Battelle
Arlington Office
2101 Wilson Boulevard, Suite 800
Arlington VA 22201-3008
Telephone (703) 528-6603
Telecopier (703) 527-5640

February 17, 1989

Ms. Mary Frankenberry
U.S. Environmental Protection Agency
Office of Toxic Substances
401 M Street, SW
Washington, DC 20460

Dear Mary:

Contract No. 68-02-4294

Enclosed is a copy of the Draft Final Report, for Task 2-22, "Quality Control Guidance," prepared under the above contract number.

If you have any questions, please call me at (703) 875-2963 or Bertram Price at (202) 457-9007.

Sincerely,

Barbara

Barbara Leczynski
Project Manager
Applied Statistics and
Computer Applications Section

BL:bs

Enclosures

cc: S. Dillman
J. Glatz
C. Stroup
P. Cross, EED Contract Monitor
E. Sterrett
L. Farmer (ltr only)

February 16, 1989

DRAFT FINAL REPORT

for

Task 2-22

QUALITY CONTROL GUIDANCE

by

Bertram Price
Anne Morris Price

Price Associates
1825 K Street, N.W.
Washington, D.C. 20006

Barbara Leczynski
Task Leader

BATTELLE
Columbus Division - Washington Operations
2101 Wilson Boulvard
Suite 800
Arlington, Virginia 22201

Contract No. 68-02-4294

Susan Dillman, Co-Task Manager
Jay Glatz, Co-Task Manager
Mary Frankenberry, Project Officer

Design and Development Branch
Exposure Evaluation Division
Office of Toxic Substances
Office of Pesticides and Toxic Substances
U.S. Environmental Protection Agency
Washington, D.C. 20460

OTS DISCLAIMER

This report was prepared under contract to an agency of the United States Government. Neither the United States Government nor any of its employees, contractors, subcontractors, or their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third part's use of or the results of such use of any information, apparatus, product, or process disclosed in this report, or represents that its use by such third party would not infringe on privately owned rights.

Publication of the data in this document does not signify that the contents necessarily reflect the joint or separate views and policies of each sponsoring agency. Mention of trade names or commercials products does not constitute endorsement or recommendation for use.

BATTELLE DISCLAIMER

This is a report of research performed for the United States Government by Battelle. Because of the uncertainties inherent in experimental or research work, Battelle assumes no responsibility or liability for any consequences of use, misuse, inability to use, or reliance upon the information contained herein, beyond any express obligations embodied in the governing written agreement between Battelle and the United States Government.

TABLE OF CONTENTS

TABLE OF CONTENTS..... iii

EXECUTIVE SUMMARY..... v

1.0 INTRODUCTION..... 1

2.0 CONCLUSIONS..... 5

3.0 INSTRUMENT RESPONSE MODEL AND ESTIMATION METHOD..... 10
3.1 The Calibration Step..... 13
3.2 Estimating Concentrations of Target Compounds.. 15
3.3 Quality Control Samples..... 18

4.0 THE QC PROGRAM..... 22
4.1 Description..... 22
4.1.1 Routine Calibration Check..... 25
4.1.2 Recovery..... 26
4.1.3 Post Analysis Review of QC Data..... 27
4.2 Discussion of QC Program Elements..... 30
4.2.1 Real Time QC..... 31
4.2.1.1 Routine Calibration Check Sample (RCC) 31
4.2.1.2 Recovery..... 33
4.2.2 Post Analysis Summary of Data Quality... 34

5.0 SIMULATION ANALYSIS..... 36
5.1 Introduction..... 36
5.2 Description of Analysis and Parameter Values... 38
5.3 Results..... 45

REFERENCES..... 52

TABLE OF CONTENTS (continued)

LIST OF TABLES

Table E-1 Probabilities of Detecting Calibration and Recovery Shifts..... viii

Table E-2 TCDD Simulation Model Parameter Values: "In Control" Case..... ix

Table 4-1 QC Procedures and Criteria for Analysis of Human Adipose Tissue Samples for PCDDs and PCDFs..... 23

Table 5-1 TCDD Simulation Model Parameter Values: "In Control" Case..... 41

Table 5-2 TCDD Simulation Model Parameter Values: RRF Shift..... 43

Table 5-3 TCDD Simulation Model Parameter Values: Recovery Shift..... 44

Table 5-4 Probabilities of Detecting Calibration and Recovery Shifts..... 46

LIST OF FIGURES

Figure 5-1 QC Tests, Decisions, and Corrective Actions..... 40

EXECUTIVE SUMMARY

The U.S. Environmental Protection Agency (EPA) analyzes adipose tissue samples collected in the National Human Adipose Tissue Survey (NHATS) to estimate and monitor exposure to environmentally persistent toxic compounds. In 1982 the program was expanded to include analysis methods for determining concentrations of polychlorinated dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs). The quality assurance project plan (QAPP) for the analysis of the 1987 samples specifies various types of quality control (QC) activities intended to assure and document data quality. Since QC costs can be significant, a study was initiated to investigate the effectiveness and costs of the current QC program as well as other QC programs that may be considered for the analysis of NHATS samples in future years.

To advance the investigation, a computer simulation model of the laboratory process for adipose tissue analysis has been developed. The model simulates events in their sequence of occurrence in the laboratory. It distinguishes batches, days required to complete a batch, and the following QC activities: initial calibration; routine calibration check at the beginning and at the end of each day; a test of absolute recovery of the internal quantitation standard (IQS) in every sample; and a

test of method recovery (MR) following the completion of each batch.

The simulation model may be applied to analyze effectiveness and costs for a variety of QC programs and data quality objectives including the QC program used for the 1987 samples. To demonstrate the analysis approach, the model has been used to address three questions concerning the current QC program for NHATS samples. A more comprehensive analysis, based on the simulation model, will be necessary to thoroughly evaluate costs and efficiency for NHATS QC program alternatives.

The three questions are:

1. What are the false positive error rates associated with the routine calibration check, the test of absolute recovery of the internal quantitation standard, and the method recovery test?

2. What are the probabilities that these three components of the QC program will detect a change in the calibration coefficient (usually referred to as the relative response factor - RRF)?

3. What are the probabilities that these three components of the QC program will detect a degradation in method recovery?

Simulation results have been developed that provide answers to the three questions. Results for TCDD are summarized in Table E-1 and discussed below. (TCDD is used as an example throughout the study wherever specificity enhances the presentation.) Briefly, the analysis indicates that:

- the routine calibration test is effective;
- the absolute IQS recovery test as currently formulated may cause positive bias in concentration estimates; and
- the relationship between tests of IQS recovery and method recovery needs better coordination with data quality objectives.

When the analytical process is "in control," (the first column of Table E-1, based on the parameter values in Table E-2), the probability of detecting a calibration failure is less than 0.01 (i.e., the false positive error rate is small). The column labeled "RRF Shift" in Table E-1 refers to an increase

Table E-1. Probabilities of Detecting Calibration and Recovery Shifts

QC Test	Status of Analytical Process		
	In Control	RFF Shift	Recovery Shift
	--------Probability of Detection--------		
Routine Calibration Check (RCC)[1]	<0.01	>0.99	<0.01
Internal Quantitation Standard (IQS)[2]	0.39	0.45	0.14
Method Recovery (MR)[3]	0.09	0.10	0.69

Notes:

1 - probability of detecting at least one RCC failure (i.e, two consecutive RCC sample failures) per batch

2 - probability of detecting at least one IQS failure per batch

3 - probability of detecting an MR failure per batch

Table E-2. TCDD Simulation Model Parameter Values: "In Control" Case

Analysis Operation	GC Resp. Parameter		Recovery		Standard Deviation		
	Intercept	Slope	TCDD	IS	Batch	Sample	Analysis
Calibration							
Std/IS	0.00	0.80	1.000	1.000	0.00	0.0000	0.0500
IS/RS	0.00	1.73	1.000	1.000	0.00	0.0000	0.0800
Field Samples							
U/IS	0.00	0.80	0.595	0.521	0.15	0.1250	0.1000
IS/RS	0.00	1.73	1.000	0.521	0.15	0.0525	0.1575
QC Samples							
Std/IS	0.00	0.80	0.595	0.521	0.15	0.0150	0.0450
IS/RS	0.00	1.73	1.000	0.521	0.15	0.0525	0.1575

Notes:

The symbols */* in the left hand column refer to ratios of areas that represent instrument responses. The numbers in each row are the parameter values used in the equation to generate an instrument response for the ratio indicated in the first column. For example, U/IS refers to the equation used to produce the ratio of areas corresponding to the concentration of TCDD in a primary sample and the concentration of the internal quantitation standard.

Std - a sample spiked with a known amount of TCDD
RS - recovery standard sample

in the RRF of 37.5 percent immediately following the initial calibration. The detection probability of the routine calibration test in this situation is greater than 0.99. The routine calibration test, therefore, is extremely effective for detecting a change of this magnitude.

When the analytical process is "in control" the internal quantitation standard (IQS) test has a probability of 0.39 for detecting failures, a large value for a false positive error rate. IQS absolute recovery test failures have two consequences. First, the batch must be reextracted and analyzed resulting in additional cost. Second, the test favors larger recovery values. Estimated concentrations of TCDD in primary tissue samples, therefore, will be biased toward larger values. These findings suggest that the IQS recovery test, which currently requires IQS recovery to be in the fixed interval between 0.40 and 1.50, should be based on a statistical interval with boundaries determined from the mean and standard deviation of the recovery estimate.

The column labeled "Recovery Shift" reflects a change from the "in control" case in both IQS absolute recovery and method recovery. IQS recovery has been increased from 0.521 for the "in control" case to 0.600 and method recovery from 1.142 to 1.500. The IQS detection probability drops from 0.39 to 0.14

because the true IQS recovery value is closer to the center of the allowable range than it was in the other two cases. The detection probability for the method recovery test has increased from 0.09 to 0.69 because the hypothetical true MR value of 1.500 is also the upper boundary of the MR test interval. (Note that the range of values defining the IQS test and the MR test are the same.) The apparent inconsistency between the detection probabilities of the two recovery tests is, in part, a consequence of the fixed interval approach to defining the recovery tests. These tests should reflect DQO's associated with applications of the data and, as indicated above, should be based on statistical characteristics of the recovery estimates.

The results discussed above reflect two sets of assumptions. The first assumptions, which are implicit in the QC program, form the basis for detecting and correcting recovery problems. These assumptions are:

1. when absolute recovery of the target compound in a primary sample declines as a result of sample processing, absolute recovery of the internal quantitation standard in the same sample also declines; and

2. when absolute recovery of the target compound in a primary sample declines as a result of sample processing, absolute recovery of that target compound in a QC sample also declines.

Under these assumptions: (i) absolute recovery of the internal quantitation standard acts as a recovery adjustment applied to estimates of target compound concentrations; and (ii) method recovery computed from QC samples is representative of method recovery in primary tissue samples. Neither assumption is easily verified and if either assumption were violated, portions of the QC program may be ineffective. It is notable that under these assumptions a change in the value of absolute recovery of the internal quantitation standard does not signal a change in method recovery.

The second set of assumptions concerns parameter values selected for the simulation that characterize recovery, variability, and the calibration curve of the analytical method. Most of the values used for the cases represented in Table E-1 were derived from data generated in a method validation study (USEPA, 1986). A few of the values, which could not be derived from the method validation data were based purely on judgement. Additional values for the parameters, determined either subjectively or from more recent data, are

needed to conduct a sensitivity analysis of the results in Table E-1 and any subsequent findings regarding alternative QC proposals.

The results presented in this report serve as one example of the type of analysis that can be conducted with the simulation model. Additional analyses using the model are needed to evaluate alternative QC programs and alternative data quality objectives. First, however, QC alternatives must be refined to ensure they are practical with respect to laboratory operating constraints and additional PCDD (PCDF) measurement data, if available, need to be analyzed to improve, if possible, estimates of model parameters.

1.0 INTRODUCTION

The U.S. Environmental Protection Agency (EPA) analyzes adipose tissue samples collected in the National Human Adipose Tissue Survey (NHATS) to estimate and monitor exposure to environmentally persistent toxic compounds. NHATS is a statistically designed program intended to represent the general U.S. adult population. In 1982 the program was expanded to include analysis methods for determining concentrations of polychlorinated dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs). A detailed quality assurance project plan (QAPP) was developed for the analysis of the 1987 samples (MRI, 1988). The QAPP specifies various types of quality control (QC) activities. These activities, which include analysis of QC samples, are intended to assure data quality and provide information that documents data quality. Since QC activities can add significant cost to an analytical program, the effectiveness and cost of alternative QC programs for the NHATS analysis is being investigated. To further that investigation, a model has been developed to analyze the effectiveness and cost of alternative QC programs. This report describes the model and its application for analyzing the QC program specified in the QAPP for PCDD's (PCDF's).

The analysis of tissue samples collected in 1987 involved 5 batches, each batch containing 12-15 composite samples. The analytical method is high resolution gas chromatography/high

resolution mass spectrometry (HRGC/HRMS). The QAPP specifies analysis of QC samples including calibration checks, controls, and spikes. The purpose of QC samples, in general, is to monitor and document the quality of data being produced. Since the unit cost of a QC sample analysis is equal to the unit cost for analyzing a primary sample, efficient allocation of QC resources in terms of types and numbers of samples is essential. Each QC sample must contribute in a measurable way to the quality of data used for estimating PCDD (PCDF) concentrations in primary samples.

To evaluate alternative QC plans from a cost-effectiveness perspective, quantitative data quality objectives (DQO's) are needed. DQO's should be associated with particular applications of the primary data. The types and numbers of QC samples necessary to achieve a DQO, then, can be assessed.

For example, the PCDD (PCDF) estimates obtained by analyzing human adipose samples may be used:

(i) to determine if the 1987 levels of PCDD (PCDF) are above or below an established standard (e.g., a standard based on health considerations); or

(ii) to determine if there is a trend in PCDD (PCDF) levels (e.g., compare 1987 results with past results).

The DQO in both of these examples may be specified as a pair of values for the Type I and Type II statistical error rates (e.g., a Type I error rate of 0.15 and a Type II error rate of 0.20) associated with statistical tests of the implied hypotheses . This approach to data quality, which focuses on error rates associated with statistical decisions based on monitoring data, is consistent with recent guidance on the development of DQO's prepared by the EPA/ORD Quality Assurance Management Staff.

The magnitudes of statistical test error rates are affected by recovery of the analytical measurement method, variability of the method, and replication. Information about recovery and variability is obtained through analyses of QC samples. An analytical response model, introduced in Section 3, is used to describe the concentration estimates produced by the HRGC/HRMS method. The model includes explicit parameters that characterize calibration, recovery, and variability. The analysis of QC program effectiveness is based on the model and these parameters. Measurement characteristics of TCDD, one of

the PCDD congeners, are used throughout this report as a specific example to enhance exposition.

The remainder of this report is presented in four sections. Conclusions are summarized in Section 2. Section 3 contains a description of the analytical response model and a discussion of the method employed for estimating concentrations of target compounds in adipose tissue. The current QC program is described in Section 4 and results of computer simulation analysis are presented in Section 5.

2.0 CONCLUSIONS

A computer simulation model of the laboratory process for adipose tissue analysis of NHATS samples has been developed. The model simulates events in their sequence of occurrence in the laboratory. It distinguishes batches, days required to complete a batch, and the following QC activities: initial calibration; routine calibration check at the beginning and at the end of each day; a test of absolute recovery of the internal quantitation standard (IQS) in every sample; and a test of method recovery (MR) following the completion of each batch. With parameter values selected to represent a particular set of laboratory characteristics, the model may be used to evaluate the effectiveness of alternative QC programs for detecting laboratory circumstances that are considered "out of control."

The model also provides information for comparing costs of QC programs. Costs depend on the total number of samples, including both primary and QC samples, that must be analyzed to complete a particular analytical program. The number of samples that must be analyzed may be greater than the minimum number specified in an analytical program plan for two reasons. First, false positive QC test results may require calibration to be repeated or primary samples to be reanalyzed. Second, an "out of control" situation that is not immediately detected by

QC tests could necessitate the reanalysis of many primary samples. Costs of alternative QC programs, therefore, can be compared by comparing the total number of sample analyses that must be conducted to complete the analytical program and achieve specified data quality objectives.

At present, the simulation model has been used to address three questions concerning the current QC program.

1. What are the false positive error rates associated with the routine calibration check, the test of absolute recovery of the internal quantitation standard, and the method recovery test?

2. What are the probabilities that these three components of the QC program will detect a change in the calibration coefficient (usually referred to as the relative response factor - RRF)?

3. What are the probabilities that these three components of the QC program will detect a degradation in method recovery?

One set of results has been developed that provides answers to the three questions. Briefly, the analysis indicates that:

- the routine calibration test is effective;
- the absolute IQS recovery test as currently formulated may cause positive bias in concentration estimates; and
- the relationship between tests of IQS recovery and method recovery needs better coordination with data quality objectives.

A discussion of each result follows.

The RCC test is extremely effective. The false positive error rate is less than 0.01. The probability of detecting changes of 40 percent or larger in the RRF is greater than 0.99.

The false positive error rate for the IQS absolute recovery QC test is approximately 0.40 when IQS recovery is slightly greater than 0.5 and method recovery is approximately 1.1. IQS recovery test failures have two consequences. First, when a failure is detected, the batch must be reextracted and analyzed resulting in additional cost. Second, since method recovery and absolute recovery are correlated, method recovery in the batches that pass the test will reflect the characteristics of

the IQS samples that pass. This test favors larger recovery values. Estimated concentrations of target analyte in primary tissue samples, therefore, will be biased toward larger values. These findings suggest that the IQS recovery test, which currently requires IQS recovery to be in the fixed interval between 0.40 and 1.50, should be based on a statistical interval with boundaries determined from the mean and standard deviation of the recovery estimate.

When IQS recovery is larger (e.g., 0.6) and method recovery is 1.5, the IQS detection probability drops to 0.14 because the IQS recovery value is closer to the center of the allowable range defining the test. The detection probability for the method recovery test in this case is 0.69 because the hypothetical MR value of 1.5 is also the upper boundary of the test interval. The apparent inconsistency between the detection probabilities of the two recovery tests is, in part, a consequence of the fixed interval approach to defining the recovery tests. These tests should reflect DQO's associated with applications of the data and, as indicated above, should be based on statistical characteristics of the recovery estimates.

The results presented in this report serve as one example of the type of analysis that can be conducted with the simulation

model. Additional analyses using the model are needed to evaluate alternative QC programs and alternative data quality objectives. First, however, QC alternatives must be refined to ensure they are practical with respect to laboratory operating constraints and additional PCDD (PCDF) measurement data, if available, need to be analyzed to improve, if possible, estimates of model parameters.

3.0 INSTRUMENT RESPONSE MODEL AND ESTIMATION METHOD

The procedure for estimating PCDD (PCDF) concentrations in human adipose tissue involves the analysis of calibration samples, quality control samples, and primary human adipose tissue samples. Each sample is fortified (spiked) with two internal standards prior to analysis: an internal quantitation standard; and a recovery standard. The internal quantitation standards are chemically almost identical to the target compounds. (For TCDD, the internal quantitation standard is a $^{13}C_{12}$ labeled version of the same analyte - $^{13}C_{12}$-2,3,7,8-TCDD. The recovery standard is $^{13}C_{12}$-1,2,3,4-TCDD.)

The analytical instrument responses associated with these samples are areas that are proportional to the analyte concentrations. The areas are summarized as two ratios: (i) the target compound area divided by the internal quantitation standard area; and (ii) the internal quantitation standard area divided by the recovery standard area. These ratios may be described mathematically as:

$$A/A' = K_0 + K_1*(C/C') + Z\sigma[K_0 + K_1*(C/C')] \qquad \text{(Equation 1)}$$

where

A is the instrument response (an area) to concentration C ;

A' is the instrument response to concentration C';

K_0 and K_1 are the intercept and slope respectively of the straight line relationship;

σ is the coefficient of variation of the response ratio; and

Z is a random deviate with distribution N(0,1).

The term, $Z\sigma[K_0 + K_1*(C/C')]$, represents a random error contribution to the instrument response which, in general, consists of three components. These are: (i) a batch component; (ii) a sample component; and (iii) an analytical replication component. The error term takes the general form:

$$\text{random error} = (Z_B\sigma_B + Z_s\sigma_s + Z_r\sigma_r)*[K_0 + K_1*(C/C')] \qquad \text{(Equation 2)}$$

where

σ_B represents batch variability;

σ_s represents sample variability;

σ_r represents analytical replication variability; and Z_B, Z_s, Z_r are standard normal variates.

There also is an instrument response relationship similar to Equations 1 and 2 for the ratio of the internal quantitation standard to the recovery standard. These response equations are discussed in more detail later.

Estimates of PCDD (PCDF) concentrations are obtained as follows:

1. Establish an instrument calibration curve (i.e., obtain estimates of K_0 and K_1).

2. Spike an adipose tissue sample with a known concentration of the internal quantitation standard;

3. Prepare (extract) the sample;

4. Add a known concentration of the recovery standard to the extract;

5. Analyze the sample and compute estimates.

Specification of these steps and subsequent details regarding the analytical method are based on the presentation in MRI, 1988.

3.1 THE CALIBRATION STEP

Calibration solutions are prepared with known concentrations of the target compound, the internal quantitation standard, and the recovery standard. Eight samples, each with different concentrations of the target compound, are used. Concentrations of the two internal standards are constant across the calibration samples. The calibration relationship for the target compound is:

$$(A_{STD}(i)/A_{IS}) = K_0 + K_1*(C_{STD}(i)/C_{IS}) + (Z_{si}\sigma_s + Z_{ri}\sigma_r)*[K_0 + K_1*(C_{STD}(i)/C_{IS})]$$

(Equation 3)

where

$A_{STD}(i)$ is the instrument response (area) corresponding to concentration $C_{STD}(i)$ of the target compound;

A_{IS} is the instrument response corresponding to concentration C_{IS} of the internal quantitation standard;

K_0, K_1 calibration parameters to be estimated; and

σ's, Z's previously defined in Equation 2.

σ_B does not enter Equation 3 since the calibration step does not involve batches.

The calibration relationship for the internal quantitation standard relative to the recovery standard is:

$$(A_{IS}(i)/A_{RS}) = L_0 + L_1*(C_{IS}(i)/C_{RS}) + (Z_{si}\sigma_s + Z_{ri}\sigma_r)*[L_0 + L_1*(C_{IS}(i)/C_{RS})]$$

(Equation 4)

C_{RS} and A_{RS} are the concentration and instrument response respectively for the recovery standard. Further discussion of this relationship is found with the discussion of recovery in Section 3.2.

Estimated values of K_0 and K_1 in Equation 3 are used to calculate target compound concentrations from analyses of primary samples. Typically, K_0 is assumed to be zero (i.e., the calibration curve goes through the origin) and K_1 is referred to as the relative response factor (RRF). K_1 may be estimated by the method of least squares. The approach used in

MRI, 1988 is a slight variation of the standard least squares approach. An RRF is calculated for each calibration sample. That is:

$$RRF(i) = (A_{STD}(i)/A_{IS}) \div (C_{STD}(i)/C_{IS}) \qquad \text{(Equation 5)}$$

If the relative standard deviation (RSD) of the RRF's is less than 0.20 (or 0.30 depending on the target compound), i.e., if

$$(\{\Sigma[RRF_{(i)} - ave\{RRF(i)\}]^2/(n-1)\}^{\frac{1}{2}} \div ave\{RRF(i)\}) \leq .20,$$

then RRF is set equal to ave{RRF(i)}. This value of RRF is, in fact, the weighted least squares estimate of K_1 when K_0 is zero and the weights are $[C_{STD}(i)/C_{IS}]^{-1}$. Throughout the ensuing discussion, the operating calibration relationship will be

$$(A/A') = RRF*(C/C') \qquad \text{(Equation 6)}$$

based on RRF as defined above.

3.2 ESTIMATING CONCENTRATIONS OF TARGET COMPOUNDS

The procedure for estimating concentrations of the target compounds in tissue samples involves steps 2, 3, 4, and 5 listed in Section 3.0. Denote the unknown concentration of the target compound and the concentration of the internal

quantitation standard by C_U and C_{IS} respectively. The internal quantitation standard is added to the sample before extraction. Denote by $\beta_1 C_U$ and $\beta_2 C_{IS}$ the concentration of these two compounds in the final extract. β_1 and β_2 represent recovery proportions. Both values are expected to be between zero and one. The estimate of the unknown concentration is:

$$\text{est}(C_U) = (A_U/A_{IS})*(C_{IS}/RRF) \qquad \text{(Equation 7)}$$

where A_U is the instrument response for the target compound of unknown concentration and RRF is the relative response factor determined in the calibration step. This estimate is potentially biased for a number of reasons, but most specifically because A_U and A_{IS} are instrument responses to concentrations of $\beta_1 C_U$ and $\beta_2 C_{IS}$ respectively rather than C_U and C_{IS}. Using Equation 6 to substitute for A_U/A_{IS} in Equation 7 demonstrates that

$$\text{est}(C_U) \doteq (\beta_1/\beta_2)\cdot C_U. \qquad \text{(Equation 8)}$$

The factor, (β_1/β_2), is method recovery (MR). β_1 represents recovery of the target compound. β_2 represents recovery of the internal quantitation standard. If $\beta_2 = \beta_1$, then MR = 1 and the estimate of C_U would be unbiased. The estimation method of Equation 7, therefore, implicitly utilizes the analytical

response for the internal quantitation standard as a recovery adjustment for estimating C_U.

β_2 may be estimated from data generated for the internal quantitation standard and the recovery standard. Equation 4 represents the calibration relationship between instrument response and concentration for these compounds. When L_0 is zero, the weighted least squares estimate of L_1, which is the relative response factor for the internal quantitation standard (RRF_{IS}), is

$$RRF_{IS} = ave\{A_{IS}/A_{RS}) \div (C_{IS}/C_{RS})\} \qquad \text{(Equation 9)}$$

with the average taken over all calibration samples. The operating calibration relationship for the internal quantitation standard relative to the recovery standard, therefore, is

$$(A_{IS}/A_{RS}) = RRF_{IS} * (C/C_{RS}). \qquad \text{(Equation 10)}$$

An estimate of β_2 is

$$est(\beta_2) = (A_{IS}/A_{RS}) \cdot (C_{RS}/C_{IS}) \div RRF_{IS}. \qquad \text{(Equation 11)}$$

Substituting the right hand side of Equation 10 with C replaced by $\beta_2 C_{IS}$, which is the concentration of the internal

quantitation standard in the extract, into Equation 11 confirms the estimating formula for β_2,

$$\text{est}(\beta_2) \doteq RRF_{IS}(\beta_2 C_{IS}/C_{RS}) \cdot (C_{RS}/C_{IS}) \div RRF_{IS} = \beta_2 \quad \text{(Equation 12)}$$

Note that the value of β_2 alone does not constitute sufficient information to assess the degree of bias in estimates of C_U, the unknown concentration of the target compound. Assuming, however, that $\beta_1 = \beta_2$ implies that estimates of C_U are unbiased. Information about β_1 and method recovery (i.e., MR = β_1/β_2) may be obtained from QC samples as discussed below.

3.3 QUALITY CONTROL SAMPLES

For purposes of this discussion, the term QC samples refers to samples of adipose tissue that originally contain, at most, the background concentration, C_0, of the target compounds. (Other types of samples are used for QC purposes also. These samples - calibration samples and routine calibration check samples - are discussed in Section 4.1.) These QC samples, then, are spiked with known concentrations of the target compounds. Unspiked samples (i.e., a spiking concentration of zero) also are included in this definition. The unspiked samples of adipose tissue are referred to as "controls."

3.3.1 Method Recovery

Spiked QC samples may be used to estimate analytical method recovery. A QC sample is analyzed by following the same procedure used for primary adipose tissue samples. The internal quantitation standard is added to the spiked sample prior to extraction and the recovery standard is added to the final extract prior to analysis. The concentration of the target compound in a QC sample will be C_{STD}, the spiking concentration, plus C_0, the background concentration. Denoting the concentrations of the target compound and the internal quantitation standard in the final extract by $\beta'_1 \cdot (C_{STD}+C_0)$ and $\beta'_2 \cdot C_{IS}$ respectively, and applying the estimation method embodied in Equation 7, yields

$$\begin{aligned} est(C_{STD}) &= est(C_{STD}+C_0) - est(C_0) \\ &= (\beta'_1/\beta'_2) \cdot C_{STD}. \end{aligned} \qquad \text{(Equation 13)}$$

Since C_{STD} is a known quantity, an estimate of the method recovery factor is:

$$est(MR) = est(C_{STD})/C_{STD}. \qquad \text{(Equation 14)}$$

Estimates of the concentration of the target compounds in primary tissue samples may be adjusted in an attempt to remove recovery bias. That is,

$$adj(C_U) = est(C_U)/(est(MR). \qquad \text{(Equation 15)}$$

Using the results of Equations 8 and 13 in Equation 15 yields

$$adj(C_U) \doteq [(\beta_1/\beta'_1)\div(\beta_2/\beta'_2)]\cdot C_U \qquad \text{(Equation 16)}$$

The adjusted estimate is unbiased if the recovery proportions of the target compound in primary samples and QC samples are equal (i.e., $\beta'_1 = \beta_1$) and the recovery proportions of the internal quantitation standard in primary samples and QC samples are equal (i.e., $\beta'_2 = \beta_2$), or if method recovery is the same in both types of samples (i.e., $\beta'_1/\beta'_2 = \beta_1/\beta_2$).

The potential differences among values of the β's result from differences in the effects of sample processing (i.e., extraction) on target compounds recently spiked into samples of adipose tissue and target compounds that were part of the sample at the time it was taken from its donor. In fact, the values of β'_1, β_2, and β'_2 may be closer to each other than to the value of β_1. Determining if β_1 and β_2 have different values than β'_1 and β'_2 cannot be resolved without extensive, complex experimentation. Even if $\beta_2 = \beta'_2$, which is likely and can be determined from QC data, the possibility that β'_1 and β_1

may differ has implications for the allocation of QC resources and the use of QC analysis results.

4.0 THE QC PROGRAM

4.1 DESCRIPTION

The 1987 NHATS samples were analyzed in five batches consisting of 12 to 15 samples per batch. For purposes of the analysis undertaken in this report, there will be five batches each consisting of 12 primary samples plus QC samples. A batch requires two days to complete - six primary samples per day plus QC samples.

The complete set of QC procedures and criteria for taking corrective action are summarized in Table 4-15 of the QAPP (MRI, 1988) which is reproduced in this report for reference purposes as Table 4-1. The current analysis of QC effectiveness focuses on a subset of the QC procedures described in the table. These are: (i) the routine calibration check; (ii) absolute recovery of the internal quantitation standard in primary samples; (iii) absolute recovery of the internal quantitation standard in QC samples; (iv) method recovery determined from QC samples; and (v) post analysis review of all QC data to evaluate method precision, constancy of recovery across batches, and constancy of recovery with respect to concentration level of the target compound. Recall that the term "QC sample" in this report is used to describe: unspiked samples of human adipose tissue with naturally occurring background levels of PCDD (PCDF) which are

Table 4-1

Table 4-15 QC Procedures and Criteria for Analysis of Human Adipose Tissue Samples for PCDDs and PCDFs[a]

Analysis event	Frequency	QC criteria	Corrective actions	Responsibility
Instrument mass calibration	Daily real time interpretation	Must demonstrate accurate mass calibration using perfluorokerosene (PFK). First activity of analysis day	Recalibration If criteria not achieved, do not proceed with analysis	MS analyst
• PCDD/PCDF analysis	First event of analysis day	Using PFK tune to a minimum resolution of 10,000 (10% valley) and optimal response and peak shape m/z 381 Adjust magnetic field to pass m/z 300 at accelerating voltage Introduce PFK through direct inlet and acquire accelerating voltage scans from 8000 to 4000 V using Incos Data System lockmass (m/z 381) identified in PFK spectrum used to update. Mass calibration ranges from 301 to 593 amu.	Refer to tuning and mass calibration procedure If criteria cannot be achieved, instrument may require maintenance	MS analyst
Column performance check	Daily real time interpretation	Must demonstrate isomer specificity for 2,3,7,8 TCDD before proceeding with analysis of calibration standard • 60-m DB-5 column, 25% resolution	Adjust column length (~0.1 m) and reanalyze performance mixture If necessary, install a new HRGC column and evaluate performance	MS analyst
Calibration standards				
• Initial calibration	Precedes initial sample analysis if routine calibration does not meet existing calibration criteria.	Analysis of six concentration calibration standards. % RSD of RRF for mean RRF for all standards ±30% for PCDD/PCDF, ±20% for TCDD/TCDF.	Prepare fresh concentration calibration standard.	MS analyst Work assignment leader
• Routine calibration	Precedes sample analysis on daily basis. Also must demonstrate calibration as last injection of each analyses day.	Measured RRF values for solution CS-7 must be within ±30% for PCDD/PCDF and ±20% for TCDD/TCDF.	Reanalyze solution CS-7 or repeat the initial calibration sequence. If calibration criteria are not met at the end of the day all samples are subject to reanalysis by HRGC/HRMS.	MS analyst Work assignment leader
Tridecane blank	Precedes sample analysis following calibration standards real time daily	Blank run should not demonstrate positive responses (< 2.5 times S/N) for the PCDDs/PCDFs	Clean injection syringe, repeat blank analysis If carryover still evident, replace syringe and/or injection liner.	MS analyst
Samples/QC samples				
• Analysis	As submitted in sample batch real time	Document respons. of internal recovery standard(s) and compare to daily calibration standard Internal recovery standard responses must be within 50% of response noted for calibration standard used to verify RRF values. Samples submitted as blinds to MS analyst.	If internal recovery standard noted to be 50% of calibration standard, reanalyze and/or check mass calibration	MS analyst
• Performance evaluation samples	Real time	Check solutions provided by QCC for measurement of accuracy, 70-130% Do not proceed with sample analysis until notified of acceptable performance by the QCC.	Reanalyze solution and/or calibration standards(s). If necessary, establish new calibration curve	QCC/QAC MS analyst Work assignment leader

Table 4-1 continued

Table 4 15 (continued)

Analysis event	Frequency	QC criteria	Corrective actions	Responsibility
• Data interpretation	Following analysis of batch sample			
• Internal (QC samples)		Accuracy (method recovery of native compounds) should be within 40 150% of spike level Recovery of internal quantitation standards should be within 40 150%	Reanalyze solution and/or calibration standards If criteria cannot be achieved on reanalysis, sample batch is subject to reextraction and analysis	Work assignment leader MS analyst QCC/QAC
• Tissue (NHAIS) samples		Recovery of internal quantitation standards should be within 40 150%.	Reanalyze solution and/or calibration standards	Work assignment leader MS analyst QCC/QAC

[a]QQOs for the analysis of PCDD and PCDF are targeted to meet similar criteria as defined for the PCDDs and PCDFs

called controls; and two other samples of adipose tissue that have been spiked with PCDD (PCDF), one at a low concentration level and the other at a high concentration level. (For TCDD, the spiking concentrations were 0 for the control, 10 pg/g, and 50 pg/g.) These three QC samples were analyzed, unidentified to the analyst and interspersed randomly among the primary samples, with each batch (Heath, 1988).

4.1.1 Routine Calibration Check

A routine calibration check (RCC) is the first and last sample analyzed each day. RCC is identical to a calibration sample with the target compound concentration at one of the lowest calibration concentrations (2.5 pg/μl for TCDD). An estimate of the relative response factors for the target compound and for the internal quantitation standard (i.e., RRF and RRF_{IS}) are determined for this sample and compared to the current operating values of these response factors (i.e., the RRF's determined from the most recent calibration). At the beginning of the day, if either RRF or RRF_{IS} differs from the current operating calibration values by ±20% for TCDD (±30% for other compounds), a new RCC sample must be prepared and analyzed before proceeding. If the second RCC fails, the initial calibration must be repeated. If the RCC at the end of the day fails twice, the initial calibration must be repeated and all

primary samples analyzed that day are subject to reanalysis (MRI, 1988; Table 4-1).

4.1.2 Recovery

Absolute recovery of the internal quantitation standard is monitored in every primary sample. This recovery factor, denoted as β_2 previously in this report, is calculated using the formula:

$$est(\beta_2) = (A_{IS}/A_{RS}) \cdot (C_{RS}/C_{IS}) \div RRF_{IS} \qquad \text{(Equation 17)}$$

The QAPP requires that $0.40 < est(\beta_2) < 1.50$. If this QC criterion is satisfied, the analysis program proceeds to the next sample. If the QC criterion is not satisfied, the sample is reanalyzed and a new estimate of β_2 is tested. If the QC criterion is not satisfied on this second attempt, the initial calibration must be repeated (i.e., new values of RRF and RRF_{IS} must be estimated).

Absolute recovery of the internal quantitation standard also is monitored in all QC samples. Method recovery is checked using the QC samples with nonzero spiking levels. The method recovery tests are conducted after all samples in a batch have been analyzed. Method recovery is determined from:

$$est(MR) = est(C_{STD})/C_{STD} \qquad \text{(Equation 18)}$$

where $est(C_{STD})$ is defined in Equation 13.

The QAPP specifies that both estimates must fall between 0.40 and 1.50 for the analysis program to continue. If the estimates of either β_2 or MR are outside the specified range, the QC sample must be reanalyzed and tested again. A second failure means the initial calibration must be repeated. If upon reanalysis the QC test fails, all primary samples in the batch are subject to reextraction and analysis.

4.1.3 Post Analysis Review of QC Data

Upon completion of the analytical program, data will have been collected on 15 QC samples, three samples from each of the five batches. The three samples from each batch are: a control sample of adipose tissue (i.e., unspiked); a low concentration spiked sample of adipose tissue (10 pg/g for TCDD); and a high concentration spiked sample of adipose tissue (50 pg/g for TCDD). These data may be used in a variety of ways to characterize the quality of data obtained from the primary samples. Following Heath, 1988, these data are used to: (i) obtain a working estimate of overall method recovery; (ii) test for method recovery differences among batches; (iii) test for method recovery dependence on concentration level; and (iv)

estimate method precision (i.e., estimate the total standard deviation of concentrations of target compounds measured in primary samples). These results are obtained using regression analysis applied to the estimated and true concentrations from QC samples.

In the basic regression model:

$$\text{est}(C_{STD}(i)) = \alpha_0 + \alpha_1 \cdot C_{STD}(i) + Z_i\sigma, \qquad \text{(Equation 18)}$$

the ordinary least squares estimate of α_1 is an approximation of method recovery assuming method recovery is independent of batch and concentration. Dependence of method recovery on batch is modeled by adding $\Sigma\tau_j \cdot B_{ji} \cdot C_{STD}(i)$ to the right side of Equation 18. B_{ji} is defined to be 1 if the ith QC sample came from the jth batch and 0 otherwise. With this addition to Equation 18, method recovery in the jth batch is equal to $\alpha_1 + \tau_j$. Dependence of method recovery on concentration is modeled by adding $\Sigma\delta_k \cdot C_{ki} \cdot C_{STD}(i)$ to the right side of Equation 18. C_{ki} is defined to be 1 if the ith QC sample has the kth spiking concentration and 0 otherwise. In this expanded equation, method recovery for a QC sample analyzed in batch j and having a true concentration equal to spiking level k would be $\alpha_1 + \tau_j + \delta_k$.

Based on the full model (i.e., Equation 18 with the terms described above added to the right side), method recovery would be independent of concentration if the δ's were all equal to 0. This can be tested as a statistical hypothesis by combining appropriate sums of squares from the expanded model and from a "restricted" model with all δ's set equal to zero to form an F ratio with 2 and 7 degrees of freedom (Chatterjee, 1977). Method recovery would be independent of batch if the τ's were all equal to 0. This hypothesis can be tested by combining the appropriate sums of squares and using an F ratio with 4 and 7 degrees of freedom.

The root mean square error calculated from the expanded model may be used as an estimate of overall method precision.

4.2 DISCUSSION OF QC PROGRAM ELEMENTS

The QC program described above consists two general components: (i) a component based on samples that may be evaluated in real time to identify and correct analytical problems when they occur; and (ii) a component that provides information about data quality only after all samples have been analyzed. The former component, as represented in the QAPP, is intended to control recovery throughout the analytical program. The latter component provides information on recovery, factors that affect recovery, and precision (Heath, 1988).

Ideally, each QC activity has a clearly defined purpose and a measurable contribution to data quality objectives. The purpose of each QC activity in the NHATS program is apparent, however the relationship of these activities to a DQO that reflects an application of the data has not been developed. The acceptance bounds in Table 4-1 for RRF's, absolute recovery, and method recovery, have not been identified with a DQO for a specific application of the data, but are treated in the QAPP as the DQO's. Therefore, the only measurable contribution to data quality that can be analyzed for this QC program is the contribution each QC activity makes to holding the RRF and recovery within the bounds specified. Whether those bounds are adequate for any specific application of the data is an open question.

4.2.1 Real Time QC

4.2.1.1 Routine Calibration Check Sample (RCC)

The RCC is employed at the beginning and end of each day to determine if the calibration relationship has changed. This is accomplished by comparing estimates of RRF and RRF_{IS} based on each RCC sample to the current operational RRF and RRF_{IS} derived previously from a calibration experiment involving eight samples at eight different concentrations. The effectiveness of this element of the QC program should be measured by the Type I and Type II statistical error rates for testing a hypothesis of no change in the RRF (i.e., the likelihood that the RCC test will indicate a significant shift in the calibration factor, RRF, when in fact no change has taken place; and the likelihood that no change will be detected when in fact a significant shift in the calibration factor has occurred).

Both types of errors have implications for the effectiveness of the QC program. Type I errors cause calibration experiments to be repeated unnecessarily which adds cost and time to the analysis program. Type II errors are at least as damaging. If a change in RRF to a larger value is undetected, concentration estimates of the target compound in primary samples will be inflated. Estimates of method recovery obtained from QC

samples will be larger than before the undetected shift occurred. If a change in RRF to a smaller value is undetected, concentrations in primary samples will be underestimated and method recovery estimates derived from QC samples also will be smaller. The Type I error rate may be reduced by expanding the allowable difference between the RRF value obtained from the RCC sample and the value of the operating RRF. The Type II error rate may be reduced by increasing the number of independent RCC samples used to compare RRF's. The number of samples required depends also on the magnitude of change in RRF that is important to detect. This quantity, which should be reflected in the DQO's for specific applications of the data, is not addressed in the QAPP.

The effectiveness of the QC program for detecting RRF shifts of approximately 40 percent has been quantified using simulation analysis. These results are presented in Section 5. Note that RRF shifts, as indicated above, may be detected not only through RCC samples, but also through QC sample estimates of method recovery. In fact, when the method recovery test fails, the first corrective action recommended is to repeat the initial calibration (Table 4-1).

4.2.1.2 Recovery

The NHATS QC program includes two forms of recovery: absolute recovery of the internal quantitation standard, which can be estimated from each primary and QC sample; and method recovery, which can be estimated only from QC samples. Absolute recovery (β_2 in the discussion found in Section 3.3) is checked in every sample to test for a shift in its value. A test for a change in method recovery is conducted with every QC sample that has a non-zero spiking level (i.e., two samples per batch, 10 samples overall). The method recovery tests are conducted at the completion of a batch. A failure that is not corrected by re-calibration requires each sample in the batch to be reextracted and analyzed (Table 4-1).

The effectiveness of both tests for uncovering calibration or recovery problems depends, in part, on the following assumptions:

1. when absolute recovery, β_1, of the target compound in a primary sample declines as a result of sample processing, absolute recovery, β_2, of the internal quantitation standard in the same sample also declines; and

2. when absolute recovery, β_1, of the target compound in a primary sample declines as a result of sample processing, absolute recovery, β'_1, of that target compound in a QC sample also declines.

(The mathematical notation employed above is the same notation defined and used in Section 3.3.)

The QC program is predicated on both assumptions. Under these assumptions: (i) absolute recovery of the internal quantitation standard, β_2, acts as an implicit recovery adjustment applied to estimates of target compound concentrations; and (iii) method recovery computed from QC samples is representative of method recovery in primary tissue samples. If either assumption were violated, portions of the QC program would produce misleading results. It is notable, however, that under these assumptions a change in the value of β_2 does not signal a change in method recovery.

4.2.2 Post Analysis Summary of Data Quality

The statistical analysis described in Section 4.1.3 of all QC samples following completion of the analytical program leads to: (i) a working estimate of overall method recovery; (ii) a test for method recovery differences among batches; (iii) a test for method recovery dependence on concentration level; and

(iv) an estimate of method precision (i.e., an estimate of the total standard deviation of concentrations of target compounds measured in primary samples). The validity of these results depends on the same assumptions underlying the validity of the real time applications of QC program elements.

Since these statistical results are not available until after all samples have been analyzed, if a problem is indicated there is little opportunity to correct the analytical process and reanalyze samples. A problem, such as a shift in recovery across batches that was not identified and corrected in real time, becomes a permanent characteristic of the data. This and other characteristics may, however, be used effectively in applications of the data. In the example cited (i.e., a shift in method recovery across batches) an estimate of method recovery for each batch could be used to adjust estimated concentrations from each primary sample prior to comparing these data to similar data from other years. Summary statistics derived from QC samples, therefore, should be retained as statements of data quality with the data from individual samples. The adequacy of the current QC program to supply this data quality information can be analyzed using the computer simulation approach described in Section 5.

5.0 SIMULATION ANALYSIS

5.1 INTRODUCTION

A computer program has been developed that simulates the laboratory process described in MRI, 1988 for analyzing the 1987 NHATS adipose tissue samples. (This computer program is referred to as the simulation model or simply as "the model" in the ensuing discussion.) The model simulates events in their sequence of occurrence in the laboratory. The model distinguishes batches, days required to complete a batch, and the following QC activities: initial calibration; routine calibration check at the beginning and at the end of each day; test of absolute recovery of the internal quantitation standard in every sample; and test of method recovery following the completion of each batch.

Laboratory measurements are generated using Equations 1 and 2 of Section 3. These equations represent the analytical instrument response, a ratio of areas, which is translated into an estimate of concentration using Equation 8 (Section 3.2). Different choices of parameter values in Equations 1 and 2 are used to generate measurements for the different types of samples addressed in the model (i.e., calibration samples, calibration check samples, primary adipose tissue samples; and QC samples).

After the parameter values are selected to represent a particular set of laboratory characteristics, the model is used to evaluate the effectiveness of alternative QC programs for detecting laboratory circumstances that are considered "out of control." For example, the parameters in the model may be adjusted to incorporate a systematic shift over time in the RRF value. Simulation results, then, would lead to an estimate of the probability of detecting the shift. The magnitude of the shift could be varied to estimate the relationship between the detection probability and the magnitude of change in the RRF. A similar analysis could be conducted for recovery. The model also may be run for a laboratory that is "in control" (i.e., no systematic shifts in the values of parameters that affect data quality) to estimate false positive rates associated with the QC activities. In general, the model can be used to evaluate the effectiveness of any QC program that has well-defined QC activities, QC test decision rules, and corrective actions.

In addition to QC effectiveness, the model provides information for comparing costs of QC programs. Costs depend on the total number of samples, including both primary and QC samples, that must be analyzed to complete a particular analytical program. The number of samples that must be analyzed may be greater than the minimum number specified in an analytical program plan for two reasons. First, false positive QC test results may require

calibration to be repeated or primary samples to be reanalyzed. Second, an "out of control" situation that is not immediately detected by QC tests could necessitate the reanalysis of many primary samples. Costs of alternative QC programs, therefore, can be compared by comparing the total number of sample analyses that must be conducted to complete the analytical program and achieve specified data quality objectives.

5.2 DESCRIPTION OF ANALYSIS AND PARAMETER VALUES

To date the simulation model has been used to investigate three questions concerning the QC program for analysis of adipose tissue to determine levels of TCDD.

1. What are the false positive error rates associated with the routine calibration check, the absolute recovery test of the internal quantitation standard, and the method recovery test?

2. What are the probabilities that these three components of the QC program will detect an increase in the RRF?

3. What are the probabilities that these three components of the QC program will detect a degradation in recovery?

The simulation model incorporates the QC tests, decision rules, and actions described in Figure 5-1. (These are based on Table 4-1 and MRI, 1988). Analytical measurements representing each type of sample are generated using Equations 1 and 2 (Section 3) with the parameters appearing in those equations replaced by appropriate values. (Parameter values for TCDD were used. TCDD is used as an example throughout the study wherever specificity enhances the presentation.)

The parameter values used to address question 1 are displayed in Table 5-1. The parameter values for this case are constant throughout the analysis of all samples in the batch. The model is used to simulate analysis of 500 batches. For each QC test, data are collected from each batch regarding the number of tests conducted and the number of tests resulting in failures. The ratio of the number of failures to the number of tests, averaged over the 500 replicate batches, is an estimate of the probability of detecting a QC failure. Since the analytical process remains "in control" in this first case (i.e., there are no changes in the parameter values defining the process), a

Figure 5-1. QC Tests, Decisions, and Corrective Actions

A: INITIAL CALIBRATION AND INITIAL CALIBRATION TEST
PASS - Begin analysis
FAIL - Repeat calibration

B: ROUTINE CALIBRATION CHECK (RCC) AT BEGINNING OF DAY
PASS - Begin analysis
FAIL - Reanalyze solution
    PASS - Begin analysis
    FAIL - Go to A

C: ROUTINE CALIBRATION CHECK (RCC) AT END OF DAY
PASS - Prepare for next day
FAIL - Reanalyze solution
    PASS - Prepare for next day
    FAIL - Go to A and reanalyze all samples from that day

D: IQS ABSOLUTE RECOVERY (ALL SAMPLES)
PASS - Continue with next sample
FAIL - Reanalyze solution
    PASS - Continue with next sample
    FAIL - Analyze an RCC sample
        PASS - Reextract batch and analyze
        FAIL - Go to A. Repeat initial calibration and reanalyze all samples from that day

E: METHOD RECOVERY (estimated concentration in spiked QC sample minus estimated concentration in QC control divided by QC spiking concentration)
PASS - Record average method recovery and continue with next batch
FAIL - Record failure and average method recovery for the batch

Table 5-1. TCDD Simulation Model Parameter Values: "In Control" Case

Analysis Operation	GC Resp. Parameter		Recovery		Standard Deviation		
	Intercept	Slope	TCDD	IS	Batch	Sample	Analysis
Calibration							
Std/IS	0.00	0.80	1.000	1.000	0.00	0.0000	0.0500
IS/RS	0.00	1.73	1.000	1.000	0.00	0.0000	0.0800
Field Samples							
U/IS	0.00	0.80	0.595	0.521	0.15	0.1250	0.1000
IS/RS	0.00	1.73	1.000	0.521	0.15	0.0525	0.1575
QC Samples							
Std/IS	0.00	0.80	0.595	0.521	0.15	0.0150	0.0450
IS/RS	0.00	1.73	1.000	0.521	0.15	0.0525	0.1575

Notes:

The symbols */* in the left hand column refer to ratios of areas that represent instrument responses. The numbers in each row are the parameter values used in the equation to generate an instrument response for the ratio indicated in the first column. For example, U/IS refers to the equation used to produce the ratio of areas corresponding to the concentration of TCDD in a primary sample and the concentration of the internal quantitation standard.

Std - a sample spiked with a known amount of TCDD
RS - recovery standard sample

QC test failure means that a false positive has occurred and the probability estimated is the false positive error rate for the QC test in question.

The parameter values used to address question 2 are displayed in Table 5-2. In this case, the RRF undergoes a change after the initial calibration. This is accomplished by using a value of K_1 equal to 0.80 in Equation 1 (Section 3) for the calibration step and a value of K_1 equal to 1.10 when generating data for all other samples. The model is used to simulate 500 batches and probabilities of detecting the change in the RRF are estimated using ratios as described above. In this case, since a shift in the RRF has taken place, the ratios estimate probabilities of correctly detecting a QC failure.

The parameter values used to address question 3 are displayed in Table 5-3. A change in method recovery from 1.142 to 1.500 has been imposed by changing values of the β's as indicated in the table.

Table 5-2. TCDD Simulation Model Parameter Values: RRF Shift

Analysis Operation	GC Resp. Parameter Intercept	GC Resp. Parameter Slope	Recovery TCDD	Recovery IS	Standard Deviation Batch	Standard Deviation Sample	Standard Deviation Analysis
Initial Calibration							
Std/IS	0.00	0.80	1.000	1.000	0.00	0.0000	0.0500
IS/RS	0.00	1.73	1.000	1.000	0.00	0.0000	0.0800
Routine Calibration Check Samples							
Std/IS	0.00	1.10	1.000	1.000	0.00	0.0000	0.0500
IS/RS	0.00	1.73	1.000	1.000	0.00	0.0000	0.0800
Subsequent Calibration							
Std/IS	0.00	1.10	1.000	1.000	0.00	0.0000	0.0500
IS/RS	0.00	1.73	1.000	1.000	0.00	0.0000	0.0800
Field Samples							
U/IS	0.00	1.10	0.595	0.521	0.15	0.1250	0.1000
IS/RS	0.00	1.73	1.000	0.521	0.15	0.0525	0.1575
QC Samples							
Std/IS	0.00	1.10	0.595	0.521	0.15	0.0150	0.0450
IS/RS	0.00	1.73	1.000	0.521	0.15	0.0525	0.1575

Notes:

The symbols */* in the left hand column refer to ratios of areas that represent instrument responses. The numbers in each row are the parameter values used in the equation to generate an instrument response for the ratio indicated in the first column. For example, U/IS refers to the equation used to produce the ratio of areas corresponding to the concentration of TCDD in a primary sample and the concentration of the internal quantitation standard.

Std - a sample spiked with a known amount of TCDD
RS - recovery standard sample

Table 5-3. TCDD Simulation Model Parameter Values: Recovery Shift

Analysis Operation	GC Resp. Parameter Intercept	GC Resp. Parameter Slope	Recovery TCDD	Recovery IS	Standard Deviation Batch	Standard Deviation Sample	Standard Deviation Analysis
Calibration							
Std/IS	0.00	0.80	1.000	1.000	0.00	0.0000	0.0500
IS/RS	0.00	1.73	1.000	1.000	0.00	0.0000	0.0800
Field Samples							
U/IS	0.00	0.80	0.900	0.600	0.15	0.1250	0.1000
IS/RS	0.00	1.73	1.000	0.600	0.15	0.0525	0.1575
QC Samples							
Std/IS	0.00	0.80	0.900	0.600	0.15	0.0150	0.0450
IS/RS	0.00	1.73	1.000	0.600	0.15	0.0525	0.1575

Notes:

The symbols */* in the left hand column refer to ratios of areas that represent instrument responses. The numbers in each row are the parameter values used in the equation to generate an instrument response for the ratio indicated in the first column. For example, U/IS refers to the equation used to produce the ratio of areas corresponding to the concentration of TCDD in a primary sample and the concentration of the internal quantitation standard.

Std - a sample spiked with a known amount of TCDD
RS - recovery standard sample

5.3 RESULTS

Results that provide answers to the three questions introduced in Section 5.2 are summarized in Table 5-4.

When the analytical process is "in control," (i.e., the process is characterized by the parameter values in Table 5-1), the probability of detecting a calibration failure is less than 0.01 and the probability of a method recovery failure is approximately 0.09. The test of absolute recovery of the internal quantitation standard has a detection probability equal to 0.39.

The absolute IQS recovery failures have two consequences. First, the batch must be reextracted and analyzed resulting in additional cost. Second, since method recovery and absolute recovery are correlated, method recovery in the batches that pass will reflect the characteristics of the IQS tests that pass. These tests favor larger recovery values. Estimated concentrations of TCDD in primary tissue samples, therefore, will be biased toward larger values. For example, MR for these parameter values is 1.142 (i.e., 0.595÷0.521 or β_1 divided by β_2 according to Equation 8 in Section 3.2). An analysis result for a sample with a below average random contribution (see Equation 4 in Section 3.2) would be likely to fail the IQS recovery test since the underlying value of β_2 is near the

Table 5-4. Probabilities of Detecting Calibration and Recovery Shifts

QC Test	Status of Analytical Process		
	In Control	RFF Shift	Recovery Shift
	--------Probability of Detection--------		
Routine Calibration Check (RCC)[1]	<0.01	>0.99	<0.01
Internal Quantitation Standard (IQS)[2]	0.39	0.45	0.14
Method Recovery (MR)[3]	0.09	0.10	0.69

Notes:

1 - probability of detecting at least one RCC failure (i.e, two consecutive RCC sample failures) per batch

2 - probability of detecting at least one IQS failure per batch

3 - probability of detecting an MR failure per batch

lower boundary (0.40) of the test interval. This below average random contribution affects recovery (see Equation 3 in Section 3.2) also, causing it to be below average. Since the IQS test is likely to result in a failure, the batch will be reextracted and analyzed again. The new set of results will be retained if the estimate of β_2 is larger than 0.40, which is assured when the random contribution is above average. An above average random contribution also ensures an above average method recovery ratio. The IQS test when β_2 is near the lower boundary of the test interval, therefore, acts as a filter, eliminating analyses of primary samples with below average recovery ratios and retaining analyses with above average recovery ratios.

The effect of the IQS recovery test described above also is observed in simulation output that summarizes recovery for all primary samples processed by the model. When method recovery is set, as above, at 1.142 and β_2 is 0.521, average method recovery estimated by the model is 1.210. The value should be 1.142. The discrepancy is due to the filtering effect of the IQS recovery test.

These findings suggest that the IQS recovery test, which currently requires IQS recovery to be in the fixed interval between 0.40 and 1.50, should be based on a statistical

interval with boundaries determined from the mean and standard deviation of the recovery estimate.

The second and third columns of Table 5-4 display probabilities that QC tests correctly detect changes in the operating characteristics of the analytical procedure. The column labeled "RRF Shift" refers to an increase in the RRF of 37.5 percent, which occurs immediately following the initial calibration. The RCC test is extremely effective for detecting a change of the stated magnitude. The detection probability is greater than 0.99. The probabilities for the IQS test and the MR test are both false positive error rates since the recovery values have not been altered from the values used for the "in control" case. The interpretation of these probabilities, therefore, is the same as that in the former case.

The column labeled "Recovery Shift" reflects a change from the "in control" case in both IQS absolute recovery and method recovery. IQS recovery has been increased from 0.521 to 0.600 and method recovery from 1.142 to 1.500. The RCC test probability for this situation is less than 0.01 as expected since it is a false positive error rate. The IQS detection probability drops to 0.14 because the IQS recovery value is closer to the center of the allowable range than it was in the other two cases. The detection probability for the method

recovery test has increased to 0.69 because the hypothetical MR value of 1.500 is equal to the upper boundary of the test interval. (Note that the range of values defining the IQS test and the MR test are the same.) The apparent inconsistency between the detection probabilities of the two recovery tests is, in part, a consequence of the fixed interval approach to defining the recovery tests. These tests should reflect DQO's associated with applications of the data and, as indicated above, should be based on statistical characteristics of the recovery estimates.

The results discussed above reflect two sets of assumptions. The first assumptions, which are implicit in the QC program, form the basis for detecting and correcting recovery problems. These assumptions are:

1. when absolute recovery of the target compound in a primary sample declines as a result of sample processing, absolute recovery of the internal quantitation standard in the same sample also declines; and

2. when absolute recovery of the target compound in a primary sample declines as a result of sample

processing, absolute recovery of that target compound in a QC sample also declines.

Under these assumptions: (i) absolute recovery of the internal quantitation standard acts as a recovery adjustment applied to estimates of target compound concentrations; and (ii) method recovery computed from QC samples is representative of method recovery in primary tissue samples. Neither assumption is easily verified and if either assumption were violated, portions of the QC program may be ineffective. It is notable, however, that under these assumptions a change in the value of absolute recovery of the internal quantitation standard does not signal a change in method recovery.

The second set of assumptions concerns values selected for parameters that characterize recovery, variability, and the calibration curve of the analytical method. Most of the values used for the case represented in Table 5-4 were derived from data generated in the method validation study of the analytical method for measuring PCDDs (PCDFs) in adipose tissue referred to previously (USEPA, 1986). A few of the values, which could not be derived from the method validation data were based purely on judgement. Additional values for the parameters, determined either subjectively or from more recent data, are needed to conduct sensitivity analyses of the results discussed

above and any subsequent findings regarding alternative QC proposals.

REFERENCES

USEPA. 1986. Analysis for Polychlorinated Dibenzo-p-Dioxins (PCDD) and Dibenzofurans (PCDF) in Human Adipose Tissue: A Method Evaluation Study. Washington, D.C. Office of Toxic Substances. EPA-560/5-86-020.

MRI. 1988. Quality Assurance Project Plan for Work Assignment 27 (Revision No. 2) Analysis of Adipose Tissue for Dioxins and Furans. Washington, D.C. U.S Environmental Protection Agency Contract No. 68-02-4252.

Heath, R.G. 1988. Two Plans (A and B) for Allocation of Quality Control Samples for Chemical Analysis of FY87 Composite Samples. Washington, D.C. U.S. Environmental Protection Agency Contract No. 68-02-4294

Chatterjee, S., B. Price. 1977. Regression Analysis By Example. New York, N.Y. John Wiley & Sons.