

An interlaboratory study evaluating the interpretation of forensic glass evidence using refractive index measurements and elemental composition

Ruthmara Corzo^{a,b}, Tricia Hoffman^a, Troy Ernst^c, Tatiana Trejos^d, Ted Berman^e, Sally Coulson^f, Peter Weis^g, Aleksandra Stryjnik^h, Hendrik Dorn^h, Edward “Chip” Pollockⁱ, Michael Scott Workman^j, Patrick Jones^k, Brendan Nytes^l, Thomas Scholz^m, Huifang Xieⁿ, Katherine Igowsky^o, Randall Nelson^p, Kris Gates^q, Jhanis Gonzalez^r, Lisa-Mareen Voss^g, Jose Almirall^{a,*}

^a Florida International University, Miami, FL, United States

^b National Institute of Standards and Technology, Gaithersburg, MD, United States

^c Michigan State Police, Grand Rapids Forensic Laboratory, Grand Rapids, MI, United States

^d West Virginia University, Morgantown, WV, United States

^e Florida Department of Law Enforcement, Orlando, FL, United States

^f Institute of Environmental Science and Research, Auckland, New Zealand

^g Bundeskriminalamt, KT 42, Wiesbaden, Germany

^h Centre of Forensic Sciences, Toronto, Canada

ⁱ Sacramento County DA's Office, Laboratory of Forensic Services, Sacramento, CA, United States

^j Missouri State Highway Patrol Crime Lab, Jefferson City, MO, United States

^k Kansas City Police Crime Laboratory, Kansas City, MO, United States

^l Microtrace LLC, Elgin, IL, United States

^m Landeskriminalamt Sachsen, Dresden, Germany

ⁿ Health Sciences Authority, Singapore

^o Minnesota Bureau of Criminal Apprehension, St. Paul, MN, United States

^p Tennessee Bureau of Investigation, Nashville, TN, United States

^q Oregon State Police Forensic Laboratory, Clackamas, OR, United States

^r Applied Spectra Incorporated, West Sacramento, CA, United States

ARTICLE INFO

Keywords:

Forensic glass analysis
Interlaboratory study
XRF
LIBS
Refractive index

ABSTRACT

Seventeen laboratories participated in three interlaboratory exercises to assess the performance of refractive index, micro X-ray Fluorescence Spectroscopy (μ XRF), and Laser Induced Breakdown Spectroscopy (LIBS) data for the forensic comparison of glass samples. Glass fragments from automotive windshields were distributed to the participating labs as blind samples and participants were asked to compare the glass samples (known vs. questioned) and report their findings as they would in casework. For samples that originated from the same source, the overall correct *association* rate was greater than 92% for each of the three techniques (refractive index, μ XRF, and LIBS). For samples that originated from different vehicles, an overall correct *exclusion* rate of 82%, 96%, and 87% was observed for refractive index, μ XRF, and LIBS, respectively. Special attention was given to the reporting language used by practitioners as well as the use of verbal scales and/or databases to assign a significance to the evidence. Wide variations in the reported conclusions exist between different laboratories, demonstrating a need for the standardization of the reporting language used by practitioners. Moreover, few labs used a verbal scale and/or a database to provide a weight to the evidence. It is recommended that forensic practitioners strive to incorporate the use of a verbal scale and/or a background database, if available, to provide a measure of significance to glass forensic evidence (i.e., the strength of an association or exclusion).

* Corresponding author.

E-mail address: almirall@fiu.edu (J. Almirall).

<https://doi.org/10.1016/j.forc.2021.100307>

Received 3 June 2020; Received in revised form 21 December 2020; Accepted 5 January 2021

Available online 12 January 2021

2468-1709/© 2021 Published by Elsevier B.V.

1. Introduction

The widespread use of glass in everyday items makes glass an important type of forensic trace evidence that is frequently encountered in crime scenes [1,2]. During the commission of a crime, glass can shatter, and fragments can transfer to an object or person. Forensic glass examiners are tasked with comparing glass fragments from a questioned (Q) source to fragments that originated from a known (K) source to determine if they can be discriminated by their physical, optical, or chemical properties. If exclusionary differences are recognized in any of the observed or measured properties, it can be concluded that the sources of the samples are distinguishable. If no exclusionary differences are recognized in these properties, it can be concluded that the sources of the samples are indistinguishable by the techniques that were applied.

In forensic casework, glass fragments are often compared by the optical property of refractive index (RI). Several studies have reported the utility of RI for the discrimination of glass samples originating from different sources [3,4]. However, a narrowing of RI distributions over time has been reported as the more limited variation in RI observed in modern flat glass has been attributed to changes and improvements in the glass manufacturing process [4,5]. As such, RI provides relatively limited discrimination capabilities for modern flat glass. The standard test method ASTM E1967 [6] outlines the procedure for the automated determination of RI using the oil immersion method. At the time of the inter-laboratory study, ASTM E1967 did not recommend a match criterion for the pairwise comparisons of the K and Q samples. However, a recent update to the standard method recommends that the average of the Q measurements is compared to the range of the K measurements, and if the Q average falls within the K range, the K and Q are considered indistinguishable [7]. Other comparison criteria have also been evaluated; Garvin and Koons reported the false exclusion rates of eight match criteria using RI measurements for five sheets of float glass [8]. The authors did not recommend a particular match criterion but rather reported that the false exclusion rates depended on the number of replicate measurements collected for the K and Q samples, so the examiner can take that into consideration when selecting a comparison criterion.

Glass fragments may also be compared by their elemental composition. Elemental analysis has been reported to provide superior discrimination capabilities compared to RI measurements [2,3,9–11]. Usually, micro X-ray Fluorescence Spectroscopy (μ XRF) or Laser Ablation-Inductively Coupled Plasma-Mass Spectrometry (LA-ICP-MS) is used for the elemental analysis of glass in forensic casework. The superior sensitivity and quantitative analysis capability of LA-ICP-MS results in excellent discrimination of glass sources [12–17]. Despite its inferior detection limits as compared to LA-ICP-MS, μ XRF also provides excellent discrimination of glass sources [3,13,14,16]. Moreover, μ XRF is a nondestructive analytical technique, whereas LA-ICP-MS requires the removal of small amounts (μ g) of the sample. Naes et al. reported on very similar discrimination for the same set of glass samples when using LA-ICP-MS, μ XRF, and Laser Induced Breakdown Spectroscopy (LIBS) [16]. In the United States, many more forensic laboratories are equipped with μ XRF in comparison to LA-ICP-MS or LIBS; thus, μ XRF is more commonly used for the elemental analysis of glass in forensic casework. A standard test method is available for μ XRF (ASTM E2926) [18]. In ASTM E2926, spectral overlay comparisons of the K and Q are recommended. If the K and Q spectra are visually indistinguishable, a semi-quantitative approach using elemental ratios can be used where the elemental ratios are compared using a range overlap and/or a ± 3 standard deviations ($\pm 3s$) comparison interval. LIBS is a relatively new tool for the elemental analysis of glass in forensic laboratories. LIBS has been reported to provide discrimination capabilities that are similar to those obtained with LA-ICP-MS and μ XRF [9,16,19,20]. However, there is currently no standard test method for the forensic analysis of glass using LIBS.

Although there are standard test methods available for the measurement of RI and elemental composition using μ XRF (as well as LA-

ICP-MS and solution ICP-MS), there is currently no consensus on how to interpret the weight of the evidence by any method. One approach to assigning weight to forensic evidence is the use of a verbal scale (also referred to as a conclusion scale or an association scale). Verbal scales differ across laboratories, but most include an exclusion, an inconclusive result, and multiple levels for the strength of an association based on the presence or absence of individual or class characteristics as well as the discrimination power of the technique(s) used. It is worth noting that the National Institute of Standards and Technology (NIST) Organization of Scientific Area Committees (OSAC) for forensic science is currently developing a standardized verbal scale for trace evidence interpretation [21]. The use of databases is prerequisite to providing a weight to forensic evidence.

Databases have been used for several purposes, including: the estimation of a technique's discrimination power and experimental error rates, which are necessary for the development of a verbal scale; the calculation of a random match probability (RMP), which is often estimated by the false inclusion rate for all pairwise comparisons within the database [15,17,22]; the calculation of a frequency of occurrence, in which the K or the Q sample is compared to each sample in the database, and the number of indistinguishable pairs is reported [22]; and the calculation of a likelihood ratio (LR) [22,23]. As previously reported in Corzo et al. [22] "a continuous likelihood ratio (LR) is defined as the probability of the evidence given H_1 (an association) divided by the probability of the evidence given H_2 (a non-association): $LR = P(E|H_1)/P(E|H_2)$. Therefore, if the LR is greater than 1, it shows support for the hypothesis of an association (H_1) and if the LR is <1 , it shows support for the hypothesis of no association (H_2). Furthermore, the larger the LR the stronger the support for H_1 , while the smaller the LR the stronger the support for H_2 ." Several researchers and operational laboratories (mostly in Europe, Australia, and New Zealand) incorporate the use of LR in the evaluation of glass evidence in casework [22–27].

The present work is part of an ongoing study by the Glass Interpretation Working Group that focuses on the development of a standardized and more objective method for evidence interpretation. This working group included an interlaboratory study, organized by the Admirall group at Florida International University, in which glass fragments were compared using RI, μ XRF, LIBS, and/or LA-ICP-MS. The LA-ICP-MS results have been presented in previous manuscripts and will not be discussed in detail here [22,23]. This manuscript summarizes the results of the RI, μ XRF, and LIBS comparisons for the three interlaboratory exercises involving the 17 participating laboratories for the first time. Particular attention is given to the evaluation of the current state of glass evidence interpretation within the forensic community. Finally, recommendations for the improvement of glass evidence interpretation are provided based on the interlaboratory study findings.

2. Materials and methods

2.1. Description of specimens submitted to participants

All glass samples included in this interlaboratory study were selected from a collection of windshield glass at Florida International University (FIU). The FIU collection includes laminated glass from 210 vehicles manufactured between 2004 and 2017. The entire FIU collection (inner and outer pane of each windshield) was previously characterized using LA-ICP-MS. A description of the LA-ICP-MS analysis conducted at FIU is provided in a previous manuscript [22]. Table 1 lists the vehicle information for the glass samples distributed to all participants for each interlaboratory exercise in blind studies. Participants also received the NIST standard reference material (SRM®) 1831 to serve as a quality control sample, as recommended in ASTM E2926.

A subset of the FIU collection (31 samples), selected based on their similarity in elemental composition, was analyzed at West Virginia University (WVU) and at NIST. From this subset of 31 samples, 22 pairs of samples originated from different vehicles (but same make, model,

Table 1

Vehicle information for glass samples submitted to participating laboratories for each interlaboratory exercise.

Interlaboratory Exercise 1					
Sample	Make	Model	Year	VIN	Remarks
K1 _{inner & outer}	Mitsubishi	Galant	2009	4A3AB36F39E024088	Same make/model/year as Q1
K2 _{inner & outer}	Subaru	Impreza	2008	JF1GE61658H503418	K2 outer same as Q2
Q1 _{outer}	Mitsubishi	Galant	2009	4A3AB36F39E029145	Same make/model/year as K1
Q2 _{outer}	Subaru	Impreza	2008	JF1GE61658H503418	Same as K2 outer
Interlaboratory Exercise 2					
Sample	Make	Model	Year	VIN	Remarks
K1 _{inner & outer}	Honda	Civic	2006	2HGFG21506H707035	K1 inner and outer same as Q1
Q1 _{inner & outer}	Honda	Civic	2006	2HGFG21506H707035	2 fragments same as K1 inner; 1 fragment same as K1 outer
Q2 _{outer}	BMW	2 Series	2014	WBA1F5C58EVV98871	Different than K1
Interlaboratory Exercise 3					
Sample	Make	Model	Year	VIN	Remarks
K1 _{inner & outer}	Honda	Civic	2007	2HGFG12607H511521	K1 outer same as Q1
Q1 _{outer}	Honda	Civic	2007	2HGFG12607H511521	Same as K1 outer
Q2 _{outer}	Mercedes	R-Class	2009	4JGCB65E59A094913	Different than K1

and year of manufacture) and were indistinguishable using LA-ICP-MS. This collection set primarily served to evaluate the complementarity of RI and elemental analysis techniques (μ XRF and LA-ICP-MS). Additionally, the subset provided a *relative* comparison of the discrimination power of RI, μ XRF, and LA-ICP-MS. Micro-XRF and LA-ICP-MS were selected for this study since these are the two elemental techniques that are typically used in casework; similarly, LIBS was excluded from this study since it is not commonly used in casework and since there is currently no standard methodology for the analysis of glass using LIBS. It should be noted that, since this subset was selected based on the chemical similarity of the samples, it does not provide an accurate estimation of the discrimination power for a *random* population of glass. For this reason, an additional randomly selected subset of 50 samples from the FIU collection was analyzed using RI and LA-ICP-MS by the Bundeskriminalamt (BKA) in Germany. This collection set provided a more accurate estimate of the discrimination power of RI and LA-ICP-MS for a random population of float glass samples. For LA-ICP-MS analysis, the following elements were measured: Li, Mg, Al, K, Ca, Fe, Ti, Mn, Rb, Sr, Zr, Ba, La, Ce, Nd, Hf, Pb, Na, and Si (used as the internal standard for quantitative analysis) [28].

2.2. Instrumentation and data analysis

A total of 11 laboratories provided RI measurements for at least one inter-laboratory test. Seven labs provided μ XRF measurements for at least one inter-laboratory test. Table 2 provides the specifications for the XRF instruments used in this inter-laboratory study. Finally, 5 labs provided LIBS measurements for at least one inter-laboratory test. Table 3 provides the instrument specifications for the LIBS instruments used in this inter-laboratory study.

Participants were provided a Microsoft Excel template spreadsheet that compared the known and questioned samples using either a range overlap or a ± 3 s match criterion, as specified in ASTM E2926 for the forensic comparison of glass samples using μ XRF. Since there is no standard that recommends a particular match criterion for RI or LIBS

Table 2

Specifications for XRF instrumentation used in this interlaboratory study.

	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
Manufacturer	EDAX	EDAX	EDAX	EDAX	EDAX	EDAX	JEOL
Model	Orbis	Eagle III	Orbis	Orbis	Orbis	Orbis	IXRF coupled to JSM6490-LV + INCA x-sight EDS
Tube Material	Rh	Rh	Rh	Mo	Rh	Rh	Rh
Capillary Optics	Mono	Mono	Poly	Poly	Poly	Mono	Mono
Spot Size (μ m)	300	100	30	30	30	1000	100
Beam Power (V)	50	50	50	50	50	50	50
Beam Current (μ A)	200	1000	150	40	150	100	1000
Acquisition Time (live s)	1300	2000	600	1200	600	500	1200

Table 3

Specifications for LIBS instrumentation used in this interlaboratory study.

	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5
Manufacturer	Applied Spectra	Applied Spectra	Applied Spectra	Applied Spectra	Applied Spectra
Model	J200	J200	J200	J200	J200
				Tandem LA-LIBS	Tandem LA-LIBS
Laser	Nd:YAG	Nd:YAG	Nd:YAG	Nd:YAG	Nd:YAG
Laser Wavelength (nm)	266	266	266	266	266
Laser Energy (mJ)	5.5	7.2	8.8	5.0	5.3
Laser Frequency (Hz)	10	10	10	10	10
Spot Size (μ m)	100	100	100	30	50
Gate Delay (μ s)	0.5	0.5	1.05	0.1	0.5

comparisons, participants were instructed to use their preferred match criterion and include their results in a separate sheet within the template spreadsheet. Pairwise comparisons between labs and pairwise comparisons for the subset of 31 samples from the FIU collection were conducted using an RStudio (version 1.0.143) routine.

3. Results and discussion

3.1. Interlaboratory exercise 1

In the first interlaboratory exercise, two known samples (K1 and K2, inner and outer windshield pane for each K) and two questioned samples (Q1 and Q2) were submitted to each participating laboratory. Three glass fragments were submitted for each specimen (i.e., three fragments for each K inner pane, each K outer pane, and each Q). All glass

fragments submitted to participants were full-thickness fragments. K2 outer and Q2 originated from the same source and should therefore be determined as an association. K1 and Q1 originated from different vehicles but these were of the same make, model, year of manufacture, and even manufactured in the same plant. Participants were instructed to analyze the samples using the procedure their laboratory normally follows for casework specimens. A total of eight laboratories collected measurements, four laboratories collected XRF measurements, and three laboratories collected LIBS measurements.

Table 4 summarizes the false exclusion and false inclusion rates for each technique. Note that these are calculated from analysis of glass from just three windshields using instrumentation at several laboratories. Two of the three windshields were from vehicles of the same make, model, and year of manufacture. Therefore, these figures do not represent false inclusion rates for these techniques for random populations of glass. Note also that comparisons between the inner K2 pane and Q2 have been excluded since these were comparisons between opposite panes of the same windshield (see Table 1), and the two panes were chemically indistinguishable. It is likely that the inner and outer pane of K2 were produced close in time within the same manufacturing plant. Since it is unknown whether these two panes are same-source or different-source panes, excluding the K2 inner/Q2 comparisons avoided overestimating the false inclusion rate.

3.1.1. Refractive index results (interlaboratory exercise 1)

Although ASTM E1967 outlines the procedure for measuring RI using the oil immersion method and a phase contrast microscope, at the time of the inter-laboratory study, the method did not recommend a match criterion for the comparison of K and Q samples. Of the eight participating labs, five used a ± 3 s match criterion; that is, the mean of the Q sample was compared to the mean ± 3 standard deviations of the K sample. Two of these five labs established a minimum standard deviation ($s = 0.00003$ or 0.000033) to reduce the risk of false exclusions in the case where the standard deviation of the measurements is very small. One lab used a fixed ± 0.0001 approach (Q mean compared to K mean ± 0.0001). One lab used a range overlap to compare the K and Q samples and the final lab used the *t*-test to compare the K and Q samples.

When comparing the outer pane of K2 with Q2 (same source), all eight labs correctly associated the pair using RI measurements, resulting in no false exclusions out of 24 same-source pairwise comparisons (8 labs \times 3 K outer vs. Q2 pairs = 24 pairwise comparisons). However, the false inclusion rate was high (34.1%). Note that although there were 144 possible different-source pairwise comparisons (8 labs \times 3 Q fragments per Q sample \times 6 possible different-source K/Q pairs), the total number

Table 4

Interlaboratory exercise 1 false exclusion and false inclusion rates for refractive index, XRF, and LIBS; the number of participating labs (N) is indicated next to each technique. Fractions in parentheses indicate the number of false exclusions/inclusions divided by the total number of pairwise comparisons.

	False exclusion (K2 _{Outer} /Q2)	False inclusion (K1/Q1, K1/Q2, K2/Q1)	Remarks
Refractive Index (N = 8)	0 (0/24)	34.1% (43/126)	38 false inclusions were comparisons between vehicles with different make; 5 false inclusions were comparisons between similar (make/model/year) vehicles
XRF (N = 4)	0 (0/12)	8.3% (6/72)	All false inclusions were comparisons between similar (make/model/year) vehicles
LIBS (N = 3)	0 (0/9)	16.7% (9/54)	All false inclusions were comparisons between similar (make/model/year) vehicles

of different-source pairwise comparisons was actually 126; this is because some labs eliminated certain K/Q pairs based on thickness measurements and, therefore, decided not to measure RI.

Most labs were unable to find reproducible differences when comparing K2 inner/outer and Q1, despite that K2 and Q1 originated from vehicles of different makes (Subaru vs. Mitsubishi) and manufactured in different years (2008 vs. 2009). On the other hand, five out of eight labs found reproducible differences when comparing K1 outer and Q1, two samples that originated from very similar vehicles (same make, model, year of manufacture, and manufacturing plant). One lab reported an inconclusive for the K1 outer/Q1 pair due to a high relative standard deviation (RSD) in the measurements. The remaining two labs reported that K1 outer/Q1 were indistinguishable; one of these labs found differences when using their match criterion (*t*-test) but decided to report an indistinguishable result because of the small standard deviation in the measurements.

3.1.2. μ XRF results (interlaboratory exercise 1)

Unlike ASTM E1967 (RI), ASTM E2926 (μ XRF) suggests three methods to compare K and Q samples: spectral overlay, range overlap of element ratios, and/or a ± 3 s match criterion using element ratios. Most labs begin with spectral overlay and, if no differences are observed, move on to a semi-quantitative approach using element ratios. For spectral overlay comparisons, ASTM E2926 suggests a signal-to-noise ratio (SNR) of at least 3 to determine whether an element is detectable. For element ratio comparisons, ASTM E2926 suggest a signal-to-noise ratio of at least 10. These values correspond to limits of detection (LOD) and limits of quantitation (LOQ), respectively, and are further described for μ XRF analysis of glass by Ernst et al. in [29].

All four participating labs correctly associated Q2 with K2 outer using μ XRF measurements, resulting in no false exclusions. The false inclusion rate was 8.3%. Two of four labs distinguished K1 outer and Q1 based on the Ca/Fe ratio. The third lab noted subtle differences between K1 outer and Q1 (also based on the Ca/Fe ratio), which they indicated suggests an exclusion, but the lab stated that in a casework scenario they would request additional K samples to confirm the differences. The fourth lab reported an indistinguishable result; however, this lab only used spectral overlay for comparisons because of poor SNR for certain elemental peaks (SNR < 10). It is worth noting that most LA-ICP-MS participants were unable to find reproducible differences for the K1 outer/Q1 pair: four out of five labs reported an association and the remaining lab reported an inconclusive result due to a high relative standard deviation in the measurements (see reference [22] for the LA-ICP-MS inter-laboratory results). However, the K1 outer/Q1 pairs that were excluded using μ XRF had a low relative standard deviation for the Ca/Fe ratio (RSD < 1.2%). The LA-ICP-MS match criterion (ASTM E2927) applies a minimum 3% RSD to reduce the risk of false exclusions, while the current μ XRF match criterion does not apply a minimum RSD [14,28]. The modified match criterion for LA-ICP-MS was selected based on the best compromise between false exclusion and false inclusion rates; the wider interval was a consequence of the better precision of the LA-ICP-MS measurements (typical RSD < 2%) [14]. However, it should be noted that a larger number of known fragments may be collected to better account for the inherent heterogeneity of the glass source, which would also reduce the risk of false exclusions.

3.1.3. LIBS results (Interlaboratory exercise 1)

There is currently no ASTM method for the forensic analysis of glass using LIBS, thus the three labs that provided LIBS measurements followed their own procedure. All labs used a ± 3 s match criterion to compare K/Q pairs using element ratios, though, the element ratios and the element emission lines used for the comparisons differed between labs. The following elements were reported by the participating labs: Si, Al, Mg, Ca, Na, K, Sr, Fe, Ti, and Ba.

All three participating labs correctly associated Q2 with K2 outer using LIBS measurements. The false inclusion rate was 16.7%. Two labs

associated K1 outer and Q1 (outer pane of similar vehicles); the remaining lab reported an inconclusive result because the Q mean was close to the ± 3 s cutoff and the standard deviation of the measurements was very small (relative standard deviation $\leq 1.8\%$). One lab could not distinguish between K1 inner and Q1, but the remaining two labs discriminated the pair.

3.1.4. Interpretation of results (Interlaboratory exercise 1)

In the case where physical and/or chemical differences were found between the K and Q sample, all labs reported an elimination (or exclusion); that is, every lab that submitted a report explicitly stated that the Q sample did not originate from the K source.

For pairs that were indistinguishable, the participating labs varied in their level of support for an association. Three labs simply reported that the K and Q sample “could not be eliminated.” Most labs used some variation of the statement “the Q sample originated from the K source, or another source with the same physical and chemical properties.” Two labs included an association scale in their report. One of these two labs incorporated an association scale with 5 levels in their report; the K2 outer/Q2 pair was given a level 3 association based on RI and μ XRF measurements. According to the scale, a level 3 association indicates that the items (i.e., K and Q) are “consistent in observed and measured physical properties and/or chemical composition” and “could have originated from the same source;” it is noted that an “individual source cannot be determined” since there are other manufactured items that would also be indistinguishable from the submitted evidence. The second lab estimated the RI frequency using a national RI database that currently includes 9,559 entries [30] and a conservatively estimated RMP for LA-ICP-MS [17]. To estimate the RMP, each sample in the database is compared to every other sample in the database. The RMP is then calculated by dividing the number of indistinguishable pairs by the total number of pairwise comparisons (with this definition, the RMP is equivalent to the false inclusion rate). However, the lab does not currently report a combined (RI and elemental) abundance in casework because RI and chemical composition cannot be assumed to be independent of each other. Finally, the lab reported their conclusion based on a conclusion scale with 5 association levels: the lab stated that the results provide “very strong support for a common source.”

3.2. Comparison of refractive Index, μ XRF, and LA-ICP-MS

As noted in the results for the first interlaboratory exercise, the K1 outer/Q1 pair (from vehicles with the same make, model, and manufacturing year) was discriminated based on RI and μ XRF measurements, while reproducible differences could not be found using LA-ICP-MS. Thus, a subset of windshield glass (31 samples) from the FIU vehicle collection, 22 pairs of which were indistinguishable by LA-ICP-MS, was selected for RI and μ XRF analysis to further explore the complementarity of RI and the two elemental techniques typically used in casework (μ XRF and LA-ICP-MS). RI was able to distinguish 3 pairs that were indistinguishable with both LA-ICP-MS and μ XRF. Micro-XRF distinguished one pair that was indistinguishable with LA-ICP-MS and RI. This pair was excluded on the basis of the Ca/Ti ratio; however, it should be noted that the two samples being compared had a small standard deviation for Ca/Ti (RSD $< 1\%$), which led to a narrow comparison interval.

Although RI was able to distinguish pairs that were indistinguishable with μ XRF and LA-ICP-MS, the elemental techniques still provided superior discrimination. The 31 samples were compared using the ASTM match criterion for LA-ICP-MS ($\pm 4s$, with a minimum of 3% RSD), the ASTM match criterion for μ XRF ($\pm 3s$), and a ± 3 s (with a minimum s of 0.00003) match criterion for RI. Note that all three match criteria are asymmetrical. That is, when comparing two samples, a different result may be obtained depending on which sample is treated as the known sample and which is treated as the questioned sample. Thus, each pair of samples was compared twice, so that each sample could be treated as the

known and the questioned sample. A sample pair was considered distinguishable if both comparisons resulted in an exclusion. Similarly, a pair was considered indistinguishable if both comparisons resulted in an association. Finally, a pair with mismatching results (one comparison resulted in an exclusion and the reverse comparison resulted in an association) was considered inconclusive. Of the 465 possible different-source pairwise comparisons ($31 \times 30 \div 2 = 465$) for this sample set, LA-ICP-MS provided 95.3% discrimination, μ XRF provided 93.3% discrimination, and RI provided 88.6% discrimination. However, it should be stressed that the discrimination power for this subset is unrealistically low since, as noted earlier, the samples in this subset were chosen based on their similarity in elemental composition. For a more realistic measure of discrimination power, the BKA analyzed a subset of 50 randomly selected samples from the FIU vehicle collection using LA-ICP-MS and RI. A modified ± 4 s match criterion [17] which is symmetrical, was used for the LA-ICP-MS comparisons, and a ± 3 s (with a minimum s of 0.00003) was used for the RI comparisons. LA-ICP-MS provided 99.8% discrimination, while RI provided 94.7% discrimination. The two techniques combined provided 99.9% discrimination (RI discriminated two pairs that were indistinguishable with LA-ICP-MS).

3.3. Interlaboratory exercise 2

In the second interlaboratory exercise, one known sample (K1, inner and outer pane) and two questioned samples (Q1 and Q2) were submitted to each participating laboratory. Three fragments were submitted for each K pane and each Q. Two Q1 fragments originated from the inner pane of K1 and one Q1 fragment originated from the outer pane of K1. The known sample was selected because the inner and outer pane were chemically distinguishable (based on LA-ICP-MS measurements). Thus, two Q1 fragments should be associated with the inner pane, but not the outer pane, and one Q1 fragment should be associated with the outer pane, but not the inner pane. The purpose of this second exercise was to evaluate whether practitioners would increase their level of support based on the presence of Q specimens that were associated with both chemically distinct K1 panes. A total of eight laboratories collected RI measurements, six laboratories collected XRF measurements, and two laboratories collected LIBS measurements.

Unlike in the first interlaboratory exercise, a case scenario was provided to participants in the second interlaboratory exercise: the hit-and-run scenario stated that a 2006 Honda Civic (K1) was found with a broken windshield; three questioned fragments (Q1) were found on the clothing of suspect 1 and three fragments (Q2) were found on the clothing of suspect 2. To better approximate realistic casework specimens, the questioned specimens were submitted as small (≈ 1 mm), irregular fragments; the known specimens were provided as full-thickness samples. Participants were asked to follow ASTM E1967 (for RI) and ASTM E2926 (for μ XRF). Table 5 summarizes the false exclusion and false inclusion rates for each technique.

Table 5

Interlaboratory exercise 2 false exclusion and false inclusion rates for refractive index, XRF, and LIBS; the number of participating labs (N) is indicated next to each technique. Fractions in parentheses indicate the number of false exclusions/inclusions divided by the total number of pairwise comparisons.

	False Exclusion (Two K1 _{Inner} /Q1 and one K1 _{Outer} / Q1)	False Inclusion (K1/ Q2, one K1 _{Inner} /Q1 and two K1 _{Outer} /Q1)	Remarks
Refractive Index (N = 8)	0 (0/24)	0 (0/72)	
XRF (N = 6)	16.7% (3/18)	0 (0/54)	All false exclusions were reported by a single lab
LIBS (N = 2)	0 (0/6)	0 (0/18)	

3.3.1. Refractive index results (Interlaboratory exercise 2)

For the RI pairwise comparisons, three labs compared the average of the Q measurements to the average K ± 3 s (one of these three labs used a minimum standard deviation of 0.0003). One lab used a *t*-test, one lab compared the Q average to the K range, one lab compared the Q average to the K average ± 2 s (minimum s of 0.0001), one lab compared the Q average to the K average ± 0.0001 , one lab compared the Q average ± 3 s to the K average ± 3 s, and the final lab used a range overlap and a Q average to K average ± 0.0001 ; for the final lab, if either of the two match criteria used resulted in an exclusion, K and Q were considered distinguishable.

All labs correctly associated two Q1 fragments with the inner K1 pane and one Q1 fragment with the outer K1 pane, resulting in no false exclusions. All labs also correctly excluded all K1/Q2 pairs, two Q1 fragments from the outer K1 pane, and one Q1 fragment from the inner K1 pane, resulting in no false inclusions.

3.3.2. μ XRF results (Interlaboratory exercise 2)

For the μ XRF comparisons using a range overlap and/or ± 3 s match criterion, four of five labs correctly associated all K1 inner/outer and Q1 pairs. The remaining lab excluded all three K1/Q1 pairs, resulting in a false exclusion rate of 16.7%. These false exclusions may have partly been the result of the small standard deviations in the K measurements for some element ratios, which led to a narrow comparison interval. The two K1 inner/Q1 pairs were excluded based on the following ratios: Ca/K (1.43% RSD), Ca/Fe (0.31% RSD), and/or Fe/Zr (5.21% RSD). The K1 outer/Q1 pair was excluded on the basis of Ca/Mg (0.62% RSD), Ca/Fe (0.26% RSD), and Fe/Zr (2.61% RSD). If a minimum 3% RSD is applied to each element ratio used for the pairwise comparisons, the ratios Ca/K, Ca/Mg, and Ca/Fe no longer result in a false exclusion. However, Fe/Zr still results in a false exclusion for two of the three Q1 fragments. Differences in the relative location of the sample during measurement and/or improper sample preparation could have also led to the false exclusions. It is worth noting that this lab was the only lab that used a Mo X-ray tube for their analyses, and the Mo X-ray lines interfere with the Zr X-ray lines. Thus, it is possible that the differences in instrumental configuration produced these differing results. All five labs correctly excluded all K/Q pairs that originated from different sources, resulting in no false inclusions.

In an effort to improve the interpretation of forensic evidence, the potential for developing a shared glass database using data (XRF element ratios) collected by different laboratories was explored. The six element ratios recommended in ASTM E2926 were used to compare the same sample across five participating laboratories; that is, K1 inner was compared to K1 inner between labs, K1 outer to K1 outer, Q1 to Q1, and Q2 to Q2. One of the six labs was excluded for the between-lab comparisons because that lab did not provide SRM 1831 data (SRM 1831 was used for normalization, discussed below). Using the ASTM ± 3 s match criterion, 99.6% of the pairwise comparisons between labs were falsely excluded. The large differences in element ratios between labs are partly due to different instrumental geometries and detector efficiencies. Trejos et al. suggested normalizing the element ratios to a standard reference material (SRM 1831) so that the between-lab ratios would be more comparable [13]. A similar approach was implemented in this study (Equation (1)).

$$\left[\frac{E_j}{E_k} \right]_{\text{normalized}} = \frac{\left[\frac{1}{n} \sum_{i=1}^n \frac{E_j}{E_k} \right]_{\text{sample}}}{\left[\frac{1}{n} \sum_{i=1}^n \frac{E_j}{E_k} \right]_{\text{SRM}}} \quad (1)$$

In Eq. (1), the normalized ratio of element *j* (E_j) to element *k* (E_k) is calculated by dividing the average E_j/E_k of the sample by the average E_j/E_k of the SRM (*n* is the number of replicate measurements). The standard deviation was estimated using a propagation of errors [31].

After normalization to SRM 1831, the ASTM ± 3 s match criterion was used for between-lab pairwise comparisons, which led to a false exclusion rate of 77.1%. Establishing a minimum 3% RSD with the normalized ratios further reduced the false exclusion rate to 71.7%. Fig. 1 shows the overall false exclusion rate (using all six ratios) as well as the relative false exclusion rate for each individual ratio using the raw ratios, the normalized ratios, and the normalized ratios with a minimum 3% RSD. After normalization, the relative false exclusion rate improved for all six element ratios; additionally, using a minimum 3% RSD improved Ca/Mg, Ca/Fe, and Ca/K (ratios that often exhibited RSDs < 1%) but did not improve Ca/Ti, Sr/Zr, and Fe/Zr (ratios that typically exhibited higher RSDs).

Although the overall false exclusion rate improved after normalization and after establishing a minimum 3% RSD with the normalized ratios, the false exclusion rate for data acquired by different laboratories, with different instrument configurations was still unacceptably high. This suggests that combining element ratio databases from multiple labs is not currently feasible. It should be noted, however, that the K1/Q1/Q2 specimens used for these comparisons were thin, irregular fragments. Therefore, shape effects and escape depth effects also played a role in the element ratio differences observed between labs. It may be possible to develop a shared database using only flat, full-thickness specimens that are infinitely thick for the elements of interest; full-thickness fragments are usually available for the known sample.

3.3.3. LIBS results (Interlaboratory exercise 2)

The LIBS element ratios and the element emission lines used for the comparisons differed between the two participating labs. The following elements were reported by the two labs: Si, Al, Mg, Ca, Na, K, Sr, Fe, Ti, and Ba.

For the pairwise comparisons, one lab compared the Q average ± 3 s to the K average ± 3 s. This lab excluded element ratios with RSDs that were considered too low or too high (all excluded ratios had RSDs < 7% or greater than 20%). The second lab compared the Q average to the K average ± 4 s. Both labs correctly associated the K1/Q1 pairs, resulting in no false exclusions. Both labs correctly excluded K/Q pairs that originated from different sources, resulting in no false inclusions.

3.3.4. Interpretation of results (Interlaboratory exercise 2)

All labs reported an elimination for K/Q pairs that exhibited reproducible physical or chemical properties. One lab calculated a likelihood ratio (LR) using the RI measurements. This lab reported that the LR provided “moderately strong support” for an exclusion of K1 and Q2.

For the second interlaboratory exercise, two Q1 fragments originated from the inner K1 pane and one Q1 fragment originated from the outer K1 pane; as mentioned earlier, the inner and outer K1 panes were chemically distinguishable. The aim of this second exercise was to evaluate whether practitioners would increase their level of support based on the presence of Q1 fragments that matched both K panes. As observed in exercise one, in the second exercise, the association level reported for the indistinguishable K/Q pairs differed between labs. One lab simply stated that two Q1 fragments were associated with K1 inner and one Q1 fragment was associated with K1 outer. One lab stated that the results “strongly suggest” that Q1 originated from K1; this lab did not include a qualifier that states other sources of glass may have similar physical and chemical composition to the known source. Most labs used some variation of the statement “the Q sample originated from the K (inner or outer) source, or another two sources with the same physical and chemical properties;” though, two labs stated that other glass sources with similar characteristics are limited. Three labs included an association scale with five levels. Two of these three labs reported a level 3 association, which states that the Q “could have originated” from the K source or a similar source. The third lab calculated a frequency for RI and an RMP for elemental data (LA-ICP-MS) but does not currently report a combined abundance since, as discussed in exercise 1, there is a lack of independence between RI and chemical composition. Using their

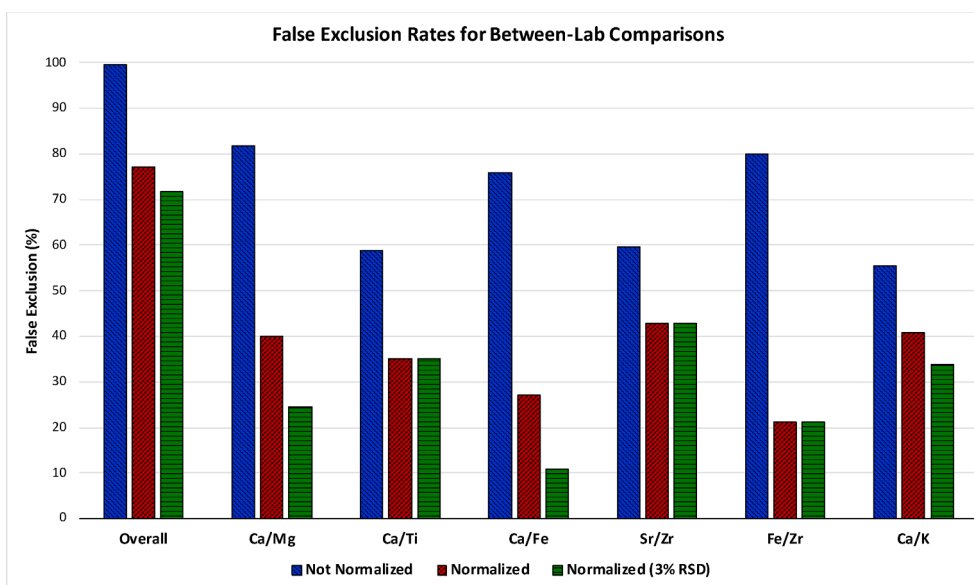


Fig. 1. Overall (all six ratios) and individual ratio false exclusion rates, using the ASTM E2926 match criterion, for between-lab comparisons of samples originating from the same source. The false exclusion rate using raw ratios (blue), SRM 1831 normalized ratios (red), and SRM 1831 normalized ratios with a minimum 3% RSD (green) is shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

conclusion scale, this lab reported that the results provided “very strong support” for an association, which is the highest level of association within a class association with “conventional characteristics.” One lab compared Q1 to a database of 661 glass samples and reported that no database sample matched Q1 on the basis of glass type (e.g., float, tempered), color, thickness, and RI. One lab compared K1 to a RI database and reported that K1 inner was indistinguishable with 2.3% of the samples in the database and K1 outer was indistinguishable with 1.2% of the samples in the database. Finally, one lab calculated a likelihood ratio using the RI measurements following the procedure outlined in Curran et al. [32]. This lab also compared Q1 to a database that included thickness and RI measurements for vehicle glass ($n = 2017$) and reported that one Q1 fragment was indistinguishable with <0.02% of the database samples and the other two Q1 fragments were indistinguishable with <5% of the database samples.

Although the majority of labs noted that Q1 included specimens that matched both K1 panes, only three labs increased their level of support based on the presence of two distinct glass sources. One lab stated that, considering the characteristics of the case, it was “unlikely” that Q1 originated from two other sources of glass with the same characteristics as the two K1 panes. The second lab stated that the presence of Q1 specimens that matched both K1 panes “gives additional significance to the findings.” Finally, the third lab is the one lab that calculated a likelihood ratio ($LR \approx 7.9 \times 10^6$). The LR calculations included the number of recovered fragments as well as the frequency of the two distinct refractive indexes (one for each of the K1 panes). Using their verbal translation scale for the LR, the lab reported that the evidence provided “extremely strong support” for an association.

3.4. Interlaboratory exercise 3

In the third interlaboratory exercise, one known sample (K1, inner and outer pane) and two questioned samples (Q1 and Q2) were submitted to each participating laboratory. Three fragments were submitted for each K pane and each Q. All three Q1 fragments originated from the K1 outer pane. A total of seven laboratories collected RI measurements, four laboratories collected XRF measurements, and three laboratories collected LIBS measurements.

A case scenario, similar to that given in the second exercise, was provided to participants in the third interlaboratory exercise: the hit-

and-run scenario stated that a 2007 Honda Civic (K1) was found with a broken windshield; three questioned fragments (Q1) were found on the clothing of suspect 1 and three fragments (Q2) were found on the clothing of suspect 2. The Q fragments were provided as small (≈ 1 mm), irregular specimens and the known fragments were provided as full-thickness specimens. Participants were asked to follow ASTM E1967 (for RI) and ASTM E2926 (for μ XRF). Table 6 summarizes the false exclusion and false inclusion rates for each technique.

3.4.1. Refractive index results (Interlaboratory exercise 3)

For the RI pairwise comparisons, two labs compared the average Q to the average $K \pm 3s$ (one lab used a minimum standard deviation of 0.0003). One lab used a t -test, one lab compared the Q average to the K average $\pm 2s$ (minimum s of 0.0001), one lab compared the Q average to the K average ± 0.0001 , one lab compared the Q average $\pm 3s$ to the K average $\pm 3s$, and the final lab used a range overlap and a Q average to K average ± 0.0001 ; for the final lab, if either of the two match criteria used resulted in an exclusion, K and Q were considered distinguishable.

All labs correctly associated Q1 to K1 outer, resulting in no false exclusions. All labs also correctly excluded all K1 inner/Q1 and K1/Q2 pairs, resulting in no false inclusions.

3.4.2. μ XRF results (Interlaboratory exercise 3)

For the μ XRF comparisons (range overlap and/or $\pm 3s$), all labs correctly associated Q1 with the outer pane of K1, resulting in no false exclusions. All labs correctly distinguished all K1 inner/Q1 and K1/Q2 pairs, resulting in no false inclusions.

As was observed in the exercise 2 results, the between-lab

Table 6

Interlaboratory exercise 3 false exclusion and false inclusion rates for refractive index, XRF, and LIBS; the number of participating labs (N) is indicated next to each technique. Fractions in parentheses indicate the number of false exclusions/inclusions divided by the total number of pairwise comparisons.

	False exclusion (K1 _{Outer} /Q1)	False inclusion (K1 _{Inner} /Q1 and K1/Q2)
Refractive Index (N = 7)	0 (0/18)	0 (0/45)
XRF (N = 4)	0 (0/12)	0 (0/35)
LIBS (N = 3)	0 (0/9)	0 (0/26)

comparisons for exercise 3 resulted in a high false exclusion rate: 100% false exclusion using the raw ratios and 94.4% false exclusion using the (SRM1831) normalized ratios with and without a minimum 3% RSD. When using the normalized ratios with a minimum 3% RSD, the false exclusion for the individual ratios Ca/Mg, Ca/Fe, and Ca/Ti improved. However, the overall false exclusion rate (94.4%), using all six ratios, did not improve.

3.4.3. LIBS results (Interlaboratory exercise 3)

The LIBS element ratios and the element emission lines used for the comparisons differed between the three participating labs. The following elements were reported by the three labs: Si, Al, Mg, Ca, Na, K, Sr, Fe, and Ti.

For the LIBS pairwise comparisons, one lab used the Q average and the K average ± 3 s, one lab used the Q average ± 3 s and the K average ± 3 s, and the third lab used the Q average and the K average ± 4 s (using only ratios with RSD < 20%). All labs correctly associated the K1/Q1 pairs, resulting in no false exclusions. All labs correctly excluded K/Q pairs that originated from different sources, resulting in no false inclusions.

3.4.4. Interpretation of results (Interlaboratory exercise 3)

All K/Q pairs that exhibited reproducible physical or chemical properties were reported as an exclusion. As observed in the previous two exercises, the association level reported for the indistinguishable K/Q pairs differed between labs. One lab simply stated that the Q1 fragments were associated with K1 outer. Most labs used some variation of the statement “the Q sample originated from the K outer source, or another source with the same physical and chemical properties;” though, two labs stated that other glass sources with similar characteristics are limited. Three labs included an association scale in their report; two labs reported a level 3 association, which states that the Q “could have originated” from the K source or a similar source, and the third lab calculated a frequency using a RI database and an RMP using an elemental database (LA-ICP-MS) and reported that there is “very strong support” that K1 outer and Q1 originated from the same source. Two labs compared Q1 to all samples in a database and reported the number of matches (i.e., the frequency of occurrence): one lab reported that the Q matched one sample in their database (n = 661) on the basis of glass type (e.g., float, tempered), color, thickness, and RI; the second lab reported no matches using their database (n = 3000) based on thickness and RI. One lab compared K1 outer to a RI and thickness database (n = 2319) and reported that K1 outer was indistinguishable with 0.9% of the database samples. Finally, one lab calculated a likelihood ratio using the RI measurements following the procedure outlined in Curran et al. [32]. The LR showed “extremely strong support” for an association between K1 and Q1. This lab also compared Q1 to a database that included thickness and RI measurements (n = 3000) and found that Q1 did not match any of the database samples.

4. Conclusions

A total of 17 laboratories participated in an interlaboratory study that included three blind exercises designed as mock case scenarios. Eleven labs provided RI measurements, seven labs provided μ XRF measurements, and five labs provided LIBS measurements. The participants were instructed to analyze and compare glass specimens and submit a report on their findings as they would in an actual case. The rate of false exclusions and false inclusions when comparing the glass specimens using RI or elemental composition (μ XRF and LIBS) was determined. Additionally, the reporting language and the use of databases to assign a significance to the evidence was evaluated to determine the current state of evidence interpretation within the forensic community.

With the exception of μ XRF in exercise 2, no false exclusions were observed for either of the three techniques in each of the three exercises.

All false exclusions in exercise 2 were reported by a single μ XRF participant. It is possible that these false exclusions were in part due to low RSDs and that a minimum RSD, similar to what is used for the LA-ICP-MS match criterion, may be warranted for μ XRF comparisons. However, considering the fact that all other μ XRF participants reported correct associations, it is possible that the differences in instrumental configuration or relative location of the sample during measurement produced these differing results.

No false inclusions were observed for either of the three techniques in exercises 2 and 3. However, a relatively high false exclusion rate was observed for the first exercise: 34.1%, 8.3%, and 16.7% for RI, μ XRF, and LIBS, respectively. In the case of μ XRF and LIBS, all false inclusions were comparisons between vehicles with the same make, model, and year of manufacture. Most of the RI false inclusions were comparisons between vehicles with a different make and year of manufacture; these pairs were readily distinguishable with μ XRF, LIBS, and LA-ICP-MS. On the other hand, most labs were able to find reproducible differences in the RI for pairs that originated from similar vehicles; these pairs were not distinguishable using LA-ICP-MS or LIBS but were distinguishable using μ XRF on the basis of the Ca/Fe ratio. To investigate the complementarity of RI and elemental composition, a set of 31 glass samples, selected based on their similarity in elemental composition, were analyzed using RI, LA-ICP-MS, and μ XRF. RI was able to distinguish 3 pairs that were indistinguishable with both LA-ICP-MS and μ XRF. Thus, glass samples found to be indistinguishable by elemental composition using a sensitive elemental analysis technique (LA-ICP-MS, μ XRF, LIBS) may benefit from additional discrimination potential from RI measurements. Therefore, RI analysis remains a viable technique in a comparative analytical scheme that includes elemental analysis.

An important consideration in forensic casework is the interpretation of the evidence. Databases are useful tools for the forensic community as they can provide information about the within-sample variability, the between-sample variability, and the discrimination power of an analytical tool. Additionally, databases can be used to assign a significance to forensic evidence (i.e., provide a quantitative measure of the rarity of a particular feature, such as RI or elemental composition) [33]. Knowledge of a technique's discrimination power as well as the commonality/rarity of certain features is necessary for the development of a verbal scale. Few labs (3 out of 17) used a verbal scale to provide a significance to an association. Additionally, few labs (3 labs) increased their level of support for an association based on the presence of questioned specimens that were indistinguishable with two distinct known sources (exercise 2). Although RI databases are readily available, they are not widely utilized within the forensic community. XRF databases, on the other hand, are not as readily available as RI databases due to inherent signal variations among instrumental configurations. The potential for developing a shared XRF database was evaluated by comparing the elemental ratios for the same glass sample analyzed by different laboratories. However, even after normalizing the ratios to a control sample, SRM 1831, the false exclusion rate for between-lab comparisons using the recommended ± 3 s criterion was unacceptably high (>70% false exclusion). This suggests that a shared database populated with elemental ratios from different labs is not currently feasible for XRF. However, it should be noted that the μ XRF measurements were collected on small, irregular glass fragments. It may still be possible to develop an XRF database comprised only of known specimens, which are usually available as full-thickness fragments. Also, laboratories may still consider the use of internal databases, in which data is collected in the same lab, with the same method and instrument over time. The rapid acquisition times of LIBS measurements (<1 min per replicate measurement) makes LIBS an attractive tool for the development of a glass database. However, a known disadvantage of LIBS is its poor precision compared to more well-established analytical tools, such as μ XRF and LA-ICP-MS. The RSDs for measurements submitted by the five participating labs ranged from <1% to greater than 50%, depending on the elemental ratios used for comparisons. Future

research efforts should focus on improving the precision of LIBS measurements and developing a standard methodology for glass analysis.

RI, μ XRF, and LIBS were shown to be fit-for-purpose for the forensic comparison of glass samples, with an overall correct association greater than 92% and an overall correct exclusion greater than 82%. It should be noted that all false inclusions (overall false inclusion of 18%) occurred in exercise one, which involved comparisons between vehicles of the same make, model, and year. There is a need to standardize the reporting language used by practitioners across different laboratories. It is recommended that practitioners use a verbal scale to provide a significance to their results when reporting their conclusions. Though, verbal scales can be subjective since they rely on the experience of the practitioner. The use of databases to estimate a frequency, RMP, and/or LR can provide a more objective approach for the interpretation of evidence. As such, practitioners should incorporate the use of databases, if available, to provide a quantitative measure of the rarity of a particular feature (e.g., RI, elemental composition); although, as noted by one lab, there is a need to address the lack of independence between different features. Future research efforts should focus on standardizing the reporting language and verbal scales used between different laboratories. There is an ongoing effort within NIST OSAC to develop an interpretation guide that aims to standardize a verbal scale that can be used by all trace evidence disciplines. This interpretation guide should encourage practitioners to incorporate a verbal scale in their casework and should lead to better agreement among practitioners. Additionally, existing RI databases can be expanded and potentially shared among practitioners. Finally, similar to the ongoing efforts to develop and share a quantitative LA-ICP-MS database, the development of a μ XRF database (and potentially a LIBS database) would also be useful to glass practitioners that currently do not perform LA-ICP-MS analysis of glass. A challenge in the development of glass databases is the change in glass formulations over time. To address this challenge, it would be necessary to continuously populate glass databases with modern glass formulations (e.g., from casework specimens submitted to a forensic laboratory) to ensure that the database reflects the relevant population of glass.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project was supported by awards no. 2015-DN-BX-K049 and 2018-DU-BX-0194 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect those of the Department of Justice or the National Institute of Standards and Technology. Names of commercial manufacturers are provided for identification purposes only, and inclusion does not imply endorsement of the manufacturer, or its products or services by any of the above organizations. The National Science Foundation (NSF) funded an INTERN award to Ruthmara Corzo through the Center for Advanced Research in Forensic Science (CARFS), award no. 1739805 to participate in an internship at the National Institute of Standards and Technology (NIST) where she is currently an NRC postdoctoral fellow.

References

- [1] J. Peterson. National Institute of Justice, U. S. G. P. Office, Washington D. C. (1974).
- [2] J.A. Buscaglia, *Elemental analysis of small glass fragments in forensic science*, Anal. Chim. Acta 288 (1-2) (1994) 17–24.
- [3] R.D. Koons, C.A. Peters, P.S. Rebert, 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. 6 (6) (1991) 451, <https://doi.org/10.1039/ja9910600451>.
- [4] R. D. Koons, J. Buscaglia, M. Bottrell and E. T. Miller. *Forensic Glass Comparisons in R. Saferstein's (ed.) Forensic Science Handbook Volume 1*, Prentice Hall, New Jersey, USA (2002).
- [5] R.D. Koons, J. Buscaglia, *Distribution of refractive index values in sheet glasses*, Foren. Sci. Commun. 3 (2001).
- [6] ASTM E1967-11, ASTM International, West Conshohocken, PA (2011).
- [7] ASTM E1967-19, ASTM International, West Conshohocken, PA (2019).
- [8] E. J. Garvin and R. D. Koons. Evaluation of match criteria used for the comparison of refractive index of glass fragments, J Forensic Sci 56 (2011) 491-500.
- [9] C.M. Bridge, J. Powell, K.L. Steele, M.E. Sigman, *Forensic comparative glass analysis by laser-induced breakdown spectroscopy*, Spectrochim. Acta Part B Atomic Spectrosc. 62 (12) (2007) 1419–1425.
- [10] D.C. Duckworth, S.J. Morton, C.K. Bayne, R.D. Koons, S. Montero, J.R. Almirall, *Forensic glass analysis by ICP-MS: a multi-element assessment of discriminating power via analysis of variance and pairwise comparisons*, J. Anal. At. Spectrom. 17 (2002) 662–668.
- [11] S. Ryland, *Discrimination of flat (Sheet) glass specimens having similar refractive indices using micro X-ray fluorescence spectrometry*, JASTEE 2 (2011) 2–12.
- [12] T. Trejos, S. Montero, J.R. Almirall, *Analysis and comparison of glass fragments by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) and ICP-MS*, Anal. Bioanal. Chem. 376 (8) (2003) 1255–1264.
- [13] T. Trejos, R. Koons, S. Becker, T. Berman, J. Buscaglia, M. Duecking, T. Eckert-Lumsdon, T. Ernst, C. Hanlon, A. Heydon, K. Mooney, R. Nelson, K. Olsson, C. Palenik, E.C. Pollock, D. Rudell, S. Ryland, A. Tarifa, M. Valadez, P. Weis, 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*, Anal. Bioanal. Chem. 405 (2013) 5393–5409.
- [14] 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.C. Pollock, D. Rudell, S. Ryland, A. Tarifa, M. Valadez, A. van Es, V. Zdanowicz, J. Almirall, *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. 28 (2013) 1270–1282.
- [15] H. Dorn, D.E. Ruddell, A. Heydon, B.D. Burton, *Discrimination of float glass by LA-ICP-MS: assessment of exclusion criteria using casework samples*, Can. Soc. Foren. Sci. J. 48 (2015) 85–96.
- [16] B.E. Naes, S. Umpierrez, S. Ryland, C. Barnett, J.R. Almirall, *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 B Atomic Spectrosc. 63 (10) (2008) 1145–1150.
- [17] P. Weis, M. Dücking, P. Watzke, S. Menges, S. Becker, *Establishing a match criterion in forensic comparison analysis of float glass using laser ablation inductively coupled plasma mass spectrometry*, J. Anal. At. Spectrom. 26 (6) (2011) 1273, <https://doi.org/10.1039/c0ja00168f>.
- [18] ASTM E2926-17, ASTM International, West Conshohocken, PA (2013).
- [19] C.M. Bridge, J. Powell, K.L. Steele, M. Williams, J.M. MacInnis, M.E. Sigman, *Characterization of automobile float glass with laser-induced breakdown spectroscopy and laser ablation inductively coupled plasma mass spectrometry*, Appl. Spectrosc. 60 (10) (2006) 1181–1187.
- [20] M.M. El-Defdar, N. Speers, S. Eiggins, S. Foster, J. Robertson, C. Lennard, *Assessment and forensic application of laser-induced breakdown spectroscopy (LIBS) for the discrimination of Australian window glass*, Foren. Sci. Int. 241 (2014) 46–54.
- [21] T. Trejos, S. Koch, A. Mehlretter, *Scientific foundations and current state of trace evidence—A review*, Foren. Chem. 18 (2020) 100223, <https://doi.org/10.1016/j.forc.2020.100223>.
- [22] T. Hoffman, R. Corzo, P. Weis, E. Pollock, A. van Es, W. Wiarda, A. Stryjnik, H. Dorn, A. Heydon, E. Hoise, S. Le Franc, X. Huifang, B. Pena, T. Scholz, J. Gonzalez, J. Almirall, *An inter-laboratory evaluation of LA-ICP-MS analysis of glass and the use of a database for the interpretation of glass evidence*, Foren. Chem. 11 (2018) 65–76.
- [23] R. Corzo, T. Hoffman, P. Weis, J. Franco-Pedroso, D. Ramos, J. Almirall, *The use of LA-ICP-MS databases to estimate likelihood ratios for the forensic analysis of glass evidence*, Talanta 186 (2018) 655–661.
- [24] A. van Es, W. Wiarda, M. Hordijk, I. Alberink, P. Vergeer, *Implementation and assessment of a likelihood ratio approach for the evaluation of LA-ICP-MS evidence in forensic glass analysis*, Sci. Justice 57 (3) (2017) 181–192.
- [25] P. Vergeer, A. van Es, A. de Jongh, I. Alberink, R. Stoel, *Numerical likelihood ratios outputted by LR systems are often based on extrapolation: When to stop extrapolating?* Sci. Just. 56 (6) (2016) 482–491.
- [26] C.G.G. Aitken, G. Zadora, D. Lucy, *A two-level model for evidence evaluation*, J. Foren. Sci. 52 (2) (2007) 412–419.
- [27] G. Zadora, *Evaluation of evidence value of glass fragments by likelihood ratio and Bayesian Network approaches*, Anal. Chim. Acta 642 (1-2) (2009) 279–290.
- [28] ASTM E2927-16e1, ASTM International, West Conshohocken, PA (2016).
- [29] T. Ernst, T. Berman, J. Buscaglia, T. Eckert-Lumsdon, C. Hanlon, K. Olsson, C. Palenik, S. Ryland, T. Trejos, M. Valadez, J.R. Almirall, *Signal-to-noise ratios in forensic glass analysis by micro X-ray fluorescence spectrometry*, X-Ray Spectrom. 43 (2014) 13–21.
- [30] S. Becker, W. Becker, N. Fritz, D. Hirschfelder, K. Holz, T. Krieghof, C. Schmid, M. Weise, *Design of a glass casework database inheriting quality assurance aspects*, Probl. Foren. Sci. 47 (2001) 73–79.

- [31] J.N. Miller, J.C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, Pearson, England, UK, 2010.
- [32] J. Curran, T. Hicks, J. Buckleton, *Forensic Interpretation of Glass Evidence*, CRC Press, Boca Raton, Florida, USA, 2000.
- [33] A. Akmeemana, P. Weis, R. Corzo, D. Ramos, P. Zoon, T. Trejos, T. Ernst, E. Pollock, E. Bakowska, C. Neumann, J.R. Almirall, Interpretation of chemical data from glass analysis for forensic purposes, *J. Chemom.* (2020), e3267, <https://doi.org/10.1002/cem.3267>.