Chapter 4
Evaluation Criteria

This chapter presents the criteria developed by EPA as a means for evaluating and selecting acceptable detection and quantitation limit approaches for use in Clean Water Act (CWA) programs. These criteria reflect EPA's careful consideration of the issues identified and discussed in Chapter 3, including EPA's needs under CWA programs. A total of six criteria were established, and are discussed in Sections 4.1 - 4.6. The six evaluation criteria are:

Criterion 1: The detection and quantitation limit approaches should be scientifically valid.
Criterion 2: The approach should address demonstrated expectations of laboratory and method performance, including routine variability.
Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance.
Criterion 4: The detection level approach should identify the signal or estimated concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in well-operated laboratories.
Criterion 5: The quantitation limit approach should identify the concentration that gives a recognizable signal that is consistent with the capabilities of the method when a method is performed by experienced staff in well-operated laboratories.
Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act (CWA), and should support state and local obligations to implement measurement requirements that area at least as stringent as those set by the Federal government.

Section 4.7 presents additional principles recommended by stakeholders commenting on EPA's assessment.

4.1 Criterion 1

Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

The concept of scientific validity is widely accepted but loosely defined. For the purposes of this evaluation, a detection/quantitation approach or methodology will be considered scientifically valid if it meets the following conditions:

- It can be (and has been) tested,
- It has been subjected to peer review and publication,
- The error rate associated with the approach or methodology is either known or can be estimated,
- Standards exist and can be maintained to control its operation (i.e., it is supported by well-defined procedures for use), and
- It has attracted (i.e., achieved) widespread acceptance within a relevant scientific community.

While EPA acknowledges that other measures could be established to demonstrate scientific validity, EPA has adopted the conditions cited because they reflect those discussed by the U.S. Supreme Court as pertaining to assessments of scientific validity when considering the admissibility of expert scientific testimony[2]. These conditions also are directly relevant to EPA's needs.

Some stakeholders supported the use of objective criteria for determining scientific validity, but questioned the appropriateness of using criteria that were designed for courts and juries to support scientific decisions made by scientific experts. EPA carefully reviewed the Court's conditions for demonstrating the scientific validity of an expert's reasoning or methodology, and believes that these conditions are appropriate for demonstrating the scientific validity of any scientific approach or methodology, including those that might be used to establish detection and quantitation limits under CWA. EPA further believes these criteria are consistent with the EPA Science Policy Council's assessment factors for evaluating the quality of scientific and technical information (EPA 100/B-03/001, June 2003), including the extent to which technical information and data are peer reviewed and appropriately tested. However, EPA is willing to consider alternative or supplemental criteria for evaluating scientific validity as it moves forward with the stakeholder process.

Stakeholders agree that detection and quantitation levels should be based on sound scientific principles, and note that low-cost and/or simple approaches should not be selected if inaccurate or unmeasurable limits may result. Stakeholders also noted that some of the conditions listed above (e.g., the condition that an approach or methodology should have attracted widespread acceptance within a relevant scientific community) have the potential for favoring concepts already adopted and required by regulatory agencies. EPA agrees that this is a valid concern, and therefore, will consider the overall validity and practicality of new approaches.

4.2 Criterion 2

Criterion 2: *The approach should address realistic expectations of laboratory and method performance, including routine variability.*

As discussed in Chapter 3 of this Assessment Document, the detection and quantitation limit(s) for an analyte in an analytical method can be established from a single-laboratory study, multiple single-laboratory studies, or an interlaboratory study.

Early methods developed by EPA under Clean Water Act programs, and nearly all methods developed by EPA under Safe Drinking Water Act programs, were developed by an EPA research laboratory in Cincinnati, Ohio with specialized experience in the analytical chemistry of drinking water. This laboratory also established method detection and quantitation limits which, in many instances, initially could not be achieved in other laboratories. Over time, however, the difficulty in achieving these limits was overcome as analysts gained experience with the use of these new methods.

Stakeholders have suggested that detection and quantitation limits be developed using data from multiple laboratories in order to account for the routine inter- and intra-laboratory variability that can occur over time. Although compliance measurements are made in single laboratories, EPA agrees that detection and quantitation limits in methods that will be widely used by many laboratories should consider

---

[2]*Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993) and *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999)

these sources of variability. For this reason, after the development in a single laboratory of a new or modified analytical method with an initial estimate of detection and quantitation limits, EPA's Office of Science and Technology evaluates and verifies these limits in multi-laboratory studies.

Voluntary consensus standards bodies (VCSBs) such as ASTM International have historically used interlaboratory studies to establish method performance. Over the past 5 to 10 years, ASTM International has been developing interlaboratory and single-laboratory approaches for detection and quantitation. Single-laboratory studies at a specialized research laboratory may produce detection and quantitation limits that are lower than those produced by studies that gather data from many laboratories that may or may not be experienced with the method. EPA believes that a realistic expectation of method and laboratory performance likely lies somewhere in between that provided by a specialized single-laboratory study and that provided by an interlaboratory study with no pre-qualification requirements. Estimates of detection and quantitation limits should consider the inherent variability of the measurement process, but not be based on the lowest common denominator, e.g., data from inexperienced or unqualified analysts and laboratories.

EPA expects that laboratories must meet some minimum standards of performance and experience with a method, and sets performance criteria in methods. Examples of such criteria include measures to demonstrate that a laboratory is producing accurate results at a concentration of interest (i.e., analysis of reference standards or spiked samples), measures to demonstrate that results are not biased by contamination (i.e., analysis of blanks), and measures to demonstrate that the laboratory can detect pollutants at low concentrations (i.e., at the method detection limit). It is likely that laboratory performance will improve (and variability will be lower) when laboratories are required to meet specified performance criteria in order to report results.

A further consideration concerning routine variability of laboratory performance is the means for rejection of outliers to more accurately estimate routine variability. True outliers can occur in laboratory data, and some means of resolving outlier issues should be included. Statistical procedures are available for the identification of candidate outlier values. Once a candidate outlier has been identified, evaluation of the value from a QA/QC perspective (e.g., some procedural error or quality control error has occurred) should be the basis of exclusion of the value from a data set. In cases where no cause for the outlier has been identified, it may reasonable to reject an outlier on statistical grounds, but every effort should be made to justify the exclusion on technical grounds.

In examining each approach against this criterion, EPA will evaluate whether the approach can be used to provide realistic expectation of laboratory performance. As part of this assessment, EPA will examine the sources of variability captured by the approach, and the degree to which the statistics that underlie the approach realistically reflect these sources of variability.

4.3 Criterion 3

Criterion 3: *The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance.*

Any approach or procedure for determining detection and quantitation limits at a single laboratory should be simple, with detailed instructions, and cost-effective to implement (i.e., it should be reliable and "laboratory-friendly"). Laboratories that use detection or quantitation procedures range from large laboratories and laboratory chains with a wide range of technical capabilities, to small laboratories operated by one or a few people with limited statistical skills. While this range of laboratory capability

places a premium on simplicity and ease, EPA agrees with stakeholders that data reliability and quality are also important. A suitable approach or procedure for detection and quantitation incorporates the right balance between the need for valid data and the need for the procedure to be simple and inexpensive to perform. EPA also believes that if a procedure is complicated, it will be prone to error in use. Similarly, if a procedure requires investment of extensive resources that cannot be billed to the client, laboratories will have a disincentive to use the procedure. Therefore, if EPA wishes to encourage development and use of innovative techniques that improve measurement performance or lower measurement costs, the Agency should consider practicality and affordability as significant, if not equal, considerations to scientific validity.

After evaluating each of the issues discussed in Chapter 3 of this document, EPA concluded that successful implementation of CWA programs depends on the ability of laboratories to easily and affordably:

- demonstrate that a method works in a particular matrix at the levels of concern (i.e., demonstrate the absence of matrix effects),
- characterize improvements in measurement capabilities in terms of detection and quantitation capabilities, and
- characterize the detection and quantitation capabilities of new methods.

A matrix effect is an interference in a measurement that is caused by substances or materials in the sample other than the analyte of interest that are not removed using the procedures in the method or other commonly applied procedures. In the context of detection and quantitation, matrix effects may manifest themselves by precluding measurements at levels as low as could be measured were the interference not present. From a practical perspective, it is not possible to test the detection and quantitation capability of an analytical method in every possible matrix in which it may be used. At a minimum, it is unlikely that EPA or any other organization or laboratory could possibly identify and obtain samples of every matrix to which the method might be applied, and even if such a feat were possible, the cost and logistics of doing so would be prohibitive.

The situation for characterizing matrix effects on detection and quantitation is similar to the situation for characterizing matrix effects on measurement performance at higher concentration levels. In the latter case, EPA typically uses one or more spiked real-world or reference matrices (e.g., reagent water, sand, diatomaceous earth) to establish QC acceptance criteria that verify performance of the method at mid-to-high concentrations. Each analytical method includes QC acceptance criteria for such real-world and reference matrix spikes, along with a suite of quality control requirements designed to verify that failures are attributable to the matrix rather than to an analytical system that is out of control. EPA would prefer to utilize detection/quantitation concepts that allow for similar characterization of detection/quantitation capabilities in representative matrices and that are supported by simple, cost-effective procedures that would allow individual laboratories to evaluate the effects of specific matrices on these capabilities on an as needed basis. Because methods approved at 40 CFR part 136 already contain a suite of quality control procedures and QC acceptance criteria that control laboratory performance, EPA believes that it is not necessary to verify detection and quantitation limits in each and every batch of each and every matrix analyzed. Rather, such testing can be done on an as-needed basis when it is suspected that matrix interferences may preclude reliable measurements at low levels.

Another consideration influencing the need for simplicity and practicality is that measurement capabilities generally improve over time. As is discussed in Section 3.1 of this document, and as has been noted by stakeholders, this is attributable to a variety of factors, including:

- increased staff experience with a given technique,
- technological upgrades or improvements in the instrumentation used for analysis, and
- development of new instrumentation or techniques that improves detection/quantitation, precision, or bias.

In each case, the improvements may not be observed across the entire laboratory community. In the case of increased staff experience, for example, it is obvious that a laboratory that specializes in one type of analysis, such as low-level mercury measurements, will develop greater experience with these analyses than a laboratory that rarely performs these measurements. Likewise, it is easy to see how one or a few laboratories that concentrate their business on a particular type of analysis might be willing to invest significant resources in new or upgraded equipment to improve performance, whereas laboratories that rarely perform such analyses would not find such upgrades to be cost-effective.

Improvements in measurement capability, including the development of new methods, may create a dynamic decision-making process, in that measurements at lower levels may allow EPA and States to identify and measure previously undetected pollutants. Such improvements offer a means for monitoring and controlling (i.e., regulating) the discharge of previously unregulated, but harmful, pollutants. Therefore, it is in the best interest of the environment for EPA to encourage the development and use of improved environmental analysis procedures and equipment by providing practical and affordable procedures for evaluating method performance.

In evaluating this criterion, EPA will favor affordable and easy-to-use approaches and procedures that allow analysts to 1) determine matrix-specific variations when necessary, based on realistic data, and 2) demonstrate lower detection and quantitation limits associated with improvements in measurement capabilities. Procedures for establishing the detection capabilities of new methods or associated with improved measurement capabilities should be practical enough to encourage such development. However, EPA recognizes that some uses for detection and quantitation limits may require a more comprehensive approach involving multiple laboratories. These procedures should specify the nature, minimum number, and concentration levels of the samples to be used, and the corrective action to be taken if the resulting detection or quantitation limit is inconsistent with the data from which it is derived.

4.4 Criterion 4

Criterion 4: *The detection level approach should estimate the theoretical concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in a well-operated laboratory.*

Any approach to establishing levels at which detection decisions are made should be capable of providing regulators, the regulated community, and data users with a high level of confidence that a pollutant reported by a well-operated laboratory as being present really is present. Historically, approaches to making detection decisions have set the criterion for detection at 99 percent confidence (i.e., with 99% confidence that the analyte concentration is greater than zero). This criterion results in the probability of a false positive i.e., that a pollutant will be stated as being present when it actually is not (this is a Type I error), of one percent. The procedure also should be capable of generating a detection level when the substance of interest is not present in a blank and/or when instrument thresholds are used

in routine operation. A well-operated laboratory is a laboratory that routinely monitors performance through QC analyses, control charts, and other measures to rapidly identify and correct deteriorating or poor performance, and with analysts experienced with method sample preparation, analysis, and detection procedures.

In evaluating this criterion, EPA will favor approaches and procedures that reflect routine analytical conditions in a well-operated laboratory.

4.5 Criterion 5

Criterion 5: *The quantitation limit approach should identify the concentration that gives a recognizable signal that is consistent with the capabilities of the method when a method is performed by experienced staff in well-operated laboratories.*

Measurement capabilities among laboratories vary depending on a number of factors, including, but not limited to, instrumentation, training, and experience. Similarly, measurement capabilities among different analytical methods vary depending on a number of factors, including the techniques and instrumentation employed and the clarity of the method itself. In evaluating different approaches to estimating quantitation limits, EPA will give preference to those approaches that strike a reasonable balance between using either state-of-the art laboratories or a highly varied community of laboratories to establish quantitation limits.

Historical approaches to recognizing laboratory capabilities in establishing detection and quantitation limits have varied between two extremes of establishing the limit in a state-of-the-art research laboratory to reflect the lowest possible limit that can be achieved, and establishing the limit based on statistical tolerance intervals calculated from a large number of laboratories with varying levels of experience, instrumentation and competence. Generally, use of the former has been employed to serve as a goal or performance standard to be met by other laboratories, whereas use of the latter treats the limit, not as a performance standard that needs to be met by each laboratory, but rather as a characterization of the performance of the capabilities of a population of laboratories at the time of method development.

Historical approaches to recognizing method capabilities also have varied between those that allow the error expressed as relative standard deviation, or RSD among low-level measurements to vary, depending on the capabilities of the method, and those that fix this error (RSD) at a specific level.

Initially, Criterion 5 stated that the *"quantitation limit should identify a concentration at which the reliability of the measured result is consistent with the capabilities of the method when a method is performed by experienced staff in a well-operated laboratory."* Reviewers from within EPA questioned the criterion's implication that measurements below a quantitation limit could be considered unreliable. A similar concern was expressed by one of the peer reviewers charged with evaluating EPA's assessment and an earlier draft of this Assessment Document. This reviewer noted that:

> *"almost all implementations of limits of quantitation have nothing to do with whether the measurements are actually quantitative,"* and that *"any level at which the instrument can be read, and at which there is a reliably estimated standard deviation is a level at which quantitation is possible"* (Rocke, 2002)

The peer reviewer suggested that Criterion 5 might be rewritten as:

> *"the quantitation limit should identify a concentration at which the instrument yields a measurable signal at least 99% of the time, and which is no smaller than the detection level. Such a quantitation limit will often be the same as the detection level."*

EPA agrees that this is a valid perspective, in that if the pollutant is identified and the analytical system produces a result (i.e., a measurable or recognizable signal), quantitation occurs. Although this interpretation of a quantitation limit has validity, implementation of such an approach would require that all values generated by an analytical system be reported, along with an estimate of the uncertainty associated with each value (e.g., the "reliably estimated standard deviation" mentioned by the peer reviewer). As noted in Section 2.3.4, several organizations, including the European Union, are developing procedures for estimating the uncertainty associated with measured results. If successful, such an approach would eliminate many of the data censoring concerns discussed in Section 3.3.5. Given the difficulty in achieving consensus on an appropriate means of establishing a quantitation limit, however, EPA believes that it would also be difficult to obtain consensus on an appropriate means for estimating the uncertainty associated with each result measured on each environmental sample. In addition, analytical chemists have used and perceive that they understand a quantitation limit to mean the lowest concentration at which an analyte can be identified and quantified with some degree of certainty. This understanding necessarily involves use of the sound judgment of a qualified analytical chemist.

Therefore, EPA will continue to monitor developments on this subject, and if appropriate, re-evaluate this issue if and when it becomes practical and widely accepted by the laboratory, regulatory, and regulated communities. In the meantime, EPA believes that the traditional approach of defining a quantitation limit at some level above the detection limit provides a data user with a reasonable degree of confidence in the measured value without requiring that laboratories develop and report individual estimates of uncertainty. Criterion 5 reflects this belief.

In evaluating the approaches, EPA will give preference to those approaches that strike a reasonable balance between using either state-of-the art laboratories or a highly varied community of laboratories to establish quantitation limits.

4.6 Criterion 6

Criterion 6: *Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government.*

The Clean Water Act requires EPA to conduct, implement, and oversee a variety of data gathering programs. As noted in Section 3.2 of this Assessment Document, these programs include, but are not limited to:

- Survey programs to establish baselines and monitor changes in ambient water quality,
- Screening studies to identify emerging concerns and establish the need for more in-depth assessment,
- Effluent guideline studies to establish technology-based standards for the control of pollutants in wastewater discharges,

- Toxicity and environmental assessment studies to establish water quality-based standards for the control of pollutants in wastewater, and
- Risk assessment studies designed to characterize and evaluate human health and environmental risks associated with various water body uses.

In addition, EPA needs to evaluate detection limit or quantitation capabilities for methods approved at 40 CFR part 136 for the following applications:

- Ambient and effluent permitting and compliance monitoring under NPDES and the pretreatment program and under State and local programs,
- Quality control in analytical laboratories, and
- Method development, promulgation, and modification.

In theory, EPA could evaluate each of these applications independently and identify a detection and quantitation limit approach that is best suited to each application, as recommended by some stakeholders commenting on EPA's assessment. In the 2003 assessment, EPA stated that this would increase confusion, record keeping burdens, and laboratory testing burdens. EPA also stated that data generated under a single procedure can be used for development of detection and quantitation limits that are applicable to more than a single use. For example, the data used to determine the capabilities of multiple laboratories using a given method may also be used to develop method-specific detection and quantitation limits. For these reasons, EPA recommended the adoption of a single pair of related detection and quantitation procedures used to address all or most Clean Water Act applications. Some stakeholders recommend the use of different approaches for different CWA applications. For example, these stakeholders would prefer a more rigorous approach to determining detection and quantitation limits for method development than for verifying laboratory performance. They would like to include a procedure that is based on a multilaboratory approach rather than a single laboratory approach to define detection and quantitation capabilities of analytical methods. EPA recognizes that the complexity and statistical rigor appropriate for a detection and quantitation approach for method development and validation would be greater than that needed for demonstrating laboratory proficiency. EPA plans to seek additional stakeholder input on whether different approaches are needed for different CWA purposes (see Chapter 6).

Although EPA prefers to identify a manageable set of detection and quantitation limit approaches to meet CWA needs, EPA believes that any reasonable approach advanced by other organizations should be acceptable for use provided it meets the needs of the specific application for which it would be used. Allowing use of detection and quantitation approaches developed by other organizations provides the stakeholder community with increased measurement options that may help reduce measurement costs or improve measurement performance for specific situations. This approach also is consistent with the intent of the National Technology Transfer and Advancement Act.

The Clean Water Act authorizes State or local governments to implement specific aspects of the Act, with the provision that they do so in a way that is at least as protective (i.e., stringent) as the national standards put forth by EPA. Therefore, this criterion is intended to ensure that any detection and quantitation limit approach adopted by the Office of Water is sufficiently clear and defined to ensure consistency with approaches adopted by State or local governments.

Finally, it is important to differentiate between detection and quantitation limit approaches and compliance evaluation thresholds. Detection and quantitation limit approaches pertain to measurement process thresholds. In contrast, compliance evaluation thresholds are used to support wastewater

discharge limits established in National Pollutant Discharge Elimination System (NPDES) or pretreatment program permits. Such limits are usually expressed as either a maximum concentration of pollutant allowed in the discharge or a maximum mass of pollutant allowed to be discharged in a specific time period.

Ideally, and in most cases, analytical methods are available to allow for detection and quantitation of pollutants at concentrations that are lower than the discharge levels needed to protect or restore the quality of the receiving water. When such measurement capability does not exist (e.g., analytical methods are not available that can reliably measure at levels necessary to protect receiving water), permitting authorities must decide how to evaluate and report pollutant concentrations at these levels. Historically, EPA has recommended that in such cases, the permitting authority include the water quality-based limit in the permit, but establish the compliance evaluation threshold at the quantitation limit of the most sensitive available method.

In examining each approach against this criterion EPA will consider 1) the applicability of various detection/quantitation approaches to the variety of data gathering decisions that must be made under the CWA, including those that do and those that do not involve compliance monitoring, and 2) the ability of the approaches to support State and local obligations for implementing the CWA. As discussed in Chapter 6, EPA believes that additional discussion about this criterion is appropriate based on negative comments from stakeholders regarding the use of a single pair of detection and quantitation limit approaches to meet all CWA needs.

4.7 Consensus Principles

Some stakeholders commenting on EPA's assessment of approaches to detection and quantitation expressed their support of a set of "consensus principles" submitted by 36 signatories representing industry and laboratory communities. EPA agrees with certain consensus principles such as the principle that detection and quantitation levels should be based on sound scientific principles, and that low-cost and/or simple approaches should not be used if invalid data will result (see Criterion 1 above). As another example, EPA incorporated routine variability, the rate of false positives, precision, and matrix effects in several criteria, and considered these aspects in its assessment of detection and quantitation concepts. Some of these consensus principles are included in the criteria discussed in this chapter. Other consensus principles have clarified or highlighted existing aspects of approaches to detection and quantitation and provide a framework for additional consideration.

For ease of consideration, the consensus principles recommended by commenters have been separated by EPA into technical and policy considerations and include:

Technical Considerations

- Detection and quantitation levels must be based on sound scientific principles. Low-cost and/or simple approaches must not be selected if inaccurate compliance determinations or unmeasurable permit limits may result.
- The definition of "quantitation" must account for both precision and bias.
- Detection limit procedures must take into account the variability and bias of method blank results.
- False positives (Type I errors), false negatives (Type II errors), and precision must all be addressed by detection concepts and reporting of analytical results for regulatory purposes.

- Precision, bias, and qualitative identification (where appropriate) must all be addressed by the definition and concepts of quantitation and by the reporting of analytical results for regulatory purposes.
- Detection limit procedures must include procedures for ongoing demonstration of sensitivity, preferably incorporated into the routine analytical quality control as a check against false negatives.
- Detection and quantitation levels must take into account routine inter- and intra-laboratory variability within a laboratory over time.
- In its procedures for establishing detection and quantitation levels, EPA must develop guidance on how to account for the effects of various matrices.

Policy Considerations

- The L_C, L_D, and L_Q are three distinct points, each of which has unique criteria that must be satisfied. For consistency with international standards, EPA must adopt the definitions of L_C (critical value), L_D (detection limit), and L_Q (quantification limit) of IUPAC (International Union of Pure and Applied Chemistry) that are being adopted by international standards organizations (e.g., the International Organization of Standardization (ISO)).
- The definitions of and procedures for determining detection and quantitation levels must take into account that quantitation levels are used as regulatory compliance levels in NPDES permits.
- EPA should specify consensus standard procedures for establishing significant figures and for rounding data.
- EPA must strive for consistency across all EPA offices (the Office of Water, Office of Research and Development, Office of Ground Water and Drinking Water, and Office of Solid Waste and Emergency Response) in defining and applying detection and quantitation levels.

Chapter 5
Assessment

This chapter summarizes EPA's assessment of various detection and quantitation limit approaches against the evaluation criteria established in Chapter 4. Assessments of detection limit approaches are presented in Section 5.1 and include an assessment of the:

- EPA method detection limit (MDL; Section 5.1.1),
- ASTM International interlaboratory detection estimate (IDE; Section 5.1.2),
- American Chemical Society (ACS) limit of detection (LOD; Section 5.1.3),
- International Organization for Standardization/International Union of Pure and Applied Chemistry (ISO/IUPAC) critical value (CRV; Section 5.1.4),
- ISO/IUPAC minimum detectable value (MDV; Section 5.1.5),
- American Council of Independent Laboratories (ACIL) Critical Value (ACIL L_C; Section 5.1.6),
- United States Geological Survey (USGS) Long-term Detection Limit (USGS LT-MDL; Section 5.1.7), and
- Inter-industry Analytical Group (IIAG) Sensitivity Test and Full-Range Validation Study (Section 5.1.8).

Assessments of quantitation limit approaches are presented in Section 5.2 and include an assessment of the:

- EPA minimum level of quantitation (ML; Section 5.2.1),
- ASTM International interlaboratory quantitation estimate (IQE; Section 5.2.2),
- ACS limit of quantitation (LOQ; Section 5.2.3), and
- ISO/IUPAC LOQ (section 5.2.4).

A brief summary of the evaluation is presented in Tables 5-1 (detection limit approaches) and 5-2 (quantitation limit approaches).

EPA's 2003 assessment of detection and quantitation limit approaches focused on approaches developed or published by ASTM International, the American Chemical Society (ACS), ISO/IUPAC, and EPA. Stakeholder commenting on the initial assessment suggested that EPA should include additional approaches in the next assessment. In addition to the initial four approaches, EPA has included three additional approaches in this Revised Assessment document. These approaches are: the long-term MDL developed by USGS, a new detection limit procedure developed by the American Council of Independent Laboratories (ACIL), and a paired approach involving a sensitivity test and full-range validation study submitted by the Petitioners (the Inter-industry Analytical Group). Several commenters advocated these as approaches that more realistically reflect measurement variability. These additional approaches are discussed and assessed in Sections 5.1.6 - 5.1.8 of this chapter.

Some stakeholders commenting on EPA's 2003 assessment believed that the evaluation criteria used by EPA were written to favor the MDL and ML over other approaches to detection and quantitation. EPA disagrees. The criteria were written to reflect EPA's needs for detection and quantitation approaches under the CWA, and it is not necessary that an acceptable approach meet all of these criteria under all conditions. Because the MDL and ML were developed to address EPA's needs, it should not be surprising that the MDL and ML procedures generally meet the criteria EPA set out to assess detection and quantitation procedures. EPA has frankly assessed the MDL and ML against these criteria and notes

that the MDL and ML procedures do not meet all of these criteria under all operating conditions (see Sections 5.1.1 and 5.2.1 below). Due to the variability and unpredictability inherent in measurement science, it is unlikely that any procedure would meet all of EPA's criteria under all conditions. However, EPA is open to further discussions with stakeholders about the appropriateness of the evaluation criteria described in Chapter 4, in particular, the issue of whether EPA should adopt different approaches for different applications, as discussed in Chapter 6.

5.1 Detection Limit Approaches

Sections 5.1.1 through 5.1.8 describe EPA's assessment of eight detection limit approaches. Each discussion is divided into two major subsections. The first subsection describes the approach and, where applicable, the procedure that supports the approach. The second subsection details EPA's assessment of the approach based on the five criteria established in Chapter 4 for evaluating detection limit approaches.

Note: Of the six assessment criteria in Chapter 4 four (Nos. 1, 2,3 and 6) pertain to both detection and quantitation limit approaches. One criterion (No. 4) pertains only to detection limit approaches, and one criterion (No. 5) pertains only to quantitation limit approaches. Therefore, the following discussion of each detection and quantitation limit approach applies only the five applicable criteria.

5.1.1 Evaluation of the MDL

Section 5.1.1.1 is an overview of the MDL approach and the procedures used to implement the approach. Section 5.1.1.2 describes EPA's assessment of the MDL against the five evaluation criteria that apply to detection limit approaches.(i.e., Criteria 1-4, and Criterion 6).

5.1.1.1 Description of the MDL Approach and Procedure

As promulgated at 40 CFR part 136, Appendix B, the MDL is defined as:

> *"the minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix containing the analyte."*

A six-step procedure is given in Appendix B, with an optional seventh step to verify the reasonableness of the MDL determined in the first six steps. The procedure is intended for use by experienced analytical chemists. A brief summary of the MDL procedure is as follows:

1. The analyst makes an estimate of the detection limit based on one of four options: the instrument signal to noise ratio; three times the standard deviation of replicate blank measurements; a break in the slope of an instrument calibration curve; or known instrument limitations.

2. The analyst prepares a volume of reagent water that is as free of the target analyte as possible (if the MDL is to be determined in reagent water).

3. The analyst prepares a sufficient volume of spiked reagent water (or of an alternate matrix) to yield seven replicate aliquots that have a concentration of the target analyte that is at least equal to or in the same concentration range as the estimated detection limit (it is recommended that the concentration of the replicate aliquots be between 1 and 5 times the estimated detection limit).

4. All of the replicate aliquots are processed through the entire analytical method.

5. The variance (S^2) and standard deviation (S) of the replicate measurements are determined, as follows:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right]$$

$$S = \sqrt{(S^2)}$$

where:

X_i = the analytical results in the final method reporting units obtained from the n sample aliquots and
Σ refers to the sum of the X values from i=1 to n, and
i =1 to n

6. The MDL is then determined by multiplying the standard deviation (S) by the Student's *t*-statistic at a 99% percentile for n-1 degrees of freedom. If seven replicates are used, the Student's *t*-value is 3.143. This information is used to calculate the MDL as follows:

$$MDL = t_{(n-1,\ 1-\alpha\ =\ 0.99)}\ (S)$$

where:

MDL = the method detection limit

$t_{(n-1,1-\alpha=.99)}$ = the Student's *t*-value appropriate for a 99% confidence level with n-1 degrees of freedom, and

S = the standard deviation of the replicate analyses.

A 95% confidence interval for the determined MDL may be calculated from percentiles of the chi square over degrees of freedom distribution (χ^2/df).

7. The optional iterative procedure to verify the reasonableness of the MDL involves spiking the matrix at the MDL that was determined in Step 6, and analyzing another seven replicates spiked at this level. The F-ratio of the variances (S^2) is determined and compared with the F-ratio found in the table,

which is 3.05. If $S^2_A/S^2_B>3.05$, the analyst is instructed to respike at the most recently calculated MDL and process the samples through the procedure starting with Step 4. If $S^2_A/S^2_B<=3.05$, then the pooled standard deviation is determined (S^2_A is the larger of the two variances). The pooled standard deviation is then used to calculate the final MDL as follows:

$$MDL = 2.681 \times S_{pooled}$$

where 2.681 is equal to $t_{(12, 1-\alpha=.99)}$.

The 95% confidence interval around the final MDL may be determined using the chi squared distribution.

The MDL procedure given at 40 CFR part 136, Appendix B is described as being applicable to 1) a wide variety of sample types, ranging from reagent water containing the analyte of interest to wastewater containing the analyte of interest, and 2) a broad variety of physical and chemical measurements.

5.1.1.2 Assessment of the MDL Against the Evaluation Criteria

The following five subsections discuss the MDL approach and procedure in the context of the five evaluation criteria that concern detection limit approaches (i.e., Criteria 1-4, and Criterion 6).

5.1.1.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

For the purposes of evaluating scientific validity, EPA is using the conditions discussed by the Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals* (1993) and *Kumho Tire Co. v. Carmichael,* (1999) (see Chapter 4, Criterion 1).

Condition 1: It can be (and has been) tested. The MDL procedure meets this condition. Over the years, as stakeholders have sought to improve upon or identify alternative procedures, the MDL has been the subject of a number of studies and comparisons, including this assessment. As a result, the MDL is one of the most widely tested detection limit procedure in the history of detection approaches. (See Appendix A for a list of literature references concerning the MDL and other detection limits.)

Critics of the MDL have noted that the detection limit produced with the MDL procedure can vary depending on the spike level used. It is true that an initial MDL may be calculated using any spike level, regardless of how high. Although a high initial spike level will result in an initially high MDL, the self-correction check in the MDL procedure requires the final spike level to be within a certain range of the reported (i.e. final) MDL. Specifically, Step 1 of the MDL procedure focuses the spiking level on the lowest concentration at which measurements can be made, and the factor of 5 requirement in Steps 3 and 4 assure that the determined MDL will be at or near this concentration. Therefore, the requirements included in Steps 1, 3 and 4 guard against an artificially high MDL being produced due to the choice of a high initial spike level. EPA also recognizes the concern that the iterative procedure in step 7, which provides a reality check on the results obtained in steps 1 - 6 is optional. EPA will consider whether additional guidance on this aspect of the procedure is needed.

In preparation for the assessment of detection and quantitation approaches, EPA tested the MDL procedure with 10 different techniques, at decreasing spike concentrations, to evaluate this concern and determine how well the procedure characterized the region of interest. Results of the study suggest that, although the calculated MDL could vary depending on the spike level used, the MDL procedure is capable of reasonably estimating the lowest level at which measurements can be made when the factor of 5 requirement is met.

One of the stakeholders commenting on EPA's 2003 assessment suggested that the MDL failed to meet this condition because EPA should have tested it in "real world" matrices. EPA does not agree with this suggestion for several reasons. First, it is not practical or possible to test detection limits in every real world matrix, and there is no consensus as to which real world matrix would represent an appropriate real world matrix for testing. Second, many real world matrices contain the target pollutant at levels well above the detection or quantitation limit, making it impossible to characterize what can and cannot be detected at low levels. In theory, the sample could be diluted to dilute the target pollutant, but in practice sample dilution would also likely dilute any interferences that might be present, thereby defeating the purpose of using a real world matrix. The current EPA approach, which exhaustively tests the MDL procedure in a reference matrix using multiple techniques and ten different concentrations that span the entire region of interest, is more than adequate to constitute "testing" of the MDL procedure. On the other hand, where data suggests that matrix interferences may significantly affect achievable quantitation and detection limits, this should be considered by a permit writer on a case by case basis.

Condition 2: It has been subjected to peer review and publication. The MDL meets this condition. Prior to promulgation by EPA, the MDL approach and supporting procedure was published by Glaser *et al.* in a peer-reviewed journal (Glaser, *et al.*, 1981). The MDL procedure has been included at 40 CFR part 136, appendix B since 1984. Values resulting from this procedure have been included, published, and tested in many analytical methods since promulgation, including methods published by EPA and other Federal agencies, and by consensus standards organizations and trade associations such as ASTM International, and APHA, AWWA, and WEF.

Condition 3: The error rate associated with the procedure is either known or can be estimated. The error rate is specified by α, with a suggested value of 0.01(1%). Therefore, the MDL meets this condition. In addition, the Step 7 of the MDL procedure suggests calculating a 95% confidence interval for the determined MDL, providing additional estimation about the uncertainty (i.e., error) of the MDL determined using the procedure.

The US Geological Survey (USGS) provided a dataset of spiked and blank sample data that EPA used to evaluate the error rate associated with the MDL. (Error rates associated with the ACIL and USGS detection limit procedures also were evaluated and are discussed in Sections 5.1.6 and 5.1.7.) Although the sample size was insufficient to conclusively demonstrate the error rate of the MDL, the results suggest the actual error rate is close to the intended 1%. In this case, the observed mean error rate was 2.9%. Readers are referred to Appendix B for a discussion of two factors affecting this estimate - relatively small sample size and some added long-term variability.

In the 2003 assessment, EPA suggested deleting the procedure for calculating the 95% confidence interval because it appeared to be rarely, if ever, used. No commenters specifically agreed with this suggestion, but several commenters responded that it should be retained. One commenter, arguing in favor of the procedure, stated that "It has long been recognized that a 95% confidence level is appropriate to establish standards and other regulatory requirements." Considering these comments, EPA now believes there is no compelling reason to remove this procedure.

Condition 4: Standards exist and can be maintained to control its operation. The MDL approach is supported by a clearly defined, published procedure to control its operation. The procedure gives the steps to be followed and instructs the analyst to use the entire measurement process. Hundreds, if not thousands, of laboratories have successfully implemented the MDL procedure since its promulgation in 1984. EPA has found that when laboratories are required to perform MDL studies as part of an interlaboratory study, the results reported by the laboratories are generally consistent. EPA has observed similar consistency in use of the MDL by laboratories required to perform the procedure to demonstrate proficiency with a method. Therefore, the MDL meets this condition.

Notwithstanding the preceding, the MDL procedure would be improved with additional guidance, particularly with respect to initial spike levels, handling outliers, the optional reasonableness step (Step 7), and multi-analyte test methods. The MDL procedure does not contain a discussion of outliers. It may be helpful to clarify that 1) results should be discarded only if the results are associated with a known error that occurred during analysis (e.g., the replicate was spiked twice) or through a statistically accepted analysis of outliers, and 2) that laboratories should not simply select the best seven results of a dataset. The optional step involves iterative testing to verify that the determined MDL is reasonable; EPA has observed that few organizations bother to perform this step. EPA also has observed that when a method involves a large number of analytes, it can be difficult to get all analytes to pass the iterative test in the same run. The MDL procedure would benefit from guidance on how and when to address each of these issues.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. The MDL meets this condition. The MDL has been used experimentally since 1980 and in a regulatory context since 1984. The MDL procedure is the most widely used and, therefore, the most widely tested detection limit procedure in the history of detection approaches. Within EPA, the MDL has been used by the Office of Research and Development, Office of Science and Technology, Office of Ground Water and Drinking Water, Office of Solid Waste, Office of Emergency and Remedial Response, and other offices. The MDL also has been used outside of EPA in methods published by ASTM International, in *Standard Methods for the Examination of Water and Wastewater*– jointly published by the American Public Health Association (APHA), the American Water Works Association (AWWA), and the Water Environment Federation (WEF), and in methods elsewhere. Although the MDL has been criticized, it is the most widely used approach of detection within the environmental chemistry community.

Stakeholders commenting on EPA's 2003 assessment of detection and quantitation procedures noted that the extent to which the MDL has been used is a result of EPA's approval and inclusion of the procedure in 40 CFR part 136, and does not necessarily demonstrate that the MDL procedure produces an accurate assessment of detection. EPA agrees that the extent of use could be attributed, in part, to promulgation of the procedure at 40 CFR part 136. For this reason, EPA has not relied on widespread use of the MDL as a sole or over-riding argument for its continued use. Rather, EPA views widespread use of the MDL as one of many factors to be considered when evaluating which concept or concepts best meet the Agency's needs under the Clean Water Act. For example, EPA agrees that the ability of a procedure to produce an accurate assessment of detection capabilities is an important consideration, and addresses this issue repeatedly throughout the assessment. In this chapter, for example, the ability of a procedure to produce an accurate assessment of detection capabilities is addressed in

- Criterion 1, condition 3, which concerns error rate,
- Criterion 1, condition 4, which concerns use of standards to control operation of the procedure,
- Criterion 2, which addresses the ability of the procedure to realistically reflect laboratory and method performance, and

- Criterion 4, which addresses the ability of the approach to identify the concentration at which users can be confident a substance reported as present is really present.

5.1.1.2.2 Criterion 2: The approach should address realistic expectations of laboratory and method performance, including routine variability.

The MDL procedure is designed to demonstrate laboratory performance with a given analytical method, and can be applied to a broad variety of physical and chemical methods. The procedure also recognizes the importance of analyst experience and explicitly directs the analyst to employ all sample processing and computation steps given in the analytical method when determining the MDL.

When the MDL procedure is followed as intended (i.e., all sample processing and analysis steps of the method that are applied to routine analyses are included in determination of an MDL), the demonstrated MDL will include some of the routine variability associated with the laboratory and the method.

Stakeholders commenting on EPA's assessment stated that, because the MDL procedure is performed in a single laboratory, on the same day, by the same analyst, in a single matrix, using a minimum of 7 replicates, the procedure does not account for all sources of variability. These commenters believe that the procedure does not address inter- or intralaboratory, long-term, concentration range, analyte/method, or matrix variability. EPA notes that the MDL procedure does not include the restrictions noted by these stakeholders (e.g., users are not restricted to use of only seven replicates; to analysis of all replicates on the same day; or to determination of MDLs only in reagent water). The MDL procedure includes, for example, instructions for determining a matrix-specific MDL and specifies that the procedure requires a complete, specific, and well-defined analytical method. However, EPA also recognizes that in practice the MDL procedure may be performed in the manner described by these comments and that doing so will limit the amount of routine variability reflected in the results.

The MDL procedure provides users with the flexibility needed for multiple applications. For example, if a laboratory desires to evaluate its performance using a single method to analyze a particularly difficult matrix over a period of time (e.g., one year), the MDL procedure allows such an evaluation. However in some cases, the MDL procedure might benefit with specific provisions for including sources of variability that may not be addressed when following the minimum requirements of the MDL procedure.

Stakeholders commenting on EPA's assessment directed most of their concern at the lack of long-term variability in the MDL procedure. These commenters pointed to the American Council of Independent Laboratories (ACIL) procedures for calculating the critical level and long term-MDL (LT-MDL) and to the US Geological Survey's (USGS) procedures for generating their LT-MDL. These procedures include the collection of blanks over a long period of time to include this source of variability. The commenters stated that the lack of long-term variability leads to underestimates of Currie's critical value (L_c), and one commenter included sets of blanks collected over 3 months to demonstrate this effect.

EPA assessed the effect of long-term variability on calculated limits by simulating multiple 7-replicate subsets from the full dataset offered by the commenter, and compared these short-term critical levels to the critical level calculated using the full data set. Although the range of days from which the sets of 7 replicates were simulated varied from between one week to greater than 3 weeks, a graphical

analysis of the data did not reveal any effect of time on the resulting L_c. The total number of blanks also did not seem to have an effect on the percentage of short-term L_c results that exceeded the overall L_c. Details of this assessment are provided in Appendix C, along with possible reasons why expected differences were not observed.

As noted in Section 3.3.3 of this RAD, a larger number of replicates will yield better estimates for standard deviations, and therefore, better estimates of Currie's L_c and EPA's analogous MDL. However the analysis performed in Appendix C demonstrates that MDLs estimating L_c based on 7 replicates are not biased low. These values are merely less precise than those based on a larger number of replicates. As noted previously, the current MDL procedure does not restrict laboratories to using 7 replicates (to the contrary, the procedure specifies a minimum of 7 replicates), nor does it restrict laboratories to performing the replicates on a single day. Laboratories that wish to perform more tests or to conduct their tests over a longer period of time should be encouraged to do so.

Due to the variability inherent in measurement science, instrumentation, and the humans conducting analyses, laboratories may routinely obtain detection limits that are lower or higher than those obtained in another laboratory. Thus, when an MDL is determined during method development, it is important to determine that MDL in more than one laboratory to ensure the MDL published in the method reflects demonstrated expectations of method performance in a community of laboratories. It is not necessary for this community to include the entire universe of all possible laboratories that might desire to practice the method. Rather, during the stages of method development and validation, this community only should include well-operated laboratories with analysts who are experienced with the techniques used in the method, and have some familiarity conducting all of the steps in the new method before generating MDLs that will be published with the new method.

In recent years, EPA's Office of Science and Technology has used single-laboratory studies to develop an initial estimate of the MDL for a new or modified method, and has verified these MDLs in interlaboratory studies or by conducting additional single-laboratory studies in other laboratories. For example, when EPA initially drafted Method 1631 for measurement of mercury, EPA estimated the MDL to be 0.05 ng/L based on results produced by a contract research laboratory. Additional single-laboratory MDL studies conducted in other laboratories suggested that the MDL should be raised to 0.2 ng/L to better reflect existing capabilities of the measurement community. During EPA's interlaboratory study, each laboratory was asked to conduct an MDL study. Every laboratory in the interlaboratory study met the MDL of 0.2 ng/L (laboratory MDLs ranged from 0.04 to 0.18 ng/L), the value published in the promulgated version of Method 1631.

The MDL procedure addresses demonstrated expectations of laboratory and method performance, including routine variability, and users should not be restricted to the minimum requirements of the MDL procedure. If the MDL procedure is employed for method development purposes, it should be performed in multiple laboratories to ensure that it adequately demonstrates expectations in a community of qualified laboratories.

5.1.1.2.3 Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance.

The MDL procedure is among the most practical and affordable procedures that have been suggested for determining detection limits because of the reasonable number of minimum replicates (seven) and the relative ease with which the spiking experiments can be designed and the resulting data

analyzed. The MDL is designed for use by a single laboratory, and can be performed by a single analyst using a single instrument. And the MDL procedure also allows MDLs from several analysts or instruments within a laboratory, or between laboratories to be pooled and provide an estimate of the range of MDLs that might be routinely expected.

Use of the optional iterative procedure would increase the number of analyses by at least seven each time the procedure is implemented. If the procedure is implemented two times in reagent water, a minimum of 14 analyses are required. If the procedure is implemented two times in an alternative matrix, EPA estimates that 17-20 analyses may be required, given the possible need to determine the background concentration of the analyte in the alternative matrix. In any of these scenarios, the entire MDL determination can be performed in a single analytical batch (most EPA methods specify batch sizes of 20 samples).

5.1.1.2.4 Criterion 4: The detection level approach should estimate the theoretical concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in a well-operated laboratory.

The MDL meets this condition as described under Section 5.1.1.2.1, Condition 3 of this document in many cases. However, EPA recognizes that there are cases where this does not hold, and that users of the MDL procedure see this as a significant problem. EPA sees merit in blank correction procedures developed by ACIL and USGS to address these cases. In future stakeholder consultations, EPA plans to discuss these and other alternative solutions to this problem.

5.1.1.2.5 Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government.

The MDL meets this criterion. The MDL has been applied to a variety of decisions under the CWA since 1984. In addition, many States and others have adopted the MDL in their own programs.

5.1.2 Evaluation of the ASTM International Interlaboratory Detection Estimate (IDE)

The interlaboratory detection estimate (IDE) was published in 1997 by ASTM International as standard D6091. The IDE was developed with support from members of the regulated industry to provide a comprehensive detection limit procedure that addressed the concerns of the regulated industry, statisticians, and analysts involved in ASTM Committee D19 on water.

A brief summary of the procedure is given in Section 5.1.2.1, and Section 5.1.2.2 presents EPA's assessment of the IDE against the five criteria established for evaluating detection limit approaches (i.e., Criteria 1-4, and Criterion 6).

5.1.2.1 Description of the IDE Approach and Procedure

ASTM Designation D 6091 is the *Standard Practice for 99 %/95 % Interlaboratory Detection Estimate (IDE) for Analytical Methods with Negligible Calibration Error.* As stated in the practice:

> *"The IDE is computed to be the lowest concentration at which there is 90 % confidence that a single measurement from a laboratory selected from the population of qualified laboratories represented in an interlaboratory study will have a true detection probability of at least 95 % and a true nondetection probability of at least 99 % (when measuring a blank sample)."*

The IDE is determined and verified using a procedure containing 5 major steps with approximately 53 substeps and conditions. The full text of the IDE procedure is available from ASTM International. The five major steps and their functions are given in Section 6 of the IDE procedure and are as follows:

1. Overview of the procedure.

2. IDE Study Plan, Design, and Protocol - in this section, the task manager (study supervisor) chooses the analyte, matrix, and analytical method. Details are given for range finding; the concentrations to be used in the study; the study protocol (ASTM Practice D 2777 is suggested); the allowable sources of variation; and the number of laboratories, analysts, and days over which the study will be conducted.

3. Conduct the IDE Study, Screen the Data, and Choose a Model - after the study data are collected and screened according to ASTM Practice D 2777, interlaboratory standard deviation (ILSD) versus concentration data are tabulated and one of three models is fit to the data. The first attempt is at fitting a constant model. If the attempt fails, a straight-line model is attempted. If the straight-line model fails, an exponential model is fitted. After fitting, the model is evaluated for reasonableness and lack of fit. If the model fails, the study supervisor determines if a subset of the data should be analyzed or if more data are needed.

4. Compute the IDE - the IDE is computed using the ILSD model selected in Step 3 to estimate the interlaboratory standard deviation at a true concentration of zero and at the IDE, using a mean recovery model to transform measured and true concentrations. The IDE is computed as a one-sided 90 % confidence upper statistical tolerance limit.

5. Nontrivial Amount of Censored Data - this section addresses the effect of "non-detects" or "less-than." Suggestions are given to see if uncensored data can be obtained from the laboratories or if the study needs to be augmented with additional data. Suggestions are given for fitting a model to data that contain less than 10 % non-detects or less-than to produce an IDE.

5.1.2.2 Assessment of the IDE Against the Evaluation Criteria

The following five subsections discuss the IDE approach and procedure in the context of the five evaluation criteria that concern detection limit approaches.

5.1.2.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. The Electric Power Research Institute provided input into the design of EPA Method 1631 and 1638 Validation Studies for the purpose of calculating IDEs and IQEs. EPRI also calculated IDEs and IQEs based on these data. These two datasets include a total of ten metal analytes and therefore do not cover a wide range of analytical techniques and methods. Other than these two datasets, EPA is not aware of any organization, including ASTM International, that has conducted a study to test the procedure as written (i.e., designed and implemented an interlaboratory study that involves estimating an initial IDE [IDE_0] and multilaboratory analyses of multiple concentrations of each matrix of interest surrounding IDE_0). Developers of the approach performed limited testing of the approach on 1) simulated data sets and 2) real-world data sets generated for other purposes. However, these real-world data sets are of limited value for testing the IDE because the concentration ranges associated with the data are above the low-level region of interest. As part of this reassessment, EPA tested a variant of the IDE procedure on single-laboratory data sets designed for characterization of an analytical method in the region of detection. Despite the lack of comprehensive testing, the procedure can be tested, and therefore meets part of this condition. Specifically, the IDE meets the condition that it can be tested, but it only partially meets the condition that it has been tested.

Condition 2: It has been subjected to peer review and publication. Although the IDE has not been published in the peer-reviewed scientific literature, the IDE has undergone extensive review and ballot by members of ASTM Committee D 19, many of whom are qualified peer reviewers. Therefore, although the IDE does not meet this condition in the sense of formal peer review and publication, it meets the intent of this condition (i.e., submission to scrutiny of the scientific community). In addition, the IDE was reviewed by four peer reviewers as part of EPA's assessment of detection and quantitation limit approaches.

Condition 3: The error rate associated with the procedure is either known or can be estimated. In theory, expert statisticians could estimate the error rate of the IDE. However, the IDE procedure is extremely complex from an analytical chemistry and statistical perspective. As a result, it is unlikely that the error rate could be estimated by the typical users of the analytical method to which it would be applied, or even by the typical developers of an analytical method. Moreover, EPA found the model selection procedure to be highly subjective, a situation likely to yield different IDEs from the same data set, depending on the staff involved in performing the calculations. In practice, such conditions make it impossible to estimate the actual error associated with the IDE. Therefore, the IDE does not meet this condition.

One of the four peer reviewers charged with evaluating EPA's assessment of detection and quantitation limit approaches concurred with EPA's assessment of the IDE, specifically stating, *"I agree that the IDE procedure as outlined is so complex as to make simple determination of error rates associated with it untenable."* (Piegorsch, 2002)

One stakeholder, however, stated that concerns about the complexity and subjectivity in the IDE (and IQE) procedures were unimportant, in part, because IDEs calculated using different models were generally very close, and in part because "user-friendly software is available that will automatically perform the IDE and IQE calculations." To consider the merit of this comment, EPA calculated single-laboratory variants of the IDE using each of the four major model types using the Episode 6000 data set, and true interlaboratory IDEs for each model type using the Method 1631 and 1638 interlaboratory study data sets. Results of these calculations, along with the RSDs between the different IDE values obtained for each analyte, are presented in Appendix B. Based on the calculated RSDs, there is a large amount of variability between the single-laboratory variants of the IDEs calculated using the different models. Generally, the IDEs calculated using the constant model were much greater than those calculated using the

other models. The hybrid model generally yielded the lowest IDEs, and the IDEs calculated using the hybrid and exponential models were quite similar for some analytes, but quite different for others. While one might hope that the variability between models would decrease if interlaboratory variability were included in the calculations (as designed), EPA found this was not the case. To the contrary, RSDs between the IDEs calculated from the interlaboratory datasets suggest that variability between model estimates appears to increase when the additional variability between laboratories is included.

To evaluate the commenters' statement that the complexity and subjectivity of the procedures was not important because the calculations can be automatically performed using "user-friendly software," EPA evaluated the two software packages offered by the commenter. One package was a DOS-based program called "QCalc" and the other was an Excel spreadsheet that calculates IDEs based on Excel functions, macros, and the Solver add-in function. EPA calculated single laboratory variants of the IDE for a random subset of 20 analytes from the Episode 6000 study using 1) the QCalc package, 2) the Excel spreadsheet, and 3) the suite of SAS programs EPA has been using to calculate IDEs as part of this assessment. To ensure that differences between results were due to the programs themselves, the same data were used for each program. Results of this comparison are provided in Appendix B to this Revised Assessment Document.

One immediate problem was that comparisons could not be made between IDEs calculated using QCalc and the other software packages for all of the models because the QCalc package only performs the IDE calculation using two of the models (exponential and hybrid). The ASTM IDE procedure suggests that one of three models be used (constant, linear, and exponential). No explanation was provided as to why the software was limited to two models instead of three, or why one of the two models (i.e., the hybrid model) used in the software was not one of the three models recommended by ASTM. (The hybrid model used in QCalc is recommended by ASTM for calculation of an IQE but not for an IDE.)

Although similarities were generally observed among the various software packages when the same model type was applied to the same set of data, EPA did observe strong differences in the values calculated using the hybrid model across the various software programs. The Excel values generated using the hybrid model were slightly higher than those determined using EPA's programs and approximately twice as high as those determined using QCalc. Possible explanations for these differences are given in Appendix C.

Perhaps the most significant problem with the assumption that use of the automated software packages alleviates the complexity and subjectivity in the IDE procedure is that the various packages do not always select the same model for the same set of data. ASTM's IDE procedure (D 6091) specifies that the fitting to the constant model should be attempted first. If this fitting fails, a straight-line model should be attempted, and if that fails, the exponential model should be fitted and evaluated for reasonableness and lack of fit. EPA's SAS programs were coded to preferentially select the constant, linear, and exponential models for the IDE, according to this scheme. However, QCalc and Excel packages each follow a different scheme. As a result, the EPA and QCalc programs selected the same model type to calculate the IDE for only 1 of the 20 analytes, the Excel and QCalc programs selected the same model type for only 6 of the 20 analytes, and the Excel and EPA programs selected the same model type for only 1 of the 20 analytes. Details and possible explanations for these underlying differences can be found in Appendix C.

Based on these differences in selecting and fitting models, it does not appear that the two available software programs remove all complexity and subjectivity from the IDE calculation. Instead, they appear to introduce new issues by using steps not included in the ASTM procedures. The results support EPA's conclusion that such conditions make it impossible to estimate the actual error associated with the IDE, and that the IDE, as currently constructed, does not meet this condition 3.

Condition 4: Standards exist and can be maintained to control its operation. The IDE approach and procedure is supported by a published procedure (standard) to control its operation. The procedure gives the steps to be followed in determining the IDE and instructs the study supervisor how to gather the data and compute an IDE.

There are several "gray areas" in the published procedure. The most significant of which is in the description of model selection. The procedure provides insufficient guidance on use of residual plots to evaluate and select models and, as a result, selection of the model may be very subjective, especially if the number of concentrations is low. The problems noted in preceding Condition 3 concerning the use of different model selection strategies among three different programs (the QCalc and the Excel software packages provided by a commenter and EPA's SAS programs) is a direct reflection of the subjective nature of model selection likely to result from the lack of guidance in the procedure. The discussion of what model to use after rejecting the exponential and linear model is also very vague. The Rocke and Lorenzato (hybrid) model is mentioned, as well as models with more than one coefficient. Much of the data evaluated by EPA have tended to suggest the exponential model, based on the statistical tests discussed. However, those data have almost always shown residual "patterns" when using this model, which would then lead to consideration of other models. In addition, fitting the constant model is never discussed in detail. Most likely, this is done by simply calculating a mean (weighted if necessary) of the variances from the different concentrations; however, such calculations are never explicitly stated.

The IDE standard gives procedures that are inconsistent with procedures in the IQE standard, even though the two approaches should be consistent for a given analyte with a given method. For example, the exponential model figures prominently in the IDE procedure, where it is one of the three main models discussed. The Rocke and Lorenzato model is not discussed in the IDE procedure, but it figures prominently in the IQE procedure. In theory, a single model should support the definition of both the detection and quantitation limits for a given analyte by a given method. As another example, the IDE procedure includes a multiplier to account for bias in estimating the true standard deviation with the sample standard deviation, but the IQE does not.

Although the IDE is supported by a published procedure, EPA found that the procedure will not adequately control its operation because of the degree of subjectivity involved in implementing the procedure and inconsistencies with its IQE counterpart. Therefore, the IDE does not meet this condition.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. The IDE was published by ASTM, International in 1997. ASTM, International is a voluntary consensus standards organization that constitutes part of the relevant scientific community, however, seven years after publication no new or revised ASTM standard has included detection limits using the IDE approach. EPA is not aware of an IDE that has been published in the open literature or in an analytical method. Thus, the IDE partially meets this criterion.

5.1.2.2.2 Criterion 2: The approach should address realistic expectations of laboratory and method performance, including routine variability.

The IDE procedure, D6091, is designed to reflect expectations of interlaboratory performance, including routine variability. The procedure contains extensive instructions for dealing with unusual conditions, including sources of variability and outliers. However, EPA studies of a single-laboratory variant of the procedure suggested that the procedure may not always work as intended. For example, model selection based upon hypothesis tests (as described in Section 6.3.3.2 of D6091) almost always indicated that the exponential model should be used, even when the data seemed to be show constant or approximately linear error, while examination of residual plot indicated "systematic behavior" (i.e., non-

random deviations from the model) for the exponential and linear models. Information about single-laboratory (or within-laboratory) variability is very important because assessments of laboratory performance is based on the variability (uncertainty) of the dat produced at that laboratory. Compliance measurements are made in a single laboratory and the results are reported with the uncertainty (variability) associated with that dataset.

Another concern with the IDE procedure is that use of the non-mandatory appendices in ASTM D 6512 to determine the fit of a model may produce results that differ from those that would be obtained by using the default procedures for testing model fit that are built into off-the-shelf statistical software, such as those used in the Excel spreadsheets discussed in Section 5.1.2.2.1. Such observations, along with the concerns described in Section 5.1.2.2.1, condition 4, lead EPA to believe that, while the IDE approach addresses demonstrated expectations of laboratory and method performance, the IDE procedure does not adequately do so. Therefore, the IDE only partially meets this criterion.

5.1.2.2.3 Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance.

The IDE procedure is designed for use by an ASTM International study supervisor or task manager and not as a procedure that a single laboratory can use to evaluate method performance. EPA is aware that ASTM Committee D 19 is developing a Within-laboratory Detection Estimate (WDE), but the WDE is presently only in the formative stages. The WDE may meet this criterion, but the IDE does not.

Regarding cost, the IDE procedure would be the most costly of the procedures that EPA has evaluated because of the time it would take to understand and implement the procedure, and requirements for: 1) estimation of IDE_0, 2) interlaboratory data, 3) extensive statistical intervention in determining the correct model, and 4) possible reanalyses if the resulting IDE does not meet the criteria in the procedure.

5.1.2.2.4 Criterion 4: The detection level approach should estimate the theoretical concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in a well-operated laboratory.

By definition, the IDE is designed to achieve "*a true detection probability of at least 95 % and a true nondetection probability of at least 99 %*." Although the 99% probability of a "true nondetection" is equivalent to the 99% confidence that the substance is actually present given in Criterion 4, ASTM International also included the simultaneous requirement for a 95% probability of a "true detection." The developers are using the IDE as a means to control the rates of both false positive and false negative results, in essence, making the IDE analogous by definition and formulaic construction to the *detection limit* (DL) defined by Currie (1968). The IDE accomplishes this goal by using a tolerance limit that increases the IDE well above the point at which the detection decision would be made. For a discussion of this issue, see Sections 3.3.6 (false positives and false negatives) and 3.3.7 (prediction and tolerance intervals) in Chapter 3 of this document.

As noted in Section 2.1 of Chapter 2 of this document, Currie (1968) used the term *detection limit* (subsequently termed the *minimum detectable value*) to refer to a true concentration that has a high probability of generating measured values greater than the critical value. That is, measurements on samples that contain concentrations equal to the *detection limit* have a high probability of exceeding the

critical value and are, therefore, unlikely to result in a decision that the substance is not detected in the sample. However, the *detection decision* is made on the basis of comparing sample measurements to the *critical value*. With regard to his definition of the *"detection limit,"* Currie (1995) states *"The single, most important application of the detection limit is for planning."*

When the allowance for false negatives and the prediction and tolerance limits are taken into account, the resulting IDE is raised to the point at which the probability of a false positive is less than .01 by several orders of magnitude. This protection against false positive results is excessive and would yield numerical values of little practical value for making the detection decision.

Although there is an estimate of Currie's Lc included in the IDE procedure, it is unclear where the detection decision is made (it really should be an ICE/IDE procedure). If one focuses on the IDE and not the Lc estimate, this criterion not met. Therefore, it is not clear whether the IDE would meet this criterion (No. 4).

5.1.2.2.5 Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government.

EPA's comparison of detection limits produced by various detection limit approaches shows that the median IDE is considerably higher than ACS, ISO/IUPAC, and EPA detection limits. Although the IDE could be applied to some decisions to be made under the CWA, it may not be appropriate for all uses. The IDE is an implementation of Currie *detection level* or *minimum detectable value*, and may in practice yield results higher than these levels. At best, the IDE only partially meets this criterion.

5.1.3 Evaluation of the ACS Limit of Detection

The limit of detection (LOD) was developed by the Committee on Environmental Improvement (CEI) of the American Chemical Society (ACS). ACS is a professional society for chemists and other scientists and the publisher of a number of scientific journals. It is not a voluntary consensus standards body (VCSB), nor does it develop or publish analytical methods. In 1978, the ACS/CEI began addressing concerns about the lack of useful standards for interlaboratory comparisons. In 1980, the Committee published its *"Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry"* (MacDougall, *et al.*, 1980), which included the approaches of the LOD and the limit of quantitation (LOQ).

5.1.3.1 Description of the ACS LOD

The 1980 *"Guidelines"* define the LOD as:

> *"... the lowest concentration of an analyte that the analytical process can reliably detect. ... The LOD in most instrumental methods is based on the relationship between the gross analyte signal S_t, the field blank S_b, and the variability in the field blank σ_b."*

and construct the formal relations using the equation:

$$S_t - S_b \geq K_d \sigma$$

where K_d is a constant. ACS recommended a minimal value of 3 for K_d. Thus, the LOD is 3σ above the gross blank signal, S_b. In the 1980 publication, the ACS stated that at $K_d = 3$, there is a 7% risk of false negatives and false positives. Given that the LOD is 3σ above the blank, however, EPA believes that the risk of false positives is somewhat less than 1%.

In 1983, the ACS Committee published *"Principles of Environmental Analysis"* (Keith *et al.*, 1983). That publication occurred after the 1981 paper on the Method Detection Limit (MDL), and ACS/CEI stated that the LOD is numerically equivalent to the MDL as S_b approaches zero. However, neither the 1980 nor 1983 ACS publications provide a specific procedure for estimating the LOD, nor do they provide a minimum number of observations needed to estimate the gross blank signal or the variability term σ_b.

5.1.3.2 Assessment of the LOD Against the Evaluation Criteria

The following five subsections discuss the LOD approach and procedure in the context of the five evaluation criteria that concern detection limit approaches (i.e., Criteria 1-4, and Criterion 6).

5.1.3.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. Testing of the ACS LOD is hampered by the lack of a supporting procedure for establishing an LOD, and a conceptual dependence on the variability associated with measuring blanks. For example, there is no detailed instructions, similar to those in the IDE and the MDL procedures, to govern the minimum number of analyses needed to characterize the variability of a blank sample. Because many environmental chemistry techniques yield a zero, or possibly even negative, value when a blank sample is analyzed, and because the LOD approach is based on the standard deviation of these results, directly testing the LOD in such techniques will yield a zero or negative value. One solution for testing is to rely on ACS' 1983 statement that the LOD is conceptually equivalent to the MDL as the blank signal approaches zero, and employ the MDL procedure as a means for indirectly testing the LOD approach. EPA believes that use of the MDL procedure is a viable means for testing the approach; therefore, the LOD meets this condition.

Condition 2: It has been subjected to peer review and publication. The LOD meets this condition because the LOD definition was published in the peer-reviewed journal *Analytical Chemistry* in 1980 and 1983.

Condition 3: The error rate associated with the procedure is either known or can be estimated. The error rates can be estimated, so the LOD meets this condition. The error rate for both false positives and false negatives is stated to be 7 % in the 1980 Analytical Chemistry article. However, EPA believes that, because the LOD is stated to be 3 times the standard deviation of replicate measurements of a blank, the

false positive rate is overstated and is actually somewhat less than 1 % whereas the false negative rate depends on the true concentration in the sample.

Condition 4: Standards exist and can be maintained to control its operation. The LOD does not meet this condition, because it lacks a clearly defined procedure for estimating the important terms required to derive it. Although it may be possible to derive LOD values from data used to derive EPA MDL values, there is no procedure giving explicit instructions on the use of replicate blanks, replicate spiked samples, or a minimum recommendation for the number of replicates.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. Because ACS does not develop and publish analytical methods, it is difficult to determine the degree of acceptance of the LOD. EPA has not specifically investigated the numbers of papers published in ACS journals that include LOD values, and EPA's literature search for detection and quantitation approaches did not uncover a large number of citations that promote the LOD in particular. However, ACS LOD values have appeared in the technical literature. Given that ACS is a relevant scientific community, and that use of the LOD has appeared in the technical literature, the LOD meets this condition.

5.1.3.2.2 Criterion 2: The approach should address realistic expectations of laboratory and method performance, including routine variability.

The LOD approach is designed to address realistic expectations of laboratory and method performance, including routine variability, and thus appears to meet this criterion. Unfortunately, ACS has not published a procedure to implement the approach. In other words, the LOD addresses demonstrated expectations of laboratory and method performance in theory, but in practice, provides no direct means for performing these demonstrations. Therefore, EPA believes the ACS LOD only partially meets this criterion.

5.1.3.2.3 Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance.

The ACS LOD approach does not meet this criterion, because it is not supported by a clearly defined procedure for establishing the LOD.

5.1.3.2.4 Criterion 4: The detection level approach should estimate the theoretical concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in a well-operated laboratory.

The 1983 publication associated the LOD with the "99% confidence level when the difference ($S_t - S_b$) > 3σ." Therefore, the LOD meets this criterion.

5.1.3.2.5 Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government.

In the absence of a procedure for determining LOD values, the ACS LOD does not meet this criterion because it cannot be used in a regulatory context unless it is assumed to be functionally equivalent to the MDL (i.e., use the MDL procedure to establish an LOD).

5.1.4 Evaluation of the IUPAC/ISO Critical Value (CRV)

The critical value (CRV) was developed by the International Union of Pure and Applied Chemistry (IUPAC) and the International Organization for Standardization (ISO). IUPAC and ISO are professional societies for chemists and other scientists. ISO develops and publishes analytical methods through its Task Groups. In 1995, Lloyd Currie of the National Institute for Standards and Technology (NIST; formerly the National Bureau of Standards) published a signature discussion of IUPAC approaches for detection and quantitation (*Pure and Appl. Chem.* 67:10, 1699-1722). Although refined during the intervening years (see Currie, L.A., *J. Radiochem. And Nuclear Chem.* 245:1, 145-156, 2000), the CRV approach remains basically as described in 1995.

5.1.4.1 Description of the ISO/IUPAC Critical Value (CRV) Approach and Procedure

The 1995 article states that the critical value (L_c) is:

"... the minimum significant value of an estimated net signal or concentration, applied as a discriminator against background noise. This corresponds to a 1-sided significance test."

For a normal distribution with known variance, L_c reduces to:

$$L_c = z_{(1-\alpha)}\sigma_0$$

where:

$1-\alpha$ is the false positive error rate, recommended at 5 % ($\alpha = 0.05$), and
σ_0 is the standard deviation at zero concentration

If σ_0 is estimated by s_0 (replicate measurements of a blank), $z_{(1-\alpha)}$ is replaced by the Student's *t*-value. For 7 replicates (6 degrees of freedom), the Student's *t*-value is 1.943, where $\alpha = 0.05$.

5.1.4.2 Assessment of the CRV Against the Evaluation Criteria

The following five subsections discuss the CRV approach and procedure in the context of the five evaluation criteria that concern detection limit approaches (i.e., Criteria 1-4, and Criterion 6).

5.1.4.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. The lack of a supporting procedure for establishing the CRV, coupled with its conceptual dependence on the variability of blank measurements makes testing of the approach difficult. For example, if blank measurements fail to produce a response, it is impossible to calculate a CRV because the standard deviation of multiple zero results is zero. One solution for testing the approach is to assume that the CRV is about equivalent to the MDL as the blank signal approaches zero, and use a slightly modified version of the MDL procedure to test the CRV approach. The slight modification involves selecting a Student's *t*-value based on $\alpha = 0.05$ instead of $\alpha = 0.01$, for n-1 degrees of freedom. EPA believes this is a reasonable assumption, and therefore, that the MDL procedure is a viable means for testing the CRV approach. Therefore, the CRV meets this condition.

Condition 2: It has been subjected to peer review and publication. The IUPAC/ISO definitions meet this criterion. Moreover, it is likely that these definitions have received greater peer review than any of the other approaches.

Condition 3: The error rate associated with the procedure is either known or can be estimated. The error rate is specified by α, with a suggested value of 0.05 (5%). Therefore, the CRV meets this condition.

Condition 4: Standards exist and can be maintained to control its operation. The CRV is defined in the various publications by Currie. However, EPA's search of the literature and the ISO web site found no standard for control of the approach. Therefore, the CRV does not meet this condition.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. Because IUPAC and ISO are international bodies, it is difficult to determine the degree of acceptance of the CRV in the U.S. and the world community. EPA has not counted the number of papers in published journals that include CRV values, but EPA's literature search for detection and quantitation approaches did not produce many citations that promote the CRV in particular. Therefore, it is difficult to determine if the CRV meets this condition.

5.1.4.2.2 Criterion 2: The approach should address realistic expectations of laboratory and method performance, including routine variability.

The CRV approach is designed to account for the variability of measurements of the blank in the context of a "chemical measurement process" (method). Unfortunately, neither ISO, IUPAC, nor Currie have published a procedure to implement the approach. As a result, the CRV addresses realistic expectations of laboratory and method performance in theory, but in practice, provides no direct means for demonstrating this performance. Therefore, the CRV partially meets this criterion.

5.1.4.2.3 Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance

The CRV approach is not supported by a clearly defined procedure for establishing a CRV. Therefore, the CRV does not meet this criterion.

5.1.4.2.4 Criterion 4: The detection level approach should estimate the theoretical concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in a well-operated laboratory.

CRV suggests $\alpha = 0.05$, resulting in $1-\alpha$ of 0.95 or 95 % probability of detection . Therefore, the CRV does not meet this criterion.

5.1.4.2.5 Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government.

In the absence of a procedure for establishing CRVs, the CRV approach does not meet this criterion because it cannot be used in a regulatory context.

5.1.5 Evaluation of the IUPAC/ISO Detection Limit

The detection limit or minimum detectable value (MDV) was developed by IUPAC/ISO and published in the same papers as the CRV (Section 5.1.4)

5.1.5.1 Description of the IUPAC/ISO Detection Limit Procedure

The 1995 publications define the minimum detectable value (detection limit) as follows:

"The Minimum Detectable Value (MDV) ... [is] *... the net signal (or concentration) of that value (L_D) for which the false negative error is β, given L_C (or α)."* (see the CRV for L_C)

For a normal distribution with known variance, L_D reduces to:

$$L_D = z_{(1-\beta)}\,\sigma_D + Lc$$

where:
z is the score variable
1-β is the false negative error rate, recommended at 5 % (β = 0.05), and
σ_D is the standard deviation at the detection limit

Earlier publications refer to the minimum detectable value as the detection limit. To avoid confusion in terminology and to help distinguish the ISO/IUPAC approach from the MDL, LOD, and CRV, the ISO/IUPAC detection limit in this assessment will be referred to as the Minimum Detectable Value, abbreviated as MDV.

5.1.5.2 Assessment of the ISO/IUPAC MDV Against the Evaluation Criteria

The following five subsections discuss the ISO/IUPAC MDV approach and procedure in the context of the five evaluation criteria that concern detection limit approaches (i.e., Criteria 1-4, and Criterion 6).

5.1.5.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. The lack of a supporting procedure for establishing the MDV makes testing of the approach difficult. However, the MDV probably can be tested using data similar to those used to generate MDL values. Therefore, the MDV meets this condition.

Condition 2: It has been subjected to peer review and publication. The IUPAC/ISO definitions meet this condition; moreover, it is likely that this definition has received greater peer review than any of the other approaches.

Condition 3: The error rate associated with the procedure is either known or can be estimated. The error rates are specified by α and β, both with suggested values of 0.05 (5 %). Therefore, the error rate is known.

Condition 4: Standards exist and can be maintained to control its operation. The MDV is defined in the various publications by Currie. However, EPA's search of the literature and the ISO web site found no standard for control of the approach. Therefore, the MDV does not meet this criterion.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. Because IUPAC and ISO are international bodies, it is difficult to determine the degree of acceptance of the MDV in the U.S. and the world community. EPA has not specifically investigated the number of papers in published journals that include MDV values, but EPA's literature search for detection and quantitation approaches did not uncover a large number of citations that promote the MDV in particular. Therefore, it is difficult to determine if the CRV meets this criterion.

5.1.5.2.2 Criterion 2: The approach should address realistic expectations of laboratory and method performance, including routine variability.

The MDV approach is designed to account for the variability of measurements of the blank in the context of a "chemical measurement process" in the sense that it is used in concert with a critical value that is based on blank measurement variability. The MDV is the true concentration that is used in the planning of method evaluation and development. The actual detection decision is made at the critical value (CRV) which is determined from measured values. The approach of a true concentration MDV and its associated allowance for false negatives is of little practical value in making the actual detection decision. Therefore, the MDV does not meet this criterion. The allowance for false negatives in a regulatory context is discussed in greater detail in Chapter 3.

5.1.5.2.3 Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance

The MDV approach is not supported by a clearly defined procedure for establishing MDV values. Therefore, the MDV does not meet this criterion.

5.1.5.2.4 Criterion 4: The detection level approach should estimate the theoretical concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in a well-operated laboratory.

The allowance for false negatives reduces the probability of false positives to a value smaller than 1% by several orders of magnitude. . This protection against false positive results is excessive and would yield numerical values of little practical value for making the detection decision. Perhaps more importantly, as noted by Currie (1995) and discussed in Section 5.1.2.2.4 of this document, the *detection decision* is made on the basis of comparing sample measurements to the *critical value.* Therefore, the MDV does not meet this criterion.

5.1.5.2.5 Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government

In the absence of a procedure for establishing MDV values, the MDV approach does not meet to meet this criterion because it cannot be used in a regulatory context.

5.1.6 Evaluation of the American Council of Independent Laboratories (ACIL) Critical Value

During the comment period on the February 2003 assessment document, the American Council of Independent Laboratories (ACIL) submitted a procedure that was developed to address errors, which are referred to as "bias", that may arise under certain conditions when estimating detection limits. The ACIL

procedure separates estimation of the detection limit into two cases; cases where analyses always produce a numeric result (i.e., even so-called "blank" samples produce a signal), and cases where tests do not always produce a numeric result (i.e., blank samples appear to produce no signal). Blanks that do not produce a signal may do so either because they really are blanks, or the instrument is suppressing the signal. For convenience, EPA refers to these as Case I and Case II, respectively. Analysis of metals with inductively coupled plasma optical emission spectroscopy (ICP-OES) is an example of ACIL Case I, and analysis of organic pollutants with gas chromatography/ mass spectrometry is an example of ACIL Case II. Although the ACIL procedure appears to be a work-in-progress, it has some interesting approaches for the use of blanks, and is similar in some respects to the USGS LT-MDL procedure.

5.1.6.1 Description of the ACIL Approach and Procedure

For Case I analyses, ACIL offers procedures for calculating a limit that approximates Currie's critical value (L_C) and procedures for calculating a limit that approximates Currie's detection limit (L_D). As discussed in Chapter 2 and noted again in Section 5.1.5 above, Currie's L_D was designed to account for the variability of measurements of the blank in the context of a "chemical measurement process" in the sense that it is used in concert with a critical value that is based on blank measurement variability. The L_D is the true concentration that is used in the planning of method evaluation and development. The actual detection decision is made at the critical value (L_C), which is determined from measured values. The approach of a true concentration L_D and its associated allowance for false negatives is of little practical value in making the actual detection decision. For this reason, EPA focused its assessment of ACIL's procedure on the ACIL version of Currie's critical value rather than the ACIL version of L_D.

For Case II analyses, ACIL suggests a procedure that does not rely on the Currie L_C and L_D framework. Instead, the procedures involve picking an initial spike value, adjusting that level up or down based on whether the analyte was detected, and spiking seven replicates at the new level.

A brief description of each procedure is provided below.

ACIL's Case I Critical Value (ACIL L_C)

As with EPA's MDL, the ACIL L_C is an attempt to approximate Currie's critical value. Whereas EPA's MDL is based on the standard deviation of blank samples spiked with low levels of the target analyte, ACIL's Case I detection limit is based on the standard deviation of the blank samples run as part of the laboratories ongoing QC program. (Because some methods will not yield a result when blanks are analyzed, ACIL's L_C procedure is accompanied by a spiked sample approach that can be used with those methods.)

Although ACIL does not formally define ACIL L_C, a footnote 2 to the procedure describes it as

> "very similar to Currie's critical level, L_C (Anal. Chem. Vol. 40 No 3, March 1968, p586). It is the level at which there is a given confidence that a result can be distinguished from the blank."

Key features of the ACIL Case I detection limit are as follows:

- The procedure relies on the use of blanks (instead of low-level spikes) to estimate standard deviation.
- When a sufficient number of blanks are used in the calculation, the mean blank result is added

into the calculation to account for high bias exhibited in the blanks.

- ACIL states that at least 7 blanks should be used, but recommends more (as many as 100). If the number of replicates is small, ACIL recommends using a tolerance interval calculation for estimating ACIL L_C. Instead of defining exactly what constitutes a "small" number of replicates, ACIL loosely defines it as fewer than 20 or 30. The confidence level for the tolerance interval also is not specified. If the tolerance level approach is used, the mean blank result is not included in the calculation (unlike the calculation used when there are more than 20 to 30 results).
- If multiple instruments are to be used for the same test and will have the same reporting limit, a minimum of 7 blank results from each instrument should be used, and the results should be combined to generate the standard deviation.
- It is acceptable (and expected) that some results will have negative values, and these negative values should not be censored. Outlier removal is allowed, using a statistically accepted test, if appropriate cautions are taken to guard against excessive or inappropriate rejection of data.
- ACIL provides a verification procedure that is based on comparing the variance of the blank results to results from a new set of blanks.
- ACIL suggests reporting all results that meet or exceed the ACIL L_C.

The formula for ACIL L_C is:

$$LC = \overline{X} + (t_{0.99,n-1} * s)$$

Where $\overline{X}$ is the mean of blank results
s is the standard deviation of blank results, and
n is the number of blank results

ACIL's Case II Detection Limit

For Case II analyses, ACIL's procedures involve picking an initial spike value, adjusting that level up or down based on whether the analyte was detected, and spiking seven replicates at the new level. Details of this procedure are as follows:

- Unlike the procedures used for methods that yield numeric results, ACIL Case II procedures would use spiked samples to determine the detection limit for methods that do not always yield numeric results.
- An initial spike value is chosen based on prior experience. (Detailed guidelines are not provided.)
- One replicate at this level is analyzed; if the analyte is detected, a new sample should be prepared at ½ the initial spike value. If the analyte is not detected at the original level, a new sample should be prepared at 2x the initial spike value. This process is repeated to find the lowest level that can be detected
- Once that level is identified, a minimum of 7 replicates spiked at the lowest level at which that analyte was detected are analyzed, and the replicates must be analyzed in three different batches. If the analyte is detected in all replicates, the Case II MDL is set to this spike value. If the analyte is not detected in all 7 replicates, at least 7 additional replicates are prepared and analyzed at twice this value. If the analyte is detected in all 7 replicates spiked at this higher concentration, the Case II MDL is set to this higher spike value. This process is repeated until the analyte is detected in all 7 replicates.

- The ACIL procedure includes a verification step that consists of spiking the reference matrix at 1 to 3 times the Case II MDL (or 1 to 4 times for multi-analyte methods) to verify that the analyte(s) can be detected. If not, the test is repeated at increasing spike levels until detection, and setting the Case II MDL to the level where the analyte(s) were first detected.
- ACIL suggests reporting all results that meet or exceed the Case II MDL.

5.1.6.2 Assessment of the ACIL L_C against the Evaluation Criteria

The following five subsections discuss the ACIL L_C approach and procedure in the context of the five evaluation criteria that concern detection limit approaches (i.e., Criteria 1-4, and Criterion 6).

5.1.6.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. Although ACIL had not conducted an exhaustive study to test the ACIL L_C, ACIL did apply data generated from member laboratories to the procedure in order to calculate ACIL L_C values. ACIL also compared those values with values produced by EPA's MDL using the same procedure. The results of these tests are included in the public docket supporting this assessment. As part of its own assessment, EPA also tested the procedure using data obtained from the U.S. Geological Survey. In this testing, EPA generated ACIL L_C values, compared those values with values produced by other procedures, and calculated error rates associated with each of the values. Given these studies, the ACIL L_C meets this condition.

Condition 2: It has been subjected to peer review and publication. The ACIL procedure was developed to support ACIL's comments on EPA's 2003, assessment, and it has been subjected to limited peer review within ACIL's member community. Although ACIL references publication of the procedure on the ACIL website, EPA made repeated attempts to locate the procedure on the website over a period of several months, and was unable to locate it. Given the limited peer review beyond the member community, and the lack of publication in a publicly accessible medium, the ACIL procedure does not meet this criterion.

Condition 3: The error rate associated with the procedure is either known or can be estimated. The ACIL procedure meets this condition. According to the formula used for estimating ACIL L_C, the error rate, is specified by α, with a suggested value of 0.01(1%). EPA was able to evaluate this error rate using a small set of data provided by the US Geological Survey. The data included spiked and blank sample results for 18 pollutants, most of which were analyzed by multiple methods, yielding 75 unique analyte/method combinations. For each combination, 25 - 52 blanks were provided. EPA used these blanks to calculate the ACIL L_C, and compared the results of individual blanks with the calculated ACIL L_C. (Details of this assessment are provided in Appendix C.) In theory, no more than 1% of the blanks should have produced a result that exceeded the ACIL L_C. Although the sample size was insufficient to conclusively demonstrate the error rate of the ACIL L_C, the results suggest the actual error rate is close to the estimate of 1%. In this case, the observed mean error rate was 1.9%, and the highest error observed for any method/analyte combination was only 3.8%. Given the small sample size, failure of a single blank could (and did) result in a 3.8% failure rate, suggesting that this study may yield an error rate that is larger than that which would be observed in a larger study. Regardless, it is clear that the ACIL L_C meets this condition because the estimated error rate is given as part of the procedure, and the actual error rate can be calculated through studies such as the one described above.

Condition 4: Standards exist and can be maintained to control its operation. The ACIL L_C is supported by a written procedure (standard) to control its operation. However, the procedure appears to be in draft form, is somewhat difficult to follow and interpret, and contains inconsistencies and ambiguities that are typical of a draft document. In particular, the instructions for Case II are not as clear or detailed as those for Case I.

As an example of the inconsistencies, a footnote to the ACIL L_C states that a tolerance interval will be a more reliable estimate of the ACIL L_C if the number of blanks is small (i.e., fewer than 20 or 30). This implies that the tolerance interval calculation and preferred ACIL L_C will converge as the number of blank results increases. However, this is not the case. The tolerance interval calculation will almost always yield a higher result than the preferred ACIL L_C calculation. The only way that the tolerance interval calculation will result in an ACIL L_C that is either lower or equal to the original ACIL L_C is when blank contamination is high (unlike the preferred ACIL L_C calculation, the tolerance interval calculation does not include the mean of the blanks). It is unclear why the reliability of one calculation compared to the other depends on the number of blank results.

An example of the ambiguities in the procedure is that the alternative calculations, such as the tolerance interval calculation, are presented as suggestions instead of requirements. This could lead to confusion, as now written, if, as ACIL recommends that, the ACIL L_C be used as a reporting limit.

A different type of ambiguity in the procedure concerns the lack of sufficient detail to ensure consistent application. For example, it is not clear exactly when the tolerance interval calculation is to be used because the procedure defines small as 20 - 30 samples. When would 20 samples be sufficient and when would 30 samples be sufficient? Moreover, the tolerance interval calculation does not specify the confidence level used. In an example, both 99% and 95% are given as possibilities. In comparison, the critical value calculated in ASTM's IDE sets the confidence level at 90%. Setting the confidence level at 99% will yield an ACIL L_C value between 11% and 37% higher than one calculated at 95%, based on the numbers of blank results for which the tolerance interval approach is suggested.

Given these problems, the current ACIL procedure does not meet this condition.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. The ACIL L_C was supported by a large number of commenters, most of whom came from the ACIL member community or the environmental laboratory community. Of note, however, is that supporters included instrument vendors, consultants, and several members of the industrial community, including the Inter-industry Analytical Group which offered its own approach to detection and quantitation and which has been highly supportive of the ASTM IDE and IQE approaches. Therefore, EPA believes that the ACIL L_C meets this condition.

5.1.6.2.2 Criterion 2: The approach should address realistic expectations of laboratory and method performance, including routine variability.

The ACIL L_C is designed to address realistic expectations of laboratory and method performance, including temporal variability, instrument variability, analyst variability, and high bias observed in blank results. Based on EPA's analysis of the ACIL L_C presented in Appendix C, EPA believes that the approach meets this criterion provided it is interpreted and applied consistently. (Concerns about the need for clarification of the procedure are described in Section 5.1.6.1, Condition 4).

5.1.6.2.3 Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance

The ACIL L_C meets this criterion. It is similar to the EPA MDL procedure, but it relies on the use of QC data generated during routine laboratory operations, thereby making it even more cost effective than the MDL.

5.1.6.2.4 Criterion 4: The detection level approach should estimate the theoretical concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in a well-operated laboratory

Footnote 2 to the ACIL procedure describes the ACIL L_C as "very similar to Currie's critical level, L_C (Anal. Chem. Vol. 40 No 3, March 1968, p586). It is the level at which there is a given confidence that a result can be distinguished from the blank." According to the formula used for estimating ACIL L_C, the error rate is specified by α, with a suggested value of 0.01(1%). This alpha value means that, if the analyte is not present in the sample, it will be reported as present (i.e., a false positive) no more than 1% of the time. In lay terms, this suggests 99% confidence that, if a substance is reported as present, it really is present. Therefore, the ACIL L_C meets this criterion.

5.1.6.2.5 Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government

If EPA's interpretation of the ACIL procedure is correct, the ACIL L_C appears to meet this criterion.

5.1.7 Evaluation of the USGS Long-term Detection Limit (USGS LT-MDL)

The USGS National Water Quality Laboratory (NWQL) began using the EPA MDL procedure in 1992. USGS NWQL has since developed a variant of the MDL called the long-term MDL (LT-MDL) that has been in routine use by the NWQL since 1999. The procedure for calculating the LT-MDL is described in Section 5.1.7.1 below. Section 5.1.7.2 describes EPA's assessment of the LT-MDL against the five evaluation criteria that concern detection limit approaches (i.e., Criteria 1 - 4, and Criterion 6).

5.1.7.1 Description of the USGS Approach and Procedure

As described in the USGS Open-File Report 99-193, the LT-MDL is a modification of the EPA MDL designed to "capture greater method variability," thereby leading to higher detection limits than those obtained using the EPA MDL procedure. As described by USGS, and noted in Chapter 2, the LT-MDL is based on many of the same fundamental assumptions as the MDL, namely:

1. Normal data distribution,
2. Constant standard deviation from the spike concentration down to zero, and
3. Best-case detection condition (because LT-MDLs typically are determined by spiking the analyte in a clean matrix, e.g., reagent water).

The LT-MDL is determined using low-level spikes of reagent water. The three primary differences between the EPA MDL and the USGS LT-MDL procedures are:

1. Larger minimum number (24) of spike samples,
2. Longer time period, and
3. Combining results from different instruments and analysts in the determination of the LT-MDL.

The USGS Open File Report does not provide an example of the exact calculation used for the LT-MDL. EPA originally presumed that the standard deviation of the results from the 24 spiked sample analyses is multiplied by the Student's *t*-value appropriate for 23 degrees of freedom (t=2.499).

However, USGS comments submitted in response to EPA's assessment of detection and quantitation approaches included a copy of a presentation from the USEPA Region 6 12th Annual Quality Assurance Conference, in Dallas, Texas in August 2002. That presentation provided significant additional information on the calculation of the LT-MDL. Specifically, the LT-MDL uses "F-pseudosigma" (F_σ) in place of S, the sample standard deviation, used in the EPA MDL calculation. F-pseudosigma is a non-parametric measure of variability that is based on the interquartile range of the data. The LT-MDL may be calculated using either the mean or median of a set of long-term blanks, or from long-term spiked sample results, such that:

$$LT\text{-}MDL = M + (t_{0.99,n-1} \times F_\sigma)$$

where:

M	=	mean or median of blank results
n	=	number of spiked sample results, and
F_σ	=	F-pseudosigma, a nonparametric estimate of variability calculated as:

$$F_\sigma = \frac{Q_3 - Q_1}{1.349}$$

where:

Q_3 and Q_1 = the 75th percentile and 25th percentile of spiked sample results, respectively.

USGS believes that the use of F_σ provides an estimate that is more robust and not influenced by outliers.

Like the EPA MDL, the LT-MDL is designed to limit the chance of a false positive result to ≤1%. However, the LT-MDL is designed to be used in conjunction with a "laboratory reporting level" (LRL) as part of an overall reporting scheme for the NWQL. As described by USGS, the LRL is set as a multiple of the LT-MDL. The multiplier varies, depending on the mean/median recovery of the analyte in the spiked samples used for the LT-MDL. If the mean or median recovery is 100%, then the multiplier is 2. At 75% mean or median recovery, the multiplier increases to 2.7, and at 50% recovery, the LRL multiplier increases to 4. In each of these cases, the multiplier is essentially equivalent to dividing twice the LT-MDL by the mean recovery (i.e., 2.7 LT-MDL ~ 2 LT-MDL/75%).

The LRL is designed to achieve a risk of ≤1% for both false negatives and false positives. The reporting scheme used at the NWQL with the LT-MDL and LRL does not censor results at the LRL, and the laboratory reports all results between the LT-MDL and the LRL with a lab-specific flag.

The USGS presentation from the 2002 meeting describes how USGS enhanced the LT-MDL procedure by using their large volume of uncensored blind laboratory blank data as a reality-check on the LT-MDL derived from spiked reagent water samples. In cases where the standard deviation used to calculate an LT-MDL based on blind blank data is significantly different (especially when greater) from the standard deviation used to calculate the spike-based LT-MDL, the blank data are used to calculate the LT-MDL. Blind blank data also are used to evaluate whether the calculated LT-MDL requires an off-set correction for blank bias, i.e, [LT-MDL = (S x Student's *t*) + median or mean blank concentration]. This offset is similar, but not identical, to the ACIL Case I procedure described in Section 2.3.3 of this document. The LT-MDL offset correction compensates for a blank distribution that is not centered at zero (an assumption in the EPA MDL procedure).

The NWQL has found that this blank bias correction to the LT-MDL is especially important for blank-limited analytes, including some metals, total organic carbon, phenol, and nutrients. In practice, the NWQL recalculates the LT-MDL annually, and compares the results between years using Levene's test of equal variance, which they have found to be less influenced by departures from normality than the F-test – an important consideration given that the LT-MDL is based on a non-parametric estimate of variability.

5.1.7.2 Assessment of the USGS LT-MDL against the Evaluation Criteria

The following five subsections discuss the USGS approach and procedure in the context of the five evaluation criteria that concern detection limit approaches (i.e., Criteria 1-4, and Criterion 6).

5.1.7.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. The LT-MDL meets this condition.

USGS has tested and used the LT-MDL since October 1998. Evaluation and use of the LT-MDL began with four methods in use by the NWQL for low-level volatiles by GC/Ms, trace metals by ICP/AES, Kjeldahl nitrogen, and phosphorus. According to the Open File Report, the LT-MDL was scheduled for testing in 17 additional methods, including semivolatile organics, organochlorine pesticides, organophosphorus pesticides, pesticides analyzed by HPLC, metals by ICP/MS, metals by GFAA, and ion chromatography.

EPA used a combination of blank and spiked data submitted by USGS to compute the USGS LT-MDL and compare it to the EPA MDL The blanks were analyzed by USGS over a period of one year and represented a combination of 78 analytes, methods, and matrices, while the spiked sample results represented 39 analytes, methods, and matrices. The analytes were all metals or wet chemistry parameters such as phosphorus and nitrate/nitrite.

Condition 2: It has been subjected to peer review and publication. The LT-MDL does not appear to meet this condition.

Information on the LT-MDL is relatively limited and EPA is not aware of additional USGS publications beyond Open File Report 99-193 and the August 2002 presentation. EPA did not identify any additional publications regarding the LT-MDL in its earlier literature search. The Open File Report itself does not provide any indication that it was subject to a peer-review process.

Condition 3: The error rate associated with the procedure is either known or can be estimated. The error rate is specified by α, with a value of 0.01(1%). Therefore, the LT-MDL may meet this condition.

In its evaluation of USGS data submitted as comments (see Appendix C) EPA found that the mean percentage of blanks results that exceeded the detection limit estimate (LT-MDL) ranged from 3.7% to 4.4%, depending on whether the mean or median blank result was used to estimate LT-MDL. These rates exceeded that of the EPA MDL Therefore, although EPA's evaluation found that the error rate for the LT-MDL exceeded the theoretical error rate designed into the procedure, the error rate can be estimated from actual data.

Condition 4: Standards exist and can be maintained to control its operation. The LT-MDL may partially meet this condition, in that the NWQL may have formal procedures in place that more fully describe the LT-MDL. However, as noted above, the information in the Open File Report does not include an explicit formula for calculation of the LT-MDL, nor are other details of the overall procedure, such as the choice of spiking levels, provided in a clear and consistent fashion. The August 2002 presentation provides critical information about the use of Fσ that is not present the Open File Report. The LT-MDL could meet this criterion, if the procedure were clearly documented by USGS.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. The LT-MDL does not meet this condition

EPA believes that the LT-MDL is only used at the NWQL. Several commenters, including ACIL, suggested that EPA examine the USGS LT-MDL more closely, specifically in regards to its inclusion of long-term variability. There is, however, no evidence in the comments that the concept has achieved a large following among laboratories or other agencies.

5.1.7.2.2 Criterion 2: *The approach should address realistic expectations of laboratory and method performance, including routine variability.*

EPA believes that the LT-MDL meets this criterion because it incorporates the variability of responses over a long time period, and where a laboratory has multiple instruments and analysts running the same analysis, it incorporates variability across instruments and analysts.

5.1.7.2.3 Criterion 3: *The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance*

The LT-MDL partially meets this criterion. However, the LT-MDL is not a detailed readily available "procedure". Also, the LT-MDL requires data collected over a 12-month period. Given that many State regulatory programs require that laboratories provide an annual demonstration of capabilities, including demonstrating their detection limits, the use of the LT-MDL would have to be limited to those laboratories that already have a year's worth of data available. Some other single-lab approach would have to be used for an initial demonstration of method performance.

5.1.6.2.4 Criterion 4: *The detection level approach should estimate the theoretical concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in a well-operated laboratory*

According to the formula used for calculating the USGS LT-MDL, the error rate is specified by α, and the LT-MDL is designed with a value of 0.01(1%). Because the method uses a nonparametric estimate of S, it may not always yield a 1% false positive rate. EPA empirical analysis indicates false positive rates in the range of 3.7% to 4.4%. This compares favorably with the performance of other methods. Thus, the LT-MDL adequately meets this criterion at least in practice.

5.1.7.2.5 Criterion 6: *Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government*

The LT-MDL may meet this criterion. The LT-MDL is designed as part of a broader reporting scheme and it is unclear that EPA, States, and local authorities would be willing or able to use results reported according to that scheme in enforcement scenarios (e.g., "flagged" data).

5.1.8 Evaluation of the Inter-industry Analytical Group (IIAG) Full-Range Validation and Sensitivity Test

In December 2002, the Inter-Industry Analytical Group (IIAG) submitted a proposal to EPA that recommends (1) a sensitivity test intended to "replace the MDL as a test of whether an individual laboratory is performing adequately," and (2) an interlaboratory validation study design intended to characterize precision and accuracy of methods used for regulatory compliance. Although their approach was received too late for consideration prior to publication in the 2003 Assessment Document, EPA provided notice of the approach, requested public comment on it, and agreed to evaluate the IIAG approach in updating the 2003 assessment. Section 5.1.8.1 describes the IIAG approach, and Section 5.1.8.2 describes EPA's assessment of the IIAG approach against the applicable evaluation criteria.

5.1.8.1 Description of the IIAG Approach and Procedure

Full Range Validation

IIAG has proposed that, EPA commit to performing interlaboratory method validation studies designed to produce a "full range" of data, including precision and accuracy, from the point of instrument detection to the upper end of the working range. IIAG has indicated that "such a full range validation will enable EPA to consider DL/QL options in light of data quality objectives without being constrained by a limited database." IIAG suggests that, at a minimum, EPA should commit to performing such full range validation studies for all new methods that it develops and that all organizations submitting new methods for EPA approval should be required to provide the full range data as well.

Sensitivity Test

IIAG also has proposed that EPA consider the use of a "sensitivity test" instead of the MDL to demonstrate that a laboratory is capable of performing according to EPA expectations at the lower range of a test method. IIAG's suggested process for developing this test is as follows:

- EPA would first identify the lowest concentration at which the entire analytical system gives a recognizable signal and acceptable calibration point.
- EPA would then select a simple dilution of that concentration, and develop QC criteria based on the test results from several laboratories performing the test at that dilution (in the same way that QC criteria are developed by EPA for initial precision and recovery demonstrations in methods

such as Method 1631).

- Laboratories could then perform an "Initial Performance Demonstration" (IPD) of their capability to achieve the desired sensitivity by (1) analyzing several replicates of the same sample dilution (using the full method), (2) using the results to compute the standard deviation, and (3) confirming that the results fall within the QC criteria range. IIAG emphasizes that the dilution level would not be considered the detection level, but rather a performance level.

IIAG further suggests that this multi-replicate IPD test would be verified on an ongoing basis. To minimize complexity, IIAG suggests that the ongoing test be conducted at the same spike level as their "Initial Performance Demonstration." IIAG did not suggest a specific frequency for conducting these ongoing tests.

Finally, IIAG suggests that EPA commit to using this IPD sensitivity test in lieu of the MDL, and that EPA express a willingness, subject to funding availability or a third party commitment, to perform testing as necessary to develop "sensitivity" QC criteria, and to modify the few existing Part 136 methods that require the MDL for IPD.

Section 5.1.8.2 below discusses EPA's evaluation of the scientific elements proposed IIAG approach.

5.1.8.2 Assessment of the IIAG Approach against the Evaluation Criteria

The following six subsections discuss the IIAG approach and procedure in the context of the six evaluation criteria. The first three criteria apply to both detection and quantitation limits, Criterion 4 applies to detection limits only, Criterion 5 applies to quantitation limits only, and Criterion 6 applies to both. Because the IIAG full range validation and sensitivity test approach applies to both types of limits, all 6 criteria are discussed below.

5.1.8.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. To EPA's knowledge, the IIAG sensitivity test approach has not been tested by any organization, including IIAG. The IIAG approach is still a rough framework, and basic details, such as the number of replicates required and the actual spiking levels to be used, still need to be specified. Testing of the approach in its current framework is possible but would be very expensive, one might have to conduct tests with multiple spiking levels and with varying numbers of replicates, for example, to be sure that the tests will reflect the final sensitivity test procedure. If the procedure were refined to describe the exact steps and requirements, it could be tested more efficiently.

IIAG's full validation study approach can be and has been tested. For example, EPA conducted a full interlaboratory validation study of Method 1631 prior to promulgating the method at 40 CFR 136. That study, which involved 12 participating laboratories, yielded an overall mean percent recovery of 93 and an overall relative standard deviation of 13 percent across all samples.

IIAG has stated that "Although the full-range interlaboratory is aimed at characterizing a method's ability to quantify rather than to detect a pollutant concentration, the study could be used to establish an interlaboratory detection level as well" and "The best solution for performing a full-range validation to establish detection and quantitation levels and precision and bias for promulgating nationwide standards and compliance levels is the ASTM IDE/IQE approach."

The ASTM IDE and IQE are constructed by fitting a model to variability versus concentration data, rather than being derived from the standard deviation of replicate measurements of a single concentration level. As discussed in Section 5.1.2 and detailed in Appendix C, EPA used data from the Episode 6000 study to compare IDEs calculated using data from all 16 concentration levels reported to IDEs calculated using data from only 5 of the concentrations (i.e., at 5, 10, 20, 40, and 80 times the standard deviation of replicate measurements of a blank sample or the lowest level at which measurements could be made). Results of the comparison are summarized in Table 9 of Appendix B to the draft TSD. The results show that the median 16-point IDE is approximately 1.3 times greater than the median 5-point IDE, indicating that data resulting from measurements of concentration levels in the region of detection and quantitation in some cases may yield lower IDE's than data from a wider range of concentration data.

EPA refers readers of this document to Sections 5.1.2.2.1, 5.2.2.2.1, 5.2.2.2.2, and Appendix B for a discussion of additional reasons why EPA believes the ambiguities and inconsistencies in IDE/IQE procedures preclude these procedures from being the best solution for performing a full range study to estimate detection and quantitation limits.

Condition 2: It has been subjected to peer review and publication. The IIAG procedure does not meet this criterion. EPA is not aware of any peer review or publication of the document in a peer reviewed journal. The IIAG document was submitted directly to EPA by the Petitioners, and EPA made the document available to the public for comment.

Condition 3: The error rate associated with the procedure is either known or can be estimated. At present the IIAG's approach consists of a proposed framework rather than a detailed procedure. It lacks key specifics, such as how many replicates would be used in the IPD phase of the test, and what spiking levels would be used. IIAG suggests that EPA would select these levels, and suggests "probably 4 - 7" for the number of replicates.

While IIAG suggests the dilution would be a simple dilution of the lowest calibration standard, offering "1/3 or ½, for example", they also state that "It is not absolutely necessary to reduce the spike level below the lowest calibration point, however, and the sensitivity test could be performed with a spike at the lowest calibrations standard instead of at a dilution of it." No guidelines are offered for which of these levels (or other levels) should be chosen, nor are guidelines offered for the number of replicates needed.

Given the lack of detail, the current framework would be subject to different interpretations by different readers or users, and the error rate associated with the procedure would vary depending on how the procedure was implemented. Because the error rate is neither known, nor can it be estimated, the IIAG approach does not meet this condition.

The IIAG procedure is a framework with interesting aspects for further consideration by the full scientific and regulatory community. EPA would be willing to work with IIAG and other stakeholders to identify the details needed to augment this framework to where it would meet this condition.

Condition 4: Standards exist and can be maintained to control its operation. As previously noted, the IIAG approach consists of a proposed framework rather than a detailed procedure framework, and lacks key details that are needed to control its operation. Given the lack of detail, the current framework does not meet this condition.

Again this procedure is a framework with interesting aspects for further consideration by the full scientific and regulatory community. EPA would be willing to work with IIAG and other stakeholders to identify the details needed to augment this framework to where it would meet this condition.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. The IIAG procedure does not meet this condition. It was suggested by a limited group of the relevant scientific community (industry firms that comprise the "Inter-industry Analytical Group" and whose wastewater discharges are regulated under the Clean Water Act), and comments on the their approach were mixed. Excluding comments submitted by IIAG itself, EPA received comments from:

- Three electric power producers whose discharges also are regulated under CWA,
- Two publicly owned wastewater treatment systems which regulates industrial discharges to their system under CWA and whose own discharges are subject to regulation under CWA,
- Two commercial environmental laboratories that utilize the methods and detection limit procedures approved at 40 CFR 136 to serve their client's needs,
- One trade council, and
- One private citizen.

All three electric power firms supported the IIAG approach. The two publicly owned treatment systems offered mixed reviews. One supported the sensitivity test and offered suggestions for further consideration; the other opposed the sensitivity test but offered limited support of the interlaboratory validation studies, suggesting that they be limited to the relatively small group of priority pollutants whose water quality based effluent limits are below the method reporting levels. Both of the environmental laboratories were opposed to the IIAG approach, and the trade council suggested that it should be used "as an alternative procedure for dischargers to implement... on a site-specific basis, at their discretion", noting that "As an alternate method, facilities would be able to deal with this on a case-by-case basis and would not need to utilize numerous laboratories to develop the more elaborate detection limits and quantitation limits that the IIAG proposes".

Given these comments, it would appear that acceptance may be widespread within the industrial discharger community, but it is not widespread among the *entire* relevant scientific community.

5.1.8.2.2 Criterion 2: *The approach should address realistic expectations of laboratory and method performance, including routine variability.*

In principle, the IIAG sensitivity test meets this criterion because it is intended to provide realistic information about laboratory and method performance, both with an initial demonstration and with follow-up demonstrations that provide information concerning routine variability. However, and as previously noted, the procedure is not sufficiently detailed to allow laboratories to meet this criterion. To clearly meet this criterion, detailed specifications to allow for consistent implementation of the procedure throughout the laboratory community need to be developed.

5.1.8.2.3 Criterion 3: *The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance*

If the IIAG framework was developed into a detailed procedure, this sensitivity approach would meet this single laboratory criterion. This could complement the IIAG full range validation study, which does not meet this criterion because it is an interlaboratory procedure. The sensitivity test, once detailed, could be performed by a single laboratory and used to evaluate method performance.

5.1.8.2.4 Criterion 4: *The detection level approach should estimate the theoretical concentration at which there is 99% confidence that the substance is actually present when the analytical method is performed by experienced staff in a well-operated laboratory*

Because the spiking level to be used in IIAG's sensitivity test is not defined it is not possible to evaluate whether that test meets or does not meet this criterion. IIAG also suggests that a full-range validation study should be used to establish an interlaboratory detection limit, and recommends use of the ASTM IDE procedure as the best means of doing so. If this is the case, the full range validation study would fail this criterion for the reasons given in Section 5.1.2.2.4 regarding the IDE.

5.1.8.2.6 Criterion 5: *The quantitation limit should identify the concentration that gives a recognizable signal that is consistent with the capabilities of a method when a method is performed by experienced staff in well-operated laboratories.*

The IIAG's proposed sensitivity test requirement is likely to meet this criterion once details regarding the procedure are specified. Depending on the spiking levels that are specified in the final procedure, however, it is very likely that the IIAG sensitivity test may not identify the *lowest* concentration at which the signal is recognizable when the method is performed by experienced staff in a well-operated laboratory.

5.1.8.2.6 Criterion 6: *Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government*

IIAG's suggested use of a full range validation study meets this criterion because such validation studies provide useful information about the performance of the method. As noted previously, EPA typically conducts interlaboratory validation studies at multiple concentrations ranges before promulgating a method for nationwide use at 40 CFR part 136. However, for the reasons discussed elsewhere in this document, EPA does not agree that data collected across the full range of the method should be used to establish detection or quantitation levels.

In the absence of a detailed procedure that could be use to fully evaluate IIAG's, it is difficult to determine if the IIAG sensitivity test meets this criterion.

5.2 Quantitation Limit Approaches

Sections 5.2.1 through 5.2.4 describe EPA's assessment of four quantitation limit approaches. Each discussion is divided into two major subsections. The first subsection describes the approach and, where applicable, the procedure that supports the approach, and the second subsection details EPA's assessment of the approach based on the five criteria established in Chapter 4 for evaluating quantitation limit approaches. These criteria are Nos. 1 -3, 5 and 6; No. 4 only is applicable to detection limits.

5.2.1 Assessment of the EPA Minimum level of Quantitation (ML)

Section 5.2.2.1 provides an overview of the ML approach and the procedures used to implement the approach. Section 5.2.2.2 contains EPA's assessment of the ML against the five evaluation criteria that concern quantitation limit approaches (i.e., Criteria 1-3, and Criteria 5 and 6).

5.2.1.1 Description of the ML Approach and Procedures

The definition of the ML includes a statement of the approach and the procedures used to establish the ML. This definition states that the ML is:

> *"the lowest level at which the entire analytical system must give a recognizable signal and acceptable calibration point for the analyte. It is equivalent to the concentration of the lowest calibration standard, assuming that all method-specified sample weights, volumes, and clean up procedures have been employed. The ML is calculated by multiplying the MDL by 3.18 and rounding the results to the number nearest to (1, 2, or 5) x 10^n, where n is an integer."*

The ML is designed to provide a practical embodiment of the quantification level proposed by Currie and adopted by IUPAC. It is functionally analogous to Currie's "determination limit" (described in Chapter 2, Section 2.1) and the American Chemical Society's Limit of Quantitation (LOQ). The LOQ is discussed in Section 5.2.3 of this chapter. Chapter 2 (Section 2.2.2) describes the ML approach in additional detail.

The first part of the ML definition (i.e., the lowest level at which the system gives a recognizable signal and acceptable calibration point for the analyte) ties the quantification limit to the capabilities of the measurement system. The second part of the ML definition provides a procedural means for establishing the ML.

The procedural component of the definition is designed to yield an ML value that equals approximately 10 times the standard deviation of replicate analyses used to determine the MDL. (The exact value corresponding to 10 times the standard deviation is rounded to avoid error that would arise from preparation of calibration standards at exact, "unrounded" concentrations.) The 3.18 multiplier is derived by dividing 10 by the value of the t-statistic for seven replicates. Laboratories that choose to perform MDL studies with more than the required minimum of seven replicates follow the instructions in appendix B of 40 CFR part 136 to select the t-statistic value for the number of replicates used. Therefore, the 3.18 multiplier for the ML calculation should be proportionally adjusted. Similarly, the Student's *t*-value is adjusted when a laboratory performs the optional iterative test described in Step 7 of the MDL procedure, or if outlier testing results in the use of less than seven replicates to establish the MDL.

5.2.1.2 Assessment of the ML against the Evaluation Criteria

The following five subsections discuss the ML approach and procedure in the context of the five evaluation criteria that concern quantitation limit approaches (i.e., Criteria 1-3, and Criteria 5 and 6).

5.2.1.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. The ML meets this condition. The ML has been used experimentally since 1979 and in the regulatory context since 1984. The ML is tested each time a

laboratory calibrates an instrument because methods that include the ML require that it be included as the lowest non-zero standard in these calibrations.

EPA also has tested the MDL and ML procedure with ten different techniques at decreasing spike concentrations to evaluate how well the MDL and ML procedures characterized the region of interest in preparation for this reassessment of detection and quantitation limit approaches. Results of the study suggest that (1) although the calculated MDL and ML could vary depending on the spike level used, the procedure was capable of reasonably estimating detection and quantitation limits when the full iterative MDL procedure was employed, and (2) the rounding process employed to determine the ML generally yielded consistent MLs even with slight variations in the calculated MDL. EPA recognizes that additional guidance may be necessary on the selection of the initial spiking level and uses of the iterative procedure.

In other words, if the procedure for establishing an ML is properly implemented for a given method, it will yield an ML value that is consistent with the approach, and this ML value can be verified (tested) by a laboratory when it calibrates the instrument used to analyze samples by the method.

One of the stakeholders commenting on EPA's 2003 assessment suggested that the ML failed to meet this condition because EPA should have tested it in "real world" matrices. EPA does not agree with this suggestion for several reasons. First, it is not practical or possible to test detection limits in <u>every</u> real world matrix, and there is no consensus as to which real world matrix would represent an appropriate real world matrix for testing. Second, many real world matrices contain the target pollutant at levels well above the detection or quantitation limit, making it impossible to characterize what can and cannot be detected at low levels. In theory, the sample could be diluted to dilute the target pollutant, but in practice sample dilution would also likely dilute any interferences that might be present, thereby defeating the purpose of using a real world matrix. The current EPA approach, which exhaustively tests the ML procedure in a reference matrix using multiple techniques and ten different concentrations that span the entire region of interest, is more than adequate to constitute "testing" of the ML procedure. On the other hand, where data suggests that matrix interferences may significantly affect achievable quantitation and detection limits, this should be considered by a permit writer on a case by case basis.

<u>Condition 2: It has been subjected to peer review and publication</u>. The ML has not been published in a peer reviewed journal. However, it was evaluated by four peer reviewers as part of EPA's assessment of detection and quantitation limits. These reviewers noted that:

> *"The MDL and ML concepts evaluated in Section 5.1.1 and 5.2.1, respectively, are shown in this evaluation to be technically sound and practical."* (Wait, 2002)
>
> *"With respect to the limit of quantitation concept, the EPA ML is as good as any of the others given..."* (Rocke, 2002)
>
> *"The MDL and ML have stood the test of time and provide a proven methodology which meets evaluation criteria stated in the TSD."* (Cooke, 2002).

In addition, the definition of the ML describes the approach and the procedures used to establish the ML. This definition is included in EPA Method 1631, which was extensively peer reviewed in accordance with EPA policies on peer review prior to publication and promulgation. Given that EPA's policies on peer review are as stringent as or more stringent than those used by many published journals, the ML has met a high standard of scientific review and scrutiny, and therefore, meets the intent of this condition.

Condition 3: The error rate associated with the procedure is either known or can be estimated. If rounding is not considered, the error can be easily estimated. The calculation is still straightforward, but tedious, when the ML rounding procedures are employed. Given these caveats, the ML partially meets this condition.

Condition 4: Standards exist and can be maintained to control its operation. The ML meets this criterion. Detailed procedures (i.e., standards) for establishing the ML are embodied in the definition.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. The ML meets this condition. The ML is analogous to the American Chemical Society's LOQ and to the ISO/IUPAC quantification limit, which suggests widespread acceptance.

5.2.1.2.2 Criterion 2: The approach should address realistic expectations of laboratory and method performance, including routine variability.

The ML procedure meets this criterion. It is designed to provide a means by which a laboratory can demonstrate performance with a method under routine laboratory operating conditions. All recently developed EPA CWA methods require that a laboratory calibrate its instrument prior to analyzing environmental samples. The ML is defined as the lowest non-zero standard in the laboratory's calibration, and therefore, reflects realistic expectations of laboratory performance with a given method under routine laboratory conditions (i.e., under conditions of routine variability).

The ML is based on the standard deviation of replicate analyses used to establish the MDL. As described in Section 5.1.1.2.2, these analyses are performed to characterize laboratory and method performance, including routine variability, at low concentrations. When a laboratory performs an MDL study with seven replicates and multiplies the results by 3.18, the laboratory has demonstrated that it can achieve expected levels of performance at the ML.

Due to the variability inherent in measurement science, instrumentation, and the humans conducting analyses, laboratories may routinely obtain limits that are lower or higher than those obtained in another laboratory. Thus, when an ML is determined during method development, it is important to determine that ML in more than one laboratory to ensure the ML published in the method reflects demonstrated expectations of method performance in a community of laboratories. It is not necessary for this community to include the entire universe of all possible laboratories that might desire to practice the method. Rather, during the stages of method development and validation, this community only should include well-operated laboratories with analysts who are experienced with the techniques used in the method, and have some familiarity conducting all of the steps in the new method before generating MDLs that will be published with the new method. See Section 5.1.1.2.2 for additional discussion of this topic.

5.2.1.2.3 Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance.

The ML meets this criterion. It is designed for use by a single laboratory. The ML can be directly determined from the MDL, which is among the most affordable of procedures for determining detection limits (see discussion in Section 5.1.1.2.3 for additional details), by a simple multiplication of the MDL and a application of a rounding procedure.

5.2.1.2.4 Criterion 5: *The quantitation limit approach should identify the concentration that gives a recognizable signal that is consistent with the capabilities of the method when a method is performed by experienced staff in well-operated laboratories.*

The ML meets this criterion. The ML can be verified in a laboratory each time it calibrates an instrument. This calibration depends on identification of a recognizable signal for the analyte. In addition, because EPA includes the ML as the low point in the calibration range, that concentration is within the capabilities of the method, as demonstrated by either multiple single-laboratory studies or a multi-laboratory study of the method.

Notwithstanding the preceding, analysis of Episode 6000 data (see appendices) produced anomalous results from two methods (EPA 502.2 and 524.2) that employ instrument thresholds. For 17% of EPA 502.2 and 49 % of EPA 524.2 analytes the calculated ML was below the concentration at which all seven spiked replicates were detected, i.e. less than the lowest MDL spike. The Episode 6000 dataset is not reflective of a typical compliance measurement or method development study because the range of concentrations studied encompassed several orders of magnitude and included concentrations well below the MDL. This atypical range was employed to push the limits of the instrumentation and the theory underlying determination of the variability of measurements.

In a qualified operating laboratory, or during a method development study, if MLs were calculated to be less than the concentration at which all seven spiked MDL replicates were detected, the laboratory would take corrective measures. When a method is developed for EPA's CWA program, each laboratory in a multi-laboratory study would consult with EPA and take corrective measures, such as calibration adjustments so that reported MDLs are above the signal threshold. In these cases, the calculation of ML = 3.18 * MDL always yields a value greater then the MDL and meets the criterion of "recognizable signal".

5.2.1.2.5 Criterion 6: *Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government.*

The ML meets this criterion. It has been used in Clean Water Act programs since 1984.

5.2.2 Assessment of the IQE

The Interlaboratory Quantitation Estimate (IQE) was published by ASTM, International in 2000 as standard D 6512. The IDE was developed with support from members of the regulated industry in an attempt to provide a comprehensive quantitation limit procedure that addresses the concerns of the regulated industry, statisticians, and analysts. A brief summary of the procedure for establishing an IQE is given in Section 5.2.2.1. Section 5.2.2.2 presents EPA's assessment of the IQE against the five criteria established for evaluating quantitation limit approaches (i.e., Criteria 1-3, and Criteria 5 and 6).

5.2.2.1 Description of the IQE Approach and Procedure

The ASTM Designation D 6512 is the *Standard Practice Interlaboratory Quantitation Estimate.* As stated in the practice:

> *"$IQE_{Z\%}$ is computed to be the lowest concentration for which a single measurement from a laboratory selected from the population of qualified laboratories represented in an interlaboratory study will have an estimated Z % relative standard deviation (Z % RSD, based on interlaboratory standard deviation), where Z is typically an integer multiple of 10, such as 10, 20, or 30, but Z can be less than 10."*

The IQE is determined and verified using a procedure containing 5 major steps with approximately 46 substeps and conditions. The full text of the IQE procedure is available from ASTM International. The 5 major steps and their functions are given in Section 6 of the IQE procedure and are summarized below:

1. Overview of the procedure.

2. IQE Study Plan, Design, and Protocol - in this section, the task manager (study supervisor) chooses the analyte, matrix, and analytical method. Details are given for the appropriate range of study concentrations; the model of recovery vs. concentration; the study protocol (ASTM Practice D 2777 is suggested); the instructions to be given to the participating laboratories, including reporting requirements; the allowable sources of variation; and the number of laboratories, analysts, measurement systems, and days over which the study will be conducted.

3. Conduct the IQE Study, Screen the Data, and Choose a Model - after the study data are collected and screened according to ASTM Practice D 2777, the interlaboratory standard deviation (ILSD) versus concentration data are tabulated and one of three models is fit to the data. The first attempt is at fitting a constant model. If the attempt fails, a straight-line model is attempted. If the straight-line model fails, a hybrid (Rocke/Lorenzato) model is fit. After fitting, the model is evaluated for reasonableness and lack of fit. If the model fails, the study supervisor determines if a subset of the data should be analyzed or if more data are needed.

4. Compute the IQE - the IQE is computed using the ILSD model selected in Step 3 to estimate the relative standard deviation as a function of concentration. The first attempt is at 10% RSD ($IQE_{10\%}$). If this attempt fails, $IQE_{20\%}$ is tried, then $IQE_{30\%}$. IQEs greater than 30% are not recommended.

5. Nontrivial Amount of Censored Data - this section of the IQE procedure addresses the effect of "non-detects" or "less-than." Suggestions are given to see if uncensored data can be obtained from the laboratories or if the study needs to be augmented with additional data. Suggestions are given for fitting a model to data that contain less than 10% non-detects or less-than to produce an IQE.

5.2.2.2 Assessment of the IQE Against the Evaluation Criteria

The following five subsections discuss the IQE approach and procedure in the context of the five evaluation criteria that concern detection limit approaches (i.e., Criteria 1-3, and Criteria 5 and 6).

5.2.2.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. The Electric Power Research Institute provided input into the design of EPA Method 1631 and 1638 Validation Studies for the purpose of calculating IDEs and IQEs. EPRI also calculated IDEs and IQEs based on these data. These two datasets include a total of ten metal analytes and therefore do not cover a wide range of analytical techniques and methods. Other than these two datasets, EPA is not aware of any organization, including ASTM, that has conducted a study to test the IQE procedure as written (i.e., designed and implemented an interlaboratory study involving multi-laboratory analysis of multiple concentrations of each matrix of interest). It has been tested by its developers using simulated data sets and on interlaboratory data sets that do not adequately characterize the low level region of interest. As part of this reassessment, EPA tested a variant of the IQE procedure on single-laboratory data sets that were designed to characterize an analytical method in the region of detection and quantitation. Despite the lack of comprehensive testing performed to date, the IQE procedure can be tested if sufficient resources are invested.

Condition 2: It has been subjected to peer review and publication. Although the IQE has not been published in the peer-reviewed scientific literature, the IQE has undergone review and ballot by members of ASTM Committee D 19, many of whom are qualified peer reviewers. Thus, the IQE meets the intent of this condition (i.e., submission to scrutiny of the scientific community). In addition, the IQE was reviewed by four peer reviewers as part of EPA's assessment of detection and quantitation limit approaches.

Condition 3: The error rate associated with the procedure is either known or can be estimated. In theory, an expert statistician could estimate the error rate of the IQE. However, the IQE procedure is extremely complex from an analytical chemistry and statistical perspective. As a result, it is unlikely that the error rate could be estimated by the staff of an environmental testing laboratory. Moreover, in attempting to follow the IQE procedure during this reassessment, EPA found the procedure to be subjective, particularly with respect to selection of an appropriate statistical model. The subjective nature of the procedure is likely to yield different IQEs from the same data set, depending on the staff involved in analyzing the data and performing the calculations. (The likelihood of this problem is illustrated in appendix B to this Assessment Document.) This subjective variability eliminates the ability to estimate the actual error associated with the IQE. Therefore, the IQE does not meet this condition.

As discussed in Section 5.2.2.1, Condition 3, regarding the IDE, one stakeholder stated that concerns about the complexity and subjectivity of the IQE (and IDE) procedures were unimportant, in part, because IQEs calculated using different models were very close, and in part, because "user friendly-software that will automatically perform the IDE and IQE calculations. EPA obtained copies of such software from the commenter and used that software to evaluate the validity of this comment. As described at length in Section 5.2.2.1, EPA concluded that 1) the subset of models used varies among the software packages, 2) the software packages do not always apply the same model to the same data sets, and 3) even if the same model is used, there is a large amount of variability between the results produced when applying the different software packages to the same set of data. Based on these differences, EPA concluded that the available software programs do not remove all complexity and subjectivity from the IQE calculations. Instead, the software programs appear to introduce new issues by using steps not included in the ASTM procedures.

Condition 4: Standards exist and can be maintained to control its operation. The IQE approach and procedure is supported by a published procedure (standard) to control its operation. The procedure gives the steps to be followed in determining the IQE and instructs the study supervisor how to gather the data and compute an IQE.

There are several "gray areas" in the published procedure. The most significant gray area is in model selection. The procedure provides insufficient guidance on the use of residual plots as a basis for selecting models and as a result, selection of the model may be very subjective, especially if the number of concentrations is low. The discussion of what model to use after rejecting the hybrid and linear models also is very vague. The exponential model is mentioned, as well as models with more than one coefficient. In addition, fitting the "constant model" is never discussed in detail. Most likely, this is done by simply calculating a mean (weighted if necessary) of the variances from the different concentrations, however such a calculation is never explicitly stated. As discussed under Condition 4 of Section 5.1.2.2.1 (scientific validity of the IDE procedure), there appear to be inconsistencies between the IDE and IQE that suggest conceptual conflicts between these two standards.

Based on these findings (along with those discussed under Criterion 2 below), the procedure is not sufficient to control operation of the IQE because of the high degree of subjectivity involved in implementing the procedure, statistical errors in the procedure, and internal inconsistencies with the IDE. Therefore, the IQE does not meet this condition.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. The IQE was published by ASTM four years ago (2000). EPA has not found an IQE in the open literature or in an analytical method, including an ASTM method.

5.2.2.2.2 Criterion 2: The approach should address realistic expectations of laboratory and method performance, including routine variability.

The IQE procedure is designed to reflect expectations of interlaboratory performance, including routine variability. The procedure contains extensive instructions for dealing with unusual conditions, including sources of variability and outliers. Based on studies of the single-laboratory variant of the procedure in which the model selection proved to be highly subjective, it is not clear that IQE procedure would demonstrate realistic expectations of laboratory and method performance.

The IQE procedure suggests attempting to fit study results to a constant, linear, or hybrid model. If all of these fail, the procedure suggests trying a different model, such as the exponential model. (The exponential model figures more prominently in the IDE procedure, where it is one of the three main models discussed, replacing the Rocke and Lorenzato model.) Although the exponential model may be appropriate for the IDE (which is not tied to a fixed RSD), it yields unacceptable results when applied to the IQE procedure. Under the exponential model, relative variability (standard deviation divided by the true concentration) does not consistently decrease with increasing concentration (i.e., as concentration increases, relative variability decreases down to a specific percentage, and then begins to increase). This is not realistic of laboratory and method performance. In addition, the exponential model will often result in having two possible values each for $IQE_{10\%}$, $IQE_{20\%}$, and $IQE_{30\%}$.

Another concern with the IQE procedure is that use of the non-mandatory appendices in ASTM D 6512 to determine the fit of a model may produce results that differ from those that would be obtained using the default procedures for testing model fit that are built into off-the-shelf statistical software, such as the Excel files discussed in Condition 3.

Given the subjectivity and confusion involved in selecting the model, EPA tried using the same data set to calculate a single-laboratory variant of the IQE with each of the available models and found that the calculated IQEs varied widely when different models were used.

Based on the problems described above, EPA believes the IQE does not meet this criterion.

5.2.2.2.3 Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance.

The IQE procedure is neither practical nor affordable in a single-laboratory context.. It is designed for use by an ASTM study supervisor or task manager and not as a procedure that a single laboratory can use to evaluate method performance. EPA is aware that ASTM Committee D 19 is contemplating development of a within-laboratory quantitation estimate (WQE), but the WQE has not been approved through an ASTM ballot, and therefore, it cannot be adequately evaluated at this time. The WQE may meet this criterion, but the IQE does not.

Regarding affordability, EPA estimates that the cost of implementing IQE procedure would be more than twice the cost of EPA's present implementation of the ML. The increased cost stems from the additional low-level data required to assure that variability versus concentration is being characterized in the region of detection and quantitation, challenges involved in applying the statistical procedures in the IQE, and because of the anticipated reanalysis and rework required if either the procedure failed to produce an IQE or if the resulting IQE failed to meet the specifications in the IQE procedure.

5.2.2.2.4 Criterion 5: The quantitation limit approach should identify the concentration that gives a recognizable signal that is consistent with the capabilities of the method when a method is performed by experienced staff in well-operated laboratories.

If the IQE were developed in an interlaboratory study that met the requirements of D 6512, the calculated IQE would likely be achievable by experienced staff in a well-operated laboratory. Therefore, the IQE meets this criterion.

However, similar to the discussion of criterion 5 for the ML (section 5.1.2.4) anomalous results occur. Analysis of episode 6000, analysis of Episode 6000 data (see appendices) produced anomalous results from two methods (EPA 502.2 and 524.2) that employ instrument thresholds. For 9% of EPA 502.2 and 59 % of EPA 524.2 analytes the calculated single-lab IQE was below the concentration at which all seven spiked replicates were detected. These results indicate that an IQE study coordinator, after calculating IQE from multi-labs results, would have calculated IQEs below the instrument threshold. The IQE procedure is silent on what happens in this case. As previously noted, the Episode 6000 dataset is not reflective of a typical compliance measurement or method development study because the range of concentrations studied encompassed several orders of magnitude and included concentrations well below the detection limit. And this dataset was not developed according to the procedures in D 6512 (the IQE).

5.2.2.2.5 Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government

There is no database of IQE values for CWA analytes that were calculated according to D 6512. These are the data with which one would compare existing CWA limits and thereby assess the effect of using IQEs as reporting and compliance limits in CWA programs.

5.2.3 Assessment of the ACS Limit of Quantitation

The Limit of Quantitation (LOQ) was developed by the Committee on Environmental Improvement of the American Chemical Society (ACS) and published in the same two papers as the LOD.

5.2.3.1 Description of the ACS LOQ Approach and Procedure

The 1983 *"Principles"* define the LOQ as:

> *"... the level above which quantitative results may be obtained with a specified degree of confidence."*

The same relationship used to define the LOD is used for the LOQ:

$$S_t - S_b \geq K_d \sigma$$

but the recommended minimal value for K_d be set at 10. Thus, the LOQ is 10σ above the gross blank signal, S_b. According to the 1983 publication, the LOQ corresponds to an uncertainty of ±30% (10σ ± 3σ). This uncertainty statement is based on σ equal to 10% of the LOQ.

Neither the 1980 nor 1983 ACS publications provide a specific procedure for estimating the LOQ, nor do they provide a minimum number of observations needed to estimate the gross blank signal or the variability term σ_b.

5.2.3.2 Assessment of the ACS LOQ Against the Evaluation Criteria

The following five subsections discuss the ACS LOQ approach and procedure in the context of the five evaluation criteria that concern detection limit approaches (i.e., Criteria 1-3, and Criteria 5 and 6).

5.2.3.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. Testing of the LOQ is hampered by the lack of a supporting procedure for establishing an LOQ, and a conceptual dependence on the variability of blank measurements. If the blank measurements fail to produce a response, it is impossible to calculate an LOQ because the standard deviation of multiple zero-value results is zero. One solution for testing the approach is to assume that the LOQ is approximately equivalent to the ML as the blank signal approaches zero. If this is a reasonable assumption, the ML procedure is a viable means for testing the LOQ approach, and the LOQ would meet this condition.

Condition 2: It has been subjected to peer review and publication. The ACS LOQ definition was published in the peer-reviewed journal *Analytical Chemistry* in 1980 and 1983. Therefore, the ACS LOQ meets this condition.

Condition 3: The error rate associated with the procedure is either known or can be estimated. The LOQ meets this condition. The definition of the LOQ specifically estimates the uncertainty associated with a concentration at the LOQ as ±30% based on 10% RSD (Kd = 10). Other choices may be made based on study requirements, policy judgments and/or specific results.

Condition 4: Standards exist and can be maintained to control its operation. The ACS LOQ lacks a clearly defined procedure for estimating the important terms required to derive it. Although it may be possible to derive ACS LOQ values from data used to derive EPA MDL values, there is no discussion of using replicate blanks, replicate spiked samples, or a minimum recommendation for the number of replicates. Therefore, the ACS LOQ does not meet this condition.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. Because the ACS does not develop and publish reference analytical methods, it is difficult to determine the degree of acceptance of the LOQ. EPA has not investigated the numbers of papers published in ACS journals that include LOQ values, but EPA's literature search for detection and quantitation approaches did not uncover a large number of citations that promote the LOQ in particular.

5.2.3.2.2 Criterion 2: The approach should address realistic expectations of laboratory and method performance, including routine variability

The LOQ approach is designed to address realistic expectations of laboratory and method performance, including routine variability, and therefore, it appears to meet this criterion. Because the ACS has not published a procedure to implement the approach, in practice the LOQ provides no direct means for demonstrating this performance. The ACS LOQ, the only partially meets this criterion.

5.2.3.2.3 Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance.

Because the ACS LOQ approach is not supported by a clearly defined procedure for establishing the LOQ, it does not meet this criterion.

5.2.3.2.4 Criterion 5: The quantitation limit approach should identify the concentration that gives a recognizable signal that is consistent with the capabilities of the method when a method is performed by experienced staff in well-operated laboratories.

Given the relationship of the ACS LOQ to the ML, EPA believes the LOQ meets this criterion for the reasons outlined in Section 5.2.1.2.4, which discusses EPA's assessment of the ML against Criterion 4 for evaluating quantitation limit approaches.

5.2.3.2.5 Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government.

In the absence of a procedure for determining LOQ values, the ACS LOQ does not meet this criterion because it cannot be used in a regulatory context. The LOQ passes this criterion only if it is assumed to be approximately equivalent to the ML (i.e., the ML procedure is used to establish an LOQ).

5.2.4 Assessment of the IUPAC/ISO Limit of Quantitation

A similar LOQ approach was developed by IUPAC/ISO and published in the same papers as the CRV and MDV (see Sections 5.1.4 and 5.1.5).

5.2.4.1 Description of the ISO/IUPAC LOQ Approach

The 1995 *"Recommendations"* define the LOQ as:

"... the ability of a CMP [chemical measurement process] *to adequately 'quantify' an analyte. The ability to quantify is generally expressed in terms of the signal or analyte (true) value that will produce estimates having a specified relative standard deviation (RSD), commonly 10 %."*

The relationship used to define the LOQ is:

$$L_Q = K_Q \times \sigma_Q$$

The recommended value for K_Q is 10. Thus, the LOQ is 10σ above the blank signal, σ_Q.

5.2.4.2 Assessment of the IUPAC/ISO LOQ Against the Evaluation Criteria

The following five subsections discuss the IUPAC/ISO LOQ approach and procedure in the context of the five evaluation criteria that concern detection limit approaches (i.e., Criteria 1-3, and Criteria 5 and 6).

5.2.4.2.1 Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Condition 1: It can be (and has been) tested. Testing of the IUPAC/ISO LOQ is hampered by the lack of a supporting procedure for establishing an LOQ, and a conceptual dependence on the variability of blank measurements. If the blank measurements fail to produce a response, it is impossible to calculate an LOQ because the standard deviation of zero is zero. One solution for testing the approach is to assume that the ISO/IUPAC LOQ is approximately equivalent to the ML as the blank signal approaches zero. If this is a reasonable assumption, the ML procedure is a viable means for testing the LOQ approach, and the ISO/IUPAC LOQ meets this condition.

Condition 2: It has been subjected to peer review and publication. The IUPAC/ISO LOQ definition has been published by Currie in the peer-reviewed journals *Pure and Appl. Chem.* in 1995; in *Anal. Chim. Acta* in 1999, in *Chemometrics and Intelligent Lab Systems* in 1997; and in *J. Radioanal. and Nuclear Chem.* in 2000. Therefore, the IUPAC/ISO LOQ meets this condition.

Condition 3: The error rate associated with the procedure is either known or can be estimated. EPA used data generated in the Episode 6000 study to estimate the error rate associated with the LOQ. The Episode 6000 results show that the median error across all analytes and analytical techniques at 10σ is approximately ±14% with approximately 95% confidence.

Condition 4: Standards exist and can be maintained to control its operation. The IUPAC/ISO LOQ lacks a clearly defined procedure for estimating the important terms required to derive it. Although it may be possible to derive IUPAC/ISO LOQ values from data used to derive EPA MDL values, there is no discussion of using replicate blanks, replicate spiked samples, or a minimum recommendation for the number of replicates. Therefore, the IUPAC/ISO LOQ does not meet this condition.

Condition 5: It has attracted widespread acceptance within a relevant scientific community. Acceptance of this approach by the scientific community is currently not known. Acceptance would be indicated by use of the LOD in ISO methods. EPA's search for detection and quantitation approaches in the open technical literature did not uncover a large number of citations that reference the LOQ. Therefore, it is difficult to determine if the ISO/IUPAC LOQ meets this condition.

5.4.2.2.2 Criterion 2: The approach should address realistic expectations of laboratory and method performance, including routine variability.

The most recent publication on the IUPAC/ISO LOQ (*J. Radioanal. and Nuclear Chem.*, op. cit.) provides insight into this issue through measurements of ^{14}C by accelerator mass spectrometry. Therefore, the IUPAC/ISO LOQ passes this criterion for at least some measurement techniques.

5.4.2.2.3 Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance.

The ISO/IUPAC LOQ approach is not supported by a clearly defined procedure for establishing the LOQ. Therefore, it does not meet this criterion.

5.4.2.2.4 Criterion 5: The quantitation limit approach should identify the concentration that gives a recognizable signal that is consistent with the capabilities of the method when a method is performed by experienced staff in well-operated laboratories.

Assuming a relationship of the IUPAC/ISO LOQ to the ML, the LOQ satisfies this criterion for the reasons outlined in Section 5.2.1.2.4, which discusses EPA's assessment of the ML against Criterion 4 for evaluating quantitation limit approaches.

5.4.2.2.5 Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support State and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government

In the absence of a procedure for determining LOQ values, the ISO/IUPAC LOQ does not meet to meet this criterion because it cannot be used in a regulatory context. The ISO/IUPAC LOQ passes only if the ML procedure is used to establish an LOQ.

Chapter 6
Findings and Next Steps

What are EPA's findings in this revised assessment?

In this revised assessment of detection and quantitation approaches, the Agency has evaluated the codified MDL procedure, the ML procedure that EPA proposed to codify in 2003, and several alternative procedures. Some of these alternative procedures were submitted to EPA during the comment period on EPA's 2003 assessment, which was detailed in the February 2003 Technical Support Document (EPA-821-R-03-005). In today's assessment, we have:

- Identified relevant procedures to include in the assessment (Chapter 2);
- Identified issues that may be relevant to the assessment from an analytical chemistry, statistical, or regulatory perspective (Chapter 3);
- Used six criteria to evaluate the ability of each procedure to support activities under the Clean Water Act (Chapter 4);
- Assessed how well each procedure meets the evaluation criteria (Chapter 5);
- With real-world data and several different procedures, calculated and compared detection and quantitation limits using, and evaluated the theoretical and practical limitations of, each procedure (Appendices B and C).

The assessment of the theoretical and practical applications of each procedure (Appendices B and C) suggests that different procedures produce different detection and quantitation limits. Observed differences are largely due to different sources of variability accounted for among the procedures. The overall assessment of each procedure against each of the evaluation criteria suggests that no single pair of detection and quantitation limit procedures perfectly meets EPA's six evaluation criteria. Although the MDL and ML procedures are closest to meeting these criteria, as discussed under EPA's next steps, we recognize that this is not the end of our consideration of future improvements to EPA procedures and/or adoption of specific alternative procedures.

In response to stakeholders who suggested that EPA clarify or revise some steps in these procedures, we proposed modest revisions to the MDL procedure and proposed to codify an ML definition and procedure in conjunction with the 2003 assessment. We also proposed to codify an existing option that allows use of other detection and quantitation procedures to develop detection and quantitation limits. Public comment on both the 2003 assessment and the proposed revisions expressed many divergent views that conflicted with the proposed modifications to the procedures. Commenters suggested that we work with stakeholders to discuss mutual concerns and possible solutions rather than proceed with the proposed revisions. Some commenters submitted detailed, alternative procedures or regulatory revisions. However, there was no agreement among these commenters as to which of the competing alternatives or revisions to adopt, and none of them fully satisfied EPA's needs under the CWA. We have therefore decided to withdraw the proposed revisions.

What are EPA's next steps?

We believe that it is appropriate to withdraw the proposed revisions, take final action on the 2003 assessment, and explore the feasibility of using a stakeholder process to facilitate a resolution of the technical and policy issues raised during the public comment period. It is in the best interest of all parties to solicit additional stakeholder input through consultations. In a *Federal Register* notice published on September 15, 2004 [69 FR 55547], we announced that a neutral party is studying the feasibility of a process by which a broad group of stakeholders would work together to define and address concerns about the way detection and quantitation limits are calculated and used to support CWA programs. This potential stakeholder process will expand the list of interested stakeholders to include state, tribal and local governments, environmental groups and other interested parties. We trust that this potential stakeholder process will address the wide variety of views held by stakeholders and may lead to recommendations for possible improvements to current EPA procedures and/or use of alternative procedures.

To facilitate open, frank and inclusive discussions, we have made every effort to ensure that this Revised Assessment Document does not prejudge the result of the potential stakeholder process. In particular, we recognize that the following stakeholder issues or suggestions provide a strong starting point for a continued dialogue with stakeholders.

Assessment Evaluation Criteria Issues

The February 2003 assessment identified and discussed six criteria the Agency used to evaluate several different approaches to detection and quantitation. The six evaluation criteria are:

Criterion 1: The detection and quantitation limit approaches should be scientifically valid.

Criterion 2: The approach should address demonstrated expectations of laboratory and method performance, including routine variability.

Criterion 3: The approach should be supported by a practical and affordable procedure that a single laboratory can use to evaluate method performance.

Criterion 4: The detection limit approach should identify the signal or estimated concentration at which there is 99% confidence that the substance is actually present (i.e., a one percent false positive rate) when the analytical method is performed by experienced staff in a well-operated laboratory.

Criterion 5: The quantitation limit approach should identify the concentration that gives a recognizable signal that is consistent with the capabilities of the method when a method is performed by experienced staff in a well-operated laboratory.

Criterion 6: Detection and quantitation approaches should be applicable to the variety of decisions made under the Clean Water Act, and should support state and local obligations to implement measurement requirements that are at least as stringent as those set by the Federal government.

Stakeholders commented that these six criteria favored the MDL and ML procedures. Some stakeholders noted instances where criterion four fails for the MDL, i.e., does not represent the limit at which there is a 99% confidence that the observed signal is not a false positive. Stakeholders also disagreed with EPA's reliance on only one detection and one quantitation procedure, the MDL and ML (see criterion six discussion at 4.6 in this document) Stakeholders suggested that different detection and

quantitation procedures with different levels of rigor be developed and applied to the disparate uses of these limits in CWA programs. Uses of these limits include verification of laboratory performance, method validation, and as a guide for reasonable bounds on values to consider for permit limits. EPA recognizes that the complexity and statistical rigor appropriate for a detection and quantitation approach for method development and validation would be greater than that needed for demonstrating laboratory proficiency. Although EPA believes that the six evaluation criteria are suitable for purposes of this assessment, they need not be the only starting point for future stakeholder evaluations of revised or alternative detection and quantitation procedures.

Technical and Policy Issues

Some of the major comments on the MDL and ML procedures that influenced our decision to withdraw the proposed rule, and to seek additional stakeholder input, include: (1) the MDL does not adequately address analytical variability or systematic error (bias); (2) a need for better guidance on the intended use of these limits in CWA programs; (3) the need for different procedures for different CWA applications, such as method development, laboratory performance checks, and permit limits. Commenters also asked for clearer guidance on specific steps in the MDL procedure, such as selection of initial spike concentrations, and use of iterative and outlier procedures.

The technical issues of analytical variability and bias attributable to blanks encompass a range of concerns. Stakeholders have suggested that detection and quantitation procedures should:

- vary in the nature and extent of statistical rigor and performance verification checks depending on the end;
- use of a calculated limit;
- account for more sources of variability, such as the variability between and within laboratories;
- require more then seven samples and collect samples over a long period of time; and
- use routine blank samples collected over long periods of time to account for background signals and temporal variability (e.g., ACIL and USGS procedures).

EPA believes these suggestions merit serious consideration, and plans to use the stakeholder process to consider ways to address them.

Conclusion

This Revised Assessment Document addresses comments and concerns received from stakeholders and peer reviewers. Based on this new information, EPA believes that discussion of alternatives or improvements to current detection and quantitation concepts or procedures and uses should continue. It is clear that there is a broad interest in improving current procedures and uses, but no consensus for a specific procedure or procedures has emerged among the laboratory, industry, regulatory or regulated communities. We look forward to further stakeholder participation in this process.