Sue Gross
OSAC Materials Subcommittee Chair
May 10, 2017

Subject: Final Determination of the Ad hoc Independent Review Panel

The Ad hoc Independent Review Panel has completed its evaluation of comments, adjudication response, and appeal received for the document listed below in accordance with RA-1800 Public Appeals.

- ASTM E2926-13 Standard Test Method for Forensic Comparison of Glass Using Micro X-ray Florescence Spectrometry - Comment adjudication completed on November 12, 2015.

A summary of the appeal, comments from the Ad hoc Panel, and final determination are as follows.

1. Summary of Appeal:

   An appeal was submitted and received on April 3, 2017 from the Legal Resource Committee members David Kaye, Ronald Reinstein, and Barry Scheck. A meeting occurred on April 20, 2017 at the OSAC meeting in Leesburg, VA to discuss the appeal. During this meeting, representatives from the materials subcommittee, legal resource committee, quality infrastructure committee, OSAC affairs, and statisticians were present and all points of the appeal were withdrawn except the comment under 10.7.3.2. The appeal is attached for reference; however, can be summarized as follows: The subcommittee did not address the comment regarding clarification of the 99.7% distribution interval in their comment adjudication.

2. Summary of Appeal:

    a. Comments from the Ad hoc Panel:

    The Ad hoc Panel participated in a conference call on April 21, 2017 where this appeal was discussed. The appeal contained opinions on the nature of the 99.7% distribution interval as written and the intent of the statement. The panel agreed that this particular item was not addressed in the comment adjudication, and should be properly adjudicated by the subcommittee. The subcommittee's response for deeming the comment non-persuasive discussed the appropriateness of the use of 3s for this type of analysis; however, it did not address the concern that referencing 99.7% of a normally distributed population could be misinterpreted by readers as a confidence interval.

    b. Final Determination:

    This appeal is sustained. The Ad hoc Panel recommends that the subcommittee re-adjudicate to address the comment regarding the sentence referencing 99.7% of a normally distributed population under section 10.7.3.2 of this document.

    In regards to the appeal itself, the Ad hoc Panel recommends – Re-Adjudication to include all elements of the comment.

Sincerely,

Kris Cano

Kris Cano

Ad hoc Panel Chair

Panel Members:

Erin Henry, QIC representative

Ted Berman, Glass representative

Larry Tang, Statistician representative

Mark Stolorow, FSSB/OSAC affairs representative

Scott Oulton
OSAC SAC Chair Chemistry/Instrumental Analysis
January 28, 2016

Susan Gross
OSAC SAC Subcommittee Chair Materials

Subject: QIC Process Control Check of Comment Adjudication

The below documents have been reviewed in accordance with RA-800 and RA-900 QIC Process Control Check of Comment Adjudication and have been approved to move forward in the Registry approval process.

- ASTM E2926: Standard Test Method for Forensic Comparison of Glass Using Micro X-ray Fluorescence (u-XRF), Comment adjudication performed on November 12, 2015

Comments and comment adjudication will be publically posted.

Sincerely,

Karin Athanas
QIC Task Group Chair
Registry Approval Comment Adjudication QIC Check

Cc: QIC Chair, OSAC Subcommittee Kavi Liaison

OSAC Registry Request Comment Adjudication Template

Document Title	ASTM E2926: Standard Test Method for Forensic Comparison of Glass Using Micro X-ray Fluorescence (u-XRF)
Requesting Subcommittee	Materials

Subcommittee Chair

Name:	Susan Gross
Affiliation:	MN BCA
Email:	sue.t.gross@state.mn.us
Phone:	651-793-2874

Subcommittee Technical Contact

Name:	Tatiana Trejos
Affiliation	Florida International University
Email:	trejost@fiu.edu
Phone:	305-348-0001

Beginning Comment Period Date	8/14/2015
End Comment Period Date	9/14/2015
Comment Adjudication Meeting Dates	The glass task group met October 27, 2015 to discuss and finalize the comments. The subcommittee met on November 12, 2015 to discuss and vote on the adjudication of comments after having what the glass task group worked on. 11 subcommittee members were present, 7 subcommittee members voted prior to the meeting for a total of 18 'yes' votes to approve the response. 2 subcommittee members were absent and did not vote.
# of Members Present	18 (11 present, 7 voted prior to meeting)

Note: This template is intended for use by all subcommittees considering a new document for addition into registry.

Categories for adjudication of negative public comment for addition to registry

Term	Definition
Not Germane	Comment is not relevant to the subject of document being considered
Editorial - review requested	There is general agreement with edit given, edit by SDO may be requested
Persuasive - review required	General agreement with negative comment given, further review by subcommittee required
Withdrawn by submitter	Comment withdrawn by submitter
Not persuasive	Justification for non persuasive rationale is indicated by committee action
Previously considered	Topic of comment was previously discussed and resolved by subcommittee
No response needed	Comment does not require a response

Resolution categories

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

To: OSAC Independent Review Panel

From: David Kaye, Ronald Reinstein, and Barry Scheck

Subject: Appeal from Adjudication of LRC-compiled Comments on ASTM E2926-13

Date: April 3, 2017

Introduction

The FSSB approved ASTM E2926–13, a "Standard Test Method for Forensic Comparison of Glass Using Micro X-ray Fluorescence (μ-XRF) Spectrometry" over objections from the FSSB Statisticians Task Group and the Legal Resource Committee (LRC). This appeal from members of the LRC is required because the adjudication process failed to address significant parts of certain comments that would have helped the subcommittee produce an improved document for the benefit of the forensic science and legal communities.

Appeal

I. Comment on § 1.3

The LRC-compiled comment on § 1.3 is that

> Section 1.3 states that "This test method does not replace knowledge, skill, ability, experience, education, or training and should be used in conjunction with professional judgment." On its face, this seems to assert that analysts can ascertain elemental composition without using any instruments or that an analyst can depart from one of the prescribed statistical rules in an ad hoc manner. The sentence should be clarified (or deleted on the theory that it goes without saying that it takes skill, experience, and judgment to perform the analysis). [1]

The response was

> This comment is non-persuasive. This is not the intended meaning of the statement. See general comments provided.

Obviously, labeling a comment as "non-persuasive" is not a sufficient response to a request for a change. There must be a reason given to believe the comment does not merit a change. The comment sought a standard that would clearly express the intent behind the sentence. Rather than consider how clear the proposed sentence is and whether a change would improve it, the subcommittee asserted that it did not intend for it to be read in the way that the

[1] Judge Reinstein did not agree with the parenthetical part of the comment. Consistent with the overall comment, he proposed a simple change to the wording to avoid ambiguity—that "judgment, training, and experience are important and must be used in conjunction with the test method, but not in place of it." He continues to believe that a sentence to this effect belongs in the standard.

comment pointed out it might be read.[2] When confronted with ambiguous language, however, the proponents of the text cannot just say that they did not intend it to be ambiguous. They could reply that it is not really ambiguous (and explain why) or make a change to eliminate the ambiguity. The response here does neither. It does not engage the issue at all.

II, Comment on Missing Words in Part 10 (Calculation and Interpretation of Results)

The following comment was not adjudicated properly:

> Full sentences (with subjects) or some other wording should be used so it is clear which tasks are mandatory, recommended, or permissible.

The comment simply asked the subcommittee to identify which steps in the procedure as enumerated are required and which are not.

The response was *not* that Part 10 already does this well enough; nor was it that more clarity would be counter-productive. Instead, the putative answer was

> This comment is non-persuasive. The should/must wording was addressed during the drafting and during the overall balloting process of the ASTM documents. See general comments provided about the ASTM process.

The general comments state that

> As per the ASTM internal guide, documents require language consistency: use the word shall when stating mandatory requirements, use the word should as advisory, use the word may to indicate optional directives, avoid use of must whenever possible. The whole process of changing shall/should/must requires new balloting. Both of the revised methods were developed keeping this in mind with thorough consideration of the practical implications on when/why to use one term over the other.

We are at a loss to understand how the fact that ASTM requires "shall" for a requirement, "should" for a recommendation, and "may" for neither, and that it eschews "must"—a word that, inexplicably, appears nine times in this ASTM standard—responds to the comment that Section 10 often omits these words—a drafting flaw that causes ambiguity.[3] Nor do we understand how the fact that an SDO formulated a standard with particular wording can be considered a response to a suggestion for improving the wording. If this were an admissible response, all subcommittees could simply say that they are putting forward an SDO-approved standard for the registry and need not reply further to comments on the merits of those standards.

III. Comment on § 10.7.3.2

Section 10.7.3.2 6 reads:

---

[2] The general comments merely state the *intended* meaning and explain that the reason for the words selected was that they are "highly recommended by the E30 ASTM committee." The general comments do not explain why or how the "highly recommended" wording found by members of the LRC to be problematic conveys the subcommittee's intended meaning.

[3] Section 10.1, for example, consists of the sentence fragment "Examine the spectrum, and identify and label the peaks." Is this directive a "shall," a "should" or a "may"?

> For each elemental ratio, compare the average ratio for the questioned specimen to the average ratio for the known specimens ±3s. This range corresponds to 99.7% of a normally distributed population. If, for one or more elements, the average ratio in the questioned specimen does not fall within the average ratio for the known specimens ±3s, it may be concluded that the samples are not from the same source.

The LRC-compiled comment, in pertinent part, is as follows:

> Is this a decision rule based on a desired 99.7% confidence interval for the true mean of the ratio in a homogenous known-glass sample? If so, it does not account for the fact that with a standard error estimated from a small sample, one needs a larger interval to achieve 99.7% confidence. In addition, the usual (and better) way to test whether two sample means are different is to use the sampling distribution of the difference between the sample means rather than the sampling distribution of only one of the sample means. Furthermore, even with the proper test statistic and distribution, the many separate tests (one for each ratio Ca/Mg, Ca/Ti, Fe/Zr, etc.) cloud the interpretation of the significance of the difference in a pair of sample means. The risk of a false exclusion for, say, ten comparisons could be ten times the nominal value of 0.003. Thus, the section should be rewritten to justify the choice of the nominal level and to indicate how the nominal level relates to the actual level.

It is not necessary to understand all the technicalities of the comment (and the subcommittee's general comments) to see that the bottom line is a request for the standard to *clarify* the fact that the even though 99.7% of the area under a normal curve lies within approximately plus-or-minus three standard deviations of its mean, "compar[ing] the average ratio for the questioned specimen to the average ratio for the known specimens ±3s"[4] across many elements *will not necessarily achieve 99.7% confidence*.

The adjudication rejects recognizing this issue in the standard, not on the ground that the matching rule will produce 99.7% confidence as promised or suggested, but rather because studies not mentioned in the standard prove that the actual confidence is "appropriate" (even if it is different than the figure in the standard):

> This comment is non-persuasive. (10.7.3.2): Research has shown that 3s is an appropriate method for elemental comparison of glass by μ-XRF based on the goal of minimizing Type I and Type II errors. See general section comments for a detailed explanation on match criteria.

This appeal does not question the judgment that the empirically established conditional error rates are "appropriate."[5] That part of the adjudication (and every other part of it) can be taken as true for the purpose of this appeal. We only contend that nothing in the adjudication of

[4] The subsection does not define "s," but the symbol often refers to a sample standard deviation as distinct from the true but unknown standard deviation (σ) of the sampling statistic. The 99.7% figure applies to a confidence interval of approximately $\pm 3\sigma$.

[5] The FSSB's Statistics Task Group disputed this conclusion. Although Appellant Kaye is a member of that group, he did not participate in the STG vote on this question and took no position on the adequacy of the studies said to establish that the conditional error probabilities of the ±3s rule are appropriately small.

comments actually responds to the request for a clarifying statement in the standard.[6] The adjudication neither adheres to the figure of 99.7%[7] nor explains why the sentence should not be modified (other than, perhaps, a general suggestion that it would be inconvenient or time-consuming to require re-balloting within ASTM).

IV. Comments Seeking Editorial Changes

A few LRC-compiled comments seeking to improve wording (but that were not needed to resolve ambiguities or to correct assertions) were met solely with the reply that

> This comment is non-persuasive. Minor editorial changes will be addressed in the document during the ASTM revision process.

A promise of future change by a different organization is not an acceptable basis for an adjudication of comments on a proposal to approve a draft in its current form. (This principle was accepted in an appeal from the adjudication of ASTM E2548-13.) The OSAC adjudication process asks for editorial as well as substantive comments. If subcommittees are unwilling to consider minor editorial changes, the process should be revised to reflect this limitation on the process for potentially revising documents before placing them on the OSAC registry of approved standards.[8]

* * *

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

---

[6] Judge Reinstein notes that the statistical content of section 10.7.3.2 is beyond his expertise.

[7] The Statistics Task Group confirmed that the standard's suggestion that the theory that a normal distribution of measurement error implies 99.7% confidence for the ±3s rule (as applied to six or seven elements) is invalid. Although it goes beyond the scope of this appeal, we believe that the adjudication process should include input from a Statistics Resource Committee so that when subcommittees adjudicate comments, they will be fully informed with respect to the statistical aspects of their standards. The Legal Resource Committee urged this enhancement at the OSAC Leadership Strategy Session held on June 22, 2016.

[8] The purely editorial changes proposed for this standard were minor. If no other procedural errors in the adjudication had occurred, the appeal panel might have able to treat them as "harmless error" not justifying further adjudication and FSSB review.

Materials Subcommittee response to LRC comments on ASTM 2330 and ASTM E2926 methods

The following section aims to address and clarify some general misconceptions reported by the LRC members in both documents (E2330 and E2926). Some general observations are provided below, while more specific answers are provided later in the document next to their respective LRC questions/comments.

1. The OSAC Materials Subcommittee would like to clarify that several of the LRC comments are beyond the scope of an ASTM **standard test method**:

A. An ASTM standard is a document that has been developed and established within the consensus principles of the society and that meets the approval requirements of ASTM procedures and regulations.

B. There are various types of ASTM standards depending on the technical content and intended use (test methods, guide, practice, etc).

C. These two ASTM methods discussed here are "standard test methods." The primary **scope of a standard test method** is to "describe a definitive procedure that produces a test result", such as identification and measurement of the elemental profiles of glass.

D. The **intended audience** of these ASTM documents is the forensic practitioners who conduct the glass analyses.

E. As per the ASTM internal guide, documents require language consistency: use the word ***shall*** when stating mandatory requirements, use the word ***should*** as advisory, use the word ***may*** to indicate optional directives, avoid use of *must* whenever possible. The whole process of changing shall/should/must requires new balloting. Both of the revised methods were developed keeping this in mind with thorough consideration of the practical implications on when/why to use one term over the other.

F. The ASTM international guide requires the use of SI units.

G. With respect to comments regarding the references for a particular statement or recommendation, the ASTM international guide has requirements for the references to be included in the standard. They read, "Include only references to publications supporting or providing needed supplementary information. Historical and acknowledgment references are not desirable." Per this requirement, background data about the value of elemental analysis of glass is beyond the scope of the standard.

H. The statement about safety concerns is a required caveat by ASTM.

I. The statement "This guide cannot replace knowledge, skill, ability acquired..." is not required by ASTM but is highly recommended by the E30 ASTM committee.

The intent of this wording appears to have been misinterpreted by the members of the Legal Resource Committee. This is standard ASTM E30 wording that was intentionally placed into all ASTM forensic standards. The purpose of the wording is to prevent people who do not have any training from picking up the standard and performing the work without the proper background (training and demonstration of competence and proficiency). Hence the wording that is used in this standard cannot replace training and experience. People who use this standard should also have the knowledge, training, and experience necessary to perform the work. (Statement provided by Fire Debris Subcommittee)

Issues of interpretation, documentation, and training raised by the LRC are **beyond the scope** of these documents. Although we recognize that some of these concerns are valid for the overall practice of trace evidence, they can't be addressed in an ASTM test method and therefore should not apply to the decision/recommendation of whether or not the test method should be included in the OSAC registry. See *Form and Style for ASTM Methods* (http://www.astm.org/COMMIT/Blue_Book.pdf).

2. The OSAC Materials Subcommittee would like to provide some **background on how the test criteria** (a.k.a match criteria) were included in the ASTM standard test methods:

A. These methods are **documents developed through a structured and rigorous consensus process that establish criteria for the analysis or methodology used during a particular examination. In the particular case of these two methods, they are designed to specify how the elemental analysis of glass is conducted for forensic comparisons (by ICP-MS or μXRF, respectively).**

B. These ASTM standards were drafted by a NIJ-funded scientific group (the Elemental Analysis Working Group, a group of 34 scientists with particular expertise in elemental analysis of glass materials (forensic glass practitioners, researchers, and statisticians). The method was then exposed to revision by ASTM subcommittee and later exposed to the main committee and balloting/review process.

C. The test criteria reported on these methods were based on inter-laboratory studies designed to minimize both type I and type II errors in the comparison of elemental data. Several test criteria were tested on these studies based on statistical methods previously reported in validation/population/survey studies. Some methods that the forensic community was using in their protocols were also included in the study - no

consensus on match criteria existed within the community at the time the inter-laboratory tests were started.

D. The inter-laboratory tests not only provided an effective way of cross-validating methods used for the elemental analysis of glass but also demonstrated which match criteria were more appropriate for elemental analysis of glass. The studies showed that the selection of test criteria was dependent on the capabilities, limitations and precision of the method of analysis. After thorough evaluation of the data derived from "worst case scenarios" the group arrived at a consensus on the best test criteria for glass examinations (by ICPMS, LAICPMS or μXRF). Decisions were made on the basis of lowest type I and type II error rates.

For example, ICP-MS and LA-ICPMS provide quantitative data with the precision of the measurements typically lower than 3%RSD, while μXRF produced semi-quantitative data with typical precision ranging from 2-25%RSD, depending on the element and its concentration.

Variability between the measurements is a combined effect of natural heterogeneity of the sample and the precision of the method. For methods with low variability of the measurements, (such as ICPMS) a narrow test criterion such as 2s or 3s produced high false exclusions. On the other hand, for a method with larger variability such as μXRF, a wide criterion of 4s would introduce an unacceptably high number of false inclusions. For a detailed description of the results the following scientific publications are provided:

a. T. Trejos, R. Koons, S. Becker, T. Berman, J. Buscaglia, M. Dueckin, T. Eckert-Lumsdon, T. Ernst, C. Hanlon, A. Heydon, K. Mooney, R. Nelson, K Olsson, C. Palenik, E. Pollock, D. Rudell, S. Ryland, A. Tarifa, M. Valadez, P. Weis and J. Almirall. Cross-validation and evaluation of the performance of methods for the elemental analysis of forensic glass by μ-XRF, ICP-MS and LA-ICP-MS, *Journal of Anal. Bional. Chem,* 2013, 405: 5393-5409
b. T. Trejos, R. Koons, P. Weis, S.Becker, T. Berman, C. Dalpe, M. Duecking, J. Buscaglia, T. Eckert-Lumsdon, T. Ernst, C. Hanlon, A. Heydon, K. Mooney, R. Nelson, K. Olsson, E. Schenk, C. Palenik, E. Chip Pollock, D.Rudell, S. Ryland, A. Tarifa, M.Valadez, A. van Es , V. Zdanowicz, and J.R. Almirall· Forensic analysis of glass by μ-XRF, ICP-MS, LA-ICP-MS and LA-ICP-OES: Evaluation of the performance of different criteria for comparing elemental composition, *Journal of Analytical Atomic Spectrometry,* 2013, 38, 1270-1282
c. Ernst, T.; Berman, T.; Buscaglia, J.; Eckert-Lumsdon, T.; Hanlon, C.; Olsson, K.; Palenik, C.; Ryland, S.; Trejos, T.; Valadez, M.; Almirall, J. R. Signal-to-noise ratios in forensic glass analysis by micro X-ray fluorescence spectrometry. *X-Ray Spectrom.* 2012, *43*, 13-21.
d. Weis, P.; Dücking, M.; Watzke, P.; Menges, S.; Becker, S. Establishing a match criterion in forensic comparison analysis of float glass using laser ablation inductively coupled plasma mass spectrometry. *J. Anal. At. Spectrom.* 2011,

26, 1273-1284.
e. Berends-Montero, S.; Wiarda, W.; de Joode, P.; van der Peijl, G. Forensic analysis of float glass using laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS): validation of a method. *J. Anal. At. Spectrom.* 2006, *21*, 1185-1193.
f. Hendrik Dorn, David E. Ruddell, Alex Heydon & Brenda D. Burton (2015)Discrimination of float glass by LA-ICP-MS: assessment of exclusion criteria using casework samples,Canadian Society of Forensic Science Journal, 48:2, 85-96.

E. The match/test criteria used for glass examinations cannot be directly applied to other materials because the selection of match criteria is not only dependent on the analytical method performance as described above but also on the natural heterogeneity of the sample. With this said, it is inappropriate to compare the case/scope/purpose of elemental analysis of bullet lead to glass. There are significant differences between the materials in terms of manufacturing, packaging, distribution, heterogeneity and chemical composition.

The use of elemental analysis for glass comparisons has a very strong foundation with dozens of scientific articles describing the heterogeneity and distribution of elements on glass panes, variability, origin, and reasoning on which elements are more discriminating/informing and how they were selected for chemical profiling/comparison of glass. A few of these publications are listed below:

a. Almirall, J. R.; Trejos, T. Advances in forensic analysis of glass fragments with a focus on refractive index and elemental analysis. *Forensic Sci. Rev. 18* **2006**, *2*, 74-96.
b. Almirall, J. R.; Trejos, T. Forensic Applications of Mass Spectrometry. In *Encyclopedia of Mass Spectrometry, 1st ed.*; Beauchemin, D.; Matthews, D., Eds.; Elsevier, **2010**; Vol. 5, pp 705-717.
c. Andrasko, J.; Maehly, A. C. The discrimination between samples of window glass by combining physical and chemical techniques. *J. Forensic Sci.* **1978**, *23*, 250-262.
d. Becker, S.; Gunaratnam, L.; Hicks, T.; Stoecklein, W.; Warman, G. The differentiation of float glass using refractive index and elemental analysis: Comparisons of techniques. *Probl. Forensic Sci.* **2001**, *47*, 80-92.
e. Berends-Montero, S.; Wiarda, W.; de Joode, P.; van der Peijl, G. Forensic analysis of float glass using laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS): validation of a method. *J. Anal. At. Spectrom.* **2006**, *21*, 1185-1193.
f. Buscaglia, J. Elemental analysis of small glass fragments in forensic science. *Anal. Chim. Acta* **1994**, *288*, 17-24.
g. Duckworth, D. C.; Baynes, C. K.; Morton, S. J.; Almirall, J. R. Analysis of variance in forensic glass analysis by ICP-MS: Variance within the method. *J. Anal. At. Spectrom.* **2000**, *15*, 821-828.

h. Duckworth, D. C.; Morton, S. J.; Baynes, C. K.; Koons, R. D.; Montero, S.; Almirall, J. R. Forensic glass analysis by ICP-MS: A multi-element assessment of discriminating power via analysis of variance and pairwise comparisons. *J. Anal. At. Spectrom.* **2002**, *17*, 662-668.
i. Hicks, T.; Monard Sermier, F.; Goldmann, T.; Brunelle, A.; Champod, C.; Margot, P. The classification and discrimination of glass fragments using non destructive energy dispersive X-ray μ fluorescence. *Forensic Sci. Int.* **2003**, *137*, 107-118.
j. Koons, R. D.; Fiedler, C.; Rawalt, R. C. Classification and discrimination of sheet and container glasses by inductively coupled plasma-atomic emission spectometry and pattern recognition. *J. Forensic Sci.* **1988**, *33*, 49-67.
k. Koons, R. D.; Peters, C. A.; Rebbert, P. S. Comparison of refractive index, energy dispersive X-ray fluorescence and inductively coupled plasma atomic emission spectrometry for forensic characterization of sheet glass fragments. *J. Anal. At. Spectrom.* **1991**, *6*, 451-456.
l. Latkoczy, C.; Becker, S.; Dücking, M.; Günther, D.; Hoogewerff, J. A.; Almirall, J. R.; Buscaglia, J.; Dobney, A.; Koons, R. D.; Montero, S.; van der Peijl, G. J.; Stoecklein, W. R.; Trejos, T.; Watling, J. R.; Zdanowicz, V. S. Development and evaluation of a standard method for the quantitative determination of eements in float glass samples by LA-ICP-MS. *J. Forensic Sci.* **2005**, *50*, 1327-1341.
m. Montero, S. Trace elemental analysis of glass by inductively coupled plasma-mass spectrometry (ICP-MS) and laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS), Florida International University, Miami, Florida, 2002.
n. Naes, B. E.; Umpierrez, S.; Ryland, S.; Barnett, C.; Almirall, J. R. A comparison of laser ablation inductively coupled plasma mass spectrometry, micro X-ray fluorescence spectroscopy, and laser induced breakdown spectroscopy for the discrimination of automotive glass. *Spectrochim. Acta, Part B* **2008**, *63*, 1145-1150.
o. Parouchais, T.; Warner, I. M.; Palmer, L. T.; Kobus, H. The analysis of small glass fragments using inductively coupled plasma mass spectrometry. *J. Forensic Sci.* **1996**, *41*, 351-360.
p. Roedel, T. C.; Bronk, H.; Haschke, M. Investigation of the influence of particle size on the quantitative analysis of glasses by energy-dispersive micro x-ray fluorescence spectrometry. *X-Ray Spectrom.* **2002**, *31*, 16-26.
q. Ryland, S. G. Discrimination of flat (sheet) glass specimens having similar refractive indices using micro X-ray fluorescence spectrometry. *Journal of the American Society of Trace Evidence Examiners* **2011**, *2*, 2-12.
r. Suzuki, Y.; Sugita, R.; Suzuki, S.; Marumo, Y. Forensic discrimination of bottle glass by refractive index measurement and analysis of trace elements with ICP-MS. *Anal. Sci.* **2000**, *16*, 1195-1198.
s. Trejos, T.; Almirall, J. R. Sampling strategies for the analysis of glass fragments by LA-ICP-MS Part I. Micro-homogeneity study of glass and its application to the interpretation of forensic evidence. *Talanta* **2005a**, *67*, 388-395.
t. Trejos, T.; Almirall, J. R. Sampling strategies for the analysis of glass fragments by LA-ICP-MS Part II: Sample size and sample shape considerations. *Talanta* **2005b**, *67*, 396-401.

u. Trejos, T.; Montero, S.; Almirall, J. R. Analysis and comparison of glass fragments by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) and ICP-MS. *Anal. Bioanal. Chem.* **2003**, *376*, 1255-1264.
v. Trejos, T.; Koons, R.; Becker, S.; Berman, T.; Buscaglia, J.; Duecking, M.; Eckert-Lumsdon, T.; Ernst, T.; Hanlon, C.; Heydon, A.; Mooney, K.; Nelson, R.; Olsson, K.; Palenik, C.; Pollock, E. C.; Rudell, D.; Ryland, S.; Tarifa, A.; Valadez, M.; Weis, P.; Almirall, J. R. Cross-validation and evaluation of the performance of methods for the elemental analysis of forensic glass by μ-XRF, ICP-MS, and LA-ICP-MS. *Anal. Bioanal. Chem.* **2013a**, *405*, 5393-5409.
w. Trejos, T.; Koons, R.; Weis, P.; Becker, S.; Berman, T.; Dalpe, C.; Duecking, M.; Buscaglia, J.; Eckert-Lumsdon, T.; Ernst, T.; Hanlon, C.; Heydon, A.; Mooney, K.; Nelson, R.; Olsson, K.; Schenk, E.; Palenik, C.; Pollock, E. C.; Rudell, D.; Ryland, S.; Tarifa, A.; Valadez, M.; van Es, A.; Zdanowicz, V.; Almirall, J. R. Forensic analysis of glass by μ-XRF, SN-ICP-MS, LA-ICP-MS and LA-ICP-OES: evaluation of the performance of different criteria for comparing elemental composition. *J. Anal. At. Spectrom.* **2013b**, *28*, 1270-1282.
x. Weis, P.; Dücking, M.; Watzke, P.; Menges, S.; Becker, S. Establishing a match criterion in forensic comparison analysis of float glass using laser ablation inductively coupled plasma mass spectrometry. *J. Anal. At. Spectrom.* **2011**, *26*, 1273-1284.
y. P.M.L. Sandercock (2000) Sample Size Considerations for Control Glass in Casework, Canadian Society of Forensic Science Journal, 33:4, 173-185
z. Hendrik Dorn, David E. Ruddell, Alex Heydon & Brenda D. Burton (2015)Discrimination of float glass by LA-ICP-MS: assessment of exclusion criteria using casework samples, Canadian Society of Forensic Science Journal, 48:2, 85-96.

We would like to stress that the overall scientific foundation of glass examinations includes aspects of transfer, persistence, and methodology validity. The principles and utility of forensic glass examinations is supported by at least 130 publications over the last 3 decades. The list provided above represents only a snapshot of the scientific support on the particular topics of sampling, homogeneity, and discrimination/variations in glass populations.

F. We are aware of Bayesian/likelihood approaches for the interpretation of glass evidence. We are not against these strategies. However, the test methods discussed here are limited to the comparison of "elemental data" to determine whether the elemental compositions of two glass samples (aka K/Q) are distinguishable or not.

What the scientists conclude based on the evaluation of the data is outside the scope of these test methods. After applying these test methods, the practitioner will need to follow an interpretation guide/standard to write a conclusion (whether they decide to use Bayesian or traditional approaches).

The test criterion is a necessary step prior to any further data interpretation. These test methods aim to standardize the way in which practitioners should conduct data analysis to determine if the elemental composition is different

or not. Nonetheless, conducting elemental analysis via either of these test methods is only one of many steps/examinations that the glass examiner must follow and later put together to evaluate the evidence and write a conclusion based on the overall glass examination (physical, optical and chemical).

It is critical to keep in mind the scope of the test method when evaluating these documents.

3. Implications of the use of different types of data in forensic science: The elemental composition of materials can be obtained by spectrochemical methods (i.e. μXRF, ICP) in three main forms:

a) qualitative
b) quantitative
c) semi-quantitative

All 3 forms of data comprise true/valid scientific information/data that can be used to determine the source of a material or make inferences about commonality, similarity or difference of chemical composition.

The decision of whether we use one form of data or another is dictated by the nature of the material, the limitations/capabilities of the technique, and the purpose of the analysis.

A. **Qualitative data:** this is, for example, a graphical representation of a spectrum that shows the identification of elements present/detected in a specimen, such as iron, calcium, etc. You can use qualitative data to determine which elements are present or absent in a sample and/or to compare if the same elements are present or absent in a comparison sample. You can overlay two spectra to compare their qualitative profile.

B. **Quantitative data:** quantitative data involves the identification of the element followed by the calculation of the absolute concentration (amount). In solid materials such as glass this is typically reported in ug/g. Calibrations are made by using certified standards of known concentration. For instance, ICPMS is a method able to generate quantitative data with excellent precision and accuracy. You can not only determine that iron and calcium are present, you can also report the actual concentration and uncertainty for each element. This data can be used to detect significant differences in the composition of the elements in comparison samples.

C. **Semi-quantitative data:** in some instances, the conditions for reliable quantitative data are not met. For example, in the case of μXRF

quantitative data requires calibration with solid standards of glass at concentrations detectable by this method and the glass fragment must have minimum thickness/shape requirements to attenuate differences in the way that the X-rays are released from the glass into the detector.

Because of the typical sample/shape of glass fragments, these limitations prevent quantitative measurements by μXRF. Instead, ratios of elements are used to compensate for those variations. The ratios of elements are selected/recommended based on similar anticipated behavior to optimize that normalization of the data. For example, these ratios of the peak areas of calcium to iron generate "numerical" data. That ratio data is not "quantitative" but due to its numerical nature it is considered "semi-quantitative" because you are comparing the relative amounts of these elements in the samples.

Semi-quantitative data allows the calculation of uncertainties and is widely accepted in the scientific community. Semi-quantitative data is the foundation of other forensic materials, such as DNA analysis.

COMMENTS BY THE OSAC LEGAL RESOURCE COMMITTEE (LRC)

TO: Materials (Trace) Subcommittee of the Chemistry-Instrumental SAC

FROM: Lynn Garcia, LRC Liaison to Chemistry-Instrumental SAC

RE: **LEGAL RESOURCE COMMITTEE (LRC) COMMENTS ON E2926-13**

The Materials (Trace) Evidence Subcommittee of the Chemistry-Instrumental SAC is proposing that ASTM E2926–13 ("Standard Test Method for Forensic Comparison of Glass Using Micro X-ray Fluorescence (μ-XRF) Spectrometry") be placed on the OSAC Registry of Standards and Guidelines.

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 technical merit and legal importance. Though the LRC is not able to assess the scientific merit of a Standard, our review encompasses whether a Standard makes a

[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)).

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

The LRC received feedback from the FSSB recently that it would be more useful for LRC members to provide consolidated comments as opposed to providing the comments of individual members and indicating which other members of the LRC join in the comments. We did not have sufficient time to attempt this before comments on E2926-13 were due. However, we have been discussing possibilities for meeting this request and will strive to make our comments as useful as possible to the FSSB and other interested readers.

Comments by LRC Member David Moran:

1.3. I object to the language "This test method does not replace knowledge, skill, ability, experience, education, or training and should be used in conjunction with professional judgment." This language strikes me as (1) unnecessary; and (2) a dodge that a bad forensic scientist could use to justify wholesale deviations from the approved method (i.e., "My professional judgment, experience and training justifies my decision to declare that the glass in this case is not excluded from the reference sample even though the variances exceeded three standard deviations").

3.5. I don't understand why determining the area under the peaks and comparing that area to the area under peaks of certain elements is considered "semi-quantitative analysis" as opposed to "quantitative analysis."

10.7.3.1 and 10.7.3.2 set out a nice quantitative method allowing the examiner to conclude that two glass specimens are not from the same source. But the standards say nothing about what the examiner should report or say if the method does not result in exclusion. The standard should explicitly say that in that case the examiner should report that the samples cannot be excluded as being from the same source, and nothing more.

Additional Comments:

1. The standard should require the examiner to attach to a written report all of the charts containing the peaks used to perform the analysis.

2. The standard should require that the examiner test several "suspect" glass samples, not just one, against the known glass, and that, wherever possible, the examiner should not know which of the "suspect" glass samples is suspected as matching the known glass. In other words, the examiner should be blinded by having someone else choose the samples to be tested without telling the examiner which is the suspected sample.

The following members of the LRC agree with comments made by David Moran:

Barry Scheck and Lynn Garcia join in David Moran's comments. David Kaye joins with the clarification that he would not foreclose the possibility of a revision to the Standard that would allow some scientifically defensible explanation of the implications of a failure to exclude. He

also questions whether the statistical decision rule for exclusions based on "peak intensity" is acceptable as written.

Comments by LRC Member David Kaye:

The current ASTM Standard Test Method contains valuable guidance for forensic science laboratories. Nevertheless, I believe that this Standard Test Method can be significantly improved and should not be included in the Registry without substantial changes. If the Standards developed or selected for inclusion on the OSAC registry are intended to represent the kind of "controlling standards" spoken of in *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993), and thus to be of the most benefit to the law, the Standard should clearly delineate what analysts must do (invariably or in specific circumstances); what they might do; and if current practices are deficient, what they should or must not do. It should indicate the limitations on conclusions that an analyst can reach and should explain and justify its choices (perhaps in a separate supporting document). It should outline the minimum content that good scientific practice dictates for written reports or other laboratory documents. A lay reader should be able to use the Standard to help determine whether a laboratory is providing scientifically reputable testimony in a given case.

Admittedly, this is a tall order, but even if one rejects this aspiration for standards placed in the OSAC registry, perhaps on the theory that a narrowly written Standard can specify enough of what should be done to make it ready for inclusion, the treatment of the matters that the Standard does address should avoid unnecessary ambiguity, should document the validity and reliability of the procedures it prescribes or recommends, and should explicitly state what it does not cover. There is little or nothing to be gained by rushing to endorse Standards that lack these features. Only Standards that accomplish these goals can fulfill the claim made in the Technical Merit Worksheet for this Standard that it is "fit for purpose" in that "this document can be used as a reference by any law enforcement agency, judges, prosecutors and defense attorneys."

1. Concerns Regarding Content

Introduction

The introductory paragraph assumes that it is necessary or desirable to reach a binary conclusion ("distinguishable" or "indistinguishable," which, in court, translates into excluded or included) when comparing two objects. Moreover, it suggests that the "the possibility that they [fragments] originated from the same source of glass" must "be eliminated" for the analysis to be useful. Neither proposition should be endorsed unequivocally.

First, neither the introduction nor the rest of the Standard explains why the examiner must use a binary classification as opposed to reporting the probability of the observed degree of similarity if the questioned fragments originated from the known glass versus that if they did not. Second, the "possibility" that different specimens have a common source never can be eliminated. The data can be quite improbable if they originated from the same source. Or, they can be much more probable under the same-source hypothesis than a different-source hypothesis. Using a sharp cutoff for exclusion carries a probability of statistical error. If a cutoff is the only permissible way to interpret the measurements, as the introduction and Part 10 intimate, this uncertainty must be acknowledged.

Part 1

Section 1.3 states that "This test method does not replace knowledge, skill, ability, experience, education, or training and should be used in conjunction with professional judgment." On its face, this seems to assert that analysts can ascertain elemental composition without using any instruments or that an analyst can depart from one of the prescribed statistical rules in an ad hoc manner. The sentence should be clarified (or deleted on the theory that it goes without saying that it takes skill, experience, and judgment to perform the analysis).

Parts 3-9

Some of the material in Parts 3-9 is descriptive, and some is prescriptive. Statements such as "Limits of detection (LOD) are dependent on several factors, including …" do not supply much guidance. How should or must limits be set? Steps that are required should be designated as such; those that are merely recommended should be phrased accordingly. There are "musts" and "shoulds" in these sections, but it is not always clear why some of the "shoulds" are not "musts" and what some other things are.

Part 10 (Calculation and Interpretation of Results)

Full sentences (with subjects) or some other wording should be used so it is clear which tasks are mandatory, recommended, or permissible. Section 10.2 could be clearer in stating that automated peak identification and purely manual identification are both acceptable (if they are). Are there other acceptable methods of manual verification of an automated determination?

What use should be made of the "visual comparison"? Can it override quantitative measurements? How should it be performed?

Section 10.7.3.1 states that "If the ranges of one or more elements in the questioned and known specimens do not overlap, it may be concluded that the specimens are not from the same source." The phrase "may be" is rather weak. Is this the recommended conclusion? Why? If the statistical properties of the "ratio ranges" are unknown, how can one know what to conclude? There may as few as 3 measurements of the questioned glass and 9 of the known one.

Section 10.7.3.2 6 adopts a 3-standard-error rule. It reads

> For each elemental ratio, compare the average ratio for the questioned specimen to the average ratio for the known specimens ±3s. This range corresponds to 99.7 % of a normally distributed population. If, for one or more elements, the average ratio in the questioned specimen does not fall within the average ratio for the known specimens ±3s, it may be concluded that the samples are not from the same source.

I have puzzled over these sentences for hours without being able to understand them. Is this a decision rule based on a desired 99.7% confidence interval for true mean of the ratio in a homogenous known glass sample? If so, it does not account for the fact that with a standard error

estimated from a small sample, one needs a larger interval to achieve 99.7% confidence. In addition, the usual (and better) way to test whether two sample means are different is to use the sampling distribution of the difference between the sample means rather than the sampling distribution of only one of the sample means. Furthermore, even with the proper test statistic and distribution, the many separate tests (one for each ratio Ca/Mg, Ca/Ti, Fe/Zr, etc.) cloud the interpretation of the significance of the difference in a pair of sample means. The risk of a false exclusion for, say, ten comparisons could be ten times the nominal value of 0.003. Thus, the section should be rewritten to justify the choice of the nominal level and to indicate how the nominal level relates to the actual level. In other words, why the 3σ rule? Is it supposed to keep the risk of a false exclusion to a low level?

Although these questions may seem technical, they are directly related to the interpretation of the results in the criminal justice system. From a legal perspective, are not false inclusions the type of error that should be guarded against more assiduously? And even if 3σ is the right rule here, why is the standard for making associations via elemental compositions in E2330 some kind of 4σ rule? Without reconciling the different standards, their value as justifications for interpretations of test results in the legal system could be jeopardized.

Assuming that the exclusion-inclusion decision is the best way to interpret the differences in "peak intensity," in discussing the interpretation of the data, the Standard needs to offer guidance about the probative value of an inclusion. Should the analyst report that the questioned fragment might have come from the known glass *or from any other glass with a similar set of elemental concentrations*? What data are there on the population distribution of these statistics? If there are none, what can or should the analyst report?

Part 11 (Precision and Bias)

These quantities should be defined within the Standard itself.

References

To show a legal audience that the Standard is based on a complete review of the scientific and statistical literature, there should be references to studies that help demonstrate the value of the testing in forensic investigations. The Standard should show how it flows from and is supported by a body of cited scientific studies. The Subcommittee commendably listed 15 papers in the Technical Merit Worksheet of 7/1/15, but readers in the legal community will not know of them. (Could this be remedied by having an appendix that justifies key choices made in the Standard placed on the registry along with the Standard? This appendix could explain which publications support which choices and how they do so. It would help prevent members of the legal community from misjudging the Standard as the kind of "ipse dixit" condemned in *General Electric v. Joiner*, 522 U.S. 136 (1997). Evaluating this document together with the body of the Standard would permit readers who are not already technical experts in the field to judge its readiness for addition to the registry. Presumably, this explanatory document would not have to be approved by ASTM.)

Concluding Comments

Recommendations about the opinions and conclusions that analysts can reach and how they should present them in reports and in court should be made in light of current thinking about the methods for interpreting and evaluating evidence across the entire domain of forensic science. The premise of the Standard is that the expert's task is to decide whether a source hypothesis is true or false. Would a likelihood ratio be a better way to express the probative value of the data? Certainly, there is an argument to that effect in the legal and forensic science literatures. See, e.g., Colin Aitken & Franco Taroni, Statistics and the Evaluation of Evidence for Forensic Science (2d ed. 2004); James M. Curran et al., Forensic Interpretation of Glass Evidence (2000); ENFSI Guideline for Evaluative Reporting in Forensic Science (2015); David H. Kaye et al., The New Wigmore: Expert Evidence (2d ed. 2011); Royal Statistical Soc'y Working Group on Statistics and the Law, Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses (2010).

In the end, the subcommittee, the SAC and the FSSB may conclude that only the categorical decision framework that is recommended in this document is acceptable for the evaluation and explanation of the evidence in court. If that is their conclusion, however, some of the reasoning behind the conclusion should be provided to assist the legal community in using the measurements wisely and fairly.

2. Drafting Problems

Introduction

Should the last sentence in the introduction read "Measuring elemental concentrations with micro X-ray fluorescence spectrometry permits high discrimination among different sources of glass"?

Part 1 (Scope)

I suggest rewording 1.1 to read as follows: "This Standard concerns μ-XRF analysis using mono- and poly- capillary optics, and an energy dispersive X-ray detector (EDS) for the determination of the concentrations of major, minor, and trace elements in glass fragments." This change would permit 1.2 to be deleted.

1.3 states that "This test method does not replace knowledge, skill, ability, experience, education, or training and should be used in conjunction with professional judgment." On its face, this seems to assert that analysts can ascertain elemental composition without using any instruments or that an analyst can depart from one of the prescribed statistical rules in an ad hoc manner. The sentence should be clarified (or deleted on the theory that it goes without saying that it takes skill, experience, and judgment to perform the analysis).

1.4 reads: "The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard." The meaning of "to be regarded as standard" is not immediately apparent. Why is it necessary to say that no units other than the International System of Units are used? Is not that apparent from reading the Standard as a whole? (Of course, if a Standard would not be expected to use these units, then referring to this choice at the outset is

helpful.) Moreover, this remark does not really describe the scope of the Standard, but only how things within its scope are expressed.

The disclaimer in section 1.5 does not seem to achieve any legal objective. It is hard to see how any reader would think that a "method . . . purport[s] to address all of the safety concerns," and stylistically, the phrase "all of the safety concerns, if any" also is awkward, and the "if any" phrase is contradicted by the fact that Section 7 does address a safety concern.

Part 10 (Calculation and Interpretation of Results)

What does it mean to "correct" sum peaks and escape peaks? That the analyst should label them as such?

The following members of the LRC agree with the comments made by David Kaye:

Barry Scheck and David Moran join in David Kaye's comments.

Ron Reinstein joins in David Kaye's comments except the comment regarding Section 1.3. Judge Reinstein believes this section should be clarified but not deleted. Judgment, training and experience are important and must be used in conjunction with the test method (but not in place of it).

Judge Reinstein and Lynn Garcia would like to emphasize the importance of David Kaye's "Concluding Comments" on page 6—this is the type of comment that the SACs and FSSB should pay particular attention to for standards that are to be included on the OSAC Registry.

Additional Comments by LRC Member Barry Scheck:

With respect to David Moran's comments, Barry Scheck would like to emphasize the term "semi-quantitative" is troubling. Either you have data or you don't and the measure uncertainty can be calculated within acceptable limits. If you are going to rely on "experience" to declare exclusions or inclusions (which seems problematical in the first place) the standard should specify when and how that would be done and any validation that justifies it.

Barry Scheck would also like to emphasize concerns expressed in the comments by David Kaye that the deficiencies in the statistical explanations are 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. The requirement of general acceptance in the scientific community, particularly among statisticians, cannot be met, nor the requirements of clearly identifying limitations and weaknesses in the methodology or an explanation of how it is "fit for purpose."

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: Materials (Trace) Subcommittee of the Chemistry-Instrumental SAC

FROM: Lynn Garcia, LRC Liaison to Chemistry-Instrumental SAC

RE: **LEGAL RESOURCE COMMITTEE (LRC) COMMENTS ON E2926-13**

The Materials (Trace) Evidence Subcommittee of the Chemistry-Instrumental SAC is proposing that ASTM E2926–13 ("Standard Test Method for Forensic Comparison of Glass Using Micro X-ray Fluorescence (μ-XRF) Spectrometry") be placed on the OSAC Registry of Standards and Guidelines.

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 technical merit and legal importance. Though the LRC is not able to assess the scientific merit of a Standard, our review encompasses whether a Standard makes a

[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)).

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

The LRC received feedback from the FSSB recently that it would be more useful for LRC members to provide consolidated comments as opposed to providing the comments of individual members and indicating which other members of the LRC join in the comments. We did not have sufficient time to attempt this before comments on E2926-13 were due. However, we have been discussing possibilities for meeting this request and will strive to make our comments as useful as possible to the FSSB and other interested readers.

Comments by LRC Member David Moran:

1.3. I object to the language "This test method does not replace knowledge, skill, ability, experience, education, or training and should be used in conjunction with professional judgment." This language strikes me as (1) unnecessary; and (2) a dodge that a bad forensic scientist could use to justify wholesale deviations from the approved method (i.e., "My professional judgment, experience and training justifies my decision to declare that the glass in this case is not excluded from the reference sample even though the variances exceeded three standard deviations").

This comment is non-persuasive. See general comments provided.

3.5. I don't understand why determining the area under the peaks and comparing that area to the area under peaks of certain elements is considered "semi-quantitative analysis" as opposed to "quantitative analysis."

"Semi-quantitative" refers to comparing ratios of peak areas. Peak areas are related to actual concentrations, which are not calculated by this method. "Quantitative" would refer to determining actual concentrations of the elements, which is not done for u-XRF examinations of glass. See general comments provided.

10.7.3.1 and 10.7.3.2 set out a nice quantitative method allowing the examiner to conclude that two glass specimens are not from the same source. But the standards say nothing about what the examiner should report or say if the method does not result in exclusion. The standard should explicitly say that in that case the examiner should report that the samples cannot be excluded as being from the same source, and nothing more.

This comment is non-persuasive. The suggestion is beyond the scope of the standard. See general comments provided.

Additional Comments:

1. The standard should require the examiner to attach to a written report all of the charts

containing the peaks used to perform the analysis.

This comment is non-persuasive. The suggestion is beyond the scope of the standard. See general comments provided.

2. The standard should require that the examiner test several "suspect" glass samples, not just one, against the known glass, and that, wherever possible, the examiner should not know which of the "suspect" glass samples is suspected as matching the known glass. In other words, the examiner should be blinded by having someone else choose the samples to be tested without telling the examiner which is the suspected sample.

This comment is non-persuasive. Each Q fragment is necessarily treated as a separate entity, while the K has multiple fragments analyzed in order to more fully characterize the known sample. This suggestion would diminish the usefulness of the comparison.

The following members of the LRC agree with comments made by David Moran:

Barry Scheck and Lynn Garcia join in David Moran's comments. David Kaye joins with the clarification that he would not foreclose the possibility of a revision to the Standard that would allow some scientifically defensible explanation of the implications of a failure to exclude. He also questions whether the statistical decision rule for exclusions based on "peak intensity" is acceptable as written.

Comments by LRC Member David Kaye:

The current ASTM Standard Test Method contains valuable guidance for forensic science laboratories. Nevertheless, I believe that this Standard Test Method can be significantly improved and should not be included in the Registry without substantial changes. If the Standards developed or selected for inclusion on the OSAC registry are intended to represent the kind of "controlling standards" spoken of in *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993), and thus to be of the most benefit to the law, the Standard should clearly delineate what analysts must do (invariably or in specific circumstances); what they might do; and if current practices are deficient, what they should or must not do. It should indicate the limitations on conclusions that an analyst can reach and should explain and justify its choices (perhaps in a separate supporting document). It should outline the minimum content that good scientific practice dictates for written reports or other laboratory documents. A lay reader should be able to use the Standard to help determine whether a laboratory is providing scientifically reputable testimony in a given case.

Admittedly, this is a tall order, but even if one rejects this aspiration for standards placed in the OSAC registry, perhaps on the theory that a narrowly written Standard can specify enough of what should be done to make it ready for inclusion, the treatment of the matters that the Standard does address should avoid unnecessary ambiguity, should document the validity and reliability of the procedures it prescribes or recommends, and should explicitly state what it does not cover. There is little or nothing to be gained by rushing to endorse Standards that lack these features. Only

Standards that accomplish these goals can fulfill the claim made in the Technical Merit Worksheet for this Standard that it is "fit for purpose" in that "this document can be used as a reference by any law enforcement agency, judges, prosecutors and defense attorneys."

1. Concerns Regarding Content

Introduction

The introductory paragraph assumes that it is necessary or desirable to reach a binary conclusion ("distinguishable" or "indistinguishable," which, in court, translates into excluded or included) when comparing two objects. Moreover, it suggests that the "the possibility that they [fragments] originated from the same source of glass" must "be eliminated" for the analysis to be useful. Neither proposition should be endorsed unequivocally.

This comment is non-persuasive. The objective of a forensic glass examination is to compare glass samples to determine if they can be discriminated using the physical, optical, and chemical properties. Ultimately, the goal of this standard test method is to determine if glass samples collected are distinguishable or indistinguishable based on the elemental composition using u-XRF. We disagree with the suggestion that an elimination is the only useful aspect of this analysis. Finding two samples indistinguishable is also useful.

First, neither the introduction nor the rest of the Standard explains why the examiner must use a binary classification as opposed to reporting the probability of the observed degree of similarity if the questioned fragments originated from the known glass versus that if they did not. Second, the "possibility" that different specimens have a common source never can be eliminated. The data can be quite improbable if they originated from the same source. Or, they can be much more probable under the same-source hypothesis than a different-source hypothesis. Using a sharp cutoff for exclusion carries a probability of statistical error. If a cutoff is the only permissible way to interpret the measurements, as the introduction and Part 10 intimate, this uncertainty must be acknowledged.

This comment is non-persuasive. See general comments provided for an explanation of the match criteria.

Part 1

Section 1.3 states that "This test method does not replace knowledge, skill, ability, experience, education, or training and should be used in conjunction with professional judgment." On its face, this seems to assert that analysts can ascertain elemental composition without using any instruments or that an analyst can depart from one of the prescribed statistical rules in an ad hoc

manner. The sentence should be clarified (or deleted on the theory that it goes without saying that it takes skill, experience, and judgment to perform the analysis).

This comment is non-persuasive. This is not the intended meaning of the statement. See general comments provided.

Parts 3-9

Some of the material in Parts 3-9 is descriptive, and some is prescriptive. Statements such as "Limits of detection (LOD) are dependent on several factors, including …" do not supply much guidance. How should or must limits be set? Steps that are required should be designated as such; those that are merely recommended should be phrased accordingly. There are "musts" and "shoulds" in these sections, but it is not always clear why some of the "shoulds" are not "musts" and what some other things are.

This comment is non-persuasive. LODs were calculated based on inter-laboratory studies following the ASTM guidelines for reporting limits of detection. See general comments section for more detail.
The should/must wording was addressed during the drafting and during the overall balloting process of the ASTM documents. See general comments provided about the ASTM process.

Part 10 (Calculation and Interpretation of Results)

Full sentences (with subjects) or some other wording should be used so it is clear which tasks are mandatory, recommended, or permissible.

This comment is non-persuasive. The should/must wording was addressed during the drafting and during the overall balloting process of the ASTM documents. See general comments provided about the ASTM process.

Section 10.2 could be clearer in stating that automated peak identification and purely manual identification are both acceptable (if they are). Are there other acceptable methods of manual verification of an automated determination?

This comment is non-persuasive. Section 10.2 states that automatic peak identification shall be manually verified by any of the three methods reported in the method.

What use should be made of the "visual comparison"? Can it override quantitative measurements? How should it be performed?

This comment is non-persuasive. Sections 10.6.2 and 10.7.2 address this comment. The visual comparison is a check to see if the samples have obvious elemental differences or to see if semi-quant comparisons are necessary. The semi-quant comparisons are used when there aren't apparent visual differences. There is not a circumstance in which they would be similar in peak intensity ratios but visually spectrally different.

Section 10.7.3.1 states that "If the ranges of one or more elements in the questioned and known specimens do not overlap, it may be concluded that the specimens are not from the same source." The phrase "may be" is rather weak. Is this the recommended conclusion? Why? If the statistical properties of the "ratio ranges" are unknown, how can one know what to conclude? There may as few as 3 measurements of the questioned glass and 9 of the known one.

This comment is non-persuasive. (10.7.3.1): XRF analysis is one of several steps within the glass analysis scheme. There may be other considerations that may prevent an examiner from a definitive elimination in specific cases. For example, if there is a single, very small fragment from a glass container that was recovered from the bottom of a shoe which was very similar to the known glass in all except one ratio, there is the possibility that the known sample was collected in a way that did not provide appropriate characterization of the known sample. Or the possibility exists that a small contaminant particle was not able to be removed or avoided in analysis. The wording in the method allows for considerations such as these.

With respect to the statistical properties of ratio range overlap, the statement that is present acknowledges that the confidence level is not directly addressed. However, research has shown that it is an appropriate method based on the goal of minimizing Type I and Type II errors. See general comments provided about the ASTM requirements for listed references.

Section 10.7.3.2 6 adopts a 3-standard-error rule. It reads

> For each elemental ratio, compare the average ratio for the questioned specimen to the average ratio for the known specimens ±3s. This range corresponds to 99.7 % of a normally distributed population. If, for one or more elements, the average ratio in the questioned specimen does not fall within the average ratio for the known specimens ±3s, it may be concluded that the samples are not from the same source.

I have puzzled over these sentences for hours without being able to understand them. Is this a decision rule based on a desired 99.7% confidence interval for true mean of the ratio in a homogenous known glass sample? If so, it does not account for the fact that with a standard error estimated from a small sample, one needs a larger interval to achieve 99.7% confidence. In addition, the usual (and better) way to test whether two sample means are different is to use the sampling distribution of the difference between the sample means rather than the sampling distribution of only one of the sample means. Furthermore, even with the proper test statistic and distribution, the many separate tests (one for each ratio Ca/Mg, Ca/Ti, Fe/Zr, etc.) cloud the interpretation of the significance of the difference in a pair of sample means. The risk of a false exclusion for, say, ten comparisons could be ten times the nominal value of 0.003. Thus, the section should be rewritten to justify the choice of the nominal level and to indicate how the nominal level relates to the actual level. In other words, why the 3σ rule? Is it supposed to keep the risk of a false exclusion to a low level?

This comment is non-persuasive. (10.7.3.2): Research has shown that 3s is an appropriate method for elemental comparison of glass by u-XRF based on the goal of minimizing Type I and Type II errors. See general section comments for a detailed explanation on match criteria.

Although these questions may seem technical, they are directly related to the interpretation of the results in the criminal justice system. From a legal perspective, are not false inclusions the type of error that should be guarded against more assiduously? And even if 3σ is the right rule here, why is the standard for making associations via elemental compositions in E2330 some kind of 4σ rule? Without reconciling the different standards, their value as justifications for interpretations of test results in the legal system could be jeopardized.

This comment is non-persuasive. Comment 1: u-XRF is just a part of the overall scheme of glass analysis. Each test assesses properties that are known to be variable among the overall population of glass. Each test is an attempt to find out if there are verifiable and distinguishable differences between two or more samples of glass. At the end, there is typically either an elimination (the Q glass did not originate from the K glass source) or the inability to distinguish the glass samples (the K glass source is a possible source of the glass). In the latter case, there are other possible sources of the Q glass. There is no identification of source using class properties. It is circumstantial evidence which may have explanations other than hypothesis that the Q glass came from the K glass source. The u-XRF testing (or any one test) cannot address comparisons to all other sources in the world, but rather (typically) only to the submitted and tested evidence. As such, it is accurate to say that based on the tests conducted, the source of K glass is a possible source of the Q glass if it can't be distinguished. False inclusions are guarded against,

but class evidence has, by its nature, the possibility of other sources being the actual source. The significance of these other possibilities is beyond the scope of this standard test method.

Comment 2: The match criteria are different because they are different tests and, as such, have different levels of precision and testing protocols. In addition, LA-ICP-MS is a quantitative method, while XRF is not. Research was conducted in order to discover which match criteria would minimize Type I and Type II errors (see general comments section).

Assuming that the exclusion-inclusion decision is the best way to interpret the differences in "peak intensity," in discussing the interpretation of the data, the Standard needs to offer guidance about the probative value of an inclusion. Should the analyst report that the questioned fragment might have come from the known glass *or from any other glass with a similar set of elemental concentrations*? What data are there on the population distribution of these statistics? If there are none, what can or should the analyst report?

Because u-XRF is only one of a series of tests conducted on glass, and the data from each test is collectively addressed by the examiner, it is inappropriate to provide guidance to report wording for glass that is indistinguishable by XRF. This is outside the scope of the method. Refer to the general comments provided for information regarding statistical methods of interpretation.

Part 11 (Precision and Bias)

These quantities should be defined within the Standard itself.

This comment is non-persuasive. Definitions of particular technical terms are addressed in other ASTM documents. The precision and bias for u-XRF of glass within this standard test method in Section 11.4 and the Appendix.

References

To show a legal audience that the Standard is based on a complete review of the scientific and statistical literature, there should be references to studies that help demonstrate the value of the testing in forensic investigations. The Standard should show how it flows from and is supported by a body of cited scientific studies. The Subcommittee commendably listed 15 papers in the Technical Merit Worksheet of 7/1/15, but readers in the legal community will not know of them.

(Could this be remedied by having an appendix that justifies key choices made in the Standard placed on the registry along with the Standard? This appendix could explain which publications support which choices and how they do so. It would help prevent members of the legal community from misjudging the Standard as the kind of "ipse dixit" condemned in *General Electric v. Joiner*, 522 U.S. 136 (1997). Evaluating this document together with the body of the Standard would permit readers who are not already technical experts in the field to judge its readiness for addition to the registry. Presumably, this explanatory document would not have to be approved by ASTM.)

This comment is non-persuasive. The intended audience of these ASTM documents is the forensic practitioners or scientists that conduct the glass analysis. See general comments section regarding reference/citation policies for ASTM methods.

Concluding Comments

Recommendations about the opinions and conclusions that analysts can reach and how they should present them in reports and in court should be made in light of current thinking about the methods for interpreting and evaluating evidence across the entire domain of forensic science. The premise of the Standard is that the expert's task is to decide whether a source hypothesis is true or false. Would a likelihood ratio be a better way to express the probative value of the data? Certainly, there is an argument to that effect in the legal and forensic science literatures. See, e.g., Colin Aitken & Franco Taroni, Statistics and the Evaluation of Evidence for Forensic Science (2d ed. 2004); James M. Curran et al., Forensic Interpretation of Glass Evidence (2000); ENFSI Guideline for Evaluative Reporting in Forensic Science (2015); David H. Kaye et al., The New Wigmore: Expert Evidence (2d ed. 2011); Royal Statistical Soc'y Working Group on Statistics and the Law, Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses (2010).

These issues have been addressed in the above comments.

In the end, the subcommittee, the SAC and the FSSB may conclude that only the categorical decision framework that is recommended in this document is acceptable for the evaluation and explanation of the evidence in court. If that is their conclusion, however, some of the reasoning behind the conclusion should be provided to assist the legal community in using the measurements wisely and fairly.

2. Drafting Problems

Introduction

Should the last sentence in the introduction read "Measuring elemental concentrations with micro X-ray fluorescence spectrometry permits high discrimination among different sources of glass"?

This comment is non-persuasive. Minor editorial changes will be addressed in the document during the ASTM revision process.

Part 1 (Scope)

I suggest rewording 1.1 to read as follows: "This Standard concerns μ-XRF analysis using mono- and poly- capillary optics, and an energy dispersive X-ray detector (EDS) for the determination of the concentrations of major, minor, and trace elements in glass fragments." This change would permit 1.2 to be deleted.

This comment is non-persuasive. Minor editorial changes will be addressed in the document during the ASTM revision process.

1.3 states that "This test method does not replace knowledge, skill, ability, experience, education, or training and should be used in conjunction with professional judgment." On its face, this seems to assert that analysts can ascertain elemental composition without using any instruments or that an analyst can depart from one of the prescribed statistical rules in an ad hoc manner. The sentence should be clarified (or deleted on the theory that it goes without saying that it takes skill, experience, and judgment to perform the analysis).

This is not the intended meaning of the statement. See general comments provided.

1.4 reads: "The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard." The meaning of "to be regarded as standard" is not immediately apparent. Why is it necessary to say that no units other than the International System of Units are used? Is not that apparent from reading the Standard as a whole? (Of course, if a Standard would not be expected to use these units, then referring to this choice at the outset is helpful.) Moreover, this remark does not really describe the scope of the Standard, but only how things within its scope are expressed.
See general comments section.

The disclaimer in section 1.5 does not seem to achieve any legal objective. It is hard to see how any reader would think that a "method . . . purport[s] to address all of the safety concerns," and stylistically, the phrase "all of the safety concerns, if any" also is awkward, and the "if any" phrase is contradicted by the fact that Section 7 does address a safety concern.

Sections 1.4 and 1.5 are required by ASTM methods.

Part 10 (Calculation and Interpretation of Results)

What does it mean to "correct" sum peaks and escape peaks? That the analyst should label them as such?

Correcting sum and escape peaks is recognizing and labeling them, as some software programs may assign the peaks to a different element.

The following members of the LRC agree with the comments made by David Kaye:

Barry Scheck and David Moran join in David Kaye's comments.

Ron Reinstein joins in David Kaye's comments except the comment regarding Section 1.3. Judge Reinstein believes this section should be clarified but not deleted. Judgment, training and experience are important and must be used in conjunction with the test method (but not in place of it).

These issues have been addressed in the above comments.

Judge Reinstein and Lynn Garcia would like to emphasize the importance of David Kaye's "Concluding Comments" on page 6—this is the type of comment that the SACs and FSSB should pay particular attention to for standards that are to be included on the OSAC Registry.

Additional Comments by LRC Member Barry Scheck:

With respect to David Moran's comments, Barry Scheck would like to emphasize the term "semi-quantitative" is troubling. Either you have data or you don't and the measure uncertainty can be calculated within acceptable limits. If you are going to rely on "experience" to declare exclusions or inclusions (which seems problematical in the first place) the standard should specify when and how that would be done and any validation that justifies it.

This comment is non-persuasive. "Semi-quantitative" is not implying partial data. The term is used because the concentrations of elements are not being quantified, but rather, the peak areas (which correspond to concentrations even when those concentrations aren't explicitly known) are used in ratios to one another. See general comments section for more details.

Barry Scheck would also like to emphasize concerns expressed in the comments by David Kaye that the deficiencies in the statistical explanations are 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. The requirement of general acceptance in the scientific community, particularly among statisticians, cannot be met, nor the requirements of clearly identifying limitations and weaknesses in the methodology or an explanation of how it is "fit for purpose."

These issues have been addressed in the above comments.

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.***