A Guide to Using SRM 2569 to Validate Analytical Methods

J. R. Sieber and J. L. Molloy
National Institute of Standards and Technology, Analytical Chemistry Division, Gaithersburg, Maryland

Introduction

Standard Reference Material (SRM) [1] 2569 Lead in Paint Films for Children's Products was developed by the National Institute of Standards and Technology in collaboration with the U.S. Consumer Product Safety Commission for the purpose of validation of test methods for determination of lead (Pb) in paints and similar coatings. This guide is intended to instruct the user who must critically evaluate a non-destructive test method, typically based on X-ray fluorescence spectrometry (XRF). The SRM consists of the paint films on three coupons of polyester. Each paint film is lacquer that contains no added Pb (Level 1), 85 mg/kg Pb (Level 2), or 314 mg/kg Pb (Level 3). Along with the three paint films, the SRM unit contains five coupons of uncoated polyester. The paint coatings were carefully characterized for mass fraction of Pb, mass of Pb per unit area, paint coating thickness, and paint coating density with values for the constituents and properties being either certified or reference values. Users should consult the certificate of analysis for official values and uncertainty estimates along with instructions for using the SRM. Users may also wish to consult a paper written by the authors describing the development and certification of SRM 2569. The paper is currently in press. After it is published, the authors will provide a reference citation on the SRM 2569 details page of the NIST online SRM catalog (available from https://www-s.nist.gov/srmors/view_detail.cfm?srm=2569).

Measurement Options

The coupons are coated on one side only with the uncoated side bearing the SRM label. For users planning to measure paint on plastic using XRF, there are several options for orientation of the coupon and the XRF spectrometer. This guide will be limited to the case of paint on plastic. The unlabeled side should be oriented toward the XRF spectrometer. The user has the option of placing additional uncoated coupons behind the SRM coupon to simulate the effects of thicker substrates. Using all five uncoated coupons that are supplied in the SRM package gives a substrate thickness of approximately 1 mm. All uncoated coupons should be in close contact with the labeled side of the SRM coupon and each other. It may be advisable to perform such tests because thicker plastic causes greater amounts of X-ray scattering, which shows up as higher spectral background and potentially higher limits of detection.

The thin coupons can be curved to conform to a cylinder as small as 1 cm diameter. It may require some care to accomplish without damaging the paint, but a user may be able to shape the coupon to other, more complex shapes having smooth curves. Do not crease or fold the coupon because that will damage the paint and void its certification. Of course, the user must consider ways to curve the coupon without having a different (usually unwanted) material within the measurement area. The user may wish to measure different surface curvatures as a means of testing the effects of non-planar surfaces on the XRF results. This option may make sense when the measured spot is relatively large. For small spots, the amount of deviation from a plane surface may be small regardless of the overall amount of curvature.

XRF instruments are capable of measuring spots having a wide range of sizes primarily circular, but in some cases oval or rectangular. The range of available spot sizes goes from tens of micrometers up to about 50 mm in diameter. SRM 2569 can be used with spot sizes from 1 mm to 50 mm. The coupons were designed to provide four or more unique measurement locations for spot diameters up to 20 mm. The user need not measure just at the center of the coupon. Measurements can be located over the entire painted surface of the coupon, but the spot should not overlap the area where the label is affixed to the opposite (uncoated) side.

Material Homogeneity

The lacquer coatings of Level 2 and Level 3 are different formulations because Level 2 contains a chlorinated resin added to increase the X-ray contrast between the paint film and the polyester substrate and Level 3 contains a higher dose of the green phthalocyanine dye added for color, but none of the resin. Unfortunately, the Level 2 paint film is

[1] Standard Reference Material® (SRM) is a registered trademark for NIST certified reference materials.

not as homogeneous as Level 3. The nature of the heterogeneity in Level 2 is such that a portion of the Pb compound added to the liquid lacquer came out of solution before curing. This portion of the Pb compound formed microscopic inclusions in the otherwise uniform paint film. Level 3 does not suffer from this problem because all Pb compound stayed in solution as the lacquer was applied. Level 1 contains no added Pb, and there is no problem with heterogeneity.

For Level 2, the result is that a small but significant fraction of measurements using an XRF with a spot size of 1 mm or smaller would measure an inclusion of elevated Pb level compared to the average mass per unit area value. With a small enough X-ray beam in a microXRF instrument, one can map the locations of the Pb compound inclusions, which were seen to be relatively few in number. The inclusions are in the size range of approximately 10 μm to 70 μm and are scattered randomly across the coupon. As an example, a set of 13,000 measurements using a spot approximately 1 mm diameter and a step size of 0.5 mm contained only about 50 measurements that were clearly affected by high Pb count rates from inclusions. Because the inclusions of Pb compound are microscopic and scattered randomly, their presence would remain undetected by users of a large measurement spot size. Conversely, users of very small spot sizes will see the effects of the inclusions as an increase in variability of measured X-ray count rate with decrease in spot size. The extreme case is microXRF, which is by definition a beam size < 1 mm, i.e., in the micrometer size range. With a 50 μm beam, it is possible for a single measurement to be located entirely on a single inclusion of Pb compound. Because the original Pb compound contained 48 % Pb by mass, the expected count rate would be thousands of times greater than the count rate from a similarly sized spot that is representative of the overall Pb mass fraction, i.e., 85 mg/kg Pb (0.0085 % Pb). This is the primary reason why SRM 2569 is not suitable for use with instruments having beam sizes < 1 mm such as microXRF. However, the 1 mm spot size available in some commercial, bench top instruments is still subject to significant variation of Pb count rate because varying numbers of Pb compound inclusions can be present within the 1 mm spot. It is difficult to decide exactly the spot size above which the inclusions have a negligible effect on variance of results because it is impossible to determine the full range of count rate increases caused by random clusters of inclusions.

Validation of a Test Method

Validation has a formal, internationally accepted definition. To put it simply, validation is the provision of objective evidence that a test method performs as required. Every aspect of the method's performance should be tested and documented. However, that is not always possible at every laboratory. Certainly, the laboratory that begins from scratch and develops or implements the method itself should check performance at every step of the process. Many laboratories do not have the expertise and resources to develop methods. They purchase the equipment and the method as a package and rely on the supplier for the expertise. In such a case, the laboratory must show only that the implementation of the method in their facility with the purchased equipment is valid and remains under statistical control. It is advisable for the laboratory to request validation documentation from the method developer. Either way, validation requires appropriate tools and complete documentation of the validation procedure and the results.

SRM 2569 can be used to test the repeatability of the measurement process. Without being able to make repeatable measurements, one cannot establish a calibration and make quantitative determinations. If the repeatability is good enough, SRM 2569 can be used to test for bias in the final results. In other words, there are two phases to validation of a test method that must be completed before the method is put into regular use. It is suggested that labs separate the two phases by considering just repeatability first, and bias second. This section will discuss ways in which SRM 2569 can be used. The choices and procedures presented below are intended as suggestions, not requirements.

Repeatability: In development of a method from scratch, an analyst would check all measurements for repeatability before calibrating the instrument. Some XRF methods, especially those intended to report the mass fraction of Pb in a thin film, utilize a number of measurements of count rates of various X-ray lines and other spectral features, including scatter of the primary X-ray beam from the specimen and spectral background at strategically chosen energies. A laboratory that purchases the instrument/method package usually can only check the repeatability of the final calculated result. Fortunately, if the final result is sufficiently repeatable, it is reasonable to conclude that the individual measurements are also sufficiently repeatable.

It may be useful to demonstrate repeatability in a number of ways. The logical first step is to choose a single location on a coupon and perform repeated measurements changing as few things as possible. The standard deviation of the set of measurements would represent the minimum variability obtainable using the available

spectrometer. Assuming the measured spot on the paint film does not change under the influence of the X-ray beam, the measured variability will represent the stability of the equipment during the time needed to complete the measurements. Because this first step is intended to show that the equipment is stable over a brief period of time, it does not matter whether the analyst chooses the Level 2 or the Level 3 coupon. If the test method uses more than one count rate measurement under more than one set of conditions, it would be ideal to obtain the measured count rates under each condition and to check the repeatability of them all. Within the limits of counting statistical uncertainty, the stability should be nearly the same under all conditions.

For the next phase of the process, it is assumed that the measured spot size is small enough that four or more unique locations can be measured on each coupon. Because Level 3 is of superior homogeneity to Level 2, Level 3 may be the better choice to demonstrate repeatability when a number of different locations are measured. It is logical to begin this phase by having a single analyst measure multiple locations to check the standard deviation of the measured values. Once the first analyst has obtained satisfactory performance, additional analysts can check their capabilities. All analysts should be instructed to perform the test procedure in its entirety. Here again, it is not accuracy that is being tested. The goal is to show that all analysts can obtain the same value within a range that is appropriate for the application of the test method.

If the analyst(s) have to this point been working with Level 3 of SRM 2569, it may be advisable to try the same tests using Level 2. Given the explanation above of the greater heterogeneity of Level 2, the analyst will obtain a set of measured values that exhibit a greater variance than the data from Level 3. When an analyst is measuring different locations, the analyst may obtain a small number of results that are much higher than the mean and outside the range expected using the data from Level 3 as a predictor of the relative standard deviation to expect from Level 2. This is to be expected because the inclusions of Pb compound (only found in Level 2) are more concentrated in Pb. It is advisable to measure a relatively high number of locations and to plot a frequency curve or histogram from the data. As the spot size decreases, the frequency plot should become more skewed to the high side because the microscopic inclusions can cause a few relatively high measured count rates for Pb. For now, the analyst should accept the data even though it makes comparisons among sets of data from different analysts or different set of locations more difficult than the data from Level 3. Characteristics of data sets from Level 2 will be discussed later.

Calibration: If the test method has not yet been calibrated, now is the time to do so, assuming the repeatability under conditions of minimum variability was acceptable. It is not intended for a laboratory to use SRM 2569 as a set of calibration standards. It is recommended that these coupons be retained for method validation. However, SRM 2569 Level 1 can be used to check the quality of the lowest mass fraction calibration standards in use by the laboratory. Level 1 contains no intentionally added Pb, and its certified value is expected to be much lower than the limits of detection required by test methods used to screen products. Spectra taken from Level 1 should show no evidence of Pb peaks unless a method's limit of detection is < 0.2 mg/kg or < 0.02 $\mu g/cm^2$. Said spectra should be accumulated using longer counting times than normal to ensure that four times the limit of detection is less than the limit of quantification required by the laboratory. High quality spectra from the Level 1 coupon can serve as benchmarks for blanks prepared or purchased by the laboratory.

Bias Detection: After the repeatability check and calibration have been completed, it is possible to make a statement about the accuracy of the test method. Accuracy is the combination of precision and bias, and accuracy is generally discussed in a qualitative manner until the uncertainty budget for the laboratory's implementation of the test method has been completed and an overall uncertainty estimate can be calculated for a given result. Therefore, the next logical step is to test for detectable bias.

Bias detection is based on comparison of a known value and its uncertainty range expressed at a confidence level of approximately 95 % to the consensus value obtained by the set of testing entities and the uncertainty interval calculated by the set of entities. For the purpose of this document, the set of testing entities consists of the analysts employed by the laboratory. For a standards development organization, the set of entities would be laboratories involved in an interlaboratory study. For a detailed discussion of bias detection, the reader can begin with NIST Special Publication 829 (available from www.nist.gov/mml/analytical/inorganic/upload/NIST_SpecialPub829.pdf). Bias detection can be made rather complex, but useful information can be obtained rather simply. It is common for a laboratory to consider the uncertainty interval of its result to be $t \cdot s$, where t is Student's variable chosen for the number of measurements, n, minus 1, and again, s is the repeatability standard deviation. In this case, it is advisable to have n be relatively high to make t relatively small. The choice of $t = 2$ implies that n is high. An alternative is

available when the laboratory can use a published standard test method that includes an estimate of the intra-laboratory repeatability obtained from an interlaboratory study as part of the method validation process. Laboratories that use an estimate of repeatability and forgo the creation of a detailed uncertainty budget are in essence ignoring a number of uncertainty sources including the uncertainties of the values of the calibration standards, and the goodness of fit of the measured data to the equation chosen as a model of the response curve. With very thin samples, the equation of a line is usually the appropriate model to choose for calibration of mass per unit area, which makes the calibration process easier and simplifies the estimation of uncertainty. One need only be certain that the highest calibration points do not appear to fall below the line. This will happen when the calibration standards are no longer behaving as thin samples and are absorbing Pb X rays.

A key element of bias detection is the choice of an estimator for the best value from a set of measurements. Choices include the mean, the median, and a long list of more complex estimators. The difficulty of making a choice increases as the frequency distribution of a set of data becomes more skewed. For data that fit a normal distribution (Gaussian, no skew), all estimators provide essentially the same value with differing estimates of the width of the distribution. For measurements within a single coupon, Level 3 of SRM 2569 should provide essentially normal data distributions because it is highly homogeneous. Level 2 will provide data sets having distributions skewed toward higher values. In such a case, the mean and median values may no longer be nearly identical because the mean is a function of the entire set of values, and the median is a function of the two values at the center of the set. Therefore, the median is the better estimator when using Level 2.

The laboratory should check for bias in measurements of both Level 2 and Level 3 because both compositions are near known action levels for Pb in products, and laboratories can learn more about the performance of their test method. Level 3 offers an easier test by being a higher mass fraction of Pb in a thicker, more homogeneous paint film. Obviously, Level 2 is more difficult because it is less homogeneous, and the thickness is half that of Level 3, which provides for a higher relative uncertainty of the thickness. In terms of mass of Pb per unit area, the value for Level 3 is approximately 6.5 times greater than the value for Level 2. It is conceivable that a laboratory may be able to obtain accurate results for Level 3, and fail to do so for Level 2 because the method lacks the sensitivity.

Statistical Control of a Test Method

As with most SRMs for chemical composition, SRM 2569 is not intended for routine use as a control chart material. The SRM coupons were not designed to withstand repeated use on a daily or more frequent basis and could degrade over time with prolonged, heavy use. In addition, it is not necessary to use a certified reference material (CRM) to demonstrate statistical control of a test method for Pb in paint. Under some circumstances, it may not even be necessary to use a paint coated material. The laboratory should consider that CRMs, such as SRM 2569, are difficult and expensive to develop. The more complex the material and its certification, the less likely it is that the certifying body can provide a long term supply of the CRM.

Statistical control does not require continual demonstration of the absence of bias. It requires demonstration that the measurement process remains unchanged from the point in time that repeatability and absence of bias were demonstrated, i.e., the original validation in the laboratory using the method. Although many companies see a significant risk in relation to the testing for Pb in their products, it is not necessary to constantly measure a CRM as a means of amassing data that can be used to convince an accreditor or a regulator. Provided the company maintains complete records of the successful validation of the method at its implementation and of the continued successful maintenance of statistical control, it is possible to provide an unbroken chain of comparisons with uncertainty estimates from the outset to the present. For added security from risk, the laboratory may find it advisable to repeat the validation of the method using the CRM during the time it takes to establish the control charts and when there are substantial changes to the process. Substantial changes may include turnover of analysts, major maintenance or replacement of equipment, lengthy disruptions to operations, and perhaps other problems and events.

Inclusion in the laboratory's quality system of the frequent use of a fragile material such as SRM 2569 can quickly lead to a disruption of the ability to continue the documented procedure when an SRM coupon becomes damaged. The certificate of analysis for SRM 2569 provides instructions for use and other information to help users avoid damage to the coupons. From this perspective, it is advisable to use a durable material available in adequate supply as a control chart material. Such a control chart reference material would be sufficiently homogeneous and not susceptible to damage by X rays. The control chart material should give measured signals or a final result similar to the paint coated materials normally tested by the laboratory. The control material should mimic the results of an

easily quantifiable sample when measured using the normal test procedures. Two choices made at NIST can serve as good examples, although one relied on another CRM for the short period of time during the development of SRM 2569. At NIST, pellets of ECRM 681k, a certified polyethylene containing 98 mg/kg Pb, were melt pressed in a steel die to a thickness of approximately 165 μm. When placed on top of the polyester sheet provided with SRM 2569 or any other Pb-free plastic substrate, these thin pieces of polyethylene contain Pb at a mass fraction of 98 mg/kg and mimic paint containing Pb at an areal density of about 1.5 $\mu g/cm^2$. For use with a low power X-ray source, these samples are more durable than SRM 2569 coupons. Weighing less than 0.1 g each, they are inexpensive to create. The second example was commercially available Pb metal films created by evaporative deposition onto thin polyester foils. These films were used as calibration standards for WDXRF analysis of paint mass of Pb per unit area. The metal films can serve well as control samples, if it is understood that they are fragile. Also, it is advisable to store them under inert gas to slow oxidation of the metal. Storage in an empty desiccator purged with N_2 (g) is a simple option. Oxidation may affect the measurement of L-series Pb lines and will certainly affect the use of M-series lines. These choices worked at NIST, but may not be optimum for other laboratories.

Comparability of Materials

Users of CRMs often must decide the extent to which validation using a given composition CRM can be extended to the range of compositions tested by the laboratory. This is an important question because certifying bodies cannot provide more than a few CRM compositions for a given application or industry. The laboratory must carefully consider the chemical and physical phenomena used in the test method and the dependence of each aspect of a measurement on the chemical and physical properties of the samples of unknown composition, the calibration standards, and the CRMs employed in validation. In this guide, the subject materials are plastic materials (various polymers) coated with a cured paint, lacquer or other polymeric material with both the substrate plastic and the thin coating possibly containing various amounts of one or more elements used in performance additives. In X-ray analysis, most polymers behave similarly, especially when the analyte element's X rays are of high energy, such as Pb L-series X rays. Therefore, it is expected that SRM 2569 with its polyester substrate will interact with the primary and secondary X rays in a similar manner to plastics more commonly employed in children's products. In XRF analyses of thin samples, the truly thin materials are those that do not absorb the X rays being measured. Given that different X-ray energies can be measured depending on the element and the line series chosen, the critical thickness of a thin film depends on the matrix composition and the X-ray energy. The more additive elements a coating contains, the thinner will be the critical thickness where absorption becomes significant. Fortunately, even in cases of polymers very dissimilar to polyester and coating compositions differing from the lacquer used for SRM 2569, X-ray fluorescence methods can be designed to compensate. The performance of the chosen compensation approach (if any) should be documented during the design and validation of the test method.

For Pb at the levels of interest associated with SRM 2569 and measuring Pb L-series lines, a number of different materials can be used as thin samples that can be accurately quantified using a single calibration. To illustrate this point, NIST created a calibration using energy dispersive XRF with a Zr secondary target source and a high-purity Ge detector. The original purpose of the calibration was to validate a set of Pb metal films deposited on thin polyester by thermal evaporation. To establish traceability to the International System of Units (SI), it was necessary to establish comparisons to NIST SRMs, and if possible, CRMs from other sources. The Pb metal films contained Pb at mass per unit area values from blank to 300 $\mu g/cm^2$. In this range, comparisons could be made to SRM 1833 Thin Glass Film on Polycarbonate for X-ray Fluorescence Spectrometry and SRM 2783 Air Particulate on Filter Media. From blank to 100 $\mu g/cm^2$, all three levels of SRM 2569, the metal films, SRM 1833, and SRM 2783 were all co-linear with a high degree of correlation. The set includes samples composed variously of Pb metal, silicate glass, mineral particles in carbonaceous matter, and organometallic compounds in cured lacquer. The 300 $\mu g/cm^2$ Pb metal film could not be included because curvature resulting from absorption was significant. However, an upper cutoff of approximately 100 $\mu g/cm^2$ is more than acceptable because the highest paint film in SRM 2569 contains Pb at an areal density of approximately 1.45 $\mu g/cm^2$.

Discussion of Data from Level 2

It may be that some users of SRM 2569 have little experience with materials that yield data sets characterized by skewed frequency distributions. With each measurement of Level 2, there is a low probability of hitting a spot that yields a high count rate. Although that probability is very low, persistent measurement (e.g. daily) will enable a user to find at least one or two locations on many of the coupons, even using a 1 mm beam spot. If the user is in the habit of measuring the minimum recommended three spots at a time, finding a high count rate spot will cause the mean of

three measurements to be biased. However, the median or middle point of the data set is much less likely to be biased because the probability of hitting hot spots in two of three unique locations is far lower.

A large set of measurements[2] was obtained from seven coupons of SRM 2569 Level 2. The measurements were made using a primary beam spot approximately 1 mm diameter and moving the coupons beneath the X-ray beam in 0.5 mm increments in a grid pattern. This form of measurement allows creation of maps of the coupons as shown in Figure 1, where Fig. 1a shows a coupon containing two small areas of high Pb counts and Fig. 1b shows an area of completely uniform counts on a second coupon. In these maps, the measured locations are not unique because they overlap when the beam is 1 mm and the step size is 0.5 mm. Therefore, the apparent number of locations yielding a high Pb count rate is inflated when a cluster of inclusions is larger than 0.5 mm across. With an understanding of the shortcomings of the data, it is possible to use it to illustrate the observations a laboratory would expect from Level 2.

The first important observation is that of approximately 13,000 individual measurements made equally divided among seven Level 2 coupons, only about 50 measurements were judged by the authors to be indicative of high Pb inclusions. Most of those inclusions gave Pb counts between 1.5 times and 3 times the median number of counts. The worst case discovered is the one shown in Fig. 1a where one region less than 2 mm across yielded Pb counts of 10 times to 20 times the median, and a second region yielded counts 5 times to 6 times the median for the coupon. The two small regions represent < 4 mm^2 out of a total measured area across seven coupons of approximately 3000 mm^2. By these measures, which are inflated by the oversampling resulting from a beam size of 1 mm diameter and a step size of 0.5 mm, the areas of elevated count rate comprise < 0.5 % of the coupons. Of the seven Level 2 coupons analyzed in this mapping experiment, the most inhomogeneous coupon observed was a single coupon that contains as many inclusions as the other six combined, resulting in 35 measurements judged to be indicative of high Pb inclusions out of approximately 1800 individual measurements. A coupon with this higher rate of Pb inclusions may be more difficult to use unless or until the user understands why and knows to handle the problem locations by using the median of the small number of measurements made during validation work.

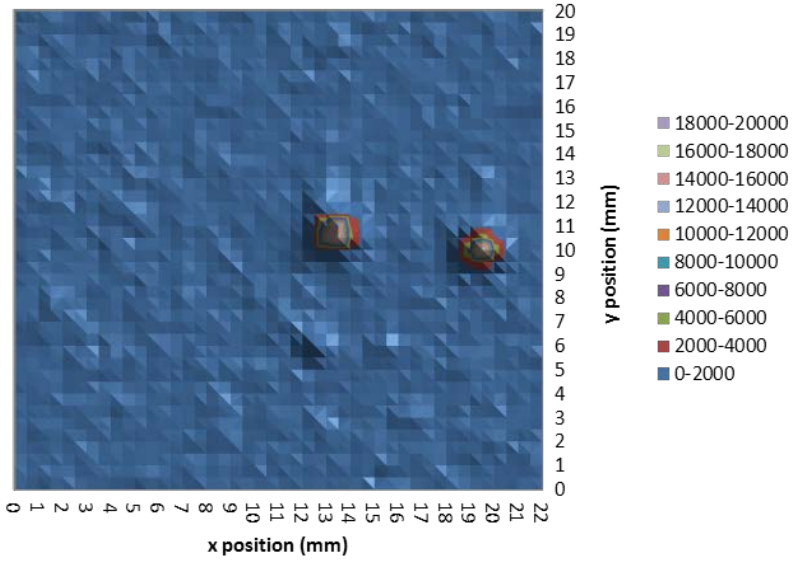
Figure 2 compares the frequency distribution of all 13,000 measurements, shown as Fig. 2a, to the distribution shown in Fig. 2b, which includes the approximately 1800 measurements from the coupon shown in Fig. 1a. In Fig. 2a, the distribution of measured counts is not symmetrical, but it is relatively uniform, and it contains the great majority of measurements with the very high count data appearing as a small up-tick above 1300 counts. The mean of the data and the median are nearly equal, which is evidence of the small influence of the high count spots on the overall distribution. In contrast, Fig. 2b shows how the data set from the coupon in Fig. 1a is clearly skewed toward high counts and yields mean and median values that differ by 76 counts (13 %). The median for this coupon equals the median of the entire data set, demonstrating that the median is unbiased even though this one coupon is significantly less homogeneous than the other six.

Figure 3 illustrates the Pb count results an analyst would obtain from each of the coupons shown in Fig. 1a and Fig. 1b, with Fig. 3a and Fig. 3c associated with the coupon in Fig. 1a and Fig. 3b and Fig. 3d associated with Fig. 1b. Applying the practice recommended in the certificate of analysis of SRM 2569 to each coupon, three measurements were combined by calculating the mean and by calculating the median. Each measurement in a set of three measurements was selected at random using a Monte Carlo approach. Sets of three were chosen until 10,000 mean and median values had been calculated. Fig. 3a and Fig. 3b are the distributions of mean values, and Fig. 3c and Fig. 3d are the distributions of median values. Looking at the distributions of mean values, it is clear that the distribution from the heterogeneous coupon (3a) is skewed toward higher count rates, and that in Fig. 3a some calculated mean values were extremely high (> 1300 counts). The mean is a function of all data in a set, and some sets include the very high count spots shown in Fig. 1a. In comparison, the median is a function of the value at the midpoint of the distribution of three measurements. Therefore, the median is less likely to be influenced by a high count measurement. This fact is shown by the less skewed and narrower distributions of median values.

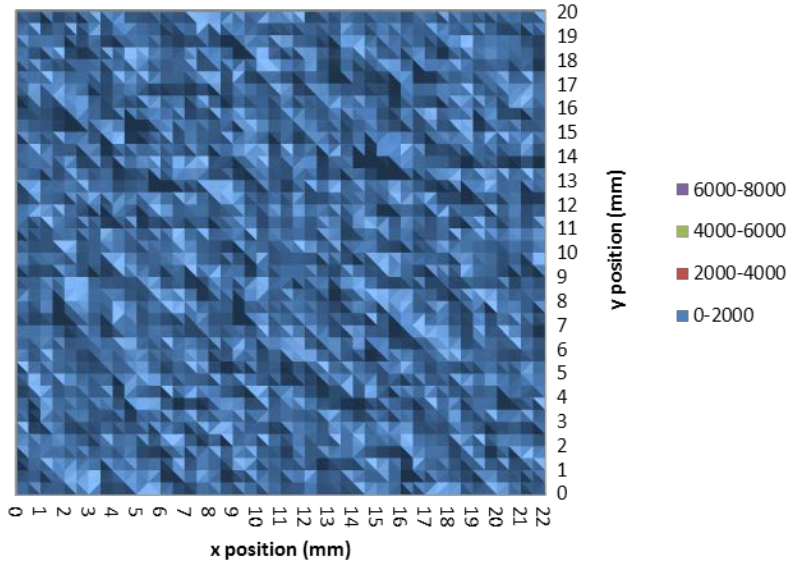
[2] Courtesy of XOS, Inc., East Greenbush, NY.

Figure 1. Subsets of data for SRM 2569 showing maps of individual measurements of a) heterogeneous and b) homogeneous subsamples,

a)



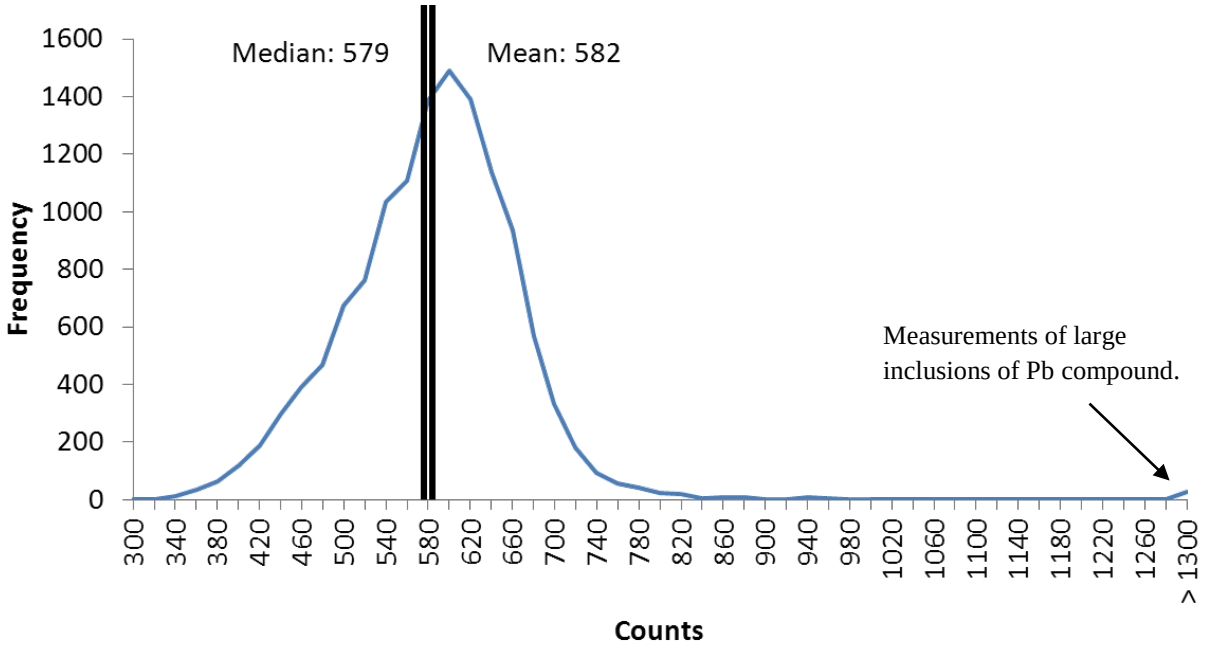
b)



The reader should consider that this illustration uses measured counts of Pb X rays rather than calculated mass per unit area or mass fraction values. It is expected that the illustration will hold for mass per unit area results because a mass per unit area result is obtained directly from the counts of Pb X rays. In contrast, mass fraction results require a number of measured counts of Pb X rays and of other responses. A distribution of mass fraction results will be a convolution of the individual distributions of all contributing measurements. As of the writing of this guide, the authors do not have a set of mass fraction results that correspond to the count rates used in this discussion.

Figure 2. Distribution of background corrected data for a) all 13,000 data points from 7 coupons, and b) data from a single coupon showing heterogeneity effects

a)



b)

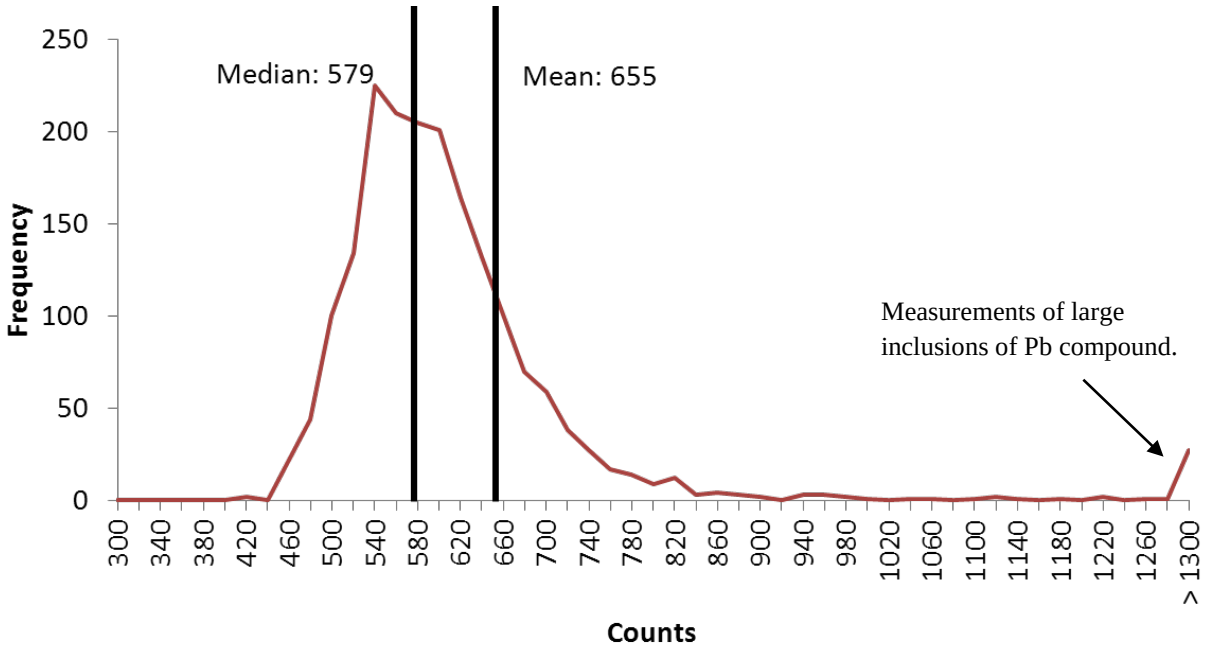
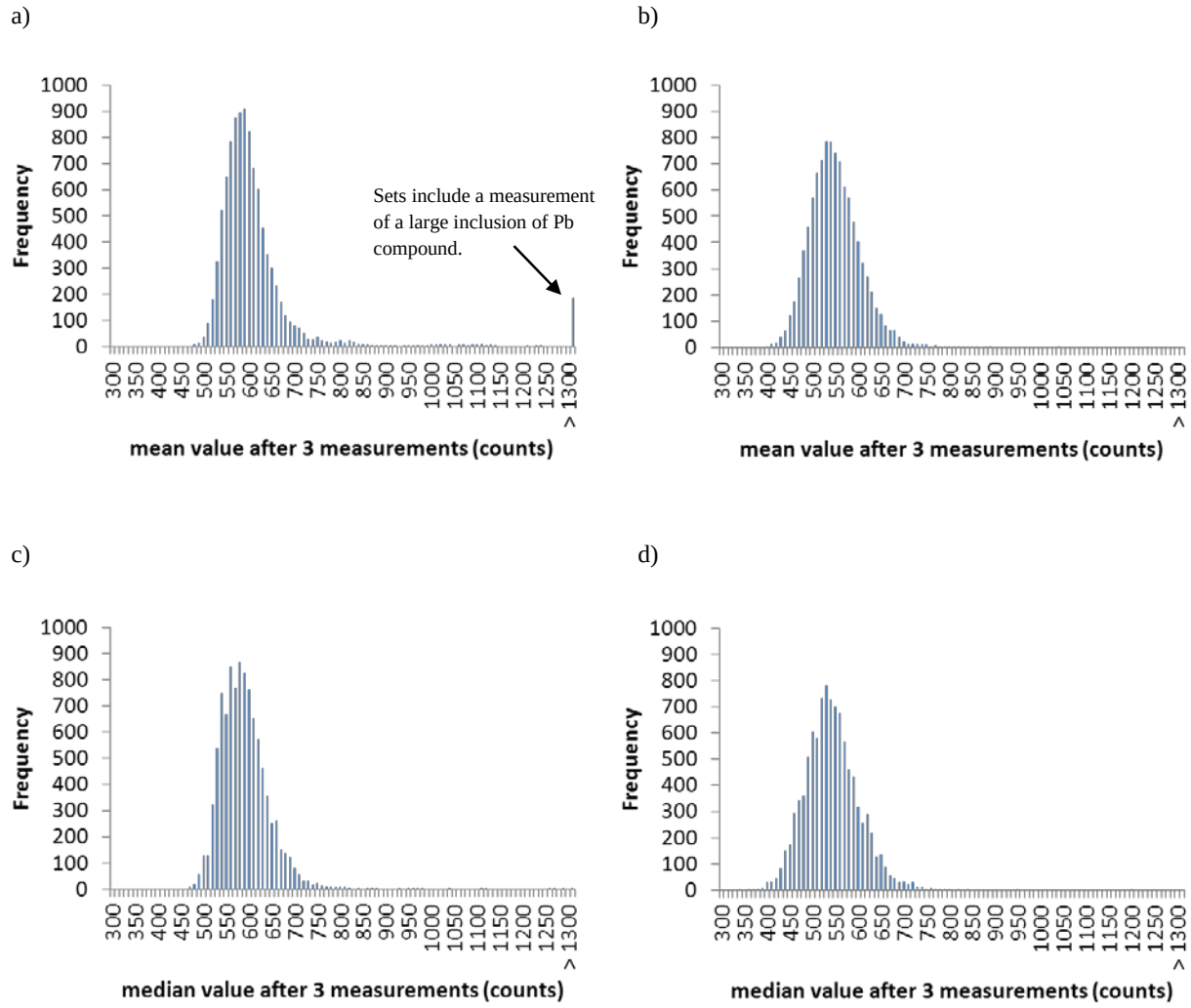


Figure 3. Subsets of data samples from coupons mapped in Figure 1 with distribution of mean values for 10,000 simulated measurements of 3 spots on a) a heterogeneous coupon and b) a homogeneous coupon. The distribution of median values for simulated measurements of 3 spots on c) a heterogeneous coupon and d) a homogeneous coupon are shown for comparison.



Measured Values *versus* Determined Results

When an analyst has obtained the final result for a constituent or property, there is typically a need to compare that result to an action limit or a known value. Meaningful comparisons require an estimate of uncertainty for at least one of the values being compared, preferably for both. To estimate uncertainty, it is necessary to understand what was actually measured and what was actually calculated on the way to the final result. To illustrate this point, it is useful to compare the process of determining the mass of Pb per unit area using XRF to the process of determining mass fraction of Pb using XRF. For mass per unit area, a calibrated instrument is used to measure the net counts of Pb X rays followed by a calculation of the mass per unit area using the calibration algorithm. The XRF determination of mass fraction is typically more complex because the method is usually designed to analyze paints of differing mass of Pb per unit area, thickness and density.

The calculation of mass fraction begins with determination of mass of Pb per unit area using the Pb X-ray counts. Calculation of the mass fraction of Pb requires knowledge of the thickness and the density of the paint, which can be obtained as part of an XRF method or by other methods. Additional XRF measurements may be needed to obtain information on other elements present to account for possible spectral interferences and to check for absorption

effects. Regardless of the details for a given method, the analysts must understand that the result for mass fraction was not determined directly; it was calculated from three determined values. All three parameters, i.e., mass per unit area, density and thickness, have associated uncertainties that contribute to the uncertainty of the mass fraction result. The consequence is that the overall relative uncertainty of a mass fraction result determined in this way is expected to be greater than the relative uncertainty of a result for mass per unit area.

Because many analysts do not understand the measurements and calculations used in these test methods, they must be certain they have all relevant information from the method developer. If the method software reports one or more values, does each have a reported uncertainty? Do the reported uncertainty estimates include all sources of uncertainty? Does the method documentation provide one or more equations showing how the final result is obtained? Equations and detailed uncertainty estimates are needed by a laboratory as part of the method validation documentation. If a complete estimate of uncertainty is not calculated, it may be more difficult to ascertain if determined values differ significantly from those listed on the SRM 2569 certificate.

Conclusions

This guide provides a thorough discussion of the recommended uses of SRM 2569 for validation of a test method. The guide is focused on X-ray methods because they are expected to be the most commonly used. The guide also points out uses of the SRM that are not recommended. Chief among these is use as a routine control chart material, which is expected to cause the coupons to deteriorate rapidly. The certificate of analysis warns that continued exposure of a location on a coupon to high-power X rays and excessive handling will result in cracking and flaking of the paint. If a laboratory includes the SRM in its quality system and proceeds to use it so much that it becomes damaged, the certification will be voided, and the laboratory will be forced to develop a new control procedure and to change its quality system.

This guide discusses the aspects of method validation and provides some suggestions of logical approaches to the use of SRM 2569. The overall process is illustrated by the flow chart in Figure 4. For both repeatability tests and bias detection, it is recommended to use Level 3 first because it contains more Pb distributed homogeneously in a thicker coating. Level 2 is expected to be more difficult to use because it is contains less Pb, is thinner, and is less homogeneous. Level 1 is useful for checking the response to a blank and for use as a benchmark for comparison to blanks from other sources. This guide provides information on the aspects of method validation, including documentation of repeatability and testing for bias. Laboratories should maintain carefully written records of all evidence from validation procedures. If the laboratory purchases the equipment and method as a package, they should obtain needed documentation from the method developer, and they should understand the measurements made and the calculations performed by the method software.

Because Level 2 is somewhat heterogeneous, this guide gives examples of the data that can be obtained from Level 2 coupons. Laboratories should use the median of multiple results rather than the mean because the median is a function of the midpoint of a data distribution and with the low probability of measuring a location containing an inclusion of Pb compound, the median is expected to be a more reliable estimator.

The National Institute of Standards and Technology provides technical and sales support for its Standard Reference Materials. A Technical Contact is assigned for every SRM, and the person's name is listed on the SRM page of the online NIST SRM catalog. As this guide is being written, Dr. John Molloy is the Technical Contact for SRM 2569 Lead Paint Films for Children's Products. At NIST, he uses X-ray fluorescence spectrometry, inductively coupled plasma optical emission spectrometry, mass spectrometry, and other tools to investigate materials and to develop SRMs. Dr. John Sieber was the Technical Project Leader for development of SRM 2569. Dr. Sieber has more than 30 years of experience with X-ray analytical methods. At NIST, he has developed dozens of SRMs for numerous industrial and clinical applications. Both gentlemen are available to assist users with technical aspects of the use of SRM 2569. It is recommended that readers contact Dr. Molloy in writing by sending an email message to john.molloy@nist.gov. General technical inquiries about other SRMs can be sent to srms@nist.gov, and sales inquiries should be sent to srminfo@nist.gov.

[Certain organizations and equipment are named in this paper for the purpose of adequately specifying experimental conditions. Such descriptions do not constitute endorsement by the National Institute of Standards and Technology, nor do they imply that the equipment and organizations are necessarily the best for the purposes described.]

Figure 4. Diagram of validation procedure