

Proposed Limit Calculations/Analyses to be used in Pilot Study

06/15/06

The following document discusses decisions that are necessary for Federal Advisory Committee on Detection and Quantitation Approaches and Uses in Clean Water Act Programs (FACDQ) to make regarding calculations and statistical analyses based on Pilot Study data. These decisions will focus on:

- Prioritizing questions that are of greatest interest to the FACDQ, including
 - a. Does a procedure limit meet each target MQO (Measurement Quality Objective)?
 - b. How is the procedure affected by design choices (concentration, replicates, etc.)?
- How should a multi-laboratory limit (i.e., a limit determined from individual single-laboratory limits) be used, and how should the limit be calculated to meet the intended use?

Introduction

As part of its evaluation of detection and quantitation limit procedures, the FACDQ decided that a study evaluating a subset of these procedures would be beneficial. Based on FACDQ decisions and Technical Work Group (TWG) recommendations, this pilot study will evaluate the ASTM Interlaboratory Detection Estimate (IDE) procedure, the ASTM Interlaboratory Quantitation Estimate (IQE) procedure, the EPA Hubaux-Vos procedure, the EPA Lowest-Concentration Minimum Reporting Level (LC-MRL) procedure, and the ACIL Proposed Procedures for Determining the Method Detection Limit (MDL) and Minimum Level (ML).

To evaluate the procedures listed above, the TWG designed a pilot study with multiple components. These are:

1. A regression design study in which laboratories will analyze multiple blind samples spiked at 12 different concentrations. Results from this design will be used to calculate the ASTM, Hubaux-Vos and LC-MRL limits and to confirm limits for all procedures;
2. A single-laboratory study in which laboratories will perform the ACIL procedure as written; and
3. An Aroclor confirmation study to generate additional results for confirming the calculated ACIL limits.

In addition, previously generated data from a Michigan Manufacturers Association (MMA) interlaboratory study will be used in place of regression study data to evaluate two PCB Aroclors for the ASTM, Hubaux-Vos and LC-MRL procedures.

FACDQ Decisions

Once the laboratories have analyzed all Pilot Study samples, results will be submitted for data analyses. Due to the tight FACDQ schedule, it will not be possible to explore every possible analysis that could potentially be performed on this data set. Therefore, data analyses will be directed at those questions of greatest interest to the FACDQ.

This document discusses potential data analyses that can be performed using the Regression Design data, Single Laboratory design data and Aroclor confirmation data. The document includes discussion of “most representative” datasets, which are a subset of data chosen to best reflect how a procedure would routinely be performed. Assessments using these most representative datasets would indicate whether the resulting limits would meet different MQO criteria (i.e., accuracy, precision, false positive rate and false negative rate). The FACDQ had previously decided that these MQO criteria reflected characteristics that the FACDQ felt were desirable for detection or quantitation limits. The document also discusses ruggedness analyses, which will indicate how sensitive a given procedure is to variations in study design decisions. The FACDQ will need to identify which analyses will be of the greatest value (i.e., determine which ruggedness scenarios that are of greatest interest, and prioritize these analyses in relation to each other and to the MQO criteria assessments). The FACDQ should also identify the method of summarizing the results of the data analyses. The summary should clearly address the questions the FACDQ has identified as being of greatest interest.

In addition, single-laboratory limits generated using the ACIL, Hubaux-Vos and LC-MRL procedures can be pooled or combined to determine multi-laboratory limits. There are many possible approaches to determining these limits. The limits can be combined to:

- a. Determine an upper limit of the single laboratory limits (estimating a value that a large percent of labs would achieve);
- b. Determine a mean or average of the single laboratory limits (estimating a value that an average laboratory would achieve); or
- c. Determine a lower limit of the single laboratory limits (estimating a value that only the best laboratories would achieve).

Determining the best approach to calculate these multi-laboratory limits should be based on how that multi-laboratory limit would be used. The FACDQ will need to determine

how these limits would be used and thereby determine how they should be calculated. The exact calculations used to determine multi-laboratory limits will be based on factors such as the exact percentile or confidence level desired (e.g., 90%, 95% or 99% for an upper limit), assumptions about the distribution of individual laboratory limits, and the type of desired limit (i.e., confidence, prediction or tolerance limits).

Because the MMA data are readily available, and are based on a design similar to that of the regression design in the Pilot Study, these data may help inform the FACDQ's decision regarding data analysis priorities. For example, the MMA data can indicate how feasible certain ruggedness analyses would be, how a most representative dataset could be generated from the data, and the validity of various assumptions regarding the data that would affect subsequent analyses.

I. Regression Design Data

A. Data Usage

1. **Laboratory Calculations.** Once all laboratory analyses have been completed, each laboratory will be instructed to calculate an LC-MRL and Hubaux-Vos L_C based on its own data. The laboratory will be given the appropriate sample IDs, and their spike levels to be used to calculate these limits. The choice of which sample results the laboratory will use to calculate the LC-MRL and Hubaux-Vos L_C will be based on the suggested spike levels supplied by the laboratory during the pre-qualification phase.
2. **Data Subsets.** Once received, the regression data from all laboratories will be divided into two groups: a "calculation" subset and a "confirmation" subset. Each subset will include five replicates for each spike level, and will be divided (either randomly or systematically) in a way such that the full amount of temporal variability will be included in both sets. For example, if the replicates are numbered in the order in which the laboratories analyzed them, the odd-numbered replicates could be used for limit calculation, and the even-numbered replicates could be used for limit confirmation.

B. Limit Calculation based on Most Representative Dataset

1. **IDE and IQE.** For each method and analyte, a single IDE and IQE will be calculated based on a "most representative" dataset. This most representative dataset will be determined by the Study Design subgroup of the TWG. This dataset should best reflect how an IDE or IQE would be calculated in a method development study. Criteria for selecting this dataset will include:
 - Which (or all) spike levels should be used;
 - Which (or all) replicate(s) for each spike level should be used; and

- Whether outliers should be removed.

For example, the IQE procedure states that a minimum of five different concentrations must be used in a study. Therefore, a study with five different concentrations may be most representative of a routine IQE study. The Study Design subgroup would then identify which five of the twelve concentrations would be most appropriate for calculating an IQE for each method and analyte.

Once the most representative dataset has been identified, an IDE, ASTM measured and true concentration Critical Values (estimates of Currie's L_C determined as part of the IDE procedure), and an IQE will be calculated using this dataset for each method and analyte, via SAS and/or Excel software.

Because the MMA PCB dataset was based on a similar design to that of the regression design pilot study, limit calculation and confirmation also will be performed using this dataset. This dataset is readily available, so these analyses can be performed prior to those using the pilot study data. Therefore, this dataset will aid the Study Design team in determining how an IDE or IQE study would typically be designed. A most representative dataset will then be created from the MMA data, and an IDE and IQE will be calculated for each Aroclor based on this dataset.

2. LC-MRL and Hubaux-Vos L_C . For each method, analyte and laboratory, a single LC-MRL and Hubaux-Vos L_C will be calculated based on a "most representative" dataset. Because the spike levels were chosen to most closely match those identified by the laboratory, the subset of data the laboratory used to determine its LC-MRL and Hubaux-Vos L_C will be considered the most representative dataset for these limits. The laboratories' calculations will be reviewed and any calculation errors will be fixed, and the limits adjusted, if necessary.

While the MMA PCB dataset will also be used to calculate an LC-MRL and Hubaux-Vos L_C for each laboratory and Aroclor, these calculations will not be done by the laboratory that performed the original analyses. Therefore, the Study Design subgroup of the TWG will identify the appropriate concentrations and replicates to use for these determinations, and will then perform the calculations.

Calculation of limits based on "most representative" datasets is summarized in Table 1.

C. Limit Confirmation (Most Representative)

1. Statistical Models/Interpolation. Once each limit has been calculated, the confirmation subset data will be used to estimate the different MQO characteristics over concentration. This will be achieved either through the use of statistical models based on results from all or some of the concentrations, or by linear interpolation between the two concentrations surrounding a limit, to estimate the MQO characteristics over a range of concentrations. The choice of which approach is most appropriate will depend on many factors, including:
 - a. Assumptions regarding the relationship between concentration and the MQO characteristic (i.e., is the relationship linear or not);
 - b. Whether this relationship is consistent over the concentration range; and
 - c. The error associated with any fitted models or other relationships.

For example, a linear regression model could be used to model recovery over concentration, and a logistic regression model could be used to model false negative rate over concentration. If linear interpolation is found to be most appropriate for estimating percent RSD, this could be done by calculating the relative standard deviations (RSDs) at the two concentrations surrounding a calculated limit. Models and/or interpolations will be fit using data from each laboratory, and using data from all laboratories.

2. Estimation of MQOs. Once the statistical models have been fit, each MQO characteristic (recovery, percent RSD, false negative rate, false positive rate) will be estimated for each applicable limit. For example, the percent RSD, mean percent recovery and false negative rate will be calculated using the model for each determined IQE. Results will be compared to assumed results and will be summarized in a table such as the one below.

IQE Summary Table – Method 300.0

Analyte	Assumed			Observed (based on model)		
	% RSD	Mean Rec.	FN Rate	% RSD	Mean Rec.	FN Rate
Chloride	20	*	1%			
Nitrate	20	*	1%			
Sulfate	20	*	1%			
Median						

* MQO to be determined by TWG

3. Confirmation of LC-MRL. For the LC-MRL, limit confirmation will be done separately for each laboratory, and will be summarized as described above separately for each laboratory.
4. Confirmation of ASTM Measured and True Concentration Critical Values, and Hubaux-Vos L_C . Both the ASTM Critical Values and Hubaux-Vos L_C will be confirmed using existing blank data submitted by the laboratory during prequalification. For the ASTM Critical Values, existing blank results submitted by all laboratories will be compared to the limit determined for the given method and analyte. For the Hubaux-Vos L_C , blank results from a given laboratory will be compared to the limit determined by that laboratory for the given method and analyte.
5. Tests of Statistical Significance. It may be possible to assess whether the estimated MQO criterion for a given procedure limit is statistically and significantly different from the target value. For example, confidence limits can be calculated for the estimated mean recovery at a specified limit, based on the fitted confirmation model. However, the statistical power of this test would depend on the fit of the model, the amount of data used to fit the model, and the type of model used. Significance tests may be more feasible for the ASTM limits, because results from all laboratories will be used. Significance tests cannot be performed if confirmation is assessed based on linear interpolation determined using results from two concentrations.

If statistical tests cannot be performed separately for each analyte and laboratory, it may be possible to assess whether a procedure is generating limits with estimated MQO characteristics that differ significantly from the target value by combining data from all analytes for a given method and/or by combining data from all laboratories. This could be done using a nonparametric test (e.g. the Wilcoxon signed-rank test).

6. Additional Summary Statistics. Additional summary statistics may also be calculated. For example, using the different models, the exact concentration at which an assumed MQO criterion is met can be estimated. From this, the percent difference between this concentration and each of the limits can be calculated.

Steps 1-6 also will be performed using the MMA PCB dataset.

Confirmation of limits calculated using the different procedures based on “most representative” datasets is summarized in Table 2.

D. Ruggedness Analyses

Once evaluation of the “most representative” limits has been completed, additional limits may be calculated using different subsets of the data. Examples of possible additional analyses include:

- Calculation of limits using different choices of spike level
- Calculation of limits using different sets of replicates
- Calculation of limits using a subset of labs (IDE and IQE only)
- Calculation of limits based on a different choice regarding outlier removal to determine if the effect is significant -- see discussion at III.M. “Pooling Data.”

Another potential analysis would be to determine LC-MRLs specific to achieving a mean recovery (of a predetermined number of results) rather than an individual recovery. Because an individual recovery is more variable than a mean recovery, the LC-MRL targets an accuracy MQO criterion that is more “difficult” to achieve than the 50-150% mean recovery MQO criterion set for the pilot study. An LC-MRL targeting a mean recovery would also be more comparable to the other procedures. However, the labs would not be able to calculate a mean-targeting LC-MRL without software or a written procedure. In addition, the spike levels used by the laboratory to determine the LC-MRL may not be suitable, because the spike levels chosen to target 50-150% recovery for an individual sample may be inappropriately high for targeting 50-150% mean recovery.

These additional limits can then be compared to the limits determined using the most representative datasets. Differences can be summarized as percent differences or Relative Percent Differences (RPDs). MQO criteria can also be calculated for the additional limits using the confirmation data models or interpolations used to assess the most representative dataset limits.

Ruggedness analyses also will be performed using the MMA PCB dataset. The MMA PCB data analyses may help inform what additional analyses are of greatest interest or value, because analyses using this dataset will be performed prior to analysis of the pilot study data.

Some example ruggedness analyses are summarized in Table 3.

E. Multi-Laboratory Limits

For the LC-MRL and Hubaux-Vos procedures, additional analyses can be performed to assess different methods of generating pooled, multi-laboratory limits.

There are three types of multi-laboratory limits that could be of interest:

- a. An upper limit of the single laboratory limits (estimating a value that a large percent of labs would achieve);
- b. Determining a mean or average of the single laboratory limits (estimating a value that an average laboratory would achieve); and
- c. Determining a lower limit of the single laboratory limits (estimating a value that only the best laboratories would achieve).

The choice of which of the above types of multi-laboratory limits is most appropriate will depend on how that limit is to be used. Once this has been determined, a method of calculating a multi-laboratory limit will be determined. For example, the FACDQ could decide that an upper limit of the single-laboratory LC-MRLs would be desired. An upper limit of the LC-MRLs could be generated by calculating the mean of the individual laboratory LC-MRLs plus 3 standard deviations of the individual laboratory LC-MRLs, or by determining the maximum or upper percentile of the individual laboratory LC-MRLs. The FACDQ could also decide that a multi-laboratory limit that represents performance of an average laboratory would be useful. This type of limit could be calculated by determining the mean or median of the single-laboratory LC-MRLs.

The pooled multi-laboratory limits would then be compared to the confirmation models or interpolations fit using data from all laboratories.

II. Single Laboratory Limits

A. Limit Calculation

Each laboratory will determine a single-laboratory limit using the ACIL procedure, and will submit all data and calculations. The results will then be reviewed for any calculation errors, and adjusted accordingly.

Calculation of limits using the ACIL procedure is summarized in Table 1.

B. Limit Confirmation

Single-laboratory limits will be confirmed using the same models or interpolations fit to confirm the laboratory's LC-MRL and Hubaux-Vos L_C limits. For uncensored methods, the single-laboratory L_C will be confirmed using all or a subset of the laboratory's blank data. Confirmation will be summarized using tables similar to the example displayed in the Regression Design discussion.

For the two aroclors analyzed using Method 608, no confirmation data will be available initially because these analytes are not included in the regression-based

pilot study. Therefore, confirmation of these aroclors will be performed using subsequent laboratory analyses (five replicate analyses at three concentrations). Models or interpolations of each MQO criterion will be fit using these results, and confirmation will be completed.

Similarly to the regression-based design limits, estimated MQO characteristics at the single-laboratory procedure limit could be compared statistically to the target value. However, there may not be enough data to perform these tests with sufficient power on a “by-laboratory” basis, and therefore comparisons may not be performed over analytes, or over laboratories, using a nonparametric test such as the Wilcoxon signed-rank test.

Confirmation of the laboratory’s MDL and ML submitted during the pre-qualification phase will also be performed.

Confirmation of limits calculated using the ACIL procedure is summarized in Table 2.

C. Ruggedness Analyses

The single-laboratory procedures include fewer concentrations and total replicates, and therefore fewer alternate design scenarios can be simulated. For this reason, it is unlikely that analyses can be performed to assess the effect of study design on the resulting limits. Some ruggedness analyses could be performed using existing blank data for uncensored methods. For example, limits could be calculated and compared using different numbers of blank results, or different time ranges.

Some example Ruggedness Analyses are summarized in Table 3.

D. Multi-Laboratory Limits

Similar to the LC-MRL and Hubaux-Vos limits, additional analyses can be used to assess different methods of calculating multi-laboratory limits. These analyses will be based on the FACDQ decision regarding the appropriate uses of multi-laboratory limits. The exact calculations used to determine multi-laboratory limits will be based on factors such as the exact percentile or confidence level desired (e.g., 90%, 95% or 99% for an upper limit), assumptions about the distribution of individual laboratory limits, and the type of desired limit (i.e., confidence, prediction or tolerance limits).

Multi-laboratory limits would then be calculated based on an approach to best target that use (such as by calculating the mean plus 3 standard deviations of the individual single-laboratory limits, if an upper limit is decided to be most

appropriate). These resulting multi-laboratory limits could then be compared to the confirmation data models or interpolations for each MQO criterion. However, if the resulting multi-laboratory limits exceed all spike levels included in the regression-based design, then the MQO characteristics estimated at that concentration may not be reliable, and therefore this comparison may not be entirely valid.

An additional analysis could involve pooling the MDLs and MLs submitted by the laboratories based on the different formulas used to pool the single-laboratory limits, and comparing the pooled MDLs and MLs to the confirmation data models or interpolations for each MQO criterion.

III. Summary Statistics

Independently of the different procedures and their limits, descriptive statistics will also be calculated using the measured results at each spike concentration. This can be used to summarize the performance of each method and laboratory in the study. Calculated descriptive statistics will include mean and median concentration and percent recovery, standard deviation and RSD of the recoveries, and the minimum and maximum concentration and percent recovery.

Table 1. Calculation of Detection and Quantitation Limits from Study Data using “Most Representative” Datasets

Limit to be Calculated	Source of Data used	Type of Limit	Comments
Hubaux and Vos- L_c	Regression Design	Single Lab	After all results have been reported, labs will be instructed to use the LC-MRL procedure and their own data calculate their own Hubaux-Vos L_c and LC-MRL. For each lab, specific sample IDs and associated spike levels will be identified that the lab should use to calculate their L_c and LC-MRL. The samples selected for use in determining the L_c and LC-MRL will be based on the spike levels the lab suggested for the LC-MRL as part of their pre-qualification package. Historical data from an MMA interlab study of Aroclor 1016 and 1260 analyses will be used to calculate lab-specific L_c and LC-MRL values for these analytes. The Study Design Team will identify the appropriate concentrations and replicates to use for these calculations.
LC-MRL	Regression Design	Single Lab	
MRL	Regression Design	Multi-lab	A multi-laboratory MRL will be determined from the individual single-laboratory LC-MRLs based on a calculation chosen by the FACDQ (for example, the MRL procedure has a default calculation of the mean plus three standard deviations of the LC-MRL values calculated above).
Interlaboratory Detection Estimate	Regression Design	Inter-lab	An IDE and an IQE will be calculated based using a “most representative dataset” identified by the Study Design Team. Historical data from an MMA interlab study of Aroclor 1016 and 1260 analyses will be used to calculate an IDE and IQE for these compounds.
Interlaboratory Quantitation Estimate	Regression Design	Inter-lab	
ASTM Measured Concentration Critical Value	Regression Design	Inter-lab	This estimate of Currie's L_c is determined as part of the IDE procedure. Accordingly, it will be calculated using the same representative dataset used to calculate the IDE.
ASTM True Concentration Critical Value	Regression Design	Inter-lab	This estimate of Currie's L_c is determined as part of the IDE procedure. Accordingly, it will be calculated using the same representative dataset used to calculate the IDE.
ACIL MDL	Single Lab	Single Lab	Each lab will use results from their own preparation and analysis of 20 spiked replicates to calculate their ACIL MDL and ACIL ML according to the ACIL procedure. Accuracy of these calculations will be confirmed and corrections will be made where necessary.
ACIL ML	Single Lab	Single Lab	
EPA Minimum Level	Single Lab	Single Lab	MDL data submitted by the laboratories as part of their bid qualification package will be used to calculate an EPA ML, based on the formula of 10x the standard deviation of the individual results used to calculate the MDL.

Table 2. Data and Approach for Confirming Calculated Limits

	Limit	False Positive Rate	False Negative Rate, Mean Percent Recovery, and Percent RSD
Single-Laboratory Limits	ACIL MDL	Compare to each lab's historical blank data and their QC blanks from FACDQ Study	Model or interpolate most values based on that lab's confirmation data (from regression design analyses).
	ACIL ML	NA	
	EPA MDL*	Each lab's historical blank data plus their QC blanks from FACDQ Study	
	EPA ML*	NA	Model or interpolate Aroclor values based on data obtained during the Aroclor confirmation study.
	Hubaux-Vos L _c	Each lab's historical blank data plus their QC blanks from FACDQ Study	
	LC-MRL	NA	
Inter-laboratory Limits	ASTM Measured Concentration Critical Value	Compare to all labs' historical blank data and all lab QC blanks from FACDQ Study	Model or interpolate these values based on entire confirmation data set (from all laboratories regression design analyses)
	ASTM True Concentration Critical Value	Compare to all labs' historical blank data and all lab QC blanks from FACDQ Study	
	ASTM IDE	NA	
	ASTM IQE	NA	
Pooled Multi-laboratory limits	ACIL MDL	Compare to all labs' historical blank data and all lab QC blanks from FACDQ Study	Model or interpolate these values based on entire confirmation data set (from all laboratories regression design analyses)
	ACIL ML	NA	
	EPA MDL*	Compare to all labs' historical blank data and all lab QC blanks from FACDQ Study	
	EPA ML*	NA	
	Hubaux-Vos L _c	Compare to all labs' historical blank data and all lab QC blanks from FACDQ Study	
	MRL	NA	

Table 3. Example Ruggedness Analyses

Design Factor	Ruggedness Analysis	Applicable Procedures
Choice of Spike Level	Compare Limits Calculated Based on Different Numbers and Ranges of Spike Levels	IDE, IQE, Hubaux-Vos, LC-MRL
Numbers of Labs	Compare Limits Calculated Based on Different Numbers of Laboratories (may require more than minimum 8 labs)	IDE, IQE, MRL, pooled multi-lab ACIL and Hubaux-Vos limits
Numbers of Replicates per Labs	Compare Limits Based on Different Numbers of Replicates per Spike Concentration	IDE, IQE, LC-MRL, Hubaux-Vos
Outlier Testing	Compare Limits Calculated With and Without Outlier Removal	IDE, IQE
Temporal Variability	Compare Limits Calculated based on Samples With Differing Time Ranges	IDE, IQE, ACIL Lc (using existing blank data)
MQO type	Calculate Limit targeting 50-150% mean recovery instead of 50-150% individual recovery (may require different spike levels)	LC-MRL