OSAC Registry Request Comment Adjudica

Document Title	ASTM E2548-11 - Standard Gu
Requesting Subcommittee	Seized Drugs
Subcommittee Chair	
Name:	Sandra Rodriguez-Cruz
Affiliation:	DEA
Email:	
Phone:	
Beginning Comment Period Date	8/14/15
End Comment Period Date	9/13/15
Comment Adjudication Meeting Dates	There were no in-person meetings. Subcommitt
# of Members Present	Voting on the adjudication comments was com
	Subcommittee met via conference call on Janu
	Second voting to address QIC requests was con

Note: This template is intended for use by all subcommittees considering a new document for a

ıtion Template

ıide for Sampling Seized Drugs for Qualitative and Quantitative Analysis

Subcommittee Technical Contact

Name:	Sandra Rodriguez-Cruz
Affiliation	DEA
Email:	
Phone:	

tee met via conference call on October 29, 2015 (17 members present).
pleted via Kavi on November 12, 2015.
ıary 14, 2016 (15 members present).
ıpleted during conference call.

ıddition into registry.

Id	Person First Name	Person Last Name	Assigned To	Origin	Category	Section	Page	Line	Subject	Comment	Proposal	Response	Status	Resolution	Disposition	Resolution Date and Vote Outcome	Company Name	Interest Category	Submission Date	Group Name	Document Name	Revision Number	Document Description	History
																								2015-08-31 18:57:52 - 172.18.0.4 - john.jones@nist.gov - Added new comment
																Resolution Date: 11/12/2015								2015-08-31 18:57:52 - 172.18.0.4 - john.jones@nist.gov - Updated submitter info
																Vote Outcome: YES-								2015-08-31 18:57:52 - 172.18.0.4 - john.jones@nist.gov - Updated comment info: section=N/A | category=Editorial | state=New | response=None
																18 votes; NO-0 votes								
																Seized Drugs subcommittee voted to address QIC re quest:								
																Resolution Date: 01/14/2016					Link to - ASTM E2548-11e1		This link will provide access to ASTM to view:	
									LRC Comments on the E2548-							**Vote Outcome:**					Standard Guide for Sampling		ASTM E2548-11e1 Standard Guide for	2015-08-31 18:57:52 - 172.18.0.4 - john.jones@nist.gov -
195	John Paul	Jones II		Public Review	N/A		1	1	11	JP uploaded LRC comments to attach them to Kavi Record -8-31-2015	These comments were submitted by the LRC on 8-30-2015 and JP Jones uploaded them into Kavi on 8-31-2015 to attach to the Kavi Record.	See Seized Drugs subcommittee responses within LRC document. **See attached**	New	See attached.	Resolved	**YES-13 votes; NO-0 votes**	NIST		8/31/15 18:57	SAC Chemistry/Instrument Analysis	Seized Drugs for Qualitative and Quantitative Analysis	0	Sampling Seized Drugs for Qualitative and Quantitative Analysis	Updated comment info: category=N/A | section=1 | item=Paragraph | page=1 | line=1 | proposal updated
										From a human factors perspective, the major danger that arises in a sampling process is the potential for inadvertent sampling bias. Sampling bias can arise unintentionally when the process of sampling relies on subjective judgment. Without being aware of it, analysts might skew the sample in a manner that causes over- or under-representation of certain items. Consequently, it is important that analysts minimize subjective judgment in the choice of items to sample. Once a sampling plan is developed, the choice of particular items to sample (within sampling categories) should result from a process that is entirely random. The standards, as currently written, do not adequately discuss this problem and how to solve it. If these standards are revised, then the Human Factors Committee recommends that further commentary on the dangers of sampling bias be incorporated. The Human Factors Committee will be happy to suggest possible language to that effect.		Not persuasive- Section 5.6.2.1 gives guidance to utilize random number generation and sampling tables. Section 5.6.2.2 addresses the "black box" method as a means to ensure random selection. Additional details for specific procedural steps regarding random sample selection should likely fall in laboratory procedures and not in a standard guide.											2015-09-16 16:32:36 - 172.18.0.4 - john.jones@nist.gov - Added new comment
																								2015-09-16 16:32:36 - 172.18.0.4 - john.jones@nist.gov - Updated submitter info
																								2015-09-16 16:32:36 - 172.18.0.4 - john.jones@nist.gov - Updated comment info: section=N/A | category=Editorial | state=New | response=None
	Human Factors	Human Factors														Resolution Date: 11/12/2015					Link to - ASTM E2548-11e1		This link will provide access to ASTM to view:	
									HFC Comments							Vote Outcome: YES-					Standard Guide for Sampling		ASTM E2548-11e1 Standard Guide for	2015-09-16 16:32:36 - 172.18.0.4 - john.jones@nist.gov -
235	Committee (HFC)	Committee (HFC)		Public Review	N/A		1	1	1 on ASTM 2548		see attached document		New	Not persuasive	Resolved	18 votes; NO-0 votes	NIST		9/16/15 16:32	Analysis	Seized Drugs for Qualitative and Quantitative Analysis	0	Sampling Seized Drugs for Qualitative and Quantitative Analysis	Updated comment info: category=N/A | section=1 | item=N/A | page=1 | line=1 | proposal updated

Categories for adjudication of negative pub

Term
Not Germane
Editorial - review requested
Persuasive - review required
Withdrawn by submitter
Not persuasive
Previously considered
No response needed

Resolution categories

Term
Resolved
Unresolved

lic comment for addition to registry

Definition
Comment is not relevant to the subject of document being considered
There is general agreement with edit given, edit by SDO may be requested
General agreement with negative comment given, further review by subcommittee required
Comment withdrawn by submitter
Justification for non persuasive rationale is indicated by committee action
Topic of comment was previously discussed and resolved by subcommittee
Comment does not require a response

Definition
Response has been adjudicated (agreed and voted on)
Response has not been adjudicated (agreed and voted on)

Final Decision of the Ad hoc Independent Review Panel for the Appeal of ASTM Standard E2548-11, Standard Guide for Sampling Seized Drug for Qualitative and Quantitative Analysis

Date:

November 22, 2016

Panel Members:

Mark Stolorow, FSSB Representative
Kris Cano, QIC Representative
Arlene Hall, QIC Representative
Mark Ruefenacht, Forensic Metrologist/Statistics
Anna Deakin, Forensic Chemist/Controlled Substances Expert

Comment on Section 5.5.1

The panel agrees with the adjudication of this comment as non-persuasive and that there is a misunderstanding of the language and an issue with the term being used in the document. The document references ASTM E1732 – Standard Terminology Relating to Forensic Science where the terms "sample" and "sampling" are defined. Although the terminology may not be clear, there is a reference to the terms used in the document. It is suggested that in the next revision of the document that these key (primary) terms are clearly defined in the body of the document. However, this does not reverse the previous determination of this comment as non-persuasive, and, consequently, the appeal of this comment is denied.

Comment on Figure 1

Comment 1: The panel agrees with the adjudication of this comment as non-persuasive. The panel commented that a drug chemist would know what is depicted in this figure and supports that this would be a management decision based on jurisdiction. The appeal to this comment is denied.

Comment 2: The panel agrees with the adjudication of this comment as non-persuasive. Sampling is defined in the referenced terminology document.

Comments on Need for Complete Reporting

The panel affirms the appeal on the two editorial comments referencing that the subject of reporting will be the subject of another document. The panel determined that the comment is not editorial and should be re-adjudicated. Referring to a document that does not currently exist is not addressing the comment or concern that the reporting is incomplete. As a possible solution, the panel suggests the option of removing the reference to reporting if there will be a subsequent document that will address reporting.

To: OSAC Independent Review Panel

From: David Kaye, Jennifer Friedman, Barry Scheck

Subject: Appeal from Adjudication of LRC-compiled Comments on ASTM E2548−11

Date: 18 October 2016

Introduction

After careful consideration, we have decided to appeal the adjudication of our comments on ASTM 2548-11. Although the majority of the standard is fit for purpose, there are certain aspects that we believe may prove problematic when applied in a forensic setting. We fully share the goal of the OSAC and the Chemistry SAC to promulgate standards that are useful to providers of forensic science services and the legal community that relies on these services. In addition, we remain impressed with the devotion and efforts of OSAC members to develop standards for the NIST registry that will be considered "gold standards"—documents that all concerned individuals always may rely on because they are based on methods that are scientifically validated and contain no technically inaccurate or ambiguous statements that might generate unnecessary controversy in court. In short, this appeal has only one motivation—to ensure that the public record establishes that the OSAC process for reviewing this standard was followed both as to form and substance, thus precluding future argument in court or elsewhere that any criticisms were not understood or not squarely addressed.

Scope of Appeal

The FSSB approved ASTM E2548−11, a superseded "Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis," over objections from the FSSB Statisticians Task Group, the HFC, and the LRC. Appreciating that reasonable people may differ regarding the gravity of these objections, we do not revisit them. However, to ensure that proper consideration is given to comments compiled by the LRC, this appeal is intended to demonstrate how the adjudication process (1) failed to address significant parts of certain comments and (2) misunderstood or misinterpreted several comments.

Subcommittee responses to comments that only reflect differences of opinion on issues raised and recognized in the adjudication are not discussed here. Because the subcommittee responded to many comments on their merits, this appeal only concerns the adjudication of three groups of comments: (1) a comment on Section 5.5.1; (2) comments on Figure 1; and (3) comments on the need for complete reporting. The analysis that follows shows that the adjudication was flawed in that it failed to respond to these comments by supplying the justification for (1) including text that implies that frequentist inference from samples yields probabilities for the values of population parameters; (2) omitting definitions of key terms such as "sampling plan," "sampling procedure," and "sampling strategy," and using such terms

instead of ones like "probability sampling" that have well-defined meanings in the statistical theory of sampling; and (3) specifying one component of what must be reported about sampling results and not specifying other, critical aspects of the sampling data and the inferences from them.

Appeal

I. Comment on Section 5.5.1

The LRC-compiled comments highlighted a common form of fallacious statistical reporting and testimony in one subsection of the guidelines:

> Section 5.5.1 speaks of "The probability that a given percentage of the population contains the drug of interest or is positive for a given characteristic." The Standard should make it clear that this posterior probability of a population parameter cannot be computed with any of the frequentist methods mentioned in the Standard.

The subsection in question states:

> Statistical approaches are applicable when inferences are made about the whole population. For example: The probability that a given percentage of the population contains the drug of interest or is positive for a given characteristic.

The purpose of the comment was to call the subcommittee's attention to the fact that frequentist hypothesis testing and confidence intervals are inherently incapable of providing the *probability* that a population parameter has a particular value or range of values and to ask the subcommittee to make the limitation on the frequentist methods referred to in ASTM E2548-11 explicit. The only way to compute "the probability that a given percentage of the population contains the drug" (without a census, in which case no probability statement would be made) is to apply Bayes' rule. The concern presented for adjudication was that, as written, the guideline suggests that frequentist statistical analyses can yield the *probability* that the population has the extrapolated percentage of the drug or percentage of positives.

The adjudication does not resolve this concern. It states:

> *Not persuasive. This comment appears to originate from a misinterpretation of the statistical language employed. The quoted text is the standard sampling hypothesis tested via frequentist sampling methods based on the hypergeometric probability distribution.*

This response does not address the fact that the frequentist statistical approaches mentioned in the guideline are inconsistent with the suggestion in section 5.5.1 that a probability can be attached to a statement of the value of a population parameter. The quoted text is not "a sampling hypothesis being tested"—it is a probability being attached to a hypothesis about the value or values of a population parameter. The subcommittee did not engage the substance of the comment—which is that probabilities cannot be assigned to population parameters "via frequentist sampling methods based on the hypergeometric distribution"—or, for that matter, via

any other distribution. Either the subcommittee intended its adjudication to deny the fundamental nature of frequentist tests,[1] which is hard to believe, or it did not understand the point of the comment made about Bayesian as opposed to frequentist hypothesis testing. Although this point is subtle, it would be unfortunate to place on the NIST registry a document that seems to contradict reference works for judges,[2] admonitions in textbooks,[3] and many other publications.[4] Wording like that in Section 5.5.1 will not make the registry a credible source of information and will undermine NIST's reputation for statistical expertise.

II. Comments on Figure 1

A series of comments pertaining to Figure 1 of ASTM E2548-11 also were not adjudicated properly. The figure is reproduced below:

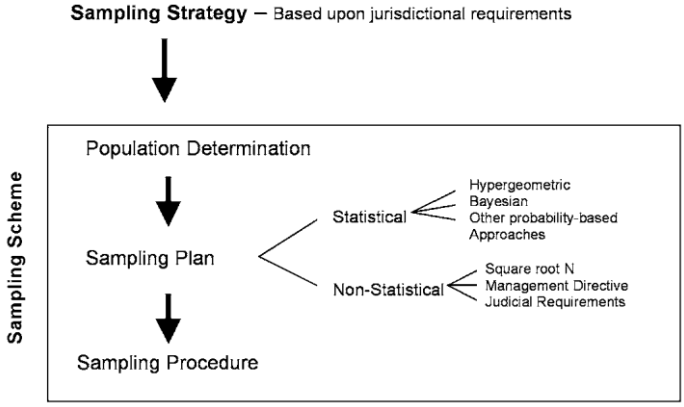

FIG. 1 Relationship of Various Levels Required in Sampling

With one arguable exception, none of the terms in the figure are defined anywhere in the document. This omission prompted the comment that

[1] On the fundamental distinction between the frequentist and Bayesian approaches in the context of drug sampling, see, for example, Colin G.G. Aitken & Franco Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists 181 (2d ed. 2004).

[2] David H. Kaye & David A. Freedman, Reference Guide on Statistics, in Reference Manual on Scientific Evidence 211, 250 (Federal Judicial Center & National Research Council Committee on the Development of the Third Edition of the Reference Manual on Scientific Evidence, 3d ed. 2011) ("According to the frequency theory of statistics, there is no meaningful way to assign a numerical probability to the null hypothesis.").

[3] E.g., D.R. Cox, Principles of Statistical Inference 42 (2006) ("a clear misconception"); Larry Wasserman, All of Statistics: A Concise Course in Statistical Inference 157 (2004) ("Warning! Do not confuse the *p*-value with $P(H_0|\text{Data})$. The *p*-value is not the probability that the null hypothesis is true.").

[4] E.g., James O. Berger & Thomas Selke, Testing a Point Null Hypothesis: The Irreconciliability of P Values and Evidence, 82 J. Am. Stat. Ass'n 112 (1987).

> The terminology within Figure 1 is confusing. It is sufficiently problematic that it rises to the level of a substantive rather than merely a stylistic concern. It seems like the middle level in the diagram is about the determination of sample size rather than a full "sampling plan." Thus, the terms "sampling procedure" and "sampling plan" should be defined. Doing so might answer questions (a)–(d) below. Clearer terminology comes from David Freedman, who wrote that "Methods for choosing samples are called 'designs.' Good designs involve the use of probability methods, minimizing subjective judgment in the choice of units to survey. Samples drawn using probability methods are called 'probability samples.'" David A. Freedman, Sampling, available on the Department of Statistics, University of California (Berkeley) website.

The thrust of the comment is plainly that definitions of key terms are absent. It proposes clearer terminology. Rather than explain why the standard contains no definitions of its terms, the response addresses only one sentence out of the four. The entire response is

> *Not persuasive. The middle level of Figure 1 is not about determination of sample size. It establishes the foundation of a plan, as sampling plans can be statistical or not. The actual sample size is determined based on laboratory requirements and jurisdiction. For example, one laboratory may only need to provide evidence that a controlled substance is present in at least 50% of the seizure population while a separate laboratory's jurisdiction may require proof of at least 90% of the population containing the drug. The required sample size would be different for these two laboratories.*

Arguing that one possible interpretation of the undefined terms is incorrect is not a response to the main point—that the terms need to be defined in the standard for users to know what it says. Indeed, the response itself demonstrates that the failure to define terms such as "sampling plan" results in a standard with indeterminate meaning. The response that "the middle level" (labeled "sampling plan") is "not about … sample size" is flatly inconsistent with the figure itself. The figure places squarely within the middle level of "sampling plan" such methods of determining sample size as "Square Root N," "Management Directives," and "Judicial Requirements," all of which can or do address sample size.[5] Likewise, the "hypergeometric," "Bayesian," and "other" approaches in this "middle level" all entail a determination of sample size.[6] The fact that the subcommittee cannot provide a consistent

[5] The "Square Root N" rule, for example, is merely a "popular rule for deciding how many containers or items … to sample for drug testing." Alan Julian Izenman, Statistical and Legal Aspects of the Forensic Study of Illicit Drugs, 16 Stat. Sci. 35, 47 (2001). If the number of units in the population is N, the rule tells a laboratory to draw a sample of size √N. The sample size √N and the more traditional √N +1 rule for acceptance sampling are statistics, but they have no justification in sampling theory. See id.

[6] E.g., R.S. Frank, S.W. Hinkley & C.G. Hoffman, Representative Sampling of Drug Seizures in Multiple Containers, 36 J. Forensic Sci. 350, 354 (1991) (explaining that "Application of the [hypergeometric] Model … consists of two basic steps or determinations: use the statistical model to determine the sample size, R, and perform the presumptive tests.").

explanation of what is included in a "sampling plan" underscores the need for a guideline that defines its terms and applies them consistently.

Furthermore, the response regarding the sentence in the comment about sample size suggests that the subcommittee misunderstood the connection made in the comment between a probability distribution (hypergeometric or otherwise) and the choice of sample size.[7] The response asserts that the sample size would differ in a jurisdiction in which the prosecution must prove that "at least 50% of the seizure population" contains a controlled substance and one in which the prosecution must establish "at least 90% of the population containing the drug." This is undoubtedly true, but what does it have to do with the meaning of terms like "sampling plan," "sampling procedure," and "sampling strategy" in Figure 1? If the sampling is random, then whatever the sample size may be, statistical theory permits valid extrapolation to the population. This is the case whether the population percentage that is of legal importance is 50%, 90%, or any other number. We are left with a figure that categorizes some approaches as "Statistical" and others as "Non-Statistical" without apparent rhyme or reason—at least, when there is no clear definition in the standard of these terms and related ones.[8] This is the substance of the comment on Figure 1, and the adjudication does not answer it.

A series of subcomments elaborate on the terminological problems with Figure 1. The response to one of them is notable. The subcomment asks

> Are there "hypergeometric sampling plans"? A sampling plan should specify the sample size (or a procedure for stopping the sampling if results on the sampled items up to that point make further testing unnecessary). For sampling a finite population without replacement, the hypergeometric probability distribution applies to sample-size computations. But how does that make the sampling plan hypergeometric? The simplest "plan," one would think, is to take a random

[7] The dichotomy between "Statistical" and "Non-Statistical" in the figure indicates that the only defining characteristic is whether sampling theory is used to determine sample size. As explained in a previous note, the "Square Root N" rule is simply a formula for picking a sample size. It is not "the foundation of a plan" for selecting which units to analyze, and using the formula does not preclude valid statistical inferences about population parameters. That is why one sentence in the comment—and the only to one to which the subcommittee has chosen to respond—reads, "It seems like the middle level in the diagram is about the determination of sample size rather than a full 'sampling plan.'"

[8] ASTM E2548-11 states that "For the purpose of this guide, the use of the term statistical is meant to include the notion of an approach that is probability-based," and the word "probability-based" also appears in figure 1. Unfortunately, "probability-based" is not defined, and it is hard to say whether "the notion of" it actually limits so-called statistical sampling to "probability sampling." Probability sampling refers to drawing a sample in which every unit in the population has a known probability of being in the sample. Methods for generating a sample that has this property justify inferences from the sample to the population, and that seems to be what the subcommittee very appropriately has in mind for "statistical sampling." But any of the types of purportedly "non-statistical" plans in Figure 1 can "include" probability sampling. For example, a sample whose size comes from the "Square root N" rule in Figure 1 can be a probability sample. The dissonance between what seems to be the intended definition of "statistical" and "nonstatistical" and the actual use of the term in Figure 1 exemplifies the need for clear definitions, applied consistently in the standard.

> sample of a size determined in light of the sample-size as computed with the hypergeometric distribution. The result is just a simple random sample.

The subcomment notes that the failure to define "sampling plan" leads to anomalous descriptions such as a "hypergeometric sampling plan" when what may be intended is simple random sampling. The adjudication responds by denying that its guideline refers to "hypergeometric" as a sampling plan and then saying that there is a literature that answers the question even if the guideline does not. The response on this point is, in full:

> *Not persuasive. The quoted text "hypergeometric sampling plans" is not found in the document. Sampling plans based on the discrete hypergeometric sampling distribution are extensively reported in the statistical and forensic literature.*

The claim that the figure does not refer to "hypergeometric sampling plans" is impossible to square with the figure. The top line of the middle row is "sampling plan——statistical——hypergeometric." Denying that the figure refers to a "sampling plan" that is "hypergeometric" is equivalent to ignoring the figure and the comment on it.

Neither does this adjudication of the comment respond to the question of why the figure refers to a sampling *plan* that is hypergeometric when it seems to be speaking of simple random sampling. Establishing a sample size and drawing inferences (both frequentist and Bayesian) from a random sample to a population parameter may involve calculations with the hypergeometric probability distribution. The literature merely explains these procedures. Figure 1, however, distinguishes between a hypergeometric "sampling plan" and a Bayesian "sampling plan." A bald statement that recourse to the literature resolves the ambiguous and puzzling juxtaposition of terms signals a misunderstanding of the question.[9]

III. Comments on the need for complete reporting

Another LRC-compiled comment not handled properly in the adjudication process concerns the limited guidance on what information to provide when reporting how sampling was done:

[9] The need for and nature of simple random sampling is not based on any particular distribution of the sample statistic. It is well known that if probability sampling is used to sample from a finite population, the probability distribution of the sample statistic (the "sampling distribution" of the number of successes in simple random samples) is hypergeometric. This mathematical fact can be used to pick a sample size to help achieve a desired standard error (or, in a Bayesian framework, to help achieve a desired posterior probability). It also can be used to form a confidence interval (or a Bayesian credible region) around the sample statistic after the sampling is done. A "sampling plan" might include one or both of these tasks. "[T]he statistical and forensic literature" on the hypergeometric distribution suggests that a sampling plan involves both tasks, and it gives the hypergeometric distribution an important role in establishing sample size. Yet, the subcommittee insists in its adjudication that "[t]he middle level of Figure 1 [on "sampling plan"] is not about determination of sample size." This contradiction is a further indication that the subcommittee has not properly adjudicated the comments about the need to define terms such as "sampling plan" in the standard and to apply those definitions consistently.

> The section on reporting needs more body. How should the results of the analyses be presented in a report and in court? Should not the full sampling plan be stated — the mechanism for drawing samples (e.g., blinded, which is to say, "black box" sampling, or selecting from numbered samples by a table of random numbers); the sample size, and the kind of sampling (SRS, stratified, etc.)? It is not enough merely to state that "the results are based on a sampling plan," as Section 8.1.1 seems to contemplate.

Section 8 of the standard is entitled "Reporting." As the suggestion for a modification indicates, this section states that certain things "must" be reported and arbitrarily omits other things.

The adjudication dismisses this concern out of hand, as an ignorable "editorial" comment:

> *Editorial. The subject of reporting will be addressed in a future Seized Drug subcommittee document.*

To describe a request to supply critical information in the section on reporting as "editorial" is to ignore the arbitrariness of a standard that says that *X* must be reported without mentioning equally important things to report. The comment asked the subcommittee to complete its work on the section on reporting before placing on the registry a document with a section on reporting. The subcommittee responded, not by questioning the value and importance of a more comprehensive section on reporting, but by promising to do the necessary work in the future.

This adjudication is improper because it fails to explain why it is unnecessary to specify *all* of the information about the sampling design and results that, at a minimum, "must" be reported. There might be a reason for requiring that one item "must" be reported and no others, but, if there is, the adjudication made no attempt to provide this reason.

A further comment on the incomplete section on reporting received the same type of nonresponse. The comment was:

> The results themselves always should be stated as in the example in Section 8.1.2 — "2 of 100 bags were analyzed and found to contain Cocaine." When probability sampling has been employed, more can be said. But what would this be? A confidence interval? With what confidence coefficient? With what definition of "confidence"? A frequentist test of a hypothesis? Explained how? A Bayesian conclusion such as "There is a probability of 90% that the weight of the cocaine in the shipment seized exceeds X" (if the analysis is Bayesian and the prior distribution and the basis for it are presented)? To be of the most benefit to analysts, lawyers, and courts, the Standard should do more to spell out the "[s]ampling information [that] shall be included in reports."

The adjudication was

> *Editorial. The subject of reporting will be addressed in a future Seized Drug subcommittee document. The quote [sic] text should not be interpreted as an*

exclusively Bayesian conclusion, as it is also the routine hypothesis tested using a hypergeometric probability based sampling procedure.

To state, without explanation, that a problem with an existing document will corrected in a future document is arbitrarily to refuse to address the existing problem. This guideline might omit the topic of reporting entirely because it is outside its scope. But that is not how guideline is written, and it is not how the comment was adjudicated. The response was, in effect, we know the subject is only addressed in part, but we will not explain why specific points are addressed at this time when others are not. It could be that the subcommittee believes that the other matters in the comment are of less moment or that the apparently incomplete reporting endorsed by the standard is somehow sufficient. If so, the adjudication process requires a presentation of such arguments rather than a promise to make changes in the future.

The second sentence of the adjudication does respond to one part of the comment. However, it is only a small part, and the incomplete response displays a failure to understand this part of the comment. The query that received a response was whether the guideline should require or approve of reporting "[a] Bayesian conclusion such as 'There is a probability of 90% that the weight of the cocaine in the shipment seized exceeds X" (if the analysis is Bayesian and the prior distribution and the basis for it are presented)?" The adjudicative response was that reporting the probability that a population parameter has certain values is a routine, frequentist procedure. As documented in notes 1 through 4, this response demonstrates a basic misunderstanding of the meaning of hypothesis tests and confidence intervals.[10] To the extent that this guideline codifies this misunderstanding, it does not belong on the registry, as it will subject OSAC, NIST, and testifying witnesses to easily avoidable criticism;[11] to the extent that the adjudication does not recognize the distinction between the Bayesian and frequentist statistical analyses that it mentions, it again fails to understand the comment.

---

[10] See also C.G.G. Aitken, Sampling—How Big a Sample?, 44 J. Forensic Sci. 750, 751 (1999) (for a "so-called confidence interval or level … no probability can be attached to the uncertain event that the interval contains θ [the population value]").

[11] For an example of this criticism in the context of estimating DNA allele frequencies from sample data, see John S. Buckleton, Duncan Taylor, Jo-Anne Bright & James M. Curran, Sampling Effects, in Forensic DNA Evidence Interpretation 181 (John S. Buckleton , Jo-Anne Bright & Duncan Taylor eds.. 2d ed. 2016):

> It is not acceptable to substitute the word *probability* for *confidence* in statements regarding confidence intervals. '… A report issued by the NRC that contains [the statement that] "the traditional 95% confidence limit, whose use implies the true value has only a 5% chance of exceeding the upper bound" must lose credibility with statisticians'. The report in question wrongly confuses a confidence interval with a probability interval. Strictly speaking, any particular confidence interval either contains the true value or it does not, but 95% of intervals should contain the true value. We cannot say, 'It is 95% probable mat this confidence interval contains the true value'. The difference appears academic but could easily lead to difficulty in court.

Id. at 184 (emphasis in original, notes omitted). We would underscore two sentences. Like the 1992 NRC report on DNA evidence, ASTM 2548-11, if understood as the subcommittee's adjudication proposes, "must lose credibility with statisticians" and "could easily lead to difficulty in court."

* * *

For the reasons stated above, the adjudication process with respect to the three sets of LRC-compiled comments listed here was flawed.

APPENDIX: COPY OF THE ADJUDICATION

COMMENTS BY THE OSAC LEGAL RESOURCE COMMITTEE (LRC)

TO: Seized Drugs subcommittee of the Chemistry-Instrumental SAC

FROM: Christopher J. Plourd, Chair, OSAC Legal Resource Committee (LRC)

RE: **OSAC LEGAL RESOURCE COMMITTEE (LRC) COMMENTS ON:** E2548-11 "Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis" (hereinafter referred to as "E2548-11").

Response from the Seized Drugs Subcommittee

The OSAC Seized Drugs subcommittee recognizes and appreciates the comments received from the Legal Resource Committee (LRC) pertaining ASTM E2548-11, Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis. The following responses attempt to address and clarify some of the issues brought to our attention by the LRC.

The document in review (E2548-11) was original published by ASTM in 2007. The document was originally published as Part IIIA of the Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) Recommendations, which are intended to assist forensic analysts and managers in the development of analytical techniques, protocols and policies. The SWGDRUG Recommendations are internationally recognized as minimum standards that may be supplemented to address unique jurisdictional laboratory requirements.

The document under consideration is neither a test method nor a prescriptive standard. Therefore, many of the comments provided by the LRC members are not considered applicable. Also, this document is not intended to encompass all other related standards, terminology, validation documentation, etc. that may assist in its application. The field of seized drug analysis is an extensive one encompassing the subjects of sampling, chemical identification, presumptive and confirmatory analytical techniques, method validation, quantitative procedures, structure elucidation, measurement uncertainty, reporting protocols, and many others. It is unrealistic and impractical to attempt to include all these subjects into one single

document, as the result would be an ineffective standard that would not be useful for practitioners. Efforts are already underway to address some of the aforementioned subjects via the publication of separate documents on the OSAC Registry.

There appears to be some misconceptions and misunderstandings regarding the OSAC process as well as the process in use by the standard development organization (SDO) under which this standard is published (ASTM). We believe that it would be useful to provide additional training regarding these procedures to not only members of the LRC but also members of other resource committees and individual discipline subcommittees.

We understand the desire for these documents to be written such that they are more comprehensible to lawyers and judges. However, many of these documents under review have already been in the forensic community for over ten years and they were originally drafted with the goal of providing useful minimum standards for seized drug analysts. In fact, there are many seized drug laboratories throughout the world that use this document (as well as the SWGDRUG Recommendations) as a foundation for their policies and procedures. It is our strong belief that attorneys and judges have the ability to consult scientists in order to interpret a scientific document, much in the same way as a scientist would consult an attorney to interpret a legal document. Serious consideration is being given to all the comments provided by the LRC members. However, we believe that making significant changes to the document in order to address the resource committee's would take away too much from the original technical and scientific intent of the document.

The Seized Drugs subcommittee of the Chemistry-Instrumental SAC is proposing to approve an existing ASTM E2548-11 standard to the OSAC Repository registry.

Our comments are primarily intended to enhance the value of the Standard to the legal community. This Standard will be most helpful if it not only helps assure high quality results in the laboratory, but also is written to show how work performed in accordance with the Standard is both well grounded in theory and data and that it is presented within the boundaries of "the knowledge and experience of [the expert's] discipline."[12] Consequently, the comments are intended to address four questions that are important to the legal reception of the Standard: (1) Is the Standard written as clearly as possible, and without undefined technical terms and symbols, so as to enable lawyers and judges to grasp the main ideas and requirements set forth? (2) Does the Standard describe in detail how the peer-reviewed and readily available scientific literature establishes the validity of the assumptions underlying the scientific tests and the interpretation of test results? (3) Does the Standard list the limitations of the tests and results and provide for expressions of the uncertainties in measurements and inferences drawn from them? (4) Does the Standard include recommendations or requirements for the creation and retention of documentation of the test and the contents of reports, including the scientific limitations of the tests and related conclusions or inferences? These are matters of both technical merit and legal importance. Although the LRC is not able to assess the scientific merit of a Standard, our review encompasses

[12] Kumho Tire Co. v. Carmichael, 526 U.S. 137, 148 (1999) (quoting Daubert v. Merrell Dow Pharms, Inc., 509 U.S. 579, 592 (1993)).

whether a Standard makes a prima facie case for the validity of the methods and legal utility of the kinds of expert opinions that a Standard contemplates.

Comment by LRC member David Kaye:

Comments of David Kaye on Placing ASTM E 2548−11 ("Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis") on the OSAC Repository of Standards and Guidelines.

The document is not ready for inclusion as an OSAC Standard. According to the QIC PowerPoint presentation of July 27, 2015, OSAC-approved documentary standards and guidelines are to "have demonstrated: Technical merit; Detailed Scope; Fitness for purpose; Uncertainty measurement and potential bias; and Method validation, as appropriate." A "standard ... specifies uniform methods, actions, practices, or processes, protocols. Compliance [is] mandatory and modified only under unusual circumstances." A "guideline strongly recommend[s] ... methods, actions, practices, or processes to consider in absence of applicable standards [or] best practices that [are] not required."

As currently drafted, this Standard does not prescribe "uniform methods, actions, practices, or processes, protocols." In addition, it does not address one crucial component that OSAC Standards should contain — advice on reporting and testifying. In theory, this component could be the subject of a separate Standard, but a more comprehensive Standard would be preferable, and there is no clear benefit to the community in placing an incomplete one, with no improvements whatsoever, on the registry.

Not Persuasive. This document is an ASTM Standard Guide. This document does not describe a prescriptive method, process or protocol. In agreement with the LRC comment, advice on reporting and testifying should be the subject of a separate document.

Therefore, detailed comments and suggestions for refining and expanding this draft follow. The first part lists concerns that should prompt significant revisions to achieve a document suitable for inclusion in the registry. The second part presents serious drafting issues that should be addressed if the subcommittee or the SAC concludes that the matters in Part I (or problems identified by other reviewers) warrant revising ASTM E 2548−11 rather than adopting it without change.

1. Concerns about Content

Section 4.1 does not adequately explain why "specific sampling strategies are not defined in this guide." (Also, the word "guide" may not be appropriate for a mandatory standard as defined in the QIC presentation.) If laws vary, specific strategies can be specified for at least a few, typical laws. Or, a rigorous strategy that would be acceptable everywhere can be presented as such — and made mandatory or designated a best practice (if it is indeed the best). Moreover, it is not obvious that the law in any U.S. jurisdiction precludes admitting probability sample data and extrapolations to population parameters. Some jurisdictions may require direct proof of total weight—no extrapolating—to establish beyond a reasonable doubt that total weight exceeds a given threshold when that quantity is an

essential element of the crime. E.g., State v. Robinson, 517 N.W.2d 336 (Minn. 1994). What sampling method should be used in these jurisdictions?

Not persuasive. This is not a document intended to present a prescriptive sampling strategy. Sampling schemes will vary and are based on jurisdictional requirements and specific case circumstances. With so many different variations throughout the US, finding a consensus strategy applicable to all jurisdictions would be difficult and most likely impossible. Even accreditation bodies do not mandate a certain plan be followed but rather that a plan followed by a laboratory meets certain criteria. It is very important to leave some flexibility as designating a specific sampling methodology will not work in some jurisdictions.

Section 4.2.2 states that "Statistically selected units shall be analyzed to meet Practice E2329 if statistical inferences are to be made about the whole population." E2329 governs the identification of material. Why would there be one set of requirements on how to identify material when an inference to a larger population is necessary and a different set of requirements when the analysis of a single unit is sufficient to prove what is necessary (e.g., "Are there any proscribed substances present?")? Furthermore, incorporating E2329 presumes that it is suitable for inclusion in the OSAC registry. The adequacy of E2329 needs to be established first. This issue recurs later in the proposed Standard. If a Standard is not self-contained, all its parts must be considered at once. This would entail providing the public with access to all the relevant ASTM documents.

Not persuasive. Section 4.2.2 quoted above has been misinterpreted by the LRC. Sampling and analysis are separate procedures in the forensic laboratory. Section 4.2.2 applies to multi-unit exhibit populations where not all units will be tested, but where the analysis conclusions will be stated in the form of a population inference. Section 4.2.2 simply states that the selected units should be analyzed as per E2329, in much the same way as a single-unit exhibit must also be analyzed as per E2329. The language quoted does not establish different identification requirements for the two different scenarios.

Reference to E2329 is unavoidable and the fact that it is not in the OSAC Registry yet is an unavoidable aspect of the process of implementing existing forensic standards in the OSAC Registry. E2329 is currently undergoing OSAC evaluation as well, but that should not be an impediment to the evaluation and inclusion of E2548.

The terminology within Figure 1 is confusing. It is sufficiently problematic that it rises to the level of a substantive rather than merely a stylistic concern. It seems like the middle level in the diagram is about the determination of sample size rather than a full "sampling plan." Thus, the terms "sampling procedure" and "sampling plan" should be defined. Doing so might answer questions (a)–(d) below. Clearer terminology comes from David Freedman, who wrote that "Methods for choosing samples are called 'designs.' Good designs involve the use of probability methods, minimizing subjective judgment in the choice of units to survey. Samples drawn using probability methods are called 'probability samples.'" David A. Freedman, Sampling, available on the Department of Statistics, University of California (Berkeley) website.

Not persuasive. The middle level of Figure 1 is not about determination of sample size. It establishes the foundation of a plan, as sampling plans can be statistical or not. The actual sample size is determined based on laboratory requirements and jurisdiction. For example, one laboratory may only need to

provide evidence that a controlled substance is present in at least 50% of the seizure population while a separate laboratory's jurisdiction may require proof of at least 90% of the population containing the drug. The required sample size would be different for these two laboratories.

a. Are there "hypergeometric sampling plans"? A sampling plan should specify the sample size (or a procedure for stopping the sampling if results on the sampled items up to that point make further testing unnecessary). For sampling a finite population without replacement, the hypergeometric probability distribution applies to sample-size computations. But how does that make the sampling plan hypergeometric? The simplest "plan," one would think, is to take a random sample of a size determined in light of the sample-size as computed with the hypergeometric distribution. The result is just a simple random sample.

Not persuasive. The quoted text "hypergeometric sampling plans" is not found in the document. Sampling plans based on the discrete hypergeometric sampling distribution are extensively reported in the statistical and forensic literature.

b. The fundamental distinction among the "statistical plans" on the far right of the figure is the divide between frequentist and Bayesian computations for sample size. Is not the frequentist approach always to use the hypergeometric distribution or a good approximation to it?

Agree. Most frequentist sampling procedures currently used throughout seized drugs laboratories are based on the hypergeometric distribution. The intent of the figure is not to limit the list to just one approach, but also to inform the user of the document as to other approaches available.

c. Suppose there is a "judicial requirement" to use a fixed percentage of the finite population as a sample size. (It might also be statutory or administrative.) Why isn't this statistical method of fixing the sample size a "statistical plan"? It is not a good idea, but it is part of a plan that uses a statistic.

Not persuasive. The essence of a statistical sampling plan is that a probability-based approach must be used for sample selection. This implies the use of random sampling. A laboratory may choose to sample a fixed number of items. However, the statistical inference on the population is only appropriate if random sampling has occurred. Simply stating a sample size (a number) does not make a selection or plan statistical.

d. Is the √N "plan" the old √N + 1 rule of thumb for determining sample size? Why should it ever be used as part of a sampling plan? See J. Muralimanohar & K. Jaianan, Determination of Effectiveness of the 'Square Root of N Plus One' Rule in Lot Acceptance Sampling Using an Operating Characteristic Curve, Quality Assurance Journal, 14(1-2): 33–37, 2011. This sample size is not based on a probability calculation, but why does that make it and the subsequent sampling "non-statistical"? If probability sampling has been conducted, statistical inferences and estimates have the same meaning regardless of how the fixed sample size was determined.

See previous response. The listed sample selection rules ($\sqrt{N}$ and $\sqrt{N} + 1$) are often used by laboratories in situations where a probability-based inference on the population is not needed. Random sampling is

not required in these situations, making these procedures non-statistical. This is very typical for laboratories doing analysis under specific jurisdictional weight-threshold penalties.

Section 5.5.1 speaks of "The probability that a given percentage of the population contains the drug of interest or is positive for a given characteristic." The Standard should make it clear that this posterior probability of a population parameter cannot be computed with any of the frequentist methods mentioned in the Standard.

Not persuasive. This comment appears to originate from a misinterpretation of the statistical language employed. The quoted text is the standard sampling hypothesis tested via frequentist sampling methods based on the hypergeometric probability distribution.

The *Guidelines on Representative Drug Sampling* noted in Section 5.5.1.2(2) contains errors corrected in the later version adopted by UNOCD in 2009. The newer document could be cited in Section 5.6.2.2 as well, since it discusses so-called "black box" methods (which are one way to draw a probability sample) more clearly than this Standard does.

Editorial.

Section 5.5.2.2 seems to say that whenever management directs it, scientific sampling can be abandoned. That is not what is intended, and the Standard should offer more guidance on when to depart from scientific methods.

Not persuasive. Section 5.5.2.2 states "Selection of a single unit from a multiple unit population may be appropriate under certain circumstances (for example, management directives and legislative or judicial requirements, or both)." There is no language suggesting users to "abandon scientific sampling." Section 5.5.2.2 is meant to address cases where it may be enough for a laboratory to simply analyze one unit. This may be due to legal, safety, and other reasons. In this case, no inference need be made as to the contents of the population.

Section 5.6.2 does not actually define the commonly misunderstood term "random sample," and this may not be the best term to use. Sometimes, "random sample" denotes a sample drawn by simple random sampling (SRS), but SRS is just the simplest form of probability sampling. The key distinction is between probability sampling, however it is achieved, and convenience sampling. Is the section intended to recommend (or require) probability sampling unless impractical? That would be good, and the section also ought to give guidance — at least an example — on when probability sampling is not practical.

Not persuasive. The intent of section 5.6.2 is to describe the selection of a random sample, not to define it. For clarification, simple random sampling (SRS) is only one type of sampling, as are systematic and stratified random sampling. On the other hand, convenience (aka opportunity) sampling is only one type of arbitrary sampling. Others examples include judgement, quota and snowball sampling. The intent of this section is not to recommend (or require) probability sampling. Whether a laboratory chooses to implement a sampling plan with the possibility for population inference is their choice, based on accreditation and jurisdictional requirements.

Section 7.1 states that "Inferences based on use of a sampling plan and concomitant analysis shall be documented." I am not sure what "concomitant analysis" means, since the analysis of the units normally will come after the sample is drawn.

Not persuasive. The adjective is fitting as written.

The section on reporting needs more body. How should the results of the analyses be presented in a report and in court? Should not the full sampling plan be stated — the mechanism for drawing samples (e.g., blinded, which is to say, "black box" sampling, or selecting from numbered samples by a table of random numbers); the sample size, and the kind of sampling (SRS, stratified, etc.)? It is not enough merely to state that "the results are based on a sampling plan," as Section 8.1.1 seems to contemplate.

Editorial. The subject of reporting will be addressed in a future Seized Drug subcommittee document.

The results themselves always should be stated as in the example in Section 8.1.2 — "2 of 100 bags were analyzed and found to contain Cocaine." When probability sampling has been employed, more can be said. But what would this be? A confidence interval? With what confidence coefficient? With what definition of "confidence"? A frequentist test of a hypothesis? Explained how? A Bayesian conclusion such as "There is a probability of 90% that the weight of the cocaine in the shipment seized exceeds X" (if the analysis is Bayesian and the prior distribution and the basis for it are presented)? To be of the most benefit to analysts, lawyers, and courts, the Standard should do more to spell out the "[s]ampling information [that] shall be included in reports."

Editorial. The subject of reporting will be addressed in a future Seized Drug subcommittee document. The quote text should not be interpreted as an exclusively Bayesian conclusion, as it is also the routine hypothesis tested using a hypergeometric probability based sampling procedure.

2. Drafting Problems

Section 1.2 states that "This guide cannot replace knowledge, skill, or ability acquired through appropriate education, training, and experience and should be used in conjunction with sound professional judgment." Either this goes without saying or it is dangerously broad. It might be misconstrued to mean that someone ostensibly conducting probability sampling, for example, can depart from the sampling plan so as to exercise a putative skill to pick the sample units based on experience and professional judgment. That would vitiate the protection probability sampling offers against bias and the value of probability sampling in enabling an estimate of precision.

Not persuasive. This is standard language in ASTM documents. It is intended to avoid use of this document by someone that does not have the appropriate education, training, etc.

Section 3.3 states that "By developing a sampling strategy and implementing appropriate sampling schemes, as illustrated in Fig. 1, a laboratory will minimize the total number of required analytical determinations, while ensuring that all relevant legal and scientific requirements are met." I think this is intended to say that "By developing and implementing a suitable sampling strategy, as illustrated in Fig.

1, a laboratory can reduce the number of analytical determinations required to meet legal and scientific requirements for producing reliable and unbiased evidence of the relevant features of a population."

Not persuasive. There isn't a discernable difference in how each of these is phrased.

Why does section 4.2 merely "recommend" that the "key points be addressed"? If these points are key, they must be addressed.

Editorial. Will be address via SDO review.

Section 4.2.1 states that "Sampling may be statistical or non-statistical." Would it be clearer to state that "Whether probability sampling is essential depends on the quantity to be estimated or determined and the nature of the population from which a sample is drawn"? The terms "statistical sampling" and "non-statistical sampling" ought to be defined. Indeed, there is a good argument that they should be abandoned, and that "probability sampling" and "convenience sampling" should replace them.

Editorial. The suggested alternate language will be evaluated through SDO review.

Not persuasive. Convenience sampling is only one type of non-statistical sampling, so its use would be incorrect.

Section 4.2.1.2 states that for inferring population characteristics from a sample, "the plan shall be statistically based and limits of the inference documented." The intended meaning could be clarified. Is systematic sampling "statistically based"? Or is "statistically based sampling" limited to probability sampling? What does "limits of the inference" mean? An interval estimate? Conditional error probabilities for a hypothesis test?

Not persuasive. Systematic sampling is a variation of simple random sampling (SRS), and therefore, allows for a population inference if employed. The general intent of this section is that only if samples are taken randomly and appropriately may an inference be made to a population, and the limits of the inference must be documented. The limits of the inference include clearly stating what was analyzed, the inferences drawn from such testing and the associated measurement uncertainty. Measurement uncertainty is generally expressed as an expanded uncertainty and includes the coverage probability (i.e. percent level of confidence).

For Section 5.2 and the sections under the heading "SAMPLING PLAN," is "population" defined in the Standard? What is an "exhibit"? What is a "unit"? The OECD and the International Statistical Institute define "unit" as follows: "A sampling unit is one of the units into which an aggregate is divided for the purpose of sampling, each unit being regarded as individual and indivisible when the selection is made." https://stats.oecd.org/glossary/detail.asp?ID=2381. If the population consists of a single unit, why speak of sampling?

In Section 5.4, what is a "bulk population"?

Not persuasive. These are commonly used terms in forensic and testing laboratories.

David Kaye
Legal Resource Committee

The following members of the LRC agree with comments made by David Kaye: Barry Scheck, Jennifer Friedman, David Moran and Ron Reinstein.

Comment by LRC Member Jennifer Friedman:

Comments from Jennifer Friedman on Proposed Standard E2548 Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis:

I do not believe this standard is ready to be published in the OSAC registry. My reasons for this opinion are set forth below.

1.2 This is a proposed standard therefore it should be described as a standard not a practice. Additionally, this seems to suggest that it would be appropriate not be follow the standard if the analysts "experience, education, or training" dictated it should not be followed.

Not persuasive. This is an ASTM Standard Guide, not a Practice as quote above; and these classifications are based on ASTM procedures.

The partial quote is extracted from standard language in ASTM documents. It is intended to avoid use of this document by someone that does not have the appropriate education, training, etc.

4.2.1 states sampling may be statistical or non-statistical. Both terms need to be clearly defined. The definition offered for statistical "probability-based" does not sufficiently define the term and there is no definition offered for non-statistical. In section 5.5.1.1 examples of statistical methods are listed. 5.5.2.1 gives examples of non-statistical methods. If this document is standard, there should be a list of which methods have been validated and thus may be employed. It should not be left to the examiner to decide.

Not persuasive. The terms statistical is defined in Note 1 within section 4.2.1. It is to be interpreted that all other situations are otherwise non-statistical.

Not persuasive. These are not testing methods. Sampling methods are based on established statistical and probability principles that allow for conclusions with a reasonable level of confidence.

4.2.2 includes by reference E2329. This may present a problem if E2329 is not incorporated into the OSAC registry,

Previously considered and discussed above.

5.6.2.2 states "random sampling of items using random table may not be practical in all cases. In these instances, an alternate sampling plan shall be designed and documented to approach random selection." This section should detail under what circumstances random sampling is not appropriate and set forth what method should be used when this occurs.

Not persuasive. This is a general document. Evidence heterogeneity from case to case and throughout jurisdictions makes it unrealistic to list all circumstances where either random or nonrandom sampling should be employed.

6.1 and 6.2 again reference E2329 which could be problematic for the reason described above.

Previously considered.

7 & 8 need to clearly set out how the results should be reported, what conclusions the analyst may make if random sampling is employed, what the limits of the opinion are so that the fact-finder may be able to appropriately assess the weight of the opinion.

Previously considered.

Jennifer Friedman
Legal Resource Committee

The following members of the LRC agree with comment made by Jennifer Friedman: Barry Scheck, David Kaye

---

Comment by LRC Member Christopher Plourd:

A probability based sampling approach for the forensic analysis of multiple containers of seized drugs is being proposed by the Seized Drugs subcommittee of the Chemistry-Instrumental Scientific Area Committee ("SAC") as an Organization of Forensic Scientific Area Committee (hereinafter "OSAC") approved standard.

This standard is identified and described as "E2548-11 *Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis".* A statistical-based chemical analytical sampling approach for seized drugs in containers should be designed to answer questions of both content weight and illicit drug identity. The standard describes a sampling method that has been used and is well understood in the forensic scientific community. (See: *Sampling of Street Drug Exhibits*, **Journal of Forensic Sciences,** 1993, 38(3): 641-648; *How many samples from a drug seizure need to be analyzed*? Journal of Forensic Sciences [2001, 46(6):1456-1461; *Bayesian Adaptive approach to estimating Sample Sizes for Seizures of Illicit Drugs*; Journal of Forensic Sciences [2012, 57(1):80-85]

From a legal perspective one critical question asked is: does existing scientific research support the E2548-11 standard? It appears that E2548-11 on the basis of the information in the Standard, the literature cited and my independent research, the technique demonstrates general-acceptance within the relevant scientific community. The subject of the standard, having been in use without any scientific

controversy for over three decades, is not a new or novel technique nor is it a novel application of an older technique or method. Therefore the E2548-11 standard would not likely be the subject of a Frye challenge. [Frye v. United States, 293 F. 1013 (D.C. App. 1923)] The directives of the standard, for the most part, are specific and clear as to what they require and or recommend.

The underlying scientific probability statements, in my opinion, are valid. The E2548-11 standard references sampling strategies and notes that the sampling procedures are divided into statistical and non-statistical. Statistical procedures (presumably hypergeometric, Bayesian, and other probability-based approaches) are to be developed by the individual laboratory. If a population opinion is given from samples tested, then the plan shall be statistically based and documented. ("4.2.1.2 If an inference about the whole population is to be drawn from a sample, then the plan shall be statistically based and limits of the inference shall be documented.") Expert testimony based on the use of this standard therefore would be based on matter that is of a type that reasonably may be relied upon by an expert in forming an opinion upon the subject to which his testimony relates.

Error rates associated with the techniques are not a concern if the E2549 standard is applied as recommended. Uncertainty associated with measurements obtained by using the standard techniques as prescribed are described and have been the subject of publications in the scientific literature.("4.2.2 Statistically selected units shall be analyzed to meet Practice E2329 if statistical inferences are to be made about the whole population.")

The techniques used by a forensic chemist using the E2548-11 standard would in my opinion not be impermissibly subjective or speculative. Any proffered testimony by an qualified expert from a laboratory that adopts and uses the E2548-11 standard who testifies as an expert witness, should be admissible under the authority of Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 113 S.Ct. 1786, 125 L.Ed.2d 469 (1993). The rigorous *Daubert* admissibility rule imposes a special gatekeeping obligation on the trial court to ensure the reliability of all expert testimony. Specifically, several procedural and substantive limitations upon the admission of expert scientific testimony are in place with *Daubert* to ensure that unreliable expertise would be excluded from the jury's consideration. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993) and *Kumho Tire Co., LTD., v. Carmichael*, 526 U.S. 137 (1999). Either under Frye or Daubert, reliable evidence should be heard by the Trier of fact under a weight of the evidence evaluation. The key question of admissibility of expert testimony under Frye/Daubert is reliability and relevance. If the methods are reliable, the theory is typically "generally accepted" in the scientific community. Use of the E2548-11 standard seems to meet the reliability test.

Of legal use and importance the E2548-11 standard address's due process-disclosure concerns because the standard requires documentation of all underlying data and assumptions.

There are no adverse appellate court rulings on the admissibility of the techniques described in the E2548-11 standard. Use of the E2548-11 as an OSAC approved standard would be expected to meet or exceed any legal admissibility requirements in jurisdictions within the United States.

Subject to an evaluation of the concerns raised by David Kaye and Ted Hunt I would recommend that the Chemistry-Instrumental Scientific Area Committee ("SAC") Approve the E2548-11standard as an Organization of Forensic Scientific Area Committee (hereinafter "OSAC") approved standard.

Christopher J. Plourd
Legal Resource Committee

The following LRC members agree with comments made by Christopher Plourd: Ron Reinstein, Lynn Garcia and Ray Miller.

Comments made by LRC member Barry Scheck:

I am in agreement with all of David Kaye's comments as well as the comments of Jennifer Friedman. I am grateful to my colleagues for their comprehensive analysis.

I would like to emphasize concerns expressed in these comments that the deficiencies in the statistical explanations offered in all of these standards is troubling and not ready for court, whether one is in a Frye or Daubert jurisdiction. These should be rejected from the OSAC Registry and, hopefully, the OSAC subcommittee and/or ASTM will revise the proposed standards to follow the template laid out in the Technical Merit Worksheets. In that connection, I cannot imagine that the requirement of general acceptance in the scientific community, particularly among statisticians, can be met, nor the requirements of clearly identifying limitations and weaknesses in the methodology or an explanation of how it is "fit for purpose."

Not persuasive. Not sure what is referred to as "deficiencies in the statistical explanations offered". There are no statistical explanations offered in this document. The general sampling guidelines discussed in this document are statistically based, when applicable, and have been used and nationally and internationally accepted by seized drug laboratories and courts for many years. Their fitness for purpose has therefore been demonstrated.

It is unrealistic to expect previously vetted and published ASTM documents to fulfill the requirements just recently established by OSAC, as they are completely separate institutions.

This comment does not offer any specific explanation as to how this document does not meet "the requirement of general acceptance in the scientific community, particularly among statisticians". It appears the document has been misinterpreted as describing a particular methodology.

Barry Scheck
Legal Resource Committee

Comments made by LRC member Ted Hunt:

GENERAL COMMENTS

David Kaye correctly observed that the July 27, 2015, QIC presentation stated that documentary standards and guidelines must demonstrate "technical merit," which includes "detailed scope," "fitness

for purpose," "uncertainty measurement and potential bias," and "method validation, as appropriate" in order to be included on the OSAC Registry.

The same presentation also stated that a standard, "specifies uniform methods, actions, practices, or processes, protocols," and that "compliance [is] recommended to be mandatory and modified only under unusual circumstances."

There is no question that this document fails to meet those requirements. This, however, is less a reflection on the substantive merit of E2548 - 11 (given its real purpose) than the present failure of the OSAC to strike a clear distinction between "standardized methods" and "consensus documentary standards" for purposes of the technical merit requirements set forth for OSAC standards and guidelines.

The only verbiage that seemed to separate these two *distinct* types of "standards" in the presentation was the inclusion of the words, "as appropriate." This cryptic reference is insufficient guidance with which to de-conflate which type of standard (high-level consensus practice standard vs. standardized method) is under consideration, and gives no guidance at all about the OSAC-required components of this second type of standard — a high-level consensus practice standard — that sets forth minimum requirements for labs to incorporate within their internally validated analytical procedures for a given technology and/or method.

There is no question that E2548 - 11 is not a "standardized method." Its scope is not detailed, its "purpose" merely states that it covers "minimum considerations" for "sampling of seized drugs for qualitative and quantitative analysis"; uncertainty of measurement is not specifically addressed; sampling bias is only addressed with the single statement, "A random sample is one selected without bias"; and method validation is not addressed at all — but that is simply because E2548 - 11 is not a "method."

Rather, E2548 - 11 appears to be a "consensus documentary standard" which — as other documentary standards do — addresses not *how* a particular laboratory's sampling method *shall* be performed, but rather *that* each laboratory engaged in sampling *shall* follow the "minimum considerations" set forth for sampling "seized drugs for qualitative and quantitative analysis."

Unfortunately, however, there is currently no OSAC guidance (that I know of) on the minimum requirements for this type of "standard" (consensus documentary standard) — other than the "as appropriate" language noted above.

Given the present absence of a distinction between "consensus documentary standards" and "standardized methods" in the OSAC nomenclature, this "standard" fails a test it was never originally designed to take.

The OSAC should remedy this omission by specifically recognizing the distinction between these two distinct types of "standards" and setting forth separate criteria by which high-level consensus documents — such as this one — are to be judged.

SPECIFIC COMMENTS

My comments above notwithstanding, E2548 - 11 can be improved upon as a "consensus documentary standard."

1. Given its proposed placement in the OSAC Registry, I think the title, "Standard Guide for Sampling ...." is confusing. The title hedges by using both the terms "standard" and "guide." Further, the document, in places, reads like a “guideline” rather than a “standard” by using the term "recommended" (4.2), (5.6.1.2), (6.1); "should" (5.6.2.1); and "may" (4.2.1.1), (5.4.2), (5.5), (5.5.2.2).

Not persuasive. This is the ASTM verbiage.

In other places, the document reads like a standard by using the term "shall" (4.2.1.2), (4.2.2), (5.2.3), (5.4), (5.6.2.2), (6.2), (7.1), (8.1); and "must" (8.1.1) (8.1.2).

The combination of this directive and permissive verbiage in the document should be re-examined if E2548-11 will go forward as a "standard."

2. 5.6.2.1 states that "For statistical approaches it is recommended that random sampling be conducted." If a statistical approach is probabilistic in nature, and is meant to support inferences drawn regarding an entire population, I don't see how sampling could validly be conducted by anything other than an appropriately chosen and appropriately random method — so I'm a bit confused by the permissive word, "recommended."

Editorial. The Seized Drugs subcommittee will request revision of the language through the SDO process. For clarification, the quoted text is found in section 5.6.1.2, not 5.6.2.1 as noted above.

3. 7.1 states, "Inferences based on the use of a sampling plan and concomitant analysis should be documented." The use of the term "concomitant" here is confusing because sampling and analysis will not literally be conducted in a "concomitant" manner — and if they were that would present its own set of problems. The word "subsequent" instead of "concomitant" seems more appropriate here.

Not persuasive. Concomitant means “associated, parallel, related, etc.”, which is the intended meaning here. The analysis is subsequent to each sampling selection, but not necessarily the complete sampling action. Concomitant is a more general term and is appropriate for this purpose.

Also, consider adding the requirement that "key assumptions" which are incorporated into the sampling plan constructed by the lab should be documented as well.

4. Consider redrafting 5.5.2.2 to make it clear (consistent with David Kaye's recommendation) that this language does not mean that the otherwise scientifically valid and reliable selection of a single unit from a multiple unit population should yield to management directives, legislative, or judicial requirements.

Previously considered. This section merely states that a laboratory may choose to sample as few as one unit from a population, but that choice comes with limits to making an inference to the population.

5. Where key "terms of art" are used within this document and are currently defined in another existing standard, then consider some direct reference to that standard (where those terms *are* defined) within the present document when the terms of art are used. If those terms are not defined within an existing standard, they should be added to the OSAC list of definitions before the standard is included on the OSAC Registry.

6. Perhaps 1.2 could be rephrased to state that this standard should be used *in conjunction with* the analyst's knowledge, skill, or ability acquired through appropriate education, training, and experience and sound professional judgment — rather than that it cannot "replace" those attributes. The guidance set forth in this document — as with all scientific standards and procedures — will always *necessarily* be executed concurrently with the judgment and discretion of a trained and experienced analyst. Neither the expert nor the method can exist or function without the other when producing scientifically sound and reliable results.

Previously considered.

Ted Hunt
Legal Resource Committee

The following LRC members agree with comments made by Ted Hunt: David Kaye and Ron Reinstein.

DISCLAIMER: ***The failure of any member of the Legal Resource committee (LRC) to provide a comment, identify a legal issue or join in another LRC comment should not be interpreted as a disagreement or endorsement of the comment, the standard or its legal sufficiency.***

COMMENTS BY THE OSAC LEGAL RESOURCE COMMITTEE (LRC)

TO: Seized Drugs subcommittee of the Chemistry-Instrumental SAC

FROM: Christopher J. Plourd, Chair, OSAC Legal Resource Committee (LRC)

RE: **OSAC LEGAL RESOURCE COMMITTEE (LRC) COMMENTS ON:** E2548-11 "Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis" (hereinafter referred to as "E2548-11").

Response from the Seized Drugs Subcommittee

The OSAC Seized Drugs subcommittee recognizes and appreciates the comments received from the Legal Resource Committee (LRC) pertaining ASTM E2548-11, Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis. The following responses attempt to address and clarify some of the issues brought to our attention by the LRC.

The document in review (E2548-11) was original published by ASTM in 2007. The document was originally published as Part IIIA of the Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) Recommendations, which are intended to assist forensic analysts and managers in the development of analytical techniques, protocols and policies. The SWGDRUG Recommendations are internationally recognized as minimum standards that may be supplemented to address unique jurisdictional laboratory requirements.

The document under consideration is neither a test method nor a prescriptive standard. Therefore, many of the comments provided by the LRC members are not considered applicable. Also, this document is not intended to encompass all other related standards, terminology, validation documentation, etc. that may assist in its application. The field of seized drug analysis is an extensive one encompassing the subjects of sampling, chemical identification, presumptive and confirmatory analytical techniques, method validation, quantitative procedures, structure elucidation, measurement uncertainty, reporting protocols, and many others. It is unrealistic and impractical to attempt to include all these subjects into one single document, as the result would be an ineffective standard that would not be useful for practitioners. Efforts are already underway to address some of the aforementioned subjects via the publication of separate documents on the OSAC Registry.

There appears to be some misconceptions and misunderstandings regarding the OSAC process as well as the process in use by the standard development organization (SDO) under which this standard is published (ASTM). We believe that it would be useful to provide additional training regarding these procedures to not only members of the LRC but also members of other resource committees and individual discipline subcommittees.

We understand the desire for these documents to be written such that they are more comprehensible to lawyers and judges. However, many of these documents under review have already been in the forensic community for over ten years and they were originally drafted with the goal of providing useful minimum standards for seized drug analysts. In fact, there are many seized drug laboratories throughout the world that use this document (as well as the SWGDRUG Recommendations) as a foundation for their policies and procedures. It is our strong belief that attorneys and judges have the ability to consult scientists in order to interpret a scientific document, much in the same way as a scientist would consult an attorney to interpret a legal document. Serious consideration is being given to all the comments provided by the LRC members. However, we believe that making significant changes to the document in order to address the resource committee's would take away too much from the original technical and scientific intent of the document.

The Seized Drugs subcommittee of the Chemistry-Instrumental SAC is proposing to approve an existing ASTM E2548-11 standard to the OSAC Repository registry.

Our comments are primarily intended to enhance the value of the Standard to the legal community. This Standard will be most helpful if it not only helps assure high quality results in the laboratory, but also is written to show how work performed in accordance with the Standard is both well grounded in theory and data and that it is presented within the boundaries of "the knowledge and experience of [the expert's] discipline."[1] Consequently, the comments are intended to address four questions that are important to the legal reception of the Standard: (1) Is the Standard written as clearly as possible, and without undefined technical terms and symbols, so as to enable lawyers and judges to grasp the main ideas and requirements set forth? (2) Does the Standard describe in detail how the peer-reviewed and readily available scientific literature establishes the validity of the assumptions underlying the scientific tests and the interpretation of test results? (3) Does the Standard list the limitations of the tests and results and provide for expressions of the uncertainties in measurements and inferences drawn from them? (4) Does the Standard include recommendations or requirements for the creation and retention of documentation of the test and the contents of reports, including the scientific limitations of the tests and related conclusions or inferences? These are matters of both

[1] Kumho Tire Co. v. Carmichael, 526 U.S. 137, 148 (1999) (quoting Daubert v. Merrell Dow Pharms, Inc., 509 U.S. 579, 592 (1993)).

technical merit and legal importance. Although the LRC is not able to assess the scientific merit of a Standard, our review encompasses whether a Standard makes a prima facie case for the validity of the methods and legal utility of the kinds of expert opinions that a Standard contemplates.

Comment by LRC member David Kaye:

Comments of David Kaye on Placing ASTM E 2548−11 ("Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis") on the OSAC Repository of Standards and Guidelines.

The document is not ready for inclusion as an OSAC Standard. According to the QIC PowerPoint presentation of July 27, 2015, OSAC-approved documentary standards and guidelines are to "have demonstrated: Technical merit; Detailed Scope; Fitness for purpose; Uncertainty measurement and potential bias; and Method validation, as appropriate." A "standard ... specifies uniform methods, actions, practices, or processes, protocols. Compliance [is] mandatory and modified only under unusual circumstances." A "guideline strongly recommend[s] ... methods, actions, practices, or processes to consider in absence of applicable standards [or] best practices that [are] not required."

As currently drafted, this Standard does not prescribe "uniform methods, actions, practices, or processes, protocols." In addition, it does not address one crucial component that OSAC Standards should contain — advice on reporting and testifying. In theory, this component could be the subject of a separate Standard, but a more comprehensive Standard would be preferable, and there is no clear benefit to the community in placing an incomplete one, with no improvements whatsoever, on the registry.

Not Persuasive. This document is an ASTM Standard Guide. This document does not describe a prescriptive method, process or protocol. In agreement with the LRC comment, advice on reporting and testifying should be the subject of a separate document.

Therefore, detailed comments and suggestions for refining and expanding this draft follow. The first part lists concerns that should prompt significant revisions to achieve a document suitable for inclusion in the registry. The second part presents serious drafting issues that should be addressed if the subcommittee or the SAC

concludes that the matters in Part I (or problems identified by other reviewers) warrant revising ASTM E 2548−11 rather than adopting it without change.

1. Concerns about Content

Section 4.1 does not adequately explain why "specific sampling strategies are not defined in this guide." (Also, the word "guide" may not be appropriate for a mandatory standard as defined in the QIC presentation.) If laws vary, specific strategies can be specified for at least a few, typical laws. Or, a rigorous strategy that would be acceptable everywhere can be presented as such — and made mandatory or designated a best practice (if it is indeed the best). Moreover, it is not obvious that the law in any U.S. jurisdiction precludes admitting probability sample data and extrapolations to population parameters. Some jurisdictions may require direct proof of total weight—no extrapolating—to establish beyond a reasonable doubt that total weight exceeds a given threshold when that quantity is an essential element of the crime. E.g., State v. Robinson, 517 N.W.2d 336 (Minn. 1994). What sampling method should be used in these jurisdictions?

Not persuasive. This is not a document intended to present a prescriptive sampling strategy. Sampling schemes will vary and are based on jurisdictional requirements and specific case circumstances. With so many different variations throughout the US, finding a consensus strategy applicable to all jurisdictions would be difficult and most likely impossible. Even accreditation bodies do not mandate a certain plan be followed but rather that a plan followed by a laboratory meets certain criteria. It is very important to leave some flexibility as designating a specific sampling methodology will not work in some jurisdictions.

Section 4.2.2 states that "Statistically selected units shall be analyzed to meet Practice E2329 if statistical inferences are to be made about the whole population." E2329 governs the identification of material. Why would there be one set of requirements on how to identify material when an inference to a larger population is necessary and a different set of requirements when the analysis of a single unit is sufficient to prove what is necessary (e.g., "Are there any proscribed substances present?")? Furthermore, incorporating E2329 presumes that it is suitable for inclusion in the OSAC registry. The adequacy of E2329 needs to be established first. This issue recurs later in the proposed Standard. If a Standard is not self-contained, all its parts

must be considered at once. This would entail providing the public with access to all the relevant ASTM documents.

Not persuasive. Section 4.2.2 quoted above has been misinterpreted by the LRC. Sampling and analysis are separate procedures in the forensic laboratory. Section 4.2.2 applies to multi-unit exhibit populations where not all units will be tested, but where the analysis conclusions will be stated in the form of a population inference. Section 4.2.2 simply states that the selected units should be analyzed as per E2329, in much the same way as a single-unit exhibit must also be analyzed as per E2329. The language quoted does not establish different identification requirements for the two different scenarios.

Reference to E2329 is unavoidable and the fact that it is not in the OSAC Registry yet is an unavoidable aspect of the process of implementing existing forensic standards in the OSAC Registry. E2329 is currently undergoing OSAC evaluation as well, but that should not be an impediment to the evaluation and inclusion of E2548.

The terminology within Figure 1 is confusing. It is sufficiently problematic that it rises to the level of a substantive rather than merely a stylistic concern. It seems like the middle level in the diagram is about the determination of sample size rather than a full "sampling plan." Thus, the terms "sampling procedure" and "sampling plan" should be defined. Doing so might answer questions (a)–(d) below. Clearer terminology comes from David Freedman, who wrote that "Methods for choosing samples are called 'designs.' Good designs involve the use of probability methods, minimizing subjective judgment in the choice of units to survey. Samples drawn using probability methods are called 'probability samples.'" David A. Freedman, Sampling, available on the Department of Statistics, University of California (Berkeley) website.

Not persuasive. The middle level of Figure 1 is not about determination of sample size. It establishes the foundation of a plan, as sampling plans can be statistical or not. The actual sample size is determined based on laboratory requirements and jurisdiction. For example, one laboratory may only need to provide evidence that a controlled substance is present in at least 50% of the seizure population while a separate laboratory's jurisdiction may require proof of at least 90% of the population containing the drug. The required sample size would be different for these two laboratories.

a. Are there “hypergeometric sampling plans”? A sampling plan should specify the sample size (or a procedure for stopping the sampling if results on the sampled items up to that point make further testing unnecessary). For sampling a finite population without replacement, the hypergeometric probability distribution applies to sample-size computations. But how does that make the sampling plan hypergeometric? The simplest “plan,” one would think, is to take a random sample of a size determined in light of the sample-size as computed with the hypergeometric distribution. The result is just a simple random sample.

Not persuasive. The quoted text “hypergeometric sampling plans” is not found in the document. Sampling plans based on the discrete hypergeometric sampling distribution are extensively reported in the statistical and forensic literature.

b. The fundamental distinction among the “statistical plans” on the far right of the figure is the divide between frequentist and Bayesian computations for sample size. Is not the frequentist approach always to use the hypergeometric distribution or a good approximation to it?

Agree. Most frequentist sampling procedures currently used throughout seized drugs laboratories are based on the hypergeometric distribution. The intent of the figure is not to limit the list to just one approach, but also to inform the user of the document as to other approaches available.

c. Suppose there is a “judicial requirement” to use a fixed percentage of the finite population as a sample size. (It might also be statutory or administrative.) Why isn’t this statistical method of fixing the sample size a “statistical plan”? It is not a good idea, but it is part of a plan that uses a statistic.

Not persuasive. The essence of a statistical sampling plan is that a probability-based approach must be used for sample selection. This implies the use of random sampling. A laboratory may choose to sample a fixed number of items. However, the statistical inference on the population is

only appropriate if random sampling has occurred. Simply stating a sample size (a number) does not make a selection or plan statistical.

d. Is the √N "plan" the old √N + 1 rule of thumb for determining sample size? Why should it ever be used as part of a sampling plan? See J. Muralimanohar & K. Jaianan, Determination of Effectiveness of the 'Square Root of N Plus One' Rule in Lot Acceptance Sampling Using an Operating Characteristic Curve, Quality Assurance Journal, 14(1-2): 33–37, 2011. This sample size is not based on a probability calculation, but why does that make it and the subsequent sampling "non-statistical"? If probability sampling has been conducted, statistical inferences and estimates have the same meaning regardless of how the fixed sample size was determined.

See previous response. The listed sample selection rules (√N and √N + 1) are often used by laboratories in situations where a probability-based inference on the population is not needed. Random sampling is not required in these situations, making these procedures non-statistical. This is very typical for laboratories doing analysis under specific jurisdictional weight-threshold penalties.

Section 5.5.1 speaks of "The probability that a given percentage of the population contains the drug of interest or is positive for a given characteristic." The Standard should make it clear that this posterior probability of a population parameter cannot be computed with any of the frequentist methods mentioned in the Standard.

Not persuasive. This comment appears to originate from a misinterpretation of the statistical language employed. The quoted text is the standard sampling hypothesis tested via frequentist sampling methods based on the hypergeometric probability distribution.

The *Guidelines on Representative Drug Sampling* noted in Section 5.5.1.2(2) contains errors corrected in the later version adopted by UNOCD in 2009. The newer document could be cited in Section 5.6.2.2 as well, since it discusses so-called "black box" methods (which are one way to draw a probability sample) more clearly than this Standard does.

Editorial.

Section 5.5.2.2 seems to say that whenever management directs it, scientific sampling can be abandoned. That is not what is intended, and the Standard should offer more guidance on when to depart from scientific methods.

Not persuasive. Section 5.5.2.2 states "Selection of a single unit from a multiple unit population may be appropriate under certain circumstances (for example, management directives and legislative or judicial requirements, or both)." There is no language suggesting users to "abandon scientific sampling." Section 5.5.2.2 is meant to address cases where it may be enough for a laboratory to simply analyze one unit. This may be due to legal, safety, and other reasons. In this case, no inference need be made as to the contents of the population.

Section 5.6.2 does not actually define the commonly misunderstood term "random sample," and this may not be the best term to use. Sometimes, "random sample" denotes a sample drawn by simple random sampling (SRS), but SRS is just the simplest form of probability sampling. The key distinction is between probability sampling, however it is achieved, and convenience sampling. Is the section intended to recommend (or require) probability sampling unless impractical? That would be good, and the section also ought to give guidance — at least an example — on when probability sampling is not practical.

Not persuasive. The intent of section 5.6.2 is to describe the selection of a random sample, not to define it. For clarification, simple random sampling (SRS) is only one type of sampling, as are systematic and stratified random sampling. On the other hand, convenience (aka opportunity) sampling is only one type of arbitrary sampling. Others examples include judgement, quota and snowball sampling. The intent of this section is not to recommend (or require) probability sampling. Whether a laboratory chooses to implement a sampling plan with the possibility for population inference is their choice, based on accreditation and jurisdictional requirements.

Section 7.1 states that "Inferences based on use of a sampling plan and concomitant analysis shall be documented." I am not sure what "concomitant analysis" means, since the analysis of the units normally will come after the sample is drawn.

Not persuasive. The adjective is fitting as written.

The section on reporting needs more body. How should the results of the analyses be presented in a report and in court? Should not the full sampling plan be stated — the mechanism for drawing samples (e.g., blinded, which is to say, "black box" sampling, or selecting from numbered samples by a table of random numbers); the sample size, and the kind of sampling (SRS, stratified, etc.)? It is not enough merely to state that "the results are based on a sampling plan," as Section 8.1.1 seems to contemplate.

Editorial. The subject of reporting will be addressed in a future Seized Drug subcommittee document.

The results themselves always should be stated as in the example in Section 8.1.2 — "2 of 100 bags were analyzed and found to contain Cocaine." When probability sampling has been employed, more can be said. But what would this be? A confidence interval? With what confidence coefficient? With what definition of "confidence"? A frequentist test of a hypothesis? Explained how? A Bayesian conclusion such as "There is a probability of 90% that the weight of the cocaine in the shipment seized exceeds X" (if the analysis is Bayesian and the prior distribution and the basis for it are presented)? To be of the most benefit to analysts, lawyers, and courts, the Standard should do more to spell out the "[s]ampling information [that] shall be included in reports."

Editorial. The subject of reporting will be addressed in a future Seized Drug subcommittee document. The quote text should not be interpreted as an exclusively Bayesian conclusion, as it is also the routine hypothesis tested using a hypergeometric probability based sampling procedure.

2. Drafting Problems

Section 1.2 states that "This guide cannot replace knowledge, skill, or ability acquired through appropriate education, training, and experience and should be used in conjunction with sound professional judgment." Either this goes without saying or it is dangerously broad. It might be misconstrued to mean that someone ostensibly

conducting probability sampling, for example, can depart from the sampling plan so as to exercise a putative skill to pick the sample units based on experience and professional judgment. That would vitiate the protection probability sampling offers against bias and the value of probability sampling in enabling an estimate of precision.

Not persuasive. This is standard language in ASTM documents. It is intended to avoid use of this document by someone that does not have the appropriate education, training, etc.

Section 3.3 states that "By developing a sampling strategy and implementing appropriate sampling schemes, as illustrated in Fig. 1, a laboratory will minimize the total number of required analytical determinations, while ensuring that all relevant legal and scientific requirements are met." I think this is intended to say that "By developing and implementing a suitable sampling strategy, as illustrated in Fig. 1, a laboratory can reduce the number of analytical determinations required to meet legal and scientific requirements for producing reliable and unbiased evidence of the relevant features of a population."

Not persuasive. There isn't a discernable difference in how each of these is phrased.

Why does section 4.2 merely "recommend" that the "key points be addressed"? If these points are key, they must be addressed.

Editorial. Will be address via SDO review.

Section 4.2.1 states that "Sampling may be statistical or non-statistical." Would it be clearer to state that "Whether probability sampling is essential depends on the quantity to be estimated or determined and the nature of the population from which a sample is drawn"? The terms "statistical sampling" and "non-statistical sampling" ought to be defined. Indeed, there is a good argument that they should be abandoned, and that "probability sampling" and "convenience sampling" should replace them.

Editorial. The suggested alternate language will be evaluated through SDO review.

Not persuasive. Convenience sampling is only one type of non-statistical sampling, so its use would be incorrect.

Section 4.2.1.2 states that for inferring population characteristics from a sample, "the plan shall be statistically based and limits of the inference documented." The intended meaning could be clarified. Is systematic sampling "statistically based"? Or is "statistically based sampling" limited to probability sampling? What does "limits of the inference" mean? An interval estimate? Conditional error probabilities for a hypothesis test?

Not persuasive. Systematic sampling is a variation of simple random sampling (SRS), and therefore, allows for a population inference if employed. The general intent of this section is that only if samples are taken randomly and appropriately may an inference be made to a population, and the limits of the inference must be documented. The limits of the inference include clearly stating what was analyzed, the inferences drawn from such testing and the associated measurement uncertainty. Measurement uncertainty is generally expressed as an expanded uncertainty and includes the coverage probability (i.e. percent level of confidence).

For Section 5.2 and the sections under the heading "SAMPLING PLAN," is "population" defined in the Standard? What is an "exhibit"? What is a "unit"? The OECD and the International Statistical Institute define "unit" as follows: "A sampling unit is one of the units into which an aggregate is divided for the purpose of sampling, each unit being regarded as individual and indivisible when the selection is made." https://stats.oecd.org/glossary/detail.asp?ID=2381. If the population consists of a single unit, why speak of sampling?

In Section 5.4, what is a "bulk population"?

Not persuasive. These are commonly used terms in forensic and testing laboratories.

David Kaye
Legal Resource Committee

The following members of the LRC agree with comments made by David Kaye:
Barry Scheck, Jennifer Friedman, David Moran and Ron Reinstein.

Comment by LRC Member Jennifer Friedman:

Comments from Jennifer Friedman on Proposed Standard E2548 Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis:

I do not believe this standard is ready to be published in the OSAC registry. My reasons for this opinion are set forth below.

1.2 This is a proposed standard therefore it should be described as a standard not a practice. Additionally, this seems to suggest that it would be appropriate not be follow the standard if the analysts "experience, education, or training" dictated it should not be followed.

Not persuasive. This is an ASTM Standard Guide, not a Practice as quote above; and these classifications are based on ASTM procedures.
The partial quote is extracted from standard language in ASTM documents. It is intended to avoid use of this document by someone that does not have the appropriate education, training, etc.

4.2.1 states sampling may be statistical or non-statistical. Both terms need to be clearly defined. The definition offered for statistical "probability-based" does not sufficiently define the term and there is no definition offered for non-statistical. In section 5.5.1.1 examples of statistical methods are listed. 5.5.2.1 gives examples of non-statistical methods. If this document is standard, there should be a list of which methods have been validated and thus may be employed. It should not be left to the examiner to decide.

Not persuasive. The terms statistical is defined in Note 1 within section 4.2.1. It is to be interpreted that all other situations are otherwise non-statistical.

Not persuasive. These are not testing methods. Sampling methods are based on established statistical and probability principles that allow for conclusions with a reasonable level of confidence.

4.2.2 includes by reference E2329. This may present a problem if E2329 is not incorporated into the OSAC registry,

Previously considered and discussed above.

5.6.2.2 states "random sampling of items using random table may not be practical in all cases. In these instances, an alternate sampling plan shall be designed and documented to approach random selection." This section should detail under what circumstances random sampling is not appropriate and set forth what method should be used when this occurs.

Not persuasive. This is a general document. Evidence heterogeneity from case to case and throughout jurisdictions makes it unrealistic to list all circumstances where either random or nonrandom sampling should be employed.

6.1 and 6.2 again reference E2329 which could be problematic for the reason described above.

Previously considered.

7 & 8 need to clearly set out how the results should be reported, what conclusions the analyst may make if random sampling is employed, what the limits of the opinion are so that the fact-finder may be able to appropriately assess the weight of the opinion.

Previously considered.

Jennifer Friedman
Legal Resource Committee

The following members of the LRC agree with comment made by Jennifer Friedman:
Barry Scheck, David Kaye

Comment by LRC Member Christopher Plourd:

A probability based sampling approach for the forensic analysis of multiple containers of seized drugs is being proposed by the Seized Drugs subcommittee of the Chemistry-Instrumental Scientific Area Committee ("SAC") as an Organization of Forensic Scientific Area Committee (hereinafter "OSAC") approved standard. This standard is identified and described as "E2548-11 *Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis".* A statistical-based chemical analytical sampling approach for seized drugs in containers should be designed to answer questions of both content weight and illicit drug identity. The standard describes a sampling method that has been used and is well understood in the forensic scientific community. (See: *Sampling of Street Drug Exhibits*, **Journal of Forensic Sciences,** 1993, 38(3): 641-648; *How many samples from a drug seizure need to be analyzed*? Journal of Forensic Sciences [2001, 46(6):1456-1461; *Bayesian Adaptive approach to estimating Sample Sizes for Seizures of Illicit Drugs*; Journal of Forensic Sciences [2012, 57(1):80-85]

From a legal perspective one critical question asked is: does existing scientific research support the E2548-11 standard? It appears that E2548-11 on the basis of the information in the Standard, the literature cited and my independent research, the technique demonstrates general-acceptance within the relevant scientific community. The subject of the standard, having been in use without any scientific controversy for over three decades, is not a new or novel technique nor is it a novel application of an older technique or method. Therefore the E2548-11 standard would not likely be the subject of a Frye challenge. [Frye v. United States, 293 F. 1013 (D.C. App. 1923)] The directives of the standard, for the most part, are specific and clear as to what they require and or recommend.

The underlying scientific probability statements, in my opinion, are valid. The E2548-11 standard references sampling strategies and notes that the sampling procedures are divided into statistical and non-statistical. Statistical procedures (presumably hypergeometric, Bayesian, and other probability-based approaches) are to be developed by the individual laboratory. If a population opinion is given from samples tested, then the plan shall be statistically based and documented. ("4.2.1.2 If an inference about the whole population is to be drawn from a sample, then the plan shall

be statistically based and limits of the inference shall be documented.") Expert testimony based on the use of this standard therefore would be based on matter that is of a type that reasonably may be relied upon by an expert in forming an opinion upon the subject to which his testimony relates.

Error rates associated with the techniques are not a concern if the E2549 standard is applied as recommended. Uncertainty associated with measurements obtained by using the standard techniques as prescribed are described and have been the subject of publications in the scientific literature.("4.2.2 Statistically selected units shall be analyzed to meet Practice E2329 if statistical inferences are to be made about the whole population.")

The techniques used by a forensic chemist using the E2548-11 standard would in my opinion not be impermissibly subjective or speculative. Any proffered testimony by an qualified expert from a laboratory that adopts and uses the E2548-11 standard who testifies as an expert witness, should be admissible under the authority of Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 113 S.Ct. 1786, 125 L.Ed.2d 469 (1993). The rigorous *Daubert* admissibility rule imposes a special gatekeeping obligation on the trial court to ensure the reliability of all expert testimony. Specifically, several procedural and substantive limitations upon the admission of expert scientific testimony are in place with *Daubert* to ensure that unreliable expertise would be excluded from the jury's consideration. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993) and *Kumho Tire Co., LTD., v. Carmichael*, 526 U.S. 137 (1999). Either under Frye or Daubert, reliable evidence should be heard by the Trier of fact under a weight of the evidence evaluation. The key question of admissibility of expert testimony under Frye/Daubert is reliability and relevance. If the methods are reliable, the theory is typically "generally accepted" in the scientific community. Use of the E2548-11 standard seems to meet the reliability test.

Of legal use and importance the E2548-11 standard address's due process-disclosure concerns because the standard requires documentation of all underlying data and assumptions.

There are no adverse appellate court rulings on the admissibility of the techniques described in the E2548-11 standard. Use of the E2548-11 as an OSAC

approved standard would be expected to meet or exceed any legal admissibility requirements in jurisdictions within the United States.

Subject to an evaluation of the concerns raised by David Kaye and Ted Hunt I would recommend that the Chemistry-Instrumental Scientific Area Committee ("SAC") Approve the E2548-11standard as an Organization of Forensic Scientific Area Committee (hereinafter "OSAC") approved standard.

Christopher J. Plourd
Legal Resource Committee

The following LRC members agree with comments made by Christopher Plourd: Ron Reinstein, Lynn Garcia and Ray Miller.

Comments made by LRC member Barry Scheck:

I am in agreement with all of David Kaye's comments as well as the comments of Jennifer Friedman. I am grateful to my colleagues for their comprehensive analysis. I would like to emphasize concerns expressed in these comments that the deficiencies in the statistical explanations offered in all of these standards is troubling and not ready for court, whether one is in a Frye or Daubert jurisdiction. These should be rejected from the OSAC Registry and, hopefully, the OSAC subcommittee and/or ASTM will revise the proposed standards to follow the template laid out in the Technical Merit Worksheets. In that connection, I cannot imagine that the requirement of general acceptance in the scientific community, particularly among statisticians, can be met, nor the requirements of clearly identifying limitations and weaknesses in the methodology or an explanation of how it is "fit for purpose."

Not persuasive. Not sure what is referred to as "deficiencies in the statistical explanations offered". There are no statistical explanations offered in this document. The general sampling guidelines discussed in this document are statistically based, when applicable, and have been used and nationally and internationally accepted by seized drug laboratories and courts for many years. Their fitness for purpose has therefore been demonstrated.

It is unrealistic to expect previously vetted and published ASTM documents to fulfill the requirements just recently established by OSAC, as they are completely separate institutions.

This comment does not offer any specific explanation as to how this document does not meet "the requirement of general acceptance in the scientific community, particularly among statisticians". It appears the document has been misinterpreted as describing a particular methodology.

Barry Scheck
Legal Resource Committee

Comments made by LRC member Ted Hunt:

GENERAL COMMENTS

David Kaye correctly observed that the July 27, 2015, QIC presentation stated that documentary standards and guidelines must demonstrate "technical merit," which includes "detailed scope," "fitness for purpose," "uncertainty measurement and potential bias," and "method validation, as appropriate" in order to be included on the OSAC Registry.

The same presentation also stated that a standard, "specifies uniform methods, actions, practices, or processes, protocols," and that "compliance [is] recommended to be mandatory and modified only under unusual circumstances."

There is no question that this document fails to meet those requirements. This, however, is less a reflection on the substantive merit of E2548 - 11 (given its real purpose) than the present failure of the OSAC to strike a clear distinction between "standardized methods" and "consensus documentary standards" for purposes of the technical merit requirements set forth for OSAC standards and guidelines.

The only verbiage that seemed to separate these two *distinct* types of "standards" in the presentation was the inclusion of the words, "as appropriate." This cryptic reference is insufficient guidance with which to de-conflate which type of standard (high-level consensus practice standard vs. standardized method) is under consideration, and gives no guidance at all about the OSAC-required components of this second type of standard — a high-level consensus practice standard — that sets forth minimum requirements for labs to incorporate within their internally validated analytical procedures for a given technology and/or method.

There is no question that E2548 - 11 is not a "standardized method." Its scope is not detailed, its "purpose" merely states that it covers "minimum considerations" for "sampling of seized drugs for qualitative and quantitative analysis"; uncertainty of measurement is not specifically addressed; sampling bias is only addressed with the single statement, "A random sample is one selected without bias"; and method validation is not addressed at all — but that is simply because E2548 - 11 is not a "method."

Rather, E2548 - 11 appears to be a "consensus documentary standard" which — as other documentary standards do — addresses not *how* a particular laboratory's sampling method *shall* be performed, but rather *that* each laboratory engaged in sampling *shall* follow the "minimum considerations" set forth for sampling "seized drugs for qualitative and quantitative analysis."

Unfortunately, however, there is currently no OSAC guidance (that I know of) on the minimum requirements for this type of "standard" (consensus documentary standard) — other than the "as appropriate" language noted above.

Given the present absence of a distinction between "consensus documentary standards" and "standardized methods" in the OSAC nomenclature, this "standard" fails a test it was never originally designed to take.

The OSAC should remedy this omission by specifically recognizing the distinction between these two distinct types of "standards" and setting forth separate criteria by which high-level consensus documents — such as this one — are to be judged.

///

///

SPECIFIC COMMENTS

My comments above notwithstanding, E2548 - 11 can be improved upon as a "consensus documentary standard."

1. Given its proposed placement in the OSAC Registry, I think the title, "Standard Guide for Sampling ...." is confusing. The title hedges by using both the terms "standard" and "guide." Further, the document, in places, reads like a "guideline" rather than a "standard" by using the term

"recommended" (4.2), (5.6.1.2), (6.1); "should" (5.6.2.1); and "may" (4.2.1.1), (5.4.2), (5.5), (5.5.2.2).

Not persuasive. This is the ASTM verbiage.

In other places, the document reads like a standard by using the term "shall" (4.2.1.2), (4.2.2), (5.2.3), (5.4), (5.6.2.2), (6.2), (7.1), (8.1); and "must" (8.1.1) (8.1.2).

The combination of this directive and permissive verbiage in the document should be re-examined if E2548-11 will go forward as a "standard."

2. 5.6.2.1 states that "For statistical approaches it is recommended that random sampling be conducted." If a statistical approach is probabilistic in nature, and is meant to support inferences drawn regarding an entire population, I don't see how sampling could validly be conducted by anything other than an appropriately chosen and appropriately random method — so I'm a bit confused by the permissive word, "recommended."

Editorial. The Seized Drugs subcommittee will request revision of the language through the SDO process. For clarification, the quoted text is found in section 5.6.1.2, not 5.6.2.1 as noted above.

3. 7.1 states, "Inferences based on the use of a sampling plan and concomitant analysis should be documented." The use of the term "concomitant" here is confusing because sampling and analysis will not literally be conducted in a "concomitant" manner — and if they were that would present its own set of problems. The word "subsequent" instead of "concomitant" seems more appropriate here.

Not persuasive. Concomitant means "associated, parallel, related, etc.", which is the intended meaning here. The analysis is subsequent to each sampling selection, but not necessarily the complete sampling action. Concomitant is a more general term and is appropriate for this purpose.

Also, consider adding the requirement that "key assumptions" which are incorporated into the sampling plan constructed by the lab should be documented as well.

4. Consider redrafting 5.5.2.2 to make it clear (consistent with David Kaye's recommendation) that this language does not mean that the otherwise scientifically valid and reliable selection of a single unit from a multiple unit population should yield to management directives, legislative, or judicial requirements.

Previously considered. This section merely states that a laboratory may choose to sample as few as one unit from a population, but that choice comes with limits to making an inference to the population.

5. Where key "terms of art" are used within this document and are currently defined in another existing standard, then consider some direct reference to that standard (where those terms *are* defined) within the present document when the terms of art are used. If those terms are not defined within an existing standard, they should be added to the OSAC list of definitions before the standard is included on the OSAC Registry.

6. Perhaps 1.2 could be rephrased to state that this standard should be used *in conjunction with* the analyst's knowledge, skill, or ability acquired through appropriate education, training, and experience and sound professional judgment — rather than that it cannot "replace" those attributes. The guidance set forth in this document — as with all scientific standards and procedures — will always *necessarily* be executed concurrently with the judgment and discretion of a trained and experienced analyst. Neither the expert nor the method can exist or function without the other when producing scientifically sound and reliable results.

Previously considered.

Ted Hunt
Legal Resource Committee

The following LRC members agree with comments made by Ted Hunt:
David Kaye and Ron Reinstein.

DISCLAIMER: ***The failure of any member of the Legal Resource committee (LRC) to provide a comment, identify a legal issue or join in another LRC comment should not be interpreted as a disagreement or endorsement of the comment, the standard or its legal sufficiency.***