

Draft Procedures Report

May 24, 2007

The TWG members were the primary authors of this document and the TWG did review and comment on this document as it was being developed. However, the TWG did not have time to fully review, discuss, and agree on this document. The TWG agreed to send this document to the FACDQ with the understanding that it is for information and decision-making purposes, but not a fully vetted document.

Procedures Report

Table of Contents

I. Executive Summary and Pilot Study Observations (p.3)

II. Introduction (p.5)

III. Evaluation (p.7)

a. 'What we need a procedure to do?' (p.7)

- i. Does the procedure provide an explicit estimate of bias at L_Q for limits that must be verifiable by labs at those limits?
- ii. Does the procedure provide an explicit estimate of precision at L_Q for limits that must be verifiable by labs at those limits?
- iii. Does the procedure provide an explicit false positive rate for L_C ?
- iv. Does the procedure provide an explicit false negative rate at L_C for the true value at L_D or L_Q that must be observed in labs at L_C for the estimated values of L_D or L_Q ?
- v. Does the procedure provide that qualitative identification criteria defined in the analytical method are met at the determined detection and quantitation limits?
- vi. Does the procedure adequately represent routine variability in lab performance?
- vii. Does the procedure perform on-going verification of estimates?
- viii. Does the procedure be capable of calculating limits using matrices other than lab reagent grade water?
- ix. Does the procedure use only data that results from test methods conducted in their entirety?
- x. Does the procedure explicitly adjust or account for situations where method blanks always return a non-zero result/response?
- xi. Does the procedure explicitly adjust or account for situations where method blanks are intermittently contaminated?
- xii. Is the procedure clearly written with enough detail so that most users can understand and implement them?
- xiii. Is the procedure cost effective?
 1. Info from post study questions with the labs?
- xiv. Does the procedure assess multi-laboratory and inter-laboratory variability when data from more than one lab is used?
- xv. Is the procedure applicable to all users and test methods?

b. Pilot Study MQOs (p.16)

- i. What did we learn from the ability of the procedures to meet the MQOs?

c. Evaluate if procedures achieved their intended purpose? (p.19)

- i. How do these results differ from above?

d. Additional Considerations or Evaluations (p.22)

- i. Implementation in CWA programs

IV. Recommendations + Rationale (p.26)

Author Initials	Name
BA	Bob Avery
BE	Brian Englert
DR	Dick Reding
JPH	John Phillips
JPL	Jim Pletl
KM	Ken Miller
KO	Ken Osborn
LL	Larry LaFleur
NT	Nan Thomey
RR	Rick Rediske
SB	Steve Bonde
SW	Steve Wendelken
TF	Tim Fitzpatrick
TWG	Technical Work Group
ZE	Zonetta English

I. Procedures Report Executive Summary (EPA)

To be completed.

Pilot Study Observations (KM)

1. The detection and quantitation limit procedures meet most of the MQOs most of the time.
2. The likelihood of a procedure limit meeting the targeted MQOs is heavily affected by what the procedure itself targets (i.e., whether the procedure MQO(s) match the FACDQ MQOs) and how it targets them (i.e., does it estimate the lowest concentration at which the procedure MQO(s) are met, or does it demonstrate that the MQO(s) can be met at a chosen concentration).
3. For some limits (false positive rate for all detection limits, RSD at the IQE20%), it would be expected that the FACDQ MQO would be met half the time, because the procedure targets the lowest concentration to achieve the FACDQ MQO. For others (false negative rate at the IDE), it would be expected that the FACDQ MQO would be met less than half the time because the procedure targets a less stringent MQO than the study. For other limits (mean recovery at the LCMRL), it would be expected that the FACDQ MQO would be met more than half the time because the procedure targets a more stringent MQO than the study, or the procedure does not target the lowest concentration to yield the FACDQ MQO.
4. For quantitation limits, the most difficult MQO to meet is the false negative rate.

False negative rates were higher when making detection decisions based on detection limit compared to making detection decisions based on instrument signal.

Unlike the other FACDQ MQOs, the false negative rate is based on two limits, the detection and the quantitation limit. As a result, a high false negative rate could be due to a biased-high detection limit or a biased-low quantitation limit.

For ASTM (IDE and IQE) limits (both single lab and interlab), false negative rates were higher when detection based on L_c compared to detection based on Y_c . This was especially true for Method 625, for which recoveries were less than 100% throughout the concentration range, and as a result the recovery-corrected L_c was greater than Y_c .

5. Mean recovery criterion rarely failed for most analytes/labs, but criterion failures occurred both on the high side (>150%) and the low side (<50%). Low failures generally occurred for a few "problem" analytes (see Method 625, 608 Endosulfans). High failures occurred mostly for Method 300.0.
6. Big differences were observed, both in performance and limit MQO success, between laboratories. Big differences were also observed between analytes. Some laboratory differences were attributable to differences in how method was applied (ex: extraction technique for Method 625).
7. For detection limits especially, variability between lab limits tended to be greater than the variability between the different procedure limits for a single lab, which may affect the

- assessment of the different procedures. For quantitation limits, the variability between lab limits tended to be greater; however, this was likely due to the different quantitation limits targeting different MQOs.
8. The level of background contamination (or other blank bias) varied widely between laboratories. Blank bias observed in blind samples was not always observed in the existing blanks (or vice versa). This was most frequently the case for Method 300.0.
 9. For uncensored methods (chiefly 200.7), the amount of existing blank data varied significantly. Therefore, the precision of the calculated false positive rates differed between laboratories. In addition, the false positive rates of interlaboratory limits were more heavily affected by some labs than others (e.g.: copper).
 10. Not all laboratories interpreted the ACIL procedure, and the SOW instructions regarding the ACIL procedure, in the same way. Therefore, some ACIL limits and limit evaluations are more representative of the written ACIL procedure than others.
 11. Detection limit procedures based on extrapolation from spiked data are more prone to unexpected false positive rates. Especially for censored methods, this is often due to the recovery vs. concentration relationship not being linear from 0 to quantitation, as assumed by the procedures. In some cases, this was due to choice of spike level; however, the false positive rates and false negative rates were not generally improved by using lower spike levels (see Study Design section - II.d iv.)
 12. False positive rate evaluation for uncensored methods is geared toward the ACIL, because data used to calculate limits are also used to determine false positive rates.
 13. For Methods 608 and 625 interlaboratory RSDs exceeded 20% throughout the concentration range for several analytes, and therefore the calculated limits would automatically fail this MQO criterion. This was due to low biases of these methods, and differences between laboratories throughout the concentration range.
 14. Based on the procedure instructions, software, and pilot study design, some procedures will yield a limit for every analyte/lab, while others will often fail to yield a limit. Therefore, MQO estimates for some procedures summarized over analytes and labs will be more heavily affected by “difficult” analytes (or labs) than MQO estimates for other procedures.
 15. Independent of procedure, the order at which the Pilot Study MQOs were met varied by method, analyte and/or lab. For uncensored methods and Method 300.0, the RSD target tended to be met at lower concentrations than the mean recovery rate MQO. However, for Methods 608 and 625, the RSD target tended to be met at higher concentrations. The false negative rate MQO was heavily affected by both the variability and bias observed (and as a result, where the recovery and RSD MQOs are met), but also by the level of blank bias.

II. Introduction (LL)

During the July 2006 meeting of the Federal Advisory Committee on Detection and Quantitation (FACDQ) and their uses in the Clean Water Act Program, the Committee asked the Technical Work Group (TWG) to develop recommendations on a procedure (or procedures) for determining detection and quantitation. These recommendations were to include the rationale used in their selection and they were to be completed for the Committee's consideration at its next meeting in May 2007. The goal of this report is to provide a response to that request.

In developing its recommendations, the TWG relied on several criteria and sources of information. In earlier FACDQ discussions, the document titled "What we need a procedure to do" was compiled and approved by the FACDQ for use in developing final recommendations to EPA. This document enumerated 15 specific criteria to be considered. In the current report, the TWG provides its interpretation of how each of the detection and/or quantitation procedures meets each criterion. These evaluations are summarized in Section III. A.

Another significant basis for the TWG recommendations is its analysis of the pilot study data (Pilot Study Report). Measurement Quality Objectives (MQOs) for false positives, false negatives, precision, and recovery were selected for use in the pilot study. Data generated as part of the study design were used to test how well each procedure performed. Where a procedure failed to achieve the desired MQO, the data were examined to determine if the failure was an artifact or limitation of the study design implementation, was due to limitations in the analytical method, or was a fundamental problem with the detection or quantitation procedure. This analysis was used to identify whether modifications to the procedure would remedy the apparent limitation. These evaluations are summarized in Section III. B.

Each detection and quantitation procedure in the pilot study was also evaluated to determine how well it achieved the authors' intended purpose. In this analysis, evaluation was not against the pilot study MQOs, but against what the procedure was designed to achieve. It should be noted that only the ACIL procedure was written with the same requirements as the pilot study MQOs and therefore the other procedures should not necessarily be expected to achieve the pilot study MQOs as tested. These evaluations are summarized in Section III. C.

The TWG also examined implementation issues in the Clean Water Act Program and/or how each of the procedures might be implemented. This was done within the framework of the policy decisions discussed below. These evaluations are summarized in Section III. D.

Based on these evaluations and the insights gained through this work, each procedure was evaluated to determine if it would meet the needs of the FACDQ. If possible, where shortcomings or limitations were observed, modifications that may eliminate those limitations are suggested. The TWG reviewed all the above information in developing its recommendations and then prepared the rationale supporting those recommendations. This is summarized in Section IV.

As noted, evaluations of the procedures were discussed in terms of how they would work in implementing the policy decisions formulated in the July 2006 FACDQ meeting. Those decisions included the need for a single laboratory detection and quantitation procedure, method promulgation (including promulgation of a national detection and quantitation limit), use of procedures to demonstrate laboratory proficiency, the need and the process described in the July 2006 policy decision for future updates of the detection and quantitation limits, and finally,

matrix effects. The TWG recognized the need to match its recommendations to those decisions in order to allow them to be effectively implemented. Appropriately, this became a key aspect in the discussion of the suitability of different procedures to meet the needs of the FACDQ.

III. Evaluation (TWG)

a. 'What we need a procedure to do'

The committee developed a list of characteristics identified by the FACDQ and evaluated these characteristics over the course of the pilot study.

i. Does the procedure provide an explicit estimate of bias at L_Q for limits that must be verifiable by labs at those limits? (JPH)

According to the characteristics chart all procedures except the ML have been identified as having this characteristic; however each procedure may meet the object in a different way. The ACIL and Consensus Group (CG) procedures specify a bias requirement (MQO) of 50-150% *mean* recovery as where the Lab Q/C procedure requires 50-150% recovery based on the *median*. All three of these procedures have periodic verification requirements. The Lab Q/C procedure requires batch verification, the CG procedure recommends batch verification, but requires monthly verification and the ACIL procedure requires quarterly verification. The LCMRL requires that the *individual* recoveries fall within 50-150%, but does not have any ongoing verification. The IQE on the other hand provides an explicit estimate for bias, but does not target or requires any specific method bias as written. Using the IQE procedure the specific bias at any concentration can be determined, but there is no on-going verification of these estimates.

ii. Does the procedure provide an explicit estimate of precision at L_Q for limits that must be verifiable by labs at those limits? (JPH)

According to the characteristics chart all procedures except the ML and maybe the LCMRL have been identified as having this characteristic; however each procedure may meet the object in a different way. The ACIL and CG procedures specify a precision *not to exceed* 20% RSD and the Lab Q/C procedure *controls at* 25% RSD. All three of these procedures have periodic verification sample requirements (MQOs). The Lab Q/C procedure requires batch verification, the CG procedure recommends batch verification, but requires monthly verification and the ACIL procedure requires quarterly verification. The IQE *targets* (but does not guarantee) a specific %RSD depending upon the estimate. For example an IQE10 targets 10% RSD, an IQE20 targets 20% RSD, an IQE30 targets 30% RSD and so on. The IQE precision is explicit in that an estimate of the %RSD is provided across the entire concentration range. The LCMRL has no explicit requirement for bias, but a bias estimate at any concentration can be derived. Neither the IQE nor the LCMRL procedures include on-going verification of their estimates.

iii. Does the procedure provide an explicit false positive rate for L_c ? (JPH)

According to the characteristics chart all procedures are identified as having this characteristic; however each procedure may meet the object in a different way. The ACIL and CG procedures require that the false positive (FP) error rate at the detection limit (MDL or L_c) not exceed 1%. Both procedures require monitoring of method blanks to ensure that the FP error rate remains below 1%, however the CG procedure does so on an on-going basis and the ACIL procedure only periodically (maximum annual evaluation). The EPA MDL and Lab QC procedures calculate a detection limit at which the false positive error rate is approximately 1%, assuming constant standard deviation and normal distribution. The IDE and Hubuax-Vos (HV) procedures target (but does not guarantee) a FP error rate of 1% at the detection limit (L_c). Neither the EPA MDL, ASTM IDE nor the HV procedure include on-going verification of FP error rate.

iv. Does the procedure provide an explicit false negative rate at L_C for the true value at L_D or L_Q that must be observed in labs at L_C for the estimated values of L_D or L_Q ? (JPH)

According to the characteristics chart all procedures are identified as having this characteristic except for the EPA MDL/ML; however each procedure targets (but does not guarantee) FN error rates. . The ACIL, CG and Lab QC procedures require that the false negative (FN) error rate for results at the quantitation limit (ML, QL or RL_C) at the detection limit (MDL or L_C) not exceed 1%. All three of these procedures have periodic FN verification requirements. The Lab Q/C procedure requires batch verification (FNQS), the CG procedure recommends batch verification (QLC), but requires monthly verification and the ACIL procedure requires quarterly verification also evaluating the QLC samples. The IDE and Hubuax-Vos (HV) procedures target (but does not guarantee) FN error rates. The HV the target FN rate is 1% and the IDE targets a 5% FN error rate. Neither the ASTM IDE nor the HV procedures include on-going verification of FN error rate.

v. Does the procedure provide that qualitative identification criteria defined in the analytical method are met at the determined detection and quantitation limits? (TF)

Lab-Determined Detection Limits (DLs) and Quantitation Limits (QLs)

ACIL DL + QL: Single Lab Pilot Tested & Multi-Lab

The procedure explicitly requires that qualitative criteria such as retention time are verified for censored methods at and perhaps below the quantitation limit. Qualitative identification is explicitly required at the initial detection estimate but is not (cannot be) verified at the calculated detection limit.

HV+ LCMRL: Single Lab Pilot Tested & Interlab Pilot Tested

Both procedures are silent on the issue of qualitative identification and the assumption is that they rely on the method to define those requirements for identification. To the extent that the LCMRL is established no lower than the lowest calibration standard, any method requirements for qualitative identification may satisfy this criterion for that procedure. H-V does not address identification of analytes at the determined detection estimate; it's possible that the detection estimate could be established well below the region at which analytes could be qualitatively identified.

IDE/IQE: Single Lab Pilot Tested & Interlab Pilot Tested

The procedures state that method criteria must be satisfied during the analysis of study sample however neither the IDE nor the IQE explicitly require that qualitative identification of analytes be verified at the determined detection or quantitation estimates.

Lab QC (Osborn) : Single Lab Non-Pilot Tested

FNQS results must meet all qualitative identification criteria as specified in the analytical SOP. The procedure does not explicitly require that qualitative criteria be satisfied, however samples analyzed in DL/QL studies are presumed to satisfy pertinent method requirements.

Consensus Group DL + QL: Single Lab Non-Pilot Tested & Multi-Lab

The procedure explicitly requires that qualitative criteria are verified for censored methods at and perhaps below the quantitation limit. Qualitative identification is explicitly required at the initial detection estimate but is not verified at the calculated detection limit.

Consensus Group L_C + L_D: Single Lab Non-Pilot Tested & Multi-Lab

The procedure explicitly requires that qualitative criteria are verified for censored methods below the quantitation limit. Qualitative identification is explicitly required at the initial detection estimate and verified on an ongoing basis.

MDL/ML: Single Lab Non-Pilot Tested

The procedure does not explicitly require that qualitative criteria be satisfied, however samples analyzed in DL/QL studies are presumed to satisfy pertinent method requirements. It's possible that the detection and/or quantitation estimate could be established well below the region at which analytes could be qualitatively identified.

MDL/ML: Multi-Lab

The procedure does not explicitly require that qualitative criteria be satisfied however, samples analyzed in DL/QL studies are presumed to satisfy pertinent method requirements.

The procedure does not explicitly require that qualitative criteria be satisfied however, samples analyzed in DL/QL studies are presumed to satisfy pertinent method requirements. It's possible that the detection and/or quantitation estimate could be established well below the region at which analytes could be qualitatively identified.

Under **Use #2 Method Promulgation**, all the Single Lab Pilot Tested procedures, the ACIL and Consensus Group Multi-Lab procedures, and the Lab QC procedure, all showed promise in addressing the weakness in the 40 CFR MDL procedure for this criterion in the region below the quantitation limit. In regards to the LT-MDL, because of incomplete documentation it is unclear how this criterion might be satisfied. In regards to the Consensus Group, the procedure should ensure analyte identification at and below quantitation limits. In regards to the MDL/ML, a weakness relative to this criterion has previously been noted and commented on. There should be a mechanism to ensure that detects reported below L_q meet qualitative identification criteria.

Under Use #3 Demonstration of Laboratory Proficiency of Detection and Quantitation

Limits, all the procedures do not guarantee that qualitative identification is always satisfied at L_c. As long as method criteria for identification are satisfied at L_q, no adverse impacts are anticipated. The LT-MDL is not sufficiently described to address this criterion.

**vi. Does the procedure adequately represent routine variability in lab performance?
(KM/BE)**

Generally, "routine variability" was assumed to encompass various sources of variability that would occur over multiple batches. This would include such sources of variability between multiple calibrations or routine maintenance practices. The ACIL and Consensus Group procedures specify that data be generated over a period of a year or more. This timeframe will be likely to capture those sources of variability that would represent routine performance. The Consensus Group procedure also includes ongoing verification of its L_c for uncensored methods and of its QL for both censored and uncensored methods. Though these procedures will quickly

detect an increase in bias or variability, they will be unable to quickly detect decreases in either bias or variability.

The EPA MDL/ML, Hubaux-Vos and LCMRL procedures do not explicitly specify the duration of data generation, the limit may only reflect within-batch variability, and therefore would not always capture the sources of variability that represent routine performance. The Lab QC procedure states that results should be spread out over multiple days and batches, which may include some, but not necessarily all, sources of variability representative of routine performance. The IDE and IQE procedures state that analyses for each laboratory should encompass multiple batches, and that analyses could be spread out over a larger period of time. While the time period of analysis would likely be shorter for these two procedures, the interlaboratory variability would encompass different calibrations by each laboratory, differences in instruments, and different analysts.

In the FACDQ Pilot Study, ASTM, OGWDW and ACIL (for censored methods only) detection limits were determined based on data generated over approximately three weeks, and compared to blank results generated over approximately six months. The false positive rates determined for these limits based on the blank data often exceeded the target 1%. This may have been due to the procedures not capturing all sources of routine variability as represented in the long-term blank data; however, other factors such as incorrect estimates of the blank bias or the small number of blanks for some laboratories may have also led to these high false positive rates.

It is also important to note that in some instances, inter-laboratory variability so greatly exceeds intra-laboratory variability that attempting to capture inter-laboratory variability by conducting tests over time may not only be complicated, but difficult.

vii. Does the procedure perform on-going verification of estimates? (DK/KO)

Procedure	Verification	Test
DL		
USEPA MDL	None	Optional Step 7 (iterative)
ACIL Uncensored	Annual Re-Estimation based on LRBs	Annual F-test of variance of old % RSD of Old LRBs vs. New LRBs
ACIL Censored	None	
CG	Determination of 1% FPR	Check LRB FPR < 2%
EBMUD	Batch by batch	FNQC +/-50%
Hubaux-Vos	None	
ASTM IDE	None	
QL		
USEPA ML	None	
ACIL Uncensored	Quarterly Analysis of LFB	Annual Determination of %RSD of LFBs, if >20%

ACIL Censored	Quarterly Analysis by LFB and	Presence / absence or LFB.
CG	Batch by batch	Mean spike recovery +/-50%
LC-MRL	Batch by batch	MRL Check +/-50%
ASTM IQE	None	

viii. Is the procedure capable of calculating limits using matrices other than lab reagent grade water? (BE)

Procedures being tested do not include evaluation in real world samples or matrices that represent real world samples.

The challenges faced in applying any procedure to real world samples, are encountered with any procedure. It is difficult to obtain real world samples with no analyte present. Even then there might be undetectable levels of analyte present. This effect is difficult to characterize for any procedure.

ix. Does the procedure use only data that results from test methods conducted in their entirety? (TF/KO)

Lab-Determined Detection Limits (DLs) and Quantitation Limits (QLs)

Both the Hubaux-Vos and the LCMRL procedures (for single lab and interlab) require or recommend that samples be carried through the entire analytical procedure in the development of DL/QLs. This requirement ensures that all likely sources of variability and bias associated with analytical methodologies are taken into account when developing DL or QL estimates.

The ACIL requires that the entire analytical procedure be used to develop detection and quantitation estimates. This requirement ensures that all likely sources of variability and bias associated with analytical methodologies are taken into account when developing DL and QL estimates.

The IDE and IQE (for single lab and interlab) require that all sources of measurement variability be taken into account including sample preparation steps.

The rest of the single lab non-pilot tested procedures and multi lab procedures clearly specify that the entire analytical methodology including preparative steps must be employed when analyzing DL/QL study samples. This requirement ensures that all likely sources of variability and bias associated with analytical methodologies are taken into account when developing DL and QL estimates.

For **Uses #2-9**, this criterion is satisfied and no adverse impacts are anticipated for all the procedures except for the LT-MDL because of incomplete documentation it is unclear how this criterion might be satisfied.

The MDL (and therefore the ML) procedure explicitly require that the analyst/ laboratory "Take a minimum of seven aliquots of the sample to be used to calculate the method detection limit and process each through the entire analytical method."

x. Does the procedure explicitly adjust or account for situations where method blanks always return a non-zero result/response? (JPH)

According to the characteristics chart most procedures are identified as having this characteristic except possibly the EPA MDL/ML and the LCMRL; however each procedure may meet the object in a different way. The ACIL, CG and Lab QC procedures all use the method blank results to calculate at least the initial estimates when non-zero results are reported. All three procedures also add the mean of the method blanks to the calculated estimates to compensate for blank bias. Both the ACIL and CG procedures also use the non-zero blank data for long term estimates, however the Lab QC procedure relies on a spiked method blank (FNQS). The IDE/IQE and LCMRL/HV procedures are regression based and take the zero intercept into account when deriving their estimates, therefore non-zero blank data and blank bias are both accounted for in their estimates. The EPA MDL/ML does not consider or use blank data in any of its estimates, however if a spiked blank is impacted by background contamination this could have some influence on the overall variability and impact the final estimates. Blank bias is not accounted for by the MDL/ML.

Use of a blank is only implied by Step 3(a) of the MDL procedure: "If the MDL is to be determined in reagent (blank) water, prepare a laboratory standard (analyte in reagent water) at a concentration which is at least equal to or in the same concentration range as the estimated method detection limit (Recommended between 1 and 5 times the estimated method detection limit).

xi. Does the procedure explicitly adjust or account for situations where method blanks are intermittently contaminated? (JPH)

The Technical Work Group agreed at their December 13, 2006 meeting that this document reflects their best effort at addressing intermittent contamination at this point in time.

The Technical Work Group also agreed to revisit this document after a procedure(s) is recommended in order to determine if the chosen procedure(s) adequately addresses intermittent contamination. Changes to the procedure(s) or this document may be made at that time based on Technical Work Group assessment of how well this issue is handled.

Definition

Intermittent blank contamination is defined as present when 5% or greater than 5% of method blank results for a given analyte have results greater than the calculated Lc.

Detection and Quantitation Limits

During the method validation studies or initial determinations of detection (LC) and quantitation limits (QL) by laboratories, there is often insufficient data to test normality in the method blank population. Therefore, it is recommended that outliers not be discarded for use unless an error or a cause for the outlier measurements has been identified and corrected. Errors and analytical artifacts should be investigated to evaluate the cause of any low-side or negative outliers as well. All method blanks associated with "reported" results shall be included when determining the frequency of intermittent blank contamination or subsequent calculations.

xii. Is the procedure clearly written with enough detail so that most users can understand and implement them? (BE)

Though fourteen labs stated that the procedures were written fairly clearly, some commented on specific areas of improvement. One lab commented that the ACIL procedure was not clearly written and that it was difficult to discern which equation should be used for various steps. Another lab expressed concern regarding the use of the term "Minimum Level" (ML) instead of "Limit of Quantitation" (LOQ) and suggested standardizing the term LOQ. Historically EPA has always stated the ML as the "minimum level of quantitation" and abbreviated it to ML.

While the majority of labs stated that the ACIL was clearly written, not all laboratories interpreted the procedure in the same way. This resulted in multiple laboratories setting the ACIL ML equal to two times the ACIL MDL, or the lowest expected result determined in the ACIL procedure. The written procedures instructs the user to set the ACIL ML to the spike level used to assess the MQOs. Other laboratories did not choose the spike level at a level appropriate to the ACIL procedure instructions; for example, several labs did not spike at a level at or above the determined ACIL MDL for uncensored methods.

There were also a number of comments on the LCMRL procedure. Eleven labs indicated that they found the LCMRL procedure to be clear and easy to perform using the software provided. Labs that had the resources to do so were able to use a computer programmer to format the data in a manner that allowed it to be imported directly into the LCMRL calculator. Labs that did not have such resources reported that the manual data entry required was very time-consuming and error-prone.

Two labs reported that they found the procedures for determining the LCMRL to be difficult to comprehend. One of these labs indicated that they found the examples to be helpful in interpreting and evaluating the calculations and graphs, but felt that the procedure lacked direction on how best to produce data for these analyses. One laboratory also suggested that if it was important that replicate analyses be analyzed on non-consecutive days, as was required in the study, this requirement should be described in the LCMRL procedure.

There needs to be more discussion as to whether difficulties with ACIL and LCMRL procedures would preclude them from being used routinely, or the specific changes that would need to be made to each to assure that they produce the intended result.

xiii. Is the procedure cost effective? (LL/RB)

Several laboratories did comment on the costs of procedures. Given that many of the procedures have fundamentally different objectives, in most instances it is not meaningful to attempt to make cost comparisons between procedures. For instance, the objective of the ACIL DL and QL procedures is to document the labs capability to achieve their stated reporting limit; not to identify the lowest concentration where they can achieve a given set of MQOs. The FACDQ needs to determine if there is a need or use for such a procedure. Also, the FACDQ has not established exactly which MQOs will be specified in the final recommendations. Not all of the procedures under consideration address all of the possible MQOs that might be specified and thus, may require some modification to meet the ultimate needs of the FACDQ. Until that is determined and/or modifications are completed, cost effectiveness cannot be addressed. Discussing costs is also difficult because some procedures can generate an estimate with as little as seven analyses but allow for improvement in that estimate by incorporating additional data (e.g. up to 100 additional data points) developed over time. If we have a means of calculating a confidence interval for the estimate, we could compare the cost of achieving an estimate at a give confidence level for one procedure vs. another (where appropriate).

However, with those caveats, some direct comparisons are possible. For instance the Hubaux-Vos/LCMRL, the Consensus procedure DL and QL and a single lab implementation of the ASTM IDE and IQE all attempt to identify the lowest concentration that a give set of MQOs can be achieve. For censored methods, the minimal data set for the H-V/LCMRL DL and QL estimates is 20 analyses; for the Corresponding estimates using the Consensus Group procedure the minimum is 14 analyses but given that the intent is to use these initial assessments as a starting point for development of a more reliable long term estimate, one would have to consider the costs of those additional measurements. However the pilot study suggests that on occasion, an additional concentration level may be required to get the necessary data set for the H-V/LCMRL and similarly, if the correct QL is not estimated for the Consensus procedure, an additional seven replicates may be required. A single lab IDE and IQE would require a minimum of 30 analyses each although careful study design could result in some overlap in the spiking concentrations that could reduce this number.

For uncensored methods, the Consensus procedure has the advantage of being able to utilize laboratory blanks which are generally performed with each batch of samples to estimate the DL. It is not possible to predict or asses the added expenses associated with getting all the analyte spikes at the lowest appropriate range in a multi-analyte analytical method because this will depend on the number of analytes and the similarities or differences in method performance with respect to those analytes. So, for censored method the costs for the Consensus procedure and the HV/LCMRL procedures are nominally the same but the single lab IDE IQE being somewhat higher. For uncensored methods, the Consensus procedure has the potential for being less costly. However, these differences are small enough that the uncertainties in how many different spikes would need to be analyzed to optimize spiking concentration could easily tip the balance one way or the other.

The Lab QC targets control of false negatives and thus would have to be coupled with another procedure which would address quantitation.

The ASTM IDE and IQE and an interlab implementation of the H-V/LCMRL are under consideration for inter-lab procedures. The minimal number of analyses is essentially the same as discussed previously for the use of these procedures in the single lab context. Thus, a minimal H-V/LCMRL inter-lab study would cost less than a minimal ASTM IDE-IQE study but this difference could be somewhat mitigated through careful design of the ASTM procedures. If the initial spiking concentrations fail to bracket the DL or QL, additional spikes would be required for both procedures but it is not possible to determine how frequently this may occur or whether it would be more likely to occur for the H-V/LCMRL vs. the ASTM IDE and IQE.

No multi-lab procedures have been proposed so no cost effectiveness evaluation is possible.

It should be noted that all procedures evaluated during the pilot study will be more costly than the current MDL.

xiv. Does the procedure assess multi-laboratory and inter-laboratory variability when data from more than one lab is used? (LL/RB)

Single laboratory procedures account for intra-laboratory (within lab) variability only, because they are determined independently. A multi-laboratory estimate pools the variability of several intra-laboratory estimates, thus accounting not only for intra-laboratory variability, but also some

inter-laboratory (between lab) variability. A true inter-laboratory estimate is performed using the same set of test samples for all laboratories. Only a true inter-laboratory estimate takes into account all inter-laboratory error including both the precision and bias components of inter-laboratory variability.

The ACIL DL + QL, HV+ LCMRL, IDE/IQE (implemented as a single lab procedure), Lab QC (Osborn), Consensus Group DL + QL and the MDL/ML all produce single lab estimates for detection and quantitation. Presumably, data from individual laboratories performing these procedures could be used for in a multi-lab procedure but to date, no such procedure has been proposed.

The IDE/IQE is the only procedure that was originally designed to be an inter-lab procedure. However, the HV+ LCMRL can be implemented as an inter-lab procedure as was tested in the pilot study. Because of the design and intent of inter-lab studies, it is hard to envision the use of the resulting DL or QL estimates in a multi-lab context.

xv. Is the procedure applicable to all users and test methods? (BE/KM)

The ACIL procedure when used correctly, performed well for all methods. In part this is because the ACIL procedure was expressly designed to meet the MQOs identified for the pilot (although it could be readily modified to meet other MQOs). In many cases failures to meet the MQOs when using the ACIL procedure were due to failure to follow the procedure correctly, especially the direction that the QL must be at least 2 times the DL. Other failures of the ACIL and other procedures were due to intermittent blank contamination problems.

A revised ACIL procedure should provide a revision that addresses and/or overcomes these problems. If recommended, the ACIL procedure should be modified to address these issues.

There were quite a few instances for which the LCMRL could not be calculated, especially for methods 625 and 608 with analytes for which recovery never reaches 100 %.

Interlab ASTM methods did not performed well for methods for which large between-laboratory variability is expected (e.g., Method 625, for which between-lab variability was often large, at least partially due to different extraction techniques).

If recommended, the ASTM procedures should be modified to address these issues

b. Pilot Study MQOs (JPH/KM/RB)

i. What did we learn from the ability of the procedures to meet the MQOs?

Pilot Study MQO Background

Measurement Quality Objectives (MQOs) for bias, precision, false positive error rate and false negative error rate were specified for use in the pilot study by the FACDQ. These MQOs were as follows;

- Bias – 50% to 150% mean recovery
- Precision – 20% or less relative standard deviation
- False Positive (FP) Error Rate – 1% or less at the L_c
- False Negative (FN) Error Rate - 1% or less at L_c for results at the L_Q

The performance of each of the procedures evaluated against each of the pilot study MQOs, for each of the analytical methods tested are summarized in the following tables;

PR.III.b.i.(Pilot Study MQO Performance for Method 300.0)

PR.III.b.i.(Pilot Study MQO Performance for Method 335.4)

PR.III.b.i.(Pilot Study MQO Performance for Method 200.7)

PR.III.b.i.(Pilot Study MQO Performance for Method 608)

PR.III.b.i.(Pilot Study MQO Performance for Method 625)

Each table lists the percentage of time that the specific procedure achieved the pilot study MQO for a particular analytical method, based on verification samples. Both interlaboratory estimates and single-lab estimates, along with results for data with statistical outliers removed and without statistical outlier's removal are presented in each table. All performance estimates were based on modeling, except for mean recovery for Method 300.0, which used interpolation. This is was done because modeling generally provided the most accurate estimates with the exception of Method 300.0 mean recovery.

The performance of all analytes, for the analytical method specified, are summarized together. So for example, Method 334.5 (Total Cyanide), which only had a single analyte will have achieved the interlaboratory MQO either 100% of the time or 0% of the time, because only a single interlaboratory estimate is available for evaluation. However, for interlaboratory estimates associated with other methods and for all single laboratory estimates greater gradations are possible, because multiple estimates were generated. The Office of Ground Water and Drinking Water (OGWDW) LCMRL could not achieve a quantification limit estimate a large percentage of the time, for Methods 608 and 625 in particular.

The intent of these tables are to provide the reader an brief overall summary of the pilot study results derived from a very large number of individual results and therefore will have a number of shortcomings. For example, the percent of time the FP error rate criteria, of less than or equal to 1% at L_c , was achieved for single-lab limits for all procedures and analytical methods are presented graphically in attachment 'PR.III.c.(Single-Lab False Positive Performance Graph)'. This is a subset of the information presented in the tables only in graphic form. Notice that a higher percentage of FP error rate success against the pilot MQO was achieved by the ASTM Y_c & L_c procedures than by the ACIL MDL procedure, for the uncensored methods (300.0, 335.4 and 200.7). However, if we look at the mean FP error rate for single laboratory limits the ACIL MDL has much lower average error rates (about 2%)

than all other procedures for method 335.4 and 200.7, as shown graphically in attachment 'PR.III.c.(Single-Lab False Positive Rate Graph)'. These are different ways of looking at the same set of data. We refer the reader to the Pilot Study Report, section II.d.viii. for more details regarding performance of each procedure against the pilot study MQOs.

Summary of Procedure Performance against the Pilot Study MQOs

ACIL-ML and MDL

The ACIL procedure is a single laboratory procedure and was evaluated as such for the pilot study. (It is possible however to generate pooled interlaboratory estimates from the single laboratory estimates.) The ACIL procedure had greater than 90% success in achieving the pilot study MQOs as evaluated against the ongoing verification samples, with the following exceptions;

- Method 300.0 FP MQO success was under 50%
- Method 200.7 FP MQO success was under 50%
- Method 335.4 FP success about 60% (outliers removed), FN success about 40% and precision MQO success only around 30%
- Method 608 FP MQO success about 75% (outliers removed), FN and precision success about 60%
- Method 625 FN success 70%, precision and bias success around 75%

False positive error rate would likely be improved for this procedure if the intermittent blank contamination protocol were included. If all labs followed the procedure accurately high success rate would also be anticipated. Finally if the on-going verification portion of the procedure were implemented success rate should also improve.

ASTM IDE, IQE20, IQE30, Yc and Lc

The ASTM procedure was designed as an interlaboratory procedure, but was also evaluated as a single laboratory procedure during the pilot study.

Interlaboratory Bias and Precision

The interlaboratory bias MQO was achieved 100% of the time for the IQE20. The interlaboratory precision MQO was achieved at least 70% of the time for methods 300.0, 335.4, 200.7 and using IQE20. The IQE20 only achieved the precision MQO 42% of the time for Method 608, only 13% of the time for method 625 and never for method 8082. Since the precision success rate of the IQE20 was substantially better than the IQE30, we would anticipate an IQE15 or IQE10 would have close to a 100% success rate, assuming the method is capable of achieving at least 20% RSD somewhere throughout the range of the method.

Interlaboratory False Positives and False Negatives

The interlaboratory FP MQO success rate was on the order of 90% for both the Lc and Yc once outliers were removed, with the exception of methods 200.7 and 608, which were near 60%. It is anticipated that additional data would result in a better estimate and more accurately predict the FP error rate. The FN MQO success rate of the Lc for results at the IQE20 was 70% or better, with the exception of Method 200.0, which was 64%. The procedural objective of the ASTM procedure is 5% FN error rate, so if this was lowered to 1% a high success rate would be achieved. A significantly improved FN success rate would be expected at both the Yc and the Lc for results at the IQE10 or IEQ15.

Single-Lab Bias and Precision

The single-lab bias MQO success rate for the IQE20 ranged from 47% to 94% depending upon the Method, with some improvement observed when statistical outliers were removed. The single-lab precision MQO success rate for the IQE20 ranged from 57% to 89% depending upon the Method, with some improvement observed when statistical outliers were removed. The single-lab bias MQO success rate at the IQE20 was always achieved for method 8082 and the precision MQO success rate was about 50%, after outlier removal.

Single-Lab False Positives and False Negatives

The single-lab FP MQO success rate for the Yc and Lc ranged from 55% to 95% depending upon the Method, once statistical outliers were removed. The single-lab FN MQO success rate at Yc for results at the IDE ranged from 68% to 83%. The IDE should be the floor for any quantitation estimate. The procedural objective of the ASTM procedure is 5% FN error rate, so if this was lowered to 1% a high success rate would be achieved. If the FN error rate at the Yc or Lc were evaluated for results at the IQE10 or IQE15 the success rate should be substantially greater.

OGWDW LCMRL and HV

The Office of Ground Water and Drinking Water (OGWDW) LCMRL and Hubaux-Vos (HV) procedures were designed as single laboratory procedures, but were also evaluated as interlaboratory procedures during the pilot study.

Interlaboratory Performance

The interlaboratory bias, precision and FN MQOs were successfully met nearly 100% of the time. The only word of caution here is that the LCMRL was unable to obtain a valid limit for a large percentage of Method 608 and 625 parameters. Wider spike ranges and/or wider procedural bias criteria would have resulted in valid limits being obtained for more analytes, but would also have most likely reduced the pilot MQO success rate. The FP MQO success rate ranged from 71% to 100% once statistical outliers were removed, with the exception of Method 200.7 which had a 54% success rate

Single-Lab Performance

The single-lab bias and precision MQO success rate was 90% to 100% for all Methods with three exceptions, precision for Method 335.4 (83% OR, 60% WO), precision for Method 608 (86% OR) and precision for Method 8082 (50% OR). This is understandable since the LCMRL does not directly control the precision of its estimates. The single-lab FP MQO success rate ranged from 49% to 89%, after outlier removal. The single-lab FN MQO success rate at the HV detection limit for results the LCMRL ranged from 88% to 98%, with the exception of Method 335.4, which had a 50% success rate. Caution is recommended when interpreting the performance results for Methods 608 and 625, due to the high frequency of not obtaining valid LCMRL results.

c. Evaluate if procedures achieved their intended purpose? (JPH/KM/RB/SW)

i. How do these results differ from above?

Procedure Objectives Background

Each procedure's intended purpose (procedural objectives or POs) for bias, precision, false positive error rate and false negative error rate are defined as part of the procedure. The POs are as follows;

ACIL-ML and MDL

- Bias – 50% to 150% mean recovery
- Precision – 20% or less relative standard deviation
- False Positive (FP) Error Rate – 1% or less at the MDL
- False Negative (FN) Error Rate - 1% or less at MDL for results at the ML

ASTM IDE, IQE20, IQE30, Yc and Lc

- Bias – None specified
- Precision – 20% RSD for IQE20, 30% RSD for IQE30
- False Positive (FP) Error Rate – 1% at the Yc or Lc
- False Negative (FN) Error Rate - 5% at the Yc or Lc for results at the IDE

OGWDW LCMRL and HV

- Bias – 50% to 150% individual recovery
- Precision – None specified
- False Positive (FP) Error Rate – 0.5% or less at the HV
- False Negative (FN) Error Rate – 0.5% or less at HV for results at the QL (LCMRL)

The performance of each of the procedures evaluated against each of the POs, for each of the analytical methods tested are summarized in the following tables;

PRIII.c.i.(Procedural Objectives Performance for Method 300.0)

PRIII.c.i.(Procedural Objectives Performance for Method 335.4)

PRIII.c.i.(Procedural Objectives Performance for Method 200.7)

PRIII.c.i.(Procedural Objectives Performance for Method 608)

PRIII.c.i.(Procedural Objectives Performance for Method 625)

Each table lists the percentage of time that the specific procedure achieved its own Procedural Objectives (POs) for a particular analytical method. Both interlaboratory estimates and single-lab estimates, along with results for data with statistical outliers removed and without statistical outlier's removal are presented in each table. All performance estimates were based on modeling, except for mean recovery for Method 300.0, which used interpolation. This is was done because modeling generally provided the most accurate estimates with the exception of Method 300.0 mean recovery.

The performance of all analytes, for the analytical method specified, are summarized together. So for example, Method 334.5 (Total Cyanide), which only had a single analyte will have achieved the interlaboratory PO either 100% of the time or 0% of the time, because only a single interlaboratory estimate is available for evaluation. However, for interlaboratory estimates associated with other methods and for all single laboratory estimates greater gradations are possible, because multiple estimates were generated. The Office of

Ground Water and Drinking Water (OGWDW) LCMRL could not achieve a quantification limit estimate a large percentage of the time, for Methods 608 and 625 in particular, so very high PO achievement rates, may be misleading for evaluating the overall performance of this procedure. We refer the reader to the Pilot Study Report, section II.d.v. for more information regarding performance of each procedure against their POs.

Summary of Procedure Performance against the Procedural Objectives

ACIL-ML and MDL

The ACIL procedure is a single laboratory procedure and was evaluated as such for the pilot study. (It is possible however to generate pooled interlaboratory estimates from the single laboratory estimates.) The ACIL Procedural Objectives were identical to the pilot study MQOs. For a summary of the performance of this procedure based on the on-going verification samples refer you to section III.b. "ACIL-ML and MDL". Another measure of success is how often the procedure was able to meet its objectives during the initial determination of the estimates. The percentage of time the ACIL procedure did not meet its precision and bias objectives are listed below. The inability to meet procedure objectives is due to two reasons.

1. The method precision and bias never achieved the target objectives at any spike concentration evaluated (failure of the method).
2. The pilot study design did not allow for iteration of spike concentrations until the objectives could be met (failure of the study).

<u>Method</u>	<u>>20% RSD</u>	<u><50% Recv</u>	<u>>150% Recv</u>	<u>Overall</u>
300.0	0%	0%	0%	0%
200.7	3.4%	0%	0%	0%
335.4	14%	0%	0%	0%
608	20%	1%	0%	21%
625	26%	17%	1%	35%

ASTM IDE, IQE20, IQE30, Yc and Lc

The ASTM procedure was designed as an interlaboratory procedure, but was also evaluated as a single laboratory procedure during the pilot study.

Interlaboratory Precision

The interlaboratory precision PO was achieved at least 70% of the time for Methods 300.0, 335.4, 200.7 and using IQE20. The IQE20 only achieved the precision MQO 42% of the time for Method 608, only 13% of the time for method 625 and never for method 8082. The interlaboratory precision PO was achieved for the IQE30 at least 73% of the time for all Methods except Method 625, which had a success rate of 44% after outlier removal. These success rates seem to be reasonable, since the POs were set at 20% and 30% RSD we might expect that half of the time the objective would be achieved.

Interlaboratory False Positives and False Negatives

The interlaboratory FP PO success rate was on the order of 90% for both the Lc and Yc once outliers were removed, with the exception of methods 200.7 and 608, which were near 60%. It is anticipated that additional data would result in a better estimate and more accurately predict the FP error rate. The FN PO success rate of the critical level for results at the IDE ranged from 63% to 86% for Yc and from 50% to 100% for Lc, with the exception of method 335.4, where the

FN PO was not achieved. A significantly improved FN success rate would be expected at both the Yc and the Lc for results at the IQE10 or IEQ15.

Single-Lab Precision

The single-lab precision PO success rate for the IQE20 ranged from 57% to 89% depending upon the Method, with some improvement observed when statistical outliers were removed. The single-lab precision PO success rate for the IQE30 ranged from 71% to 89% depending upon the Method, with some improvement observed when statistical outliers were removed. These success rates seem to be reasonable, since we would anticipate the PO to be achieved at least half of the time.

Single-Lab False Positives and False Negatives

The single-lab FP PO success rate for the Yc and Lc ranged from 55% to 95% depending upon the Method, once statistical outliers were removed. The single-lab FN PO success rate at Yc for results at the IDE ranged from 71% to 89%. The single-lab FN PO success rate at Lc for results at the IDE ranged from 41% to 87%.

OGWDW LCMRL and HV

The Office of Ground Water and Drinking Water (OGWDW) LCMRL and Hubaux-Vos (HV) procedures were designed as single laboratory procedures, but were also evaluated as interlaboratory procedures during the pilot stud

Interlaboratory Performance

The interlaboratory bias and FN POs were successfully met 100% of the time, especially after statistical outliers were removed. The only word of caution here is that the LCMRL was unable to obtain a valid limit for a large percentage of Method 608 and 625 parameters. The FP PO success rate prior to outlier removal ranged from 0% to 43%, except for Method 300.0 (100%). Following the removal of statistical outliers FN PO success improved substantially to the range of 71% to 100% except for Method 200.7 (18%).

Single-Lab Performance

The single-lab bias PO success rate was 100% for all Methods when an LCMRL could be determined. The single-lab FP MQO success rate ranged from 35% to 89%, generally showing slight improvement after outlier removal. The single-lab FN PO success rate at the HV detection limit for results the LCMRL ranged from 86% to 98%, with the exception of Method 335.4, which had a 40-50% success rate. Caution is recommended when interpreting the performance results for Methods 608 and 625, due to the high frequency of not obtaining valid LCMRL results.

d. Additional Considerations or Evaluations. (DR/RR/BA)

The pilot study was limited to three candidate procedures for the calculation of detection and quantitation limits (DLs/QLs). The Technical Work Group's (TWG) task was to complete all assignments and studies within the resources and timeframes provided. Due to these constraints, it was not possible to evaluate all candidate procedures, nor was it possible to hybridize the best portions of any existing procedure(s) into a composite procedure(s) and test the hybrids. This section describes additional features or guidance that would help EPA implement a new approach at 40 CFR Part 136 to calculating DLs and QLs. This section also describes additional work to be completed before proposing any new DL or QL procedure(s).

Pooling Single-lab Data into National Limits

The calculation procedures piloted in 2006 included: (1) a single laboratory (the ACIL) procedure, (2) an interlaboratory (IDE/IQE) procedure, and (3) a single laboratory (H-V/LCMRL) procedure conducted in an interlaboratory mode. In the single laboratory mode each lab prepared and ran customized spikes to calculate their detection and quantitation limit for each pollutant in the study. In the interlaboratory mode, each laboratory analyzed the same set of spikes. These spikes were designed by the TWG study coordinators comprised of members of the TWG, and prepared by an independent vendor.

The IDE and IQE calculation software pools results from each participating laboratory to obtain a detection and quantitation limit for each pollutant in the study. The H-V/LCMRL procedure calculates a detection and quantitation limit for each pollutant and from each laboratory in the study. These single lab results are pooled to obtain one limit each for detection and quantitation. For the ACIL procedure, each lab calculates their detection and quantitation limit for each pollutant in the study. However, there is no standard way to pool these single lab results to obtain a detection and quantitation limit for each pollutant in the study. A standardized pooling procedure might be used to set national limits or performance benchmarks.

As part of a new approach to calculating detection and quantitation limits, if EPA approves a procedure(s) that allows laboratories to individually calculate their detection and quantitation limits, EPA also might specify a standard way to pool single laboratory results into national limits. The IDE/IQE procedures have built in ways to pool single lab results into national limits. For the pilot, H-V/LCMRL data was pooled in a way modeled on the IDE/IQE. Procedures, such as the ACIL, Consensus Group and MDL/ML do not have this feature. However, these single lab procedures provide a relative picture of each lab's performance in an interlaboratory study, and verification of each lab's capability to conduct single-lab, routine analyses, which is very important to the laboratory community.

Consistency Across Programs

Many federal and state programs reference 40 CFR Part 136 Appendix B for the determination of Method Detection Limits (MDLs), which in turn is linked to the minimum level (ML). The charter and charge to the FACDQ is to consider uses of DLs and QLs in Clean Water Act programs. It is not known if other programs will accept any new MDL and ML procedures recommended by the FACDQ. While every effort should be made to have a uniform DL/QL procedure across all environmental monitoring programs, the TWG recommends that the promulgation at Appendix B of new DL/QL procedures move forward with or without CAP. Programs that do not accept the change in procedure(s) will have to decouple (unCAP) their reference to 40 CFR Part 136 Appendix B.

There are several accreditation programs recognized in the United States that have DL/QL standards. Promulgation of the new DL/QL procedure will require compliance with all laboratories performing analytical tests under the CWA program. We recommend notifying these accreditation programs to prepare for a change to their standards related to DL/QL requirements.

Improve Clarity of the Procedures

In the pilot several laboratories did not follow directions, or used concentrations for the initial spike level that were either too high or too low. A longer study would have corrected some of these issues as a reiterative process would have identified the correct spike concentration. However, some of these errors may be due to the procedure as currently written. We recommend that the final procedure be clearer, provide examples, and have outreach/training available.

Matrix Effects

Because the study used blank water as the matrix, it is unknown how a change in matrix might have affected the results. That is, would the performance (not just the calculated limit) of the procedure be affected by choice of matrix? For example, would the ability to meet the study MQOs, or provide ongoing laboratory verification of performance data have been compromised by matrix effects, e.g., intermittent blank contamination?

Specifying New or Changing Existing Limits

Calculating new limits with procedures that require several days or months of data to establish routine and verifiable performance presents significant challenges to laboratories, especially small laboratories. Many accreditation programs recommend or require at least annual verification of a lab's DLs or QLs, or upon a significant change, such as a new or modified instrument or analyst. In setting national limits for new pollutants or new monitoring technologies or methods, EPA conducts interlaboratory studies to develop performance specs. These studies are costly, time-consuming, and require extensive peer review. Additional time to establish a long term DL or QL: increases this burden, hinders approval of less costly, more accurate or more sensitive methods, or delays monitoring with methods for pollutants of emerging concern. The TWG could recommend ways to address these problems.

We believe EPA also would welcome guidance on creative, less burdensome yet defensible ways to develop new national limits for new or existing pollutants or analytical technologies.

Other Recommendations

To execute a logical and uniform implementation of new DL or QL procedures the following should be considered:

1. The recommended procedure(s) must be promulgated for compliance uses.
2. All new analytical methods and updated methods must have a DL/QL for each analyte.
3. The promulgated procedure(s) must be used to demonstrate performance against a required limit (e.g. permit limits), or to establish the performance of a new method or analyte. Users may use any part 136 analytical methods for CWA uses, but to demonstrate their DLs/QLs they must use the promulgated DL/QL procedure. Thus when organizations, such as ASTM, AOAC, Standard Methods, USGS publish new methods it may difficult (in cases where there is no clearly equivalent EPA method) to compare and evaluate their published QLs/DLs, if they do not use the promulgated procedure.

4. Methods already promulgated without DL/QL limits identified may pose a problem with the new DL/QL procedure determinations and existing permit language.
5. A table of DL/QL identifying minimum requirements may be promulgated and updated as needed. All labs must meet or have limits below the table DL/QL numbers.
6. Recognition that various federal and state programs (i.e. reasonable potential, Great Lakes Initiative, etc.) may require determination and reporting of the lowest DL/QL limits independent of the table described in item 5 above.
7. FACDQ's recommendations for a new promulgated DL/QL procedure(s) likely will represent significant changes to current CWA practices.
 - a. We need innovative, burden reducing and defensible ways to develop new limits and provide ongoing verification of a laboratory's ability to achieve those limits.
 - b. To replace existing MDLs or MLs with new limits could require extensive, primary data from new interlaboratory studies. For future multi-laboratory determinations should we recommend use of secondary data from DMRs if single lab DL/QL determinations are listed on the reports as "information for possible future setting of new national limits?"

i. Implementation in CWA programs (RR)

The Great Lakes Initiative (GLI) was developed in an effort to harmonize disparate water quality standards between the Great Lakes states and was designed to specifically address the unique environmental and socioeconomic issues in the Great Lakes basin. The program initiated in 1989 by the EPA and the Great Lakes States to further address the environmental concerns identified in the Great Lakes Toxic Substances Control Agreement. The GLI was intended to provide a forum for the Great Lakes States and the EPA to develop uniform water quality criteria, anti-degradation policies, and implementation procedures for the Great Lakes Basin so as to create an even playing field for all industries in the region. Following publication of the guidance in 1995, states had to modify their water quality rules in order to be at least as stringent as the GLI standards, and submit their programs to EPA for approval. The guidance included water quality criteria for a number of Persistent Bioaccumulative and Toxic (PBTs) chemicals. PBT pollutants pose risks because they are toxic in very small quantities, persist in ecosystems, bioaccumulate in food chains, and can travel great distances (e.g., in air or water, or in foodstuffs, equipment or products). While much progress has been made to reduce loadings, PBTs continue to threaten human and ecosystem health, and since these pollutants transfer readily across air, water, and land, and span boundaries of programs, geography, and generations, environmental agencies increasingly recognize that they need to go beyond single-media control approaches. Several of the PBT chemicals have criterion values well below nominal detection limits in EPA approved methods. For example, the criterion in the GLI for protection of wildlife for 2,3,7,8-TCDD is $3.1E-9$ ug/L, compared to a detection limit of $10E-6$ ug/L for Method 1613. For PCBs, the interim criterion announced by U.S. EPA in 1997 for protection of human health (following a court challenge of the original criterion) was $2.6E-5$ ug/L, compared to a method detection limit for Aroclor 1242 of 0.65 ug/L for Method 608. For a number of pollutants, GLI criteria remain lower than nationally recommended water quality criteria. Because EPA announced in 2000 a ban on mixing zones for new discharges, and a phaseout for existing discharges in the Great Lakes Basin, there is increasing need to confirm that discharges are either not exceeding low water quality based effluent limitations for the PBT chemicals of concern or are being lowered in the direction of eventually meeting these limits. In order to accurately monitor and assess the impact

of PBTs on the Great Lakes system, it is critical that scientists, regulators, and industry have analytical methods that can correctly measure these chemicals at low levels.

III. Recommendations and Rationale

To be completed.