXA} as being statistically significant ($p<0.05$) needs a change of at least 0.18 if there are no sample biases, and this increases to 0.20 and 0.24 for biases of 1.75 and 4.0 taxa respectively. For EQI_{ASPT}, the equivalent required changes for biases of 0, 2.0 and 4.0 taxa are 0.10, 0.13 and 0.15 respectively.

Although correcting for sample biases gives only a small increase in estimated EQI values based on ASPT, it now correctly increases the uncertainty in the true EQI_{ASPT} values, making it considerably harder to detect changes. The effect of a bias of 1.75 taxa (compared to no biases) is to reduce the percentage of the simulated matched sites detected as having statistically significant changes in biological condition from 40% to 33% when based on EQI_{TAXA}, but to nearly halve it, from 30% to 18%, when based on EQI_{ASPT}.

Allowing the sample bias to have been 4.0 taxa rather than the current permitted level of 1.75 taxa would have resulted in nearly 40% fewer sites being "very likely" (i.e. >90% probability) to have statistically significant changes in either EQI_{TAXA} or EQI_{ASPT} .

There is no obvious value of bias at which the ability to detect differences in EQI values shows an abrupt change and hence might be used as a critical bias limit. There is a fairly smooth decrease in the power to detect change and the size of the sample processing biases.

These analyses suggest that there would undesirable loss of statistical power to detect change in biological condition if the tolerable sample analytical bias was allowed to increase from 1.75 to a level of over 3 taxa per sample.

There appears to be considerable reduction in the size of change in EQI, especially for EQI_{ASPT}, which can be detected with high statistical confidence if the sample analytical errors could be completely eliminated. However, this is not a practical option for the Environment Agency. **The gain in power to detect change obtained by reducing the sample analytical bias from 1.75 to 1.0 taxa is not great**.

We recommend that the Environment Agency continue to aim to achieve and maintain a gross level of missed taxa of no more than 2.0 per sample, which has been equivalent to a sample analytical bias of 1.75 taxa.

7 FACTORS ASSOCIATED WITH POOR PERFORMANCE IN SAMPLE PROCESSING

The aim of this section is to assess what factors, if any, seem to be associated with problems in processing macro-invertebrate samples and which lead to relatively large numbers of taxa being missed and an unacceptable degree of under-estimation of the number of taxa present.

7.1 Availability of Audited Samples in 1995

This analyses requires data on the difference between the number of taxa originally recorded for a sample by the Agency biologist and the number actually present in a sample, as recorded by IFE in their primary audit of the sample. This difference represents the net under-estimation of number of taxa present, and is termed the sample bias in RIVPACS III+ terminology (Clarke *et al*. 1997).

The analysis was based on all 1995 GQA sites which had a sample subjected to a primary audit in 1995 and for which site identifiers in the IFE audit data base and the IFE Quinquennial Survey Database (QSD) could be reliably matched. In total, 481 GQA samples were given a primary audit by IFE in 1995 (Gunn *et al*. 1996a, 1996 c-i). Of these, 393 could be reliably linked to sites in the QSD database (mostly through their 9-digit site codes) and these were used to relate sample biases to environmental characteristics of sites.

It is worth noting that there is still room for improvement in the consistency with which completely standard names and codes for sites and rivers are used on the sample audit forms.

7.2 Sample Processing Biases in Relation to Site Quality

Furse *et al.* (1995) have already shown, in their analyses of the 1990 and 1992 audited samples, that, perhaps surprisingly, the average net under-estimation of the number of taxa present (i.e. biases) does not generally increase with the taxon richness of the sample. They concluded that the sampling bias for single season samples could be assumed to be a constant, except for samples with less than five taxa recorded in which case the average number of taxa missed was less and on average, only about one taxon per sample.

Table 7.1 shows the bias for samples in 1995 in relation to the biological condition of a site as represented by its Ecological Quality Index (EQI) value based on number of BMWP taxa, namely EQI_{TAXA}. This re-enforces the conclusion of Furse *et al.*(1995) that the sample processing biases appears to be independent of the quality of site, except perhaps for very poor quality sites (GQA grades e/f).

Table 7.2 gives the average bias amongst this set of 393 audited samples for sites in each NRA/Environment Agency Region. Regional differences in sample processing performance in 1995 are likely to affect apparent correlations of size of sample processing errors with the values of site environmental variables, as these are known to vary between Regions. A Kruskal-Wallis one-way analysis of variance (ANOVA) of ranks of sample bias in relation to Region indicated that there were some statistically significant differences in sample biases between Regions in 1995 ($p=0.026$). However, there are insufficient samples to warrant completely separate analyses for each Region.

Table 7.1 Sample processing biases amongst GQA sites in 1995 in relation to site quality as measured by EQI_{TAXA}.

EQI_{TAXA}	number of audit samples available	net underestimation of number of taxa present in the sample	
		mean	maximum
<0.4	17	0.65	3
0.40-0.49	27	1.85	8
0.50-0.59	39	1.51	6
0.60-0.69	43	1.79	5
0.70-0.79	41	2.10	7
0.80-0.89	45	1.36	7
0.90-0.99	61	2.23	6
1.00-1.09	55	2.00	7
1.10-1.19	37	1.49	5
1.20-1.29	16	1.69	5
>1.29	12	1.75	5
overall	**393**	**1.76**	**8**

Table 7.2 Average sample processing biases in 1995 in each NRA/Environment Agency Region based on the total of 393 audited samples matched to 1995 GQA sites.

Region in 1990	number of samples	average bias
Anglian	44	2.16
Northumbrian	16	1.00
North-West	49	2.31
Midlands	41	1.90
Southern	59	1.19
South-West	23	1.34
Thames	51	1.88
Welsh	51	1.71
Wessex	23	1.52
Yorkshire	36	2.00
England and Wales	**393**	**1.76**

7.3 Sample Processing Biases in Relation to Environmental Characteristics of Sites

To assess the relationship between a site's quality and its environmental characteristics, Davy-Bowker *et al.* (2000), in R&D Technical Report E103 of this project, grouped the GQA sites into six categories on the basis of their value for a particular environmental variable. It seemed sensible and was convenient to use the same groupings of sites for each environmental variable in this study of sample processing biases in relation to environmental characteristics; and this has been done (Table 7.3).

Table 7.3 Average sample processing biases (number of samples in brackets) for 1995 audit samples from sites in England and Wales classified into six categories of each RIVPACS environmental variable. r_S denotes Spearman's rank correlation between sample bias and the site's value for the variables; p_K denotes statistical significance level of Kruskal-Wallis one-way ANOVA of ranks of sample bias on the variable categories; *, denote significance at the p<0.05, <0.01 respectively, ns = not significant (p>0.05).**

							p_K	r_S
Altitude (m)	<16 1.86 (104)	16-36 1.46 (76)	37-64 1.80 (84)	65-99 1.54 (70)	100-200 1.88 (48)	>200 3.64 (11)	*	0.04
Slope (m/km)	<1.1 1.92 (93)	1.1-2.2 1.79 (85)	2.3-4.4 1.84 (79)	4.5-9.1 1.51 (80)	9.2-25 1.45 (47)	>25 3.11 (9)	*	-0.07
Discharge class	1 1.68 (204)	2 1.84 (45)	3 1.60 (48)	4-5 1.77 (52)	6-7 2.36 (33)	8-10 1.82 (11)	ns	0.06
Distance from source (km)	<5.0 1.38 (95)	5.0-7.9 1.74 (66)	8.0-12.5 1.83 (87)	12.6-24 1.80 (59)	24.1-84 2.20 (76)	>84 1.50 (10)	ns	0.14 **
Stream width (m)	<2.3 1.61 (80)	2.3-3.5 1.67 (82)	3.6-5.3 1.81 (81)	5.4-9.5 1.73 (66)	9.6-29 2.11 (71)	>29 1.23 (13)	ns	0.08
Stream depth (cm)	<12 1.49 (72)	12-16 1.93 (73)	17-23 1.59 (83)	24-36 1.92 (83)	37-132 1.92 (72)	>132 1.60 (10)	ns	0.10 *
Alkalinity mg/l $CaCO_3$	<61 1.99 (68)	61-123 1.76 (71)	124-182 1.65 (92)	183-227 1.64 (81)	228-284 1.82 (65)	>284 1.88 (16)	ns	-0.01
%boulders/ cobbles	0-4 1.71 (99)	5-13 1.70 (87)	14-30 1.72 (72)	31-51 1.80 (71)	52-76 2.00 (53)	77-100 1.64 (11)	ns	0.03
%pebbles/ gravel	0-20 1.57 (84)	21-33 2.14 (71)	34-43 1.68 (78)	44-55 2.09 (78)	56-76 1.50 (76)	77-100 1.41 (17)	ns	-0.05
%Sand	0-2 1.77 (87)	3-6 1.68 (85)	7-11 1.52 (65)	12-20 1.70 (76)	21-43 2.15 (67)	44-100 1.85 (13)	ns	0.06
%Silt/Clay	0-1 2.07 (82)	2-5 1.65 (62)	6-13 1.81 (80)	14-34 1.55 (74)	35-95 1.69 (86)	96-100 1.78 (9)	ns	-0.07
Mean substratum (phi units)	-7.8:-5.0 2.03 (67)	-4.9:-3.1 1.67 (72)	-3.0:-1.4 1.75 (75)	-1.3:1.5 1.87 (85)	1.6:7.6 1.54 (85)	7.7:8 1.78 (9)	ns	-0.03

There is some suggestion that the processing errors tend to be greater for those few samples taken from sites which are either at high altitude (i.e. >200m) or on steep slopes (i.e. >25m km^{-1}) (Table 7.3). These tendencies for higher biases in steep sloped sites can be partly explained by regional co-differences in sample processing errors and site slope (the highest slopes occurred mostly in North West and Welsh Regions). However, the association between high bias and high altitude still existed after eliminating regional differences, using a generalised ANOVA. This association with altitude is investigated further below.

There was a positive and statistically significant correlation (r_S=0.14, p=0.004) between bias size and distance from source of the site; this was not dependent on any regional co-differences. As distance from source increase, so discharge tends to increase; although there was no overall (ANOVA) relationship with discharge, sample processing biases appeared to be higher for river sites with discharges classes 6-7 (Table 7.3).

The statistically significant (p=0.039) rank correlation between bias size and stream depth was partly related to their joint covariance with Region; allowing for regional differences reduced the size (r_S=0.09) and significance (p=0.064) of their rank correlation.

Perhaps, surprisingly, the type of river bed substratum at a site did appear to have any consistent effect on the size of the sample processing errors and the rate of missing taxa in the sample.

Why were there larger sample processing errors at sites with the highest altitudes and steepest slopes ? Table 7.4 shows that most (7 out of 9) of the steepest sloped sites were also at relatively high altitude (i.e. >100m) so associations with high slope are inseparable from those with high altitude. However, high altitude sites occur across a range of slopes and all have higher than average sample processing biases.

Table 7.4 Average sample processing biases (number of samples in brackets) for 1995 audit samples from sites in England and Wales in relation to site altitude (m) and slope (m km^{-1}).

	Slope (m km^{-1})						
Altitude (m)	<1.1	1.1-2.2	2.3-4.4	4.5-9.1	9.2-25	>25	Overall
<16	1.89 (61)	1.95 (20)	2.20 (10)	0.88 (8)	1.25 (4)	5.00 (1)	1.86 (104)
16-36	2.13 (15)	1.40 (25)	1.60 (20)	0.71 (14)	1.00 (2)		1.46 (76)
37-64	1.50 (10)	1.95 (19)	1.84 (25)	2.10 (20)	1.11 (9)	1.00 (1)	1.80 (84)
65-99	2.43 (7)	1.69 (16)	1.75 (12)	1.33 (24)	1.00 (11)		1.54 (70)
100-200		2.80 (5)	1.91 (11)	1.73 (11)	1.31 (16)	3.00 (5)	1.88 (48)
>200			3.00 (1)	3.67 (3)	3.80 (5)	3.50 (2)	3.64 (11)
Overall	1.92 (93)	1.79 (85)	1.84 (79)	1.51 (80)	1.45 (47)	3.11 (9)	1.76 (393)

After taking a sample, it is either sorted live or preserved back in the laboratory. The method of sorting may influence the sample processing biases. Table 7.5(a) shows the sorting method(s) used by each Region for the 393 audit samples from the 1995 GQA survey analysed here, together with the average processing biases for samples sorted by each method. Although the average bias is slightly higher for live-sorted samples, there are no large differences between the two methods and no consistency across Regions in terms of which method gives the greatest precision. However, there are no instances where direct comparisons can be made between sorting performances for at least twenty live and twenty preserved samples in any given Region. Table 7.5(b) shows the average bias in relation to sorting method for sites in each altitude class for each Region where both methods were used. Sorting method does not explain the higher processing errors associated with high altitude sites.

Table 7.5 Average sample processing biases (number of samples in brackets) for 1995 audit samples from sites in England and Wales in relation to (a) Region and sorting method (whether done on the live or preserved sample), (b) Region, site altitude (m) and sorting method.

(a) Region	Sorting method	
	Live	Preserved
Anglian	2.19 (43)	1.00 (1)
Northumbrian		1.00 (16)
North West		2.31 (49)
Midlands	1.70 (27)	2.29 (14)
Southern		1.14 (58)
South West		1.35 (23)
Thames	1.88 (49)	2.00 (2)
Welsh		1.71 (51)
Wessex	1.38 (8)	1.60 (15)
Yorkshire		2.00 (36)
Grand Total	1.92 (127)	1.68 (265)

(b) Region	Sorting method	Altitude.						Overall
		<16	16-36	37-64	65-99	100-200	>200	
Anglian	Live	2.00 (19)	2.14 (14)	2.78 (9)	1.00 (1)			2.19 (43)
	Preserved	1.00 (1)						1.00 (1)
Midlands	Live		0.80 (5)	1.71 (7)	2.11 (9)	2.00 (5)	1.00 (1)	1.70 (27)
	Preserved	3.00 (5)	2.00 (1)	4.00 (1)	2.25 (4)	0.67 (3)		2.29 (14)
Thames	Live	2.00 (6)	1.63 (8)	1.47 (15)	2.00 916)	3.25 (4)		1.88 (49)
	Preserved		1.00 (1)		3.00 (1)			2.00 (2)
Wessex	Live	2.00 (1)	0.00 (2)	2.00 (2)	0.00 (1)	2.50 (2)		1.38 (8)
	Preserved	1.00 (2)	1.60 (5)	6.00 (1)	0.75 (4)	1.67 (3)		1.60 (15)
Anglian	Mostly live	1.95 (20)	2.14 (14)	2.78 (9)	1.00 (1)			2.16 (44)
Northumbrian	All Preserved	0.00 (1)		0.71 (7)	1.25 (4)	1.50 (4)		1.00 (16)
North West	All Preserved	2.67 (15)	2.44 (9)	0.83 (6)	1.40 (5)	2.09 (11)	5.33 (3)	2.31 (49)
Midlands	both	3.00 (5)	1.00 (6)	2.00 (8)	2.15 (13)	1.50 (8)	1.00 (1)	1.90 (41)
Southern	Mostly Preserved	1.36 (28)	1.06 (17)	1.22 (9)	0.25 (4)	2.00 (1)		1.19 (59)
South West	All Preserved	1.25 (4)	0.33 (3)	1.60 (5)	0.83 (6)	2.00 (3)	3.00 (2)	1.35 (23)
Thames	Mostly live	2.00 (6)	1.56 (9)	1.47 (15)	2.06 (17)	3.25 (4)		1.88 (51)
Welsh	All Preserved	1.79 (14)	1.14 (7)	2.17 (12)	0.14 (7)	1.67 (6)	3.40 (5)	1.71 (51)
Wessex	both	1.33 (3)	1.14 (7)	3.33 (3)	0.60 (5)	2.00 (5)		1.52 (23)
Yorkshire	All Preserved	1.88 (8)	1.00 (4)	2.30 (10)	2.75 (8)	1.33 (6)		2.00 (36)

7.4 Summary

The size of sample processing errors does not appear to depend on the quality of the site (i.e. its value for EQI_{TAXA});,except for very poor quality sites with GQA grades e/f, which have very few taxa. This agrees with Furse *et al.* (1995) who found no relationship between sample biases and the taxonomic richness of the sample.

Sample processing errors were examined in relation to the environmental characteristics of the sites. Those few samples from sites at high altitudes (i.e. >200m) and/or with steep slopes (i.e. >25m/km) had, on average, larger processing errors. This was partly associated with the higher sample errors in North West Region.

No other environmental characteristics, including substratum sediment type, appeared to be associated with higher than average sample processing errors.

Whether a sample was sorted live or after being preserved did not seem to consistently influence the size of sample processing errors across all Regions in general. However no single Region processed adequate numbers of **both** live and preserved samples for meaningful comparisons of sorting performances to be made.

It is recommended that the Agency improves their consistency in the use completely standard names and codes for each site and river used on the sample audit forms and in biological, environmental and chemical databases. Spatial and temporal comparisons of site data, or co-analysis of, say, chemical and biological data, are strongly compromised by failure to adhere to this strategy.

8 REFERENCES

Biological Monitoring Working Party (1978) *Assessment and Presentation of the Biological Quality of Rivers in Great Britain*. The final report of the Biological Monitoring Working Party , December 1978. Standing Technical Advisory Committee on Water Quality, Freshwater Monitoring Group report number STACWQ FWMG/79/23.

Chesters, R K (1980) Biological Monitoring Working Party. The 1978 national testing exercise. *Department of the Environment, Water Data Unit Technical Memorandum*, **19**, 1-37.

Clarke, R T (in press) Uncertainty in estimates of biological quality based on RIVPACS In *Assessing the Biological Quality of Fresh Waters. RIVPACS and Similar Techniques*, edited by J F Wright, D W Sutcliffe & M T Furse, pp39-54, Ambleside: Freshwater Biological Association.

Clarke, R T, Cox, R, Furse, M T, Wright, J F & Moss, D (1997) *River InVertebrate Prediction and Classification System with Error Assessments. User Manual.* R&D Technical Report E26. 64pp + appendices, Bristol: Environment Agency.

Clarke, R T, Furse, M T & Davy-Bowker, J (1999) *Analysis of 1995 Survey Data. Phase 2 Post-survey Appraisal. Unit II: Changes in Biological Condition*. R&D Technical E101, 130pp, Bristol: Environment Agency

Clarke, R T, Furse, M T & Wright, J F (1992) *Testing and Further Development of RIVPACS. A Comparison of Single, Paired and 3 Season Combined Macro-invertebrate Samples for the Biological Banding of River Quality*. A preliminary R&D report (243/2/Y) to the National Rivers Authority.

Clarke, R T, Furse, M T & Wright, J F (1994) *Testing and Further Development of RIVPACS. Phase II: Aspects of Robustness*. An interim R&D report (243/7/Y) to the National Rivers Authority.

Commission of the European Communities (1998) Amended proposal for a Council Directive establishing a framework for Community action in the field of water policy (COM(97) 49 final, Brussels, 17-02-98, COM(1998) 76 final, 57pp.

Davy-Bowker, J, Clarke, R T, Furse, M T & Gravelle, M J (2000) *Analysis of 1995 Survey Data. Phase 2 Post-survey Appraisal. Unit I: Taxon distribution Studies*. R&D Technical Report E103, Bristol: Environment Agency

Department of the Environment and the Welsh Office (1971) *Report of a River Pollution Survey of England and Wales* 1970. Volume 1, London: HMSO.

Department of the Environment and the Welsh Office (1972) *River Pollution Survey of England and Wales. Updated 1972. River Quality*, London: HMSO.

Department of the Environment and the Welsh Office (1975) *River Pollution Survey of England and Wales. Updated 1973. River Quality and Discharges of Sewage and Industrial Effluents*, London: HMSO.

Department of the Environment and the Welsh Office (1978) *River Pollution Survey of England and Wales. Updated 1975. River Quality and Discharges of Sewage and Industrial Effluents*, London: HMSO.

Department of the Environment and The Welsh Office (1986) *River Quality in England and Wales 1985. A Report of the 1985 Survey*, London: HMSO.

Dines, R A & Murray-Bligh, J A D (in press) Quality assurance and RIVPACS. In *Assessing the Biological Quality of Fresh Waters. RIVPACS and Similar Techniques*, edited by J F Wright, D W Sutcliffe & M T Furse, pp71-78, Ambleside: Freshwater Biological Association.

Environment Agency (1996a*) Procedure for Quality Assurance for RIVPACS Compatible Macro-invertebrate Samples Analysed to the Taxonomic Level Needed for the BMWP Score System.* Quality Management Systems for Environmental Monitoring: Biological Techniques, BT003, draft version . Bristol: Environment Agency.

[Note several draft versions of this report were prepared. The version quoted in this report was the first draft version . The first non-draft version, which is still in current usage in November 1999, was Version 1.0 (1999)].

Environment Agency (1996b*) Procedure for Collecting and Analysing Macro-invertebrate Samples for GQA Surveys*. Quality Management Systems for Environmental Monitoring: Biological Techniques, BT002. Bristol: Environment Agency.

Environment Agency (1996c) *The Quality of Rivers in England and Wales (1990 to 1995).* Unpublished Report, November 1996. Bristol: Environment Agency.

Environment Agency (1997a) *Procedures for Collecting and Analysing Macro-invertebrate Samples*. Quality Management Systems for Environmental Monitoring: Biological Techniques, BT001. Bristol: Environment Agency.

[Note several draft versions of this report were prepared. The version quoted in this report was draft 0.5 and was produced in 1996. The version available in 1995 was draft 0.4 . The first non-draft version was 1.0 (1997). The version in current usage in November 1999 is Version 2.0 (1999)].

Environment Agency (1997b) *The Quality of Rivers and Canals in England and Wales 1995*. Bristol: Environment Agency.

Extence, C A, Balbi, D M & Chadd, R P (in press) River flow indexing using benthic macroinvetebrates: a framework for setting biological flow objectives. *Regulated Rivers.*

Furse, M.T. (1995) *The Faunal Richness of Headwater Streams. Stage 4 - Development of a Conservation Strategy*, National Rivers Authority R&D Note 455, Bristol: National Rivers Authority.

Furse, M T, Clarke, R T, Winder, J M, Symes, K L, Blackburn, J H, Grieve, N J & Gunn, R J M (1995) *Biological Assessment Methods: Controlling the Quality of Biological Data. Package 1 The Variability of Data Used for Assessing the Biological Condition of Rivers*. A report to the National Rivers Authority, 139pp.

Furse, M T & Symes, K L (1997) *Operational Mechanisms for the Protection and Enhancement of Headwaters. Scoping Study*. Environment Agency R&D Technical Report E25, 125pp, Bristol: Environment Agency.

Furse, M T, Winder, J M, Symes, K L, Clarke, R T, Gunn, R J M, Blackburn, J H & Fuller, R M (1993) *The Faunal Richness of Headwater Streams. Stage 2 - Catchment Studies.* National Rivers Authority R&D Note 221, Bristol: National Rivers Authority.

Gunn R J M, Blackburn J H, Winder J M, Wiggers, R, Symes K L, Davy-Bowker, J & Wright J F (1999) *An Audit of Performance in the Analysis of Biological Samples in 1998. Environment Agency: Primary Audit*. A report to the National Rivers Authority, 112pp.

Gunn R J M, Blackburn, J H, Winder J M, Wright J F & Symes K L (1996a) *An Audit of Performance in the Analysis of Biological Samples in 1995. NRA Anglian Region*: *Primary Audit*. A report to the National Rivers Authority, Anglian Region, 74pp.

Gunn R.J.M., Blackburn J.H., Winder J.M., Wright J.F. & Symes K.L. (1996b) *An Audit of Performance in the Analysis of Biological Samples in 1995. NRA Anglian Region*: *AQC Audit*. A report to the National Rivers Authority, Anglian Region. 48pp.

Gunn R J M, Blackburn J H, Winder J M, Wright J F & Symes K L (1996c) *An Audit of Performance in the Analysis of Biological Samples in 1995. NRA Northumbria & Yorkshire Region*: *Primary Audit*. A report to the National Rivers Authority, Northumbria & Yorkshire Region, 73pp.

Gunn R J M, Blackburn J H, Winder J M, Wright J F & Symes K L (1996d) *An Audit of Performance in the Analysis of Biological Samples in 1995*. NRA North West Region: *Primary Audit*. A report to the National Rivers Authority, North West Region, 74pp.

Gunn R J M, Blackburn J H, Winder J M, Wright J F & Symes K L (1996e) *An Audit of Performance in the Analysis of Biological Samples in 1995. NRA Severn-Trent Region*: *Primary Audit*. A report to the National Rivers Authority, Severn-Trent Region, 75pp.

Gunn R J M, Blackburn J H, Winder J M, Wright J F & Symes K L (1996f) *An Audit of Performance in the Analysis of Biological Samples in 1995. NRA Southern Region: Primary Audit*. A report to the National Rivers Authority, Southern Region, 72pp.

Gunn R J M, Blackburn J H, Winder J M, Wright J F & Symes K L (1996g) *An Audit of Performance in the Analysis of Biological Samples in 1995*. NRA *South Western Region: Primary Audit*. A report to the National Rivers Authority, South Western Region, 75pp.

Gunn R J M, Blackburn J H, Winder J M, Wright J F & Symes K L (1996h) *An Audit of Performance in the Analysis of Biological Samples in 1995. NRA Thames Region: Primary Audit*. A report to the National Rivers Authority, Thames Region, 72pp.

Gunn R J M, Blackburn J H, Winder J M, Wright J F & Symes K L (1996i) *An Audit of Performance in the Analysis of Biological Samples in 1995. NRA Welsh Region: Primary Audit*. A report to the National Rivers Authority, Welsh Region, 74pp.

Gunn R J M, Blackburn J H, Winder J M, Wright J F, Symes K L & Davy-Bowker, J (1998) *An Audit of Performance in the Analysis of Biological Samples in 1997. Environment Agency: Primary Audit*. A report to the National Rivers Authority, 112pp.

Gunn, R J M., Wright, J F, Blackburn, J H & Furse, M T (1991) *Quality Audit of Samples for the 1990 River Quality Survey, NRA Wessex Region*. A report to the National Rivers Authority, Wessex Region, 45 pp.

Hawkes, H A (1997) Origin and development of the Biological Monitoring Working Party score system. *Water Research*, **32**, (3), 964-968.

Hickin, N E (1967) *Caddis Larvae*. London: Hutchinson, 476pp.

Holmes, N.T.H., Newman, J.R., Chadd, S., Rouen, K.J., Sharp, L. and Dawson, F.H. (1999) Mean Trophic Rank: A users' manual. R&D Technical Report No E38. Bristol: Environment Agency.

Humphrey, C L, Storey, A W & Thurtell, L (in press) AUSRIVAS: operator sample processing errors and temporal variability – implications for model sensitivity. In *Assessing the Biological Quality of Fresh Waters. RIVPACS and Similar Techniques*, edited by J F Wright, D W Sutcliffe & M T Furse, pp143-163, Ambleside: Freshwater Biological Association.

Kelly, M.G. (1998) Use of the trophic diatom index to monitor eutrophication in rivers. *Water Research*, **32**, 236-242

Kinley, R D & Ellis, J C (1991) *The Application of Statistical Quality Control Methods to Macroinvertebrate Data*, Unpublished report to the National Rivers Authority, Thames Region, WRc Report N^{o} CO 2905, Medmenham: Water Research Centre.

National Rivers Authority (1991) *The Quality of Rivers, Canals and Estuaries in England and Wales. Report of the 1990 Survey*, Bristol: National Rivers Authority.

National Water Council (1981) *River Quality: The 1980 Survey and Future Outlook*, London: National Water Council.

Rutt, G P, Weatherly, N S & Ormerod, S J (1990) Relationships between the physiochemistry and macroinvertebrates of British upland streams: the development of modelling and indicator systems for predicting fauna and detecting acidity. *Freshwater Biology*, **24**, 463-480.

Sweeting, R A, Lowson, D, Hale P. & Wright, J F (1992) 1990 biological assessment of rivers in the UK. In *River Water Quality - Ecological Assessment and Control*, edited by P J Newman, M A Piavaux and R A Sweeting, 319-326, Luxembourg: Commission of the European Communities, publication number EUR 14606 EN-FR.

van Dijk, P (1994) *Analytical Quality Control for Macroinvertebrate Enumeration*, National Rivers Authority R&D Note 331, Bristol: National Rivers Authority.

Walley, W J, Fontama, V N & R. W. Martin R W (1998). Applications of artificial intelligence in river quality surveys. Environment Agency R&D Technical Report E52, Bristol: Environment Agency.

Walley, W J & Martin, R W (1997). *Distribution of Macroinvertebrates in English and Welsh Rivers Based on the 1995 Survey*. Environment Agency R&D Technical Report E12, Bristol: Environment Agency.

Walley W J & Martin, R W (1998). Applications of artificial intelligence in river quality surveys. Environment Agency R&D Project Record, E1/i621/6, Bristol: Environment Agency.

Williams, P, Biggs, J, Whitfield, M, Corfield, A, Fox, G & Adare, K (1998*). Biological techniques of still water quality assessment. 2. Method development.* Environment Agency R&D Technical Report E56, Bristol: Environment Agency.

Wright, J F, Clarke, R T, Gunn, R J M, Blackburn, J H & Davy-Bowker, D (1999) *Testing and Further Development of RIVPACS Phase 3. Development of New RIVPACS Methodologies. Stage 1*. Environment Agency R&D Technical Report E71, Bristol: Environment Agency.

Wright, J F, Furse, M T & Armitage, P D (1993) RIVPACS - a technique for evaluating the biological quality of rivers in the UK, *European Water Pollution Control*, **3**, 15-25.

Wright, J F, Furse, M T, Clarke, R T & Moss, D (1991) *Testing and Further Development of RIVPACS*. An interim R&D report (243/1/Y) to the National Rivers Authority.