Designation: E3243 – 21

Standard Specification for Field Detection Equipment and Assays Used for Fentanyl and Fentanyl-Related Compounds[1]

This standard is issued under the fixed designation E3243; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

INTRODUCTION

One consequence of the widespread use of synthetic opioids is that first responders and other personnel increasingly encounter them in the field. Within this class of compounds are fentanyl and fentanyl-related compounds (some of which are referred to as analogs or analogues), which can present significant safety hazards to first responders if proper protocols and PPE are not used. Thus, the ability to detect fentanyl and fentanyl-related compounds reliably, and take appropriate protective measures, is crucial. Evaluation of equipment and assays for field detection of fentanyl and fentanyl-related compounds is necessary to assess if a system meets or exceeds performance metrics for the intended application and end-user. Since fentanyl and fentanyl-related compounds are often mixed with cutting agents or other drugs, it is also important to assess the effects of these compounds on equipment and assay performance. The performance assessments described in this specification will determine the potential for false-positive and false-negative results, a lower bound on the probability of detection (POD), and potential impacts of other substances such as common diluents (for example, cutting agents) and other drugs (for example, heroin) that are commonly mixed with fentanyl.

The performance assessments presented herein are laboratory evaluations. Laboratory evaluations with trained personnel are recommended to establish the best-case performance for a system without confounding performance issues that might arise during field testing (for example, lack of user training or environmental conditions). Laboratory evaluations also serve to eliminate systems that have deficiencies or limitations, before extensive cost and effort are expended for field testing for the specific intended application.

This specification is a companion standard to Test Method E3290. This specification describes a statistical testing approach to quantify performance and also defines test sample compositions and amounts. However, it does not provide details for sample preparation, specific protocols for conducting testing of different types of instruments and assays, or reporting. These details are described in Test Method E3290.

The statistical approach used in this specification ties performance of an instrument or assay to a specified lower confidence bound (LCB) on the POD at a known confidence level (CL). Testing is conducted to establish a system's performance over a set of possible performance outcomes of LCB $\geq$ 0.85 and CL $\geq$ 80 %. The 0.85/80 % LCB/CL is the minimum performance level that can be considered as 'pass' for an instrument or assay. Testing results that do not achieve or exceed the minimum level of 0.85/80 % LCB/CL have failed to pass the performance established by this specification. Testing results can produce performance levels above the minimum 0.85/80 % LCB/CL and users shall report the highest LCB at or above 0.85 at the highest CL achieved.

The performance level established in this specification requires that testing include all the test modules (TMs) given in Annex A1 that are applicable to the detection technology used. (See Table 2 and Table 3. An 'X' denotes technology not applicable for the conditions of the Test Module.) The TMs include sets of samples containing specific numbers and types of target or non-target compounds and are used to measure the performance of the detection technology. The user (that is, an agency directing testing or a testing entity) is ultimately responsible for deciding the number and extent of sample types that will be tested, based on the following: (*1*) Desired level of performance (POD and CL). Testing more samples can result in higher POD and CL as illustrated in Table 4 through Table 6.

(*2*) The variety of samples types to be tested is summarized in 4.10. Two different testing tiers may be employed as described in 4.12: Tier 1 (all 14 different sample types are tested in each applicable TM) vs. Tier 2 (only the first four of the different sample types are tested in each applicable TM).

[1] This specification is under the jurisdiction of ASTM Committee E54 on Homeland Security Applications and is the direct responsibility of Subcommittee E54.01 on CBRNE Detection and Decontamination.

Current edition approved May 1, 2021. Published July 2021. DOI: 10.1520/E3243-21.

1. Scope

1.1 *General:*

1.1.1 This specification provides system designers, manufacturers, integrators, procurement personnel, end-users, practitioners, and responsible authorities a common set of parameters to match the capabilities of chemical detection tools with user needs for their specific application.

1.1.2 This specification describes required test sample compositions, amounts, and a statistically-based testing approach to be used for evaluating the performance of field fentanyl and fentanyl-related detection equipment and assays as described in Test Method E3290. This specification does not address the estimation of limit of detection.

1.1.3 This specification is not meant to provide for all uses. Manufacturers, purchasers, and end-users will need to determine specific requirements including, but not limited to, use by hazardous material (HAZMAT) teams; use in explosive or other hazardous environments or atmospheres; use with personal protective equipment (PPE); use by firefighters, law enforcement officers, or FEMA Urban Search & Rescue teams, special electromagnetic compatibility needs, extended storage periods, and extended mission time. These specific requirements may or may not be generally applicable to all chemical detection systems.

1.2 *Operational Concepts*—Chemical detection systems are used to detect or identify chemical hazards to support short-term tactical decision-making to protect responders and the public. The system should provide low false-positive and false-negative rates. Uses of these systems include survey, surveillance, and screening of samples, particularly during a response to a suspected fentanyl or fentanyl-related compound. A field-deployable system should withstand the rigors associated with uses including, but not limited to, operation and storage in high and low temperatures, shock and vibration, radio frequency interference, and rapid changes in operating temperature and humidity. Note that this specification does not address testing the potential impact of the rigors associated with use of systems in the field.

1.2.1 *Units*—When creating multicomponent test samples for TM 2, TM3, and TM4, all % compositions are stated as weight/volume percent (mg/mL) for both solid and liquids.

1.3 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.4 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E2771 Terminology for Homeland Security Applications

E3131 Specification for Nucleic Acid-Based Systems for Bacterial Pathogen Screening of Suspicious Visible Powders

E3290 Test Method for Establishing Performance of Equipment and Assays for Field Detection of Fentanyl and Fentanyl-Related Compounds

2.2 *Other Standards:*

18 USC 178 Definitions[3]

Eurachem/CITAC Guide CG 4:2012 Quantifying Uncertainty in Analytical Measurement, Third edition[4]

ISO 17034 General requirements for the competence of reference material producers[5]

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[3] Available from U.S. Government Printing Office, Superintendent of Documents, 732 N. Capitol St., NW, Washington, DC 20401-0001, http://www.access.gpo.gov.

[4] Available from the Eurachem organization, https://www.eurachem.org/index.php/publications/guides/quam.

[5] Available from International Organization for Standardization (ISO), ISO Central Secretariat, Chemin de Blandonnet 8, CP 401, 1214 Vernier, Geneva, Switzerland, https://www.iso.org.

3. Terminology

3.1 *Definitions:*

3.1.1 *accuracy, n*—closeness of agreement between a test result and the accepted reference value. **E2771**

3.1.2 *analog, n*—equivalent to analogue.

3.1.3 *analogue, n*—a substance not specifically designated for control in Drug Enforcement Agency (DEA) Schedules I through V may still be subject to the Controlled Substances Act (CSA) as a controlled substance analogue. A controlled substance analogue is a substance not otherwise approved by the Food and Drug Administration (FDA) or scheduled under the CSA that has (*1*) a chemical structure substantially similar to that of a controlled substance in Schedule I or II, or (*2*) an actual or intended effect that is "substantially similar to or greater than the stimulant, depressant, or hallucinogenic effect...of a controlled substance in schedule I or II." A substance that meets those criteria and is intended for human consumption is treated as a controlled substance in Schedule I. Synthetic opioids related to fentanyl may qualify as controlled substance analogues.[6]

3.1.4 *assay, n*—quantitative or qualitative test used to determine the presence or absence of a chemical compound.

3.1.5 *bulk sample, n*—total sample amount (including the sum of target and non-target compounds) that is visible to the naked eye.

3.1.5.1 *Discussion*—The amount of sample available for testing bulk samples according to this standard is >1 µg and <10 mg.

3.1.6 *compound, n*—chemical substance under evaluation in this specification.

3.1.7 *confidence interval (CI), n*—range of values created using a procedure that, when repeated many times, on distinct data sets, generated from the same underlying stochastic process, will bracket the true measure of performance, such as POD and the proportion of times stated. **E3131**

3.1.8 *confidence level (CL), n*—probability value associated with a CI; the percentage of intervals that can be expected to include the true population parameter in the long run.

3.1.9 *fentanyl-related substances, n*—fentanyl-related substances include any substance that is structurally related to fentanyl (see Annex A2). **E3131**

3.1.10 *lower confidence bound (LCB), n*—lowest value of a one-sided CI created using a procedure that when repeated many times, on distinct data sets, generated from the same underlying stochastic process, will include the true measure of performance a proportion of times equal to the stated probability. **E3131**

3.1.10.1 *Discussion*—The LCB ensures that the POD attains a satisfactory value for the CL selected and determines the minimum number of samples that shall be analyzed.

3.1.11 *measurement process, n*—process used to detect a material or determine if a system or instrument performs as intended. **E3131**

3.1.12 *non-target compound, n*—collection of diluents/cutting agents, dyes/colorants, and other drugs that are not target fentanyl or fentanyl-related substances.

3.1.12.1 *Discussion*—Non-target compounds should not give a positive test result for fentanyl or fentanyl-related substances; this is considered a false positive result. Ideally, all non-target compounds should result in negative detection results for fentanyl and fentanyl-related compounds.

3.1.13 *precursor, n*—a chemical compound used to synthesize fentanyl and fentanyl-related compounds.

3.1.14 *probability of detection (POD), n*—proportion of positive analytical outcomes for a qualitative method for a given matrix containing a target compound at a given concentration.

3.1.14.1 *Discussion*—Note that, for non-target compounds, the POD can be defined as the proportion of negative (for fentanyl and fentanyl-related compounds) analytical outcomes for a qualitative method for a given matrix at a given concentration. This allows the performance outcomes to be applicable to the non-target compounds (TM4) case and still use the same LCB and CL measures of performance to limit the number of false positives appropriately. Alternatively, one could instead set an upper confidence bound on the probability of a false positive (proportion of false-positive analytical outcomes for a qualitative method for a given matrix at a given concentration of non-target compounds) with an appropriate CL using a similar method as that presented in Annex A3, except using the upper limit for the upper confidence bound. This definition differs slightly from that in Specification E3131, making it more general.

3.1.15 *reach-back support, n*—service that allows equipment users in the field to be in real-time contact with subject matter experts (SMEs) to provide advice on assessment of analysis results, including, but not limited to identification of one or more substances in an unknown sample.

3.1.15.1 *Discussion*—This service is often available 24 h a day/7 days per week via paid subscription through an equipment (instrument or assay) manufacturer. For example, a field chemical detection instrument like a Raman spectrometer may be used to scan an unknown sample suspected of containing narcotics. The user can send the resulting data to reach-back support for review, interpretation, or both. The SMEs providing the support can identify poor quality data and recommend settings for a successful rescan, perform advanced data analysis techniques to identify chemicals present in the sample, and quickly provide decision-support capabilities and guidance in response to end-user requests and inquiries. Reach-back services are commonly used for aiding in the interpretation of ambiguous data and for confirmation of positive results (especially those with which decisions of consequence may be taken).

3.1.16 *reference material, n*—substance sufficiently homogenous and stable with respect to one or more specified properties that has been established to be fit for its intended use in the measurement process; properties can be quantitative or qualitative. **ISO 17034**

[6] See https://crsreports.congress.gov/product/pdf/LSB/LSB10404.

3.1.17 *sensitivity, n*—change in the response of a measuring instrument divided by the corresponding change in the stimulus. **Eurachem/CITAC Guide CG 4:2012**

3.1.18 *specificity/selectivity, n*—ability of a measurement procedure to determine accurately and specifically the analyte of interest in the presence of other components in the sample matrix under the stated conditions of the test. **Eurachem/CITAC Guide CG 4:2012**

3.1.19 *target compound, n*—any of a collection of fentanyl and fentanyl-related compounds.

3.1.20 *test event, n*—a test event describes the TMs, the samples within those TMs that are included in the test event, the total number of samples to be analyzed from each TM, and whether trace or bulk sample amounts will be used.

3.1.21 *test module (TM), n*—set of samples with particular characteristics used to establish the performance of the detection technology.

3.1.21.1 *Discussion*—TMs include target compound samples at three different percent sample compositions (≥95 %, 10 %, and 1 %), non-target samples that do not contain fentanyl or fentanyl-related compounds, and fentanyl synthesis precursors or other compounds used for synthesis.

3.1.22 *test sample, n*—amount and identity of a particular substance (target and non-target compounds) prepared for testing.

3.1.23 *testing tier, n*—level of effort, including the selection of samples and number of experimental tests, involved in determining the performance of an instrument or assay.

3.1.23.1 *Discussion*—Testing involves either all 14 samples from each applicable TM or the first four samples from each applicable TM.

3.1.24 *trace sample, n*—total sample amount, including target and non-target compounds, ≤1 µg, used for testing.

3.1.24.1 *Discussion*—In this specification, a trace sample is intended to represent a sample not visible to the naked eye, a situation commonly encountered by first responders. For practical purposes, it is assumed that the sizes represented by bulk and trace samples sufficiently cover the range of situations encountered in the field where the testing equipment may be used.

3.2 *Acronyms:*

3.2.1 *ATR*—attenuated total reflectance.

3.2.2 *CDC*—Centers for Disease Control.

3.2.3 *CSA*—Controlled Substances Act.

3.2.4 *CI*—confidence interval.

3.2.5 *CL*—confidence level.

3.2.6 *DEA*—Drug Enforcement Agency.

3.2.7 *FDA*—Food and Drug Administration.

3.2.8 *FTIR*—Fourier transform infrared (also known as Fourier transform infrared spectroscopy).

3.2.9 *GC*—gas chromatography.

3.2.10 *GC/MS*—gas chromatography/mass spectrometry.

3.2.11 *HAZMAT*—hazardous material team.

3.2.12 *HPMS*—high pressure mass spectrometry.

3.2.13 *IMS*—ion mobility spectrometry.

3.2.14 *ISO*—International Organization for Standardization.

3.2.15 *LCB*—lower confidence bound.

3.2.16 *MS*—mass spectrometry.

3.2.17 *NIOSH*—National Institute for Occupational Safety and Health.

3.2.18 *POD*—probability of detection.

3.2.19 *PPE*—personal protective equipment.

3.2.20 *SERS*—surface enhanced Raman spectroscopy.

3.2.21 *SME*—subject matter expert.

4. Overview

4.1 See Table 1 and its footnotes for an overview of planning and conducting an instrument or assay evaluation. Table 1 and the eight associated footnotes provide a summary of all aspects of this specification and include references to Test Method E3290.

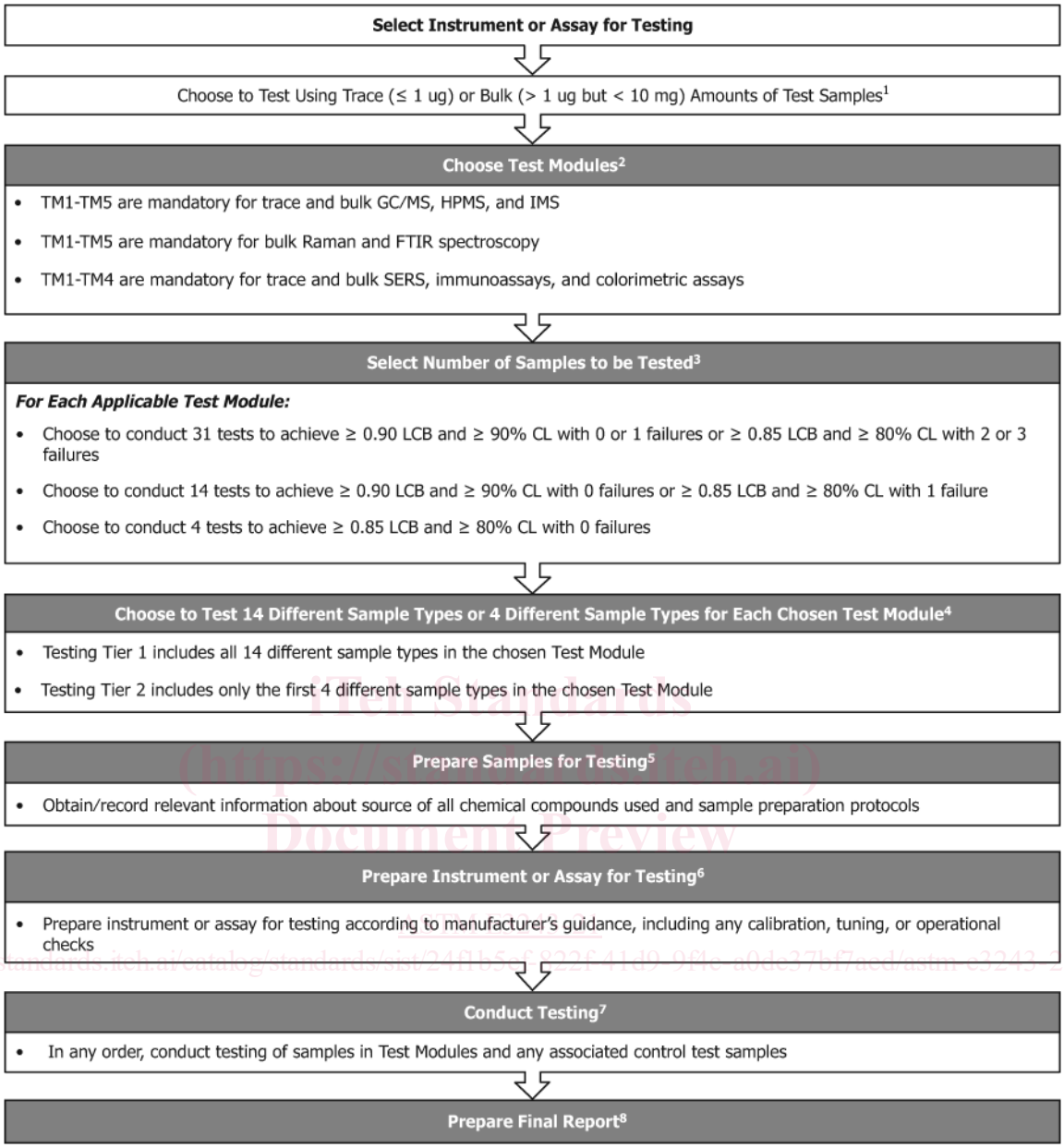
4.2 Bulk and trace sample amounts reflect two different scenarios commonly encountered in the field. Trace sample amounts are typical of samples that may be acquired by swabbing or wiping a surface, a person's hands, the knot or outer surface of a bag suspected of containing drugs, and so forth. Bulk sample amounts reflect samples that are visible to the naked eye that are often encountered in the field. It should be noted that results obtained from sampling the outside of packaging materials do not necessarily reflect what is actually inside the packaging.

4.3 There are currently several types of detection technologies used for field detection of fentanyl and fentanyl-related compounds including gas chromatography/mass spectrometry (GC/MS), mass spectrometry alone (that is, high-pressure mass spectrometry (HPMS)), ion mobility spectrometry (IMS), Fourier transform infrared spectroscopy (FTIR), Raman spectroscopy, surface enhanced Raman spectroscopy (SERS), colorimetric assays, and immunoassays.

4.4 Not all TMs listed in Annex A1 are appropriate for every technology/product type. While bulk detection technologies cannot detect trace quantities (≤1 µg total sample amount), most trace detection technologies can test a bulk sample (>1 µg and <10 mg total sample amount) with appropriate sample preparation/dilution. In general, trace detection technologies that can also test bulk samples with proper dilution include GC/MS, HPMS, and IMS. Bulk-only detection technologies include Raman and FTIR spectroscopy. While SERS, immunoassays, and colorimetric assays can detect bulk amounts of sample, they also may be capable of detecting trace amounts of sample. In general, SERS, immunoassays, and colorimetric assays cannot identify specific chemicals, so TM 5 is not applicable.

4.5 Depending on the type of technology, the samples in each of the five TMs can be analyzed as "trace" samples (<1 µg of sample used for testing) or bulk samples (>1 µg and <10 mg of sample used for testing). The exact amount of sample used for testing is according to instrument or assay manufacturer

TABLE 1 Overview of Specification E3243: Planning an Instrument or Assay Evaluation



[1] This is the total amount of all target (fentanyl and fentanyl-related compounds) and non-target sample components combined in a final prepared sample. Prepare and mix all test samples as described in Test Method E3290. The methodology in this specification applies to trace, bulk, ≥95 % fentanyl and fentanyl-related compounds, and multi-component test samples.

[2] TM1, TM2, and TM3 only require that a detection technology indicates that any fentanyl or fentanyl-related compound is present. No specific compound identification or quantification is necessary. TM4 requires that a detection technology does not exceed an acceptable number of false-positive results for fentanyl or fentanyl-related compounds in a given number of tests. See 3.1.14 and 5.2 for further clarification on the number of allowed false positives. TM5 requires a detection technology to correctly identify the specific compound. No quantification is necessary. All TMs are shown in Annex A1.

[3] The minimum passing LCB and CL considered in this specification are 0.85 and 80 %, respectively. These minimum LCB and CL requirements can be met or exceeded with 4, 14, or 31 test samples for each chosen TM. See Table 4 through Table 6 in Section 5 for details. Determine the number of samples to be tested prior to a test event and do not increase it once testing begins. For example, if it is decided to conduct four tests (anticipating no failures) to achieve an LCB of 0.85 with 80 % CL and a failure occurs, additional tests beyond the four planned cannot be conducted. Testing must start over. Testing results that do not meet or exceed the 0.85/80 % LCB/CL have failed to achieve the minimum performance level in this specification.

[4] The purpose of Testing Tier 2 is to reduce testing burden by significantly reducing the variety of compounds that must be procured and mixed. Although Testing Tier 2 only contains four different samples from each TM, the total number of tests conducted can still be chosen. See 4.14 for further clarification.

[5] Prepare and mix all test samples following the methodology described in Test Method E3290. Record protocols and all calculations for any laboratory operations (for example, weighing and mixing). Identify any solvents used (including water) for those detection technologies that dissolve or suspend a sample prior to analysis as standard practice. Identify suitable positive and negative control samples that will be used during testing based on manufacturer guidance.

[6] Document protocols and results as described in Test Method E3290.

[7] As described in Test Method E3290 and as recommended by the manufacturer and when available, test a positive and negative control sample at least once per 10 samples analyzed or at least once each day that testing is performed, whichever is greater.

TABLE 1 *Continued*

[8] As described in Test Method E3290: In the final report, state which test event was conducted (Maximum Tier 1, Maximum Tier 2, Moderate Tier 1, Moderate Tier 2, or Minimal as described in 4.14), which TMs and test samples were included, and any control samples that were tested. Save all associated raw data, data analysis, and reporting outputs provided to the user by the instrument or assay after testing a sample. Report two testing results for field detection products that include reach-back technical support: (*1*) the instrument or assay direct readout result obtained immediately after a test event and (*2*) the result provided by reach-back support. Reach-back support must provide a verbal or written result within 60 min of obtaining the data for each sample or the result counts as a failed test. It is acceptable to plan with the reach-back provider for when you will be sending data. Include in the final report the written report from the reach-back provider, which references and documents the verbal result that was provided. For field-detection products that do not include reach-back support, report only one testing result (the instrument or assay direct readout result obtained immediately after a test). Results are qualitative only (not quantitative) and reported as: (*1*) a sample contains fentanyl or fentanyl-related compounds, (*2*) a sample does not contain fentanyl or fentanyl-related compounds, or (*3*) a sample contains the fentanyl synthesis precursor or related compound as specifically identified in TM5.

recommendation, but within the requirements above. Further guidance can be found in Test Method E3290. Table 2 shows the TMs that shall be used to analyze trace samples with the detection technologies considered in this specification. In Table 2, an 'X' denotes technology not applicable for the conditions of the Test Module.

4.6 Table 3 shows the TMs that shall be used to analyze bulk samples with the detection technologies considered. In Table 3, an 'X' denotes technology not applicable for the conditions of the Test Module.

4.7 Depending on the capabilities of the instrument or assay, up to five TMs containing multiple samples each are used to evaluate performance (see Table 2, Table 3, and Annex A1, Tables A1.1-A1.5).

4.8 TM1, TM2, and TM3 only require that a detection technology indicates that any fentanyl or fentanyl-related compound is present. No specific compound identification or quantification is necessary.

4.8.1 See Annex A2 for a list of fentanyl and fentanyl-related compounds. All instrument or assay results that indicate the presence of any of these compounds, including any compounds listed in TM1, are considered a successful test when testing samples from TM1, TM2, and TM3.

4.9 TM4 requires that a detection technology does not exceed an acceptable number of false positive results in a given number of tests for fentanyl or fentanyl-related compounds (see 3.1.12).

4.10 TM5 requires a detection technology to correctly identify the specific compound. No quantification is necessary. Colorimetric assays, immunoassays, and SERS typically cannot identify a specific compound, so TM5 is generally not applicable.

4.11 5.2 and Tables 4, 5, and 6 describe in detail the number of tests and number of allowable failed results needed to demonstrate compliance with the requirements in this specification.

4.12 Two different testing tiers are also defined: Testing Tier 1 requires that all 14 different samples in each chosen TM are tested and Testing Tier 2 requires that only samples 1 to 4 in each chosen TM are tested, thus reducing the cost and time burden of procuring and preparing a much larger number of different sample mixtures. A testing entity may still choose to test multiple replicates of those four samples to achieve a higher LCB and CL. See 4.14 for further clarification. TMs contain 14 different test samples, which are simply referred to as samples. A sample number is assigned to each sample within a TM. A sample with a lower number (for example, No. 1) is more important (more commonly encountered) than a sample with a higher number.

4.12.1 The testing entity or agency requesting testing determines whether Testing Tier 1 or Testing Tier 2 with every TM the instrument or assay can be applied to will be used based on the coverage of different sample compositions desired for testing.

4.13 The number of samples chosen to be tested for each TM depends on the desired LCB for the POD and CL selected.

4.13.1 To achieve the lowest acceptable LCB and CL (0.85 and 80 %, respectively), four samples shall be tested a single time from each of the applicable TMs with no failures observed.

4.13.2 To allow for achieving a higher LCB with a CL equal to or higher than 80 %, more tests are needed. It is recommended that users of this specification select the largest number of tests practical for their application, as this will lead to an LCB value that is closer to the calculated POD.

TABLE 2 Applicability of Detection Technologies by Test Module for Trace Sample Amounts (≤1 µg)

Technology Class	Test Module 1 >95 % Target (0.95 µg)	Test Module 2 10 % Target (100 ng)	Test Module 3 1 % Target (10 ng)	Test Module 4 Non-Targets (1 µg)	Test Module 5 Precursors (1 µg)
GC/MS	√	√	√	√	√
MS	√	√	√	√	√
IMS	√	√	√	√	√
Raman	X[A]	X	X	X	X
SERS	√	√	√	√	X
FTIR	X	X	X	X	X
Immunoassays	√	√	√	√	X
Colorimetric assays	√	√	√	√	X

[A] An 'X' denotes technology not applicable for the conditions of the Test Module.

TABLE 3 Applicability of Detection Technologies by Test Module for Bulk Sample Amounts (>1 µg to <10 mg)

Technology Class	Test Module 1 >95 % Target (>1 µg to 10 mg)	Test Module 2 10 % Target (100 ng to 1 mg)	Test Module 3 1 % Target (10 ng to 0.1 mg)	Test Module 4 Non-Targets (>1 µg to <10 mg)	Test Module 5 Precursors (>1 µg to <10 mg)
GC/MS	√	√	√	√	√
MS	√	√	√	√	√
IMS	√	√	√	√	√
Raman	√	√	√	√	√
SERS	√	√	√	√	X[A]
FTIR	√	√	√	√	√
Immunoassays	√	√	√	√	X
Colorimetric assays	√	√	√	√	X

[A] An 'X' denotes technology not applicable for the conditions of the Test Module.

Conducting 14 or 31 tests on samples from each applicable TM are two alternatives that should be considered, as they provide the smallest number of tests required for achieving higher LCB and CL values using the statistical approach in this specification. Testing results that do not meet or exceed the 0.85/80 % LCB/CL have failed to achieve the minimum performance level established in this specification.

4.13.3 The statistical approach employed in this specification requires that the number of tests conducted is determined prior to a test event and that this number remains fixed once testing starts. Because it is not possible to know the number of failures that will be observed prior to testing, and the number of failures determines the LCB for a fixed CL, a larger number of tests allows for the possibility of achieving a desired LCB and CL even if some failures are observed.

4.14 Each of the possible test events that may be performed is summarized below.

4.14.1 *Maximum Testing Tier 1:*

4.14.1.1 A total of 31 tests of samples from TM1, TM2, TM3, TM4, and TM5 (TM5 as applicable depending on the capability of the detection technology) as follows: three replicates of Sample Numbers 1 to 3 (nine tests) and two replicates of Sample Numbers 4 to 14 (22 tests).

4.14.2 *Maximum Testing Tier 2:*

4.14.2.1 A total of 31 tests of samples from TM1, TM2, TM3, TM4, and TM5 (TM5 as applicable depending on the capability of the detection technology) as follows: eight replicates of Sample Numbers 1 to 3 (24 tests) and seven replicates of Sample Number 4 (seven tests).

4.14.3 *Moderate Testing Tier 1:*

4.14.3.1 A total of 14 tests of samples from TM1, TM2, TM3, TM4, and TM5 (TM5 as applicable depending on the capability of the detection technology) as follows: Sample Numbers 1 to 14, each tested once.

4.14.4 *Moderate Testing Tier 2:*

4.14.4.1 A total of 14 tests of samples from TM1, TM2, TM3, TM4, and TM5 (TM5 as applicable depending on the capability of the detection technology) as follows: four replicates of Sample Numbers 1 to 2 (8 tests) and three replicates of Sample Numbers 3 to 4 (6 tests).

4.14.5 *Minimal Testing*—Minimal testing only requires testing four samples from TM1, TM2, TM3, TM4, and TM5, therefore, there is no difference between Tier 1 and Tier 2 testing.

4.14.5.1 Four tests of samples from TM1, TM2, TM3, TM4, and TM5 (as applicable depending on the capability of the detection technology) as follows: Sample Numbers 1 to 4, each tested once.

5. Statistical Considerations for Testing

5.1 The testing approach described herein uses a score CI,[7] to define the number of tests that shall be conducted so that a desired POD estimate, as indicated by the value of a given LCB in a one-sided interval with a specified CL, can be achieved under some conditions for an instrument or assay. In this specification, testing shall be conducted to establish instrument or assay performance over a set of possible performance outcomes with LCB ≥ 0.85 and CL ≥ 80 %. Results from selecting and executing one of the testing plans presented in 4.14 of this specification can be used to determine if the chosen level of statistical performance has been met or exceeded. Testing results that do not meet or exceed the 0.85/80 % LCB/CL have failed to achieve the minimum performance level established in this specification.

5.2 The following is the minimum number of tests that shall be conducted for each of the five applicable TMs. See 4.14 and Table 1 for further clarification.

5.2.1 To achieve an LCB ≥ 0.85 and CL ≥ 80 % at the chosen test concentration:

5.2.1.1 If minimal testing is chosen (4 tests per TM), no failures shall be observed;

5.2.1.2 If moderate testing is chosen (14 tests per TM), no more than a single failed result shall be observed; or

5.2.1.3 If maximum testing is chosen (31 tests per TM), no more than three failed results shall be observed.

5.3 Only instrument or assay results indicating a clear presence or absence of fentanyl or a fentanyl-related compound count toward the total number of tests. Inconclusive results that may be caused by instrument or assay failures should be treated like control sample failures (identify the cause of the failed test result, fix it, include it in the final report, and ensure the system is performing correctly before testing is continued or restarted).

5.3.1 In general, only testing plans with zero, one, two, or three failed results and LCB ≥ 0.85 and CL ≥ 80 % are

[7] Agresti, A., Coull, B.A. "Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions." *The American Statistician*, Vol. 52, No. 2, May 1998, pp. 119-126.