STP1618 On Detection Limits

Fundamentals of the ASTM International Committee D22 Practice for Determination of the Detection Limit of a Well-Behaved Analytical Method

M. J. Brisson[1] and Harry L. Rook[2]

[1]Savannah River National Laboratory, Aiken, SC 29808, USA <ORCID logo>http://orcid.org/XXXX-XXXX-XXXX-XXXX

[2]National Institute of Standards and Technology, Moretown, VT 05660, USA (deceased) <ORCID logo>http://orcid.org/XXXX-XXXX-XXXX-XXXX

CITATION

Manuscript received November 29, 2018; accepted for publication May 20, 2019

ABSTRACT

The concept of detection limits dates back to the 1960s, with the best-known treatise being Lloyd Currie's 1968 definition of "detection limit" in terms of a "critical level." The concept is important for scientists, policymakers, instrument manufacturers, and other stakeholders. It remains a source of controversy, however, because although the theory is based in statistics, it is applied by analytical practitioners. This is not always compatible . Over the past half-century, there have been many formulations, some developed for particular programs, and with these formulations, a cornucopia of terms and abbreviations, presenting challenges for a laboratory dealing with the requirements of different programs. Currie's principles were solidified in the 1995 International Union of Pure and Applied Chemistry recommendations and International Organization for Standardization 11843 series. These principles have been embraced in Europe, but not in the United States. One possible reason is that the concepts focus on underlying statistics and not on practical application. The goal of ASTM Committee D22 on Air Quality is to begin with the smallest common denominator, a well-behaved air quality method, and then to standardize a practice that has statistical rigor; will be understood and used by the analytical community; and, perhaps most important, will produce consistent results. Committee D22 hopes to expand on this admittedly narrow focus in the future. This paper describes fundamental concepts of the Committee D22 methodology, based on Currie's 1968 and 1999 papers. It uses a one-sided test of significance to affirm or reject the null hypothesis (analyte not detected). Rejection of the null is used to determine a value above the mean background that represents detection of the analyte signal with a stated probability of repeatability. It evaluates all significant error sources and uses the magnitude and variance of the background and combined error to estimate the detection limit.

Keywords

detection limit, statistics, probability estimate, analytical chemistry, data reporting

INTRODUCTION

Advances in analytical laboratory instrumentation from the mid-twentieth century to date have allowed for the quantification of ever-smaller quantities of material. This required the formation of some concept to describe the capability of an analytical method to detect trace-level quantities with a desired degree of reliability. Thus was born the concept of "detection limit" (DL), the minimum value at which one has a defined level of confidence that an analyte has been detected. The reliability is based on a given probability of false positives, or of false positives and false negatives. Among the earliest attempts at defining this concept were those made by Altshuler and Pasternack[1] and Currie.[2] The impacts of properly defining the concept are significant. Statistical formulations are used to determine DLs; however, these formulations are used by laboratory professionals who often do not fully understand how the formulations are derived, and thus may apply them improperly. The resulting data are used for making decisions and defining public policy. Even in 1968, Currie noted that different literature definitions of DL could yield results that varied by up to three orders of magnitude.[2]

Adding to the confusion, a multiplicity of ways to define and calculate detection limits have developed over the past 50 years. For example, the ASTM International terminology directory had 63 entries for "detection limit" or terms that contained those words.[3] The International Organization for Standardization (ISO) had 84 entries in its terminology database.[3] In some cases, the decision as to which definition and formulation to use is dictated by regulation or by convention. In other cases, the decision may be based on whether one wants to control for false positives, or false negatives, or both.

A related term, "quantification limit" or "limit of quantification,"[*]

---

[*] The phrase "lower limit of quantification" is sometimes used to distinguish from an "upper limit of quantification" when the signal is above what the instrument can handle or is above the highest point of calibration.

was originally referred to by Currie as the "determination limit,"[2] although he later recommended using "quantification limit". [4] The difference in these is widely understood to be the difference in *detecting* an analyte with a given level of confidence versus *quantifying* the amount of analyte with a certain level of precision and bias (e.g., expressed in terms of relative standard deviation or RSD). The expression that the quantification limit is 10 standard deviations above the mean blank signal is widely used.[2]

Making matters worse, different programs have coined different terms to describe their variation on the core concepts. It is easy to see that DL and limit of detection describe the same concept, and likewise for quantification limit and limit of quantification. Now, however, the cornucopia of terms includes "contractor-required detection limit," "practical quantitation limit," "estimated quantitation limit," "minimum detectable activity," and "minimum level of quantitation." It can become challenging when a single laboratory is required to follow different guidelines and use different terms to report data for different programs.

In 2015, ASTM Committee D22 on Air Quality was in the process of updating definitions in its terminology standard, ASTM D1356-17, *Standard Terminology Relating to Sampling and Analysis of Atmospheres*,[5] and received repeated negatives for revisions to "limit of detection." This led to the recognition that there were multiple definitions in different Committee D22 standards that could yield different results for similar air-quality test methods. As a result, in August 2016, an initial Conference on Detection Limits was sponsored by Committee D22, with participation from members of Committee E11 on Quality and Statistics. At this conference, a consensus among the members of Committee D22 was reached that the committee needed a unified consensus practice for determination of the DL for air-quality methods.

GOALS OF THE COMMITTEE D22 PRACTICE

One of the key goals behind the effort to develop a unified consensus practice in Committee D22 is to go back to the first principles as espoused by Currie. [2,4,6,7] These principles are best described using the three key terms defined by Currie. The first of these is the *critical value* or L_c, the value at which the probability of a false positive (or type I error) is α. In mathematical terms, the critical value is the smallest value L_C such that $\Pr\left(\hat{L} > L_C \middle| L = 0\right) \leq \alpha$, where L is the true signal and $\hat{L}$ is the measured value, or estimate, of the signal. This tests the null hypothesis of $H_o\colon L = 0$. The second is the *detection limit* or L_d, which controls for both α and β, the probability of a false negative (or type II error). In mathematical terms, this is the minimum value of L_D such that $\Pr\left(\hat{L} > L_C \middle| L = L_D\right) \geq 1 - \beta$, testing the alternative hypothesis H_a: $L = L_D$. The third is the *quantitation limit* or L_q, defined in terms of a specified RSD.[6] The standard practice currently being developed by Committee D22 does not address L_q; this will be addressed in a later standard practice.

One school of thought, referred to by Currie as the "signal/noise" school,[6] controls only for α, which effectively results in a value of 50% for β and is equivalent to L_c.[2] This concept is followed in the definition and calculation of "method detection limit" promulgated by the U.S.

Environmental Protection Agency's Office of Water.[8] A second school of thought seeks to control for both α and β based on hypothesis testing as described previously. This is the basis for L_d as defined by Currie, and for the standard practice proposed in Committee D22. It is important for air-quality measurements to have a level of confidence regarding false negatives as well as for false positives.

A second key goal is to provide laboratory practitioners with a more understandable approach while maintaining sufficient statistical rigor. Although it is neither necessary nor expected for laboratory practitioners, users of laboratory data, or policy makers to understand the statistical theory behind the calculations used to determine detection and quantification limits, it is desirable for laboratory practitioners to understand the process well enough to follow it correctly and identify unexpected results that may require investigation. All parties (statisticians, laboratory practitioners, and data users) benefit when this is the case.

Another key goal is to improve harmonization with the global approach as adopted by the International Union of Pure and Applied Chemistry (IUPAC)[4] and ISO.[9] The ISO 11843 series defines L_c based on α = 0.05, L_d based on $\alpha = \beta = 0.05$, and L_q based on a RSD of 10%. In Europe, the IUPAC/ISO approach is widely followed; however, the EU focuses more on expressing uncertainty, through the Guide for Uncertainty in Measurements (GUM)[10] for every measurement than on DLs *per se*. Indeed, IUPAC recommends that both the estimated value of the measured quantity and its uncertainty should always be reported, even when the estimate is below the critical value.[6] This approach, however, is not widely followed in the United States possibly because the IUPAC/ISO concepts are entirely statistically based without considering any operational aspects. Also, some industries within the United States have sought an approach that is specific to their needs. Examples of this include the definition of "method detection limit" by the EPA's Office of Water[8] and development of standards in ASTM Committee D19 on Water.[11–13] The current Committee D22 effort hopes to bridge this gap, at least in part, by providing a practice more in line with the IUPAC/ISO approach while also considering operational aspects and the needs of the laboratory practitioner.

Committee D22 also wanted to promote consistency in how these limits are determined, which in turn promotes consistency in the final results. An important result of using a single standard practice for estimating the DL is that the author of a scientific study cannot choose which method would yield the result most favorable to that study's premise. An example of the lack of consistency among current DL methodologies can be found in an EPA study conducted in 2004.[14] This study compared DLs from four different approaches: a simplified approach, created by this report for comparison purposes, based on a condensed version of the EPA MDL and the American Chemical Society (EPA/ACS DL); the ISO critical value (ISO CRV);[9] the ISO minimum detectable value (ISO MDV);[9] and a single-laboratory version of the Committee D19 Interlaboratory Detection Estimate (SL-IDE).[12] Table 1 shows the results for six metals measured by inductively coupled plasma emission spectroscopy using EPA Method

1620.[15] Note that the EPA/ACS DL value generally fell between the ISO CRV (lowest) and ISO MDV values, whereas the SL-IDE values were highest.

TABLE 1 Comparison of DLs (all values in μg/L) for the EPA Episode 6000 Dataset[14]

Analyte	EPA/ACS DL	ISO CRV	ISO MDV	ASTM SL-IDE
Silver	4.907	3.588	6.495	10.668
Sodium	69.53	49.595	97.649	138.768
Thallium	0.512	0.651	1.406	1.153
Tin	3.670	2.019	3.143	3.932
Titanium	4.777	4.453	8.050	5.376
Vanadium	7.344	4.207	8.359	10.630

The standard practice currently being developed by Committee D22 is intended as a starting point for the development of additional standards. In this effort, we intentionally limited the scope of applicability to "well-behaved" analytical methods, recognizing that to have hope of achieving consensus based on these principles, we needed to begin with a small subset of the air-quality measurement realm. Once consensus is achieved on such a subset, we can then seek to expand to a broader scope of air-quality measurements and add a procedure for limit of quantification. It is also a longer-range goal for Committee D22 to collaborate with Committee E11 in developing a manual with more detailed how-to guidance for laboratory practitioners and data users.

LIMITATIONS OF THIS STANDARD PRACTICE

As noted previously, given the contentious nature of the subject and the fractured approach within the United States, Committee D22 felt that some simplifications were necessary for our initial attempt at a standard practice as a basic building block to obtain consensus. One simplification was to limit the scope of the practice to "well-behaved" analytical methods. Such methods are those shown to be consistent over time, in full statistical control, and that have a normal, lognormal, or gamma distribution. If there are issues with repeatability, then it is not a well-behaved method. Reasons for this can include the existence of outliers for unknown reasons; a larger variance than would be expected from the signal; or data that do not agree with other analytical methods.

Committee D22 also chose to address the concept of detection limit, consistent with how Currie defines it [2,4,6,7] and not the quantification limit, which will be addressed in a future practice. This practice also does not consider the topic of censored data, or results that are not reported because they are lower than a laboratory's quantification limit.[2] This is a subject of active discussion within Committees D22 and E11 and several papers presented in this

---

[2] There can be both lower and upper quantification limits, with data below the lower limit being "left-censored" and data above the upper limit being "right-censored". The focus here is on the lower limit.

compendium. It has impacts on the very nature of how laboratory data are reported. At least one future workshop on this subject is planned by Committees D22 and E11.

SUMMARY OF PROCEDURE

The first step in the procedure is to estimate the mean of the background signal, $\bar{B}$, using the same signal region as that used for the measurement of the specific analyte of interest. The number of background measurements, n_B, can vary between 5 and 20, depending on the level of statistical assurance needed for the given study. A table is provided to assist in determining the proper number of measurements. In most cases, 7–10 measurements will provide the best balance between statistical power and cost. For n_B measurements, $\bar{B}$ is calculated using equation (1). Although some measurements used to calculate $\bar{B}$ may be zero or negative, it is important to ensure that $\bar{B}$ itself is not zero or negative; otherwise, this procedure is not applicable.

$$\bar{B} = \frac{1}{n_B}\sum_{i=1}^{n_B} B_i. \tag{1}$$

The next step is to determine the variance of the background (V_B), the standard deviation (s_B) and the coefficient of variation (u_B), using equations (2)–(4).

$$V_B = \frac{1}{n_B-1}\sum_{i=1}^{n_B}(B_i - \bar{B})^2, \tag{2}$$

$$s_B = \sqrt{V_B}, \text{ and} \tag{3}$$

$$u_B = s_B/\bar{B}. \tag{4}$$

The next step is to determine, or assume, the distribution to be used for background measurements and for determination of L_c based on the 95th or 99th percentile of the background distribution. As noted previously, the distribution should be normal, lognormal, or gamma. In the case of a gamma distribution, none of the individual measurements may be negative. Data should be continuous rather than counting data (such as for particles). Also, the measurement process should not produce repeated values of zero.

At this point there are three possible cases to consider. These cases are summarized in table 2.

TABLE 2 Cases for determination of L_c

Case	Background Distribution	Background Measurements	u_B
1	Normal	Some may be slightly negative	--
2	Normal	None may be negative	<1
3	Gamma	None may be negative	≥1

In Cases 1 and 2, L_c can be determined as an estimated upper percentile (such as the 95th or 99th percentile) of the background distribution using equation (5), where $t_{n_B-1}(\alpha)$ is the α critical value from the t-distribution with $n_B - 1$ degrees of freedom:

$$L_c = \bar{B} + s_B t_{n_B-1}(\alpha). \tag{5}$$

In Case 3, L_c is determined as an estimated upper percentile of the background distribution, assumed to be (two-parameter) gamma, using the simple moment estimations in equations (6) and (7) for shape (*k*) and scale (*q*), as follows:

$$\boldsymbol{\hat{k} = \bar{B}^2 / V_B}, \text{ and} \tag{6}$$

$$\boldsymbol{\hat{q} = V_B / \bar{B}}. \tag{7}$$

These estimates can be used with the GAMMA.INV function in Excel or the qgamma function in R to compute the 95th or 99th percentile.

Once L_c has been determined, L_d is the value such that repeated measurements of samples with concentration at that value will exceed L_c with probability of $1 - \beta$. The distribution of these repeated measurements can be assumed to be normal because of their distance from zero. On the basis of this assumption, an initial estimate of L_d may be computed using equation (8), where DL_0 is the initial estimate of L_d:

$$\boldsymbol{DL_0 = \bar{B} + s_B\left[t_{n_B-1}(\alpha) + t_{n_B-1}(\beta)\right] = L_c + s_B t_{n_B-1}(\beta)}. \tag{8}$$

Then, *i* measurements of samples are spiked at a value at or near DL_0. The mean ($\boldsymbol{\bar{x}}$) of these measurements is calculated and then the variance (V_{dl}) and standard deviation (s_{dl}), using equations (9) and (10):

$$\boldsymbol{V_{DL} = \frac{1}{n_{DL}-1}\sum_{i=1}^{n_{DL}}(x_i - \bar{x})^2}, \text{ and} \tag{9}$$

$$\boldsymbol{s_{DL} = \sqrt{V_{DL}}}. \tag{10}$$

Finally, L_d is calculated using equation (11):

$$\boldsymbol{L_d = CV + s_{DL} t_{n_{DL}-1}(\beta)}. \tag{11}$$

CONCLUSION AND A PATH FORWARD

This discussion provides the fundamentals of the draft Committee D22 standard practice based on the most recently balloted version. Members of Committee D22 have been overwhelmingly supportive of the development of this practice based on their votes, but they also have identified some additional needs to be addressed by the standard, such as the inclusion of one or more examples and elucidation of additional details in the procedure steps.

Issues such as these are part of the normal process of standards development. It is currently envisioned that this practice should be finalized by the end of 2019.

As previously discussed, this practice is intended to be the first in a series of documents intended to provide a consistent methodology and offer additional guidance to the laboratory practitioner, while ensuring the necessary level of statistical rigor so that data users can have confidence in the decisions that they make with the data provided by air-quality methods. Additional documents that could be developed include the following:

- A version of this practice with a broader scope, including less well-behaved methods. The current practice will be a starting point for this broader practice.
- A practice on limits of quantification.
- A guidance manual with more "how-to" information, and possibly more background on the statistical theories behind the calculations.
- A guide on reporting of laboratory data that among other things addresses the issue of left-censored data. This effort will be challenging because of the need to balance the needs of users, who generally would prefer that no data are ever censored, versus requirements of accreditation bodies to perform such censoring, and the legitimate concerns of laboratories regarding the potential by some users to misinterpret such data.

In all of these instances, Committee D22 will focus on air-quality measurements because that is the committee's scope. All of these issues, however, are much broader than air-quality measurements. Committee E11 has recognized the broader need and has been cooperating with Committee D22 in part as a starting point toward addressing that broader need. Although it is likely not reasonable to expect a one-size-fits all approach for the broader universe of analytical measurements, some increased level of consistency and reduction of the cacophony of terms and disparate procedures are sorely needed.

ACKNOWLEDGMENTS

The authors gratefully acknowledge John Carson, Tom Bzik, and other members of Committee E11 for their support and assistance on this project.

The authors gratefully acknowledge the members of Committee D22, particularly those who made constructive comments to improve the quality of the Detection Limits standard practice.

The authors gratefully acknowledge the reviewers of this manuscript.

REFERENCES

1. B. Altshuler and B. Pasternack, "Statistical Measures of the Lower Limit of Detection of a Radioactivity Counter," *Health Physics* 9, no. 3 (March 1963): 293–298.

2. L. A. Currie, "Limits for Qualitative Detection and Quantitative Determination: Application to Radiochemistry," *Analytical Chemistry* 40, no. 3 (March 1968): 586–593, DOI:10.1021/ac60259a007.

3. M. Brisson, "The Trouble with Detection Limits as We Know Them Today" (presentation, ASTM International Committee D22 Conference on Detection Limits, Raleigh, NC, August 29, 2016).

4. L. A. Currie, "Nomenclature in Evaluation of Analytical Methods Including Detection and Quantification Capabilities (IUPAC Recommendations 1995)," *Pure Applied Chemistry* 67, no. 10 (October 1995): 1699–1723, DOI:10.1351/pac199567101699.

5. *Standard Terminology Relating to Sampling and Analysis of Atmospheres*, ASTM D1356-17 (West Conshohocken, PA: ASTM International, 2017).

6. L. A. Currie, "Detection and Quantification Limits: Origins and Historical Overview," *Analytica Chimica Acta* 391, no. 2 (May 1999): 127–134, DOI:10.1016/S0003-2670(99)00104-X.

7. L. A. Currie, "Detection: Overview of Historical, Societal, and Technical Issues," in *Detection in Analytical Chemistry*, ed. L. Currie (Washington, DC: American Chemical Society, 1988), 1–62, DOE:10.1021/bk-1988-0361.ch001.

8. *Definition and Procedure for the Determination of the Method Detection Limit, Revision 2*, EPA 821-R-16-006 (Washington, DC: U.S. Environmental Protection Agency, December 2016).

9. *Capability of Detection–Part 1: Terms and Definitions*, ISO 11843-1 (Geneva, Switzerland: International Organization for Standardization, July 1, 1997).

10. *Evaluation of Measurement Data—Guide to the Expression of Uncertainty In Measurement*, JCGM 100:2008 (Paris, France: Joint Committee for Guides in Metrology, September 2008).

11. N. Grams, "ASTM International Interlaboratory Detection Estimate (IDE) and Interlaboratory Quantitation Estimate (IQE)" (presentation to the NELAC Institute, February 2011).

12. *Standard Practice for 99 %⁄95 % Interlaboratory Detection Estimate (IDE) for Analytical Methods with Negligible Calibration Error*, ASTM D6091-07(2014) (West Conshohocken, PA: ASTM International, approved January 15, 2014).

13. *Standard Practice for Interlaboratory Quantitation Estimate*, ASTM D6512-07(2014) (West Conshohocken, PA: ASTM International, approved January 15, 2014).

14. *Revised Assessment of Detection and Quantitation Approaches*, EPA-821-B-04-005, (Washington, DC: U.S. Environmental Protection Agency, October 2004).

15. *Method 1620: Metals by Inductively Coupled Plasma Atomic Emission Spectroscopy and Atomic Absorption Spectroscopy*, EPA 821-B-89-100 (Washington, DC: U.S. Environmental Protection Agency, September 1989).