


How to use and how not to use certified reference materials in industrial chemical metrology laboratories

John R. Sieber 

Chemical Sciences Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20878, USA

(Received 18 March 2020; accepted 23 March 2020)

As a producer of certified reference materials (CRMs), NIST faces high demand for Standard Reference Materials (SRMs). The demand is exacerbated by widespread misuse of CRMs. When should one use CRMs? When should one not use CRMs? Must labs always use NIST SRMs? How can labs demonstrate analytical capabilities for their accreditation scopes? Why so many questions? Standards developers, laboratory accreditors, and laboratory staff must be able to understand these topics with respect to quality systems in compliance with ISO/IEC 17025. They must calibrate and validate test methods and document traceability to the International System of Units (SI). Many people working in laboratory accreditation and under the umbrella of a quality system do not fully understand what these things are, let alone the language of chemical metrology. On average, they have little training in analytical chemistry, elemental analysis, and reference material development. It is hoped this paper will impress upon the reader the need for understanding how CRMs can be best used in the laboratory. This paper provides a brief background on the above problems and then looks at some of the support and reference information provided by NIST to metals and mining industries labs, commercial CRM producers, and accrediting bodies. The concepts and guidance apply broadly to chemical metrology and fundamental analytical chemistry. The paper includes examples (some from X-ray fluorescence spectrometry) to illustrate concepts. © *National Institute of Standards and Technology, on behalf of Secretary of Commerce, 2020. Published by Cambridge University Press on behalf of the International Centre for Diffraction Data. This is a work of the US Government and is not subject to copyright protection within the United States.* [doi:10.1017/S0885715620000202]

Key words: materials science, XRF, certified reference materials

I. STATE OF AFFAIRS IN APPLIED CHEMICAL METROLOGY

To put the uses of certified reference materials (CRMs) into a modern perspective, consider that accreditation is now a *de facto* requirement to do business. For business purposes, it began with ISO 9001 more than 30 years ago. ISO is the International Organisation for Standardization, which is one of the leading organizations for development of voluntary consensus standards for a wide range of topics, including laboratory organization practices and standard test methods. Laboratories find compelling business reasons to adopt an ISO/IEC 17025 (2017) quality system. ISO/IEC 17034 (2016) and the ISO Guide 30 series are required for organizations that produce CRMs, who may also need to comply with ISO/IEC 17025. However, these documents are difficult to digest, especially when there is little experience within the analytical community. The difficulty is compounded when there is little or no experience or guidance in how and when to best use CRMs. During two ASTM International/National Institute of Standards and Technology workshops in 2004 and 2014, it was shown that roughly four of five auditors from accrediting bodies came from the world of physical

metrology, not chemical metrology. Unfortunately, these quality system auditors were not equipped to work with laboratories doing analytical chemistry.

With the combination of a low level of analytical chemistry expertise by auditors and management of accrediting bodies and by quality assurance officers at industrial companies, some audit requirements have been implemented that are incorrect for chemistry labs. The staff of labs undergoing audits typically do not know how and when to challenge requirements. They may also be dissuaded from doing so by their quality officer because there is a desire to avoid jeopardizing the accreditation.

In seeking guidance, lab staff may look to the organizations from which they obtain CRMs. After all, CRMs are an important and visible tool in the maintenance of quality performance of test methods. Commercial CRM producers are not necessarily equipped to help customers because they may have a relatively low level of knowledge of the aspects of chemical metrology that apply to laboratory quality systems and uses of CRMs. A recent, informal survey of certificates of analysis issued by commercial reference materials producers uncovered shortcomings, including incorrect treatments of uncertainty, mixing of certified and non-certified values in a single reporting table, lack of rigorous statistical approaches, and lack of indications of overall understanding of the fundamental concepts of CRM development: measurand, traceability, uncertainty, and validation.

Author to whom correspondence should be addressed. Electronic mail: john.sieber@nist.gov

Empirical observations appear to show that organizations involved in industrial chemical metrology and the support of industry labs are weak in the expertise needed to perform method validation and to utilize the statistical tools that support demonstration of the quality of test results. Consequently, unnecessary and expensive work is done in the name of quality. As a direct result, NIST Standard Reference Materials (SRMs) have come to be seen as a panacea to satisfy accreditors and clients. If a lab uses NIST SRMs in calibrations, they are likely to get fewer questions and challenges from auditors because auditors and the manufacturers' clients believe that is the way to get a quality calibration and to ensure the entire lab operation obtains quality results. This is simply not true, nor is it possible in all cases, because NIST cannot provide SRMs for all analytical needs.

II. PURPOSES FOR CRMs

At NIST, CRMs are seen as providing higher-order references with values for the true amounts of constituents and properties. Certified values are estimates of the true values because the values are determined based on testing sufficient to elucidate biases in individual test methods, or at least among the methods (May *et al.*, 2000). Typically, that involves using multiple, independent test methods and examining how well the results from those methods agree. Independent test methods do not share the measurement method, and preferably, they do not share methods of sample preparation. In addition, quality assurance materials are analyzed by each method to critically evaluate the individual method.

In the ISO/IEC 17034 system, it is allowed to certify values for method-dependent definitions of a measurand. The measurand is simply what has been measured. For example, an analytical instrument may measure X-ray photons, for the example of X-ray fluorescence spectrometry (XRF), but the test method determines the mass fraction of an element, which is the measurand. A method-dependent value for a measurand would be obtained if only one test method is used for the assignment of the certified value for the reference material. Using a single test method, it is not possible to test for bias in a way that ensures the test result is an estimate of the true value.

CRMs are primarily tools for validation of test methods, the idea is to use a CRM as if it is a typical sample. See the right side of Figure 1, which maps out how a laboratory result can be traced back to the SI. The test results are evaluated to see if they agree with the certified value. It is recognized that CRMs are also frequently used as calibration standards. See the left side of Figure 1. The CRM use is unavoidable, and even necessary when using test methods such as XRF and spark optical emission spectrometry (Spark-OES). These methods are frequently used with solid form materials, e.g., alloys. Both techniques are typically used to determine 10 elements or more in alloys, and both typically require tens of calibration standards to fully calibrate such a method.

CRMs provide values with traceability links to units of the International System (SI), including the kilogram, the mole, and other common units. The SI is not the only standards system that may be used, but it is the world standard system for chemical metrology.

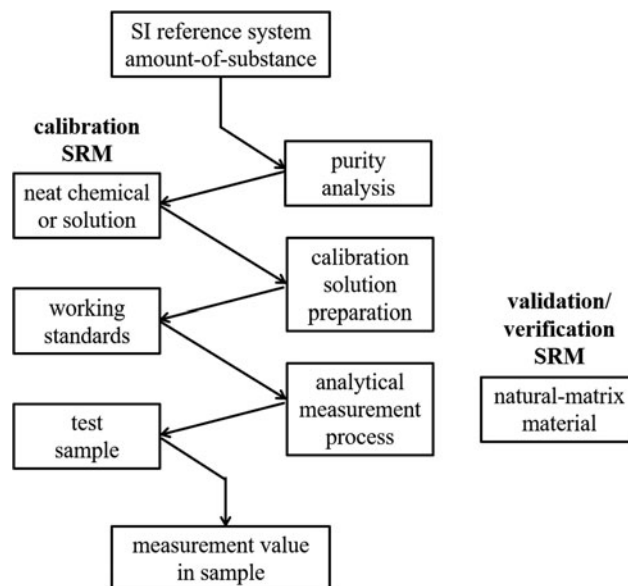


Figure 1. Traceability chain in chemical metrology showing points at which CRMs are used for calibration or validation. The SI is the International System of Units for metrology.

CRMs are often artifacts, that is they are exemplars of real materials with extraordinary homogeneity, often much better than is typically produced for the normal, industrial use of a material. However, many materials do have as-manufactured homogeneity that is fit for the purpose of certification.

For reasons that will be explained below, one should know that CRMs are rare, expensive, and require long development times. NIST cannot guarantee the continued or continuous supply of any SRM. It may go out of stock with no notice to customers, and it may not be renewed. It may be put on sales restriction due to technical issues, e.g., stability testing, which may result in withdrawal of the material. NIST has a complex system for evaluation of the fitness of materials for their purpose and for the needs and potential impacts for SRMs. SRM users are a critical source of feedback to the NIST Quality System for Measurement Services. Their feedback helps define the needs for reference materials (RMs) and reference methods in areas of both existing and emerging technologies.

III. PURPOSES FOR RMs OF OTHER TYPES

The term reference material, in the global lexicon, covers all types of RMs with CRMs being a subset. At NIST, the two categories are separate and are called Standard Reference Materials and Reference Materials. SRMs have at least one certified value based on the NIST criteria for certification, and RMs are materials that have no certified values. When there are certified and non-certified values in the same COA, it is required by ISO Guide 31 (2015) to keep them separate. The goal is to prevent confusion by users of the CRM.

Note that in-house RMs can be developed in a single lab or developed by a third party under contract to the lab needing them. There are a number of reasons to use a RM instead of a CRM. Most reasons are based on the amount of information necessary for the intended use and the lower investment needed to develop an RM compared to a CRM.

The first use for a RM may be in process control, that is for use of control charts or control limits to maintain statistical control of a measurement process, i.e., a test method. For this purpose, the RM need not have a certified value. One can simply obtain a value for the measurand from a freshly implemented and validated calibration of the process to be controlled. Multiple measurements provide a mean value as the nominal target value, and the repeatability standard deviation is used to calculate warning and control limits. The primary need is to have a material in good supply and with a high level of homogeneity fit for the purpose, meaning the standard deviation from measurement to measurement, day to day, and piece to piece is low enough to provide the required statistical control and not inflate the contribution of process control to the overall uncertainty of test results.

The same requirements apply to using a RM for drift correction. The need is for a homogeneous material that provides a stable signal measurable with high precision. Drift correction options may use more than one material to control drift of sensitivity and baseline. There is no need to certify the material. Simply obtain a value for the signal at time zero and compare that to future measurements to perform drift correction, if it is necessary.

A third use for an RM is instrument conditioning, that is measurements of a material to warm up and to stabilize the instrument response for a particular matrix. It is surprising how often labs will use expensive CRMs for this purpose. An excellent example comes from inert gas fusion and combustion test methods. People have been known to take a \$1000 bottle of an SRM and run six, eight, or even 10 samples just for this purpose. That is a great waste because they do not need a certified material. They only need something with sufficient homogeneity to demonstrate the instrument is working correctly with the repeatability precision necessary for quantitative determinations of the species being measured. In the metals industry, ASTM International Committee E01 on Analytical Chemistry of Metals, Ores, and Related Materials has made progress toward ending this behavior by including statements in the standard test methods that this should not be done.

Calibration is another common use for RMs. There is some debate about that, especially among quality system accreditors and advisory panels, who subscribe to the idea that the only way to get accurate calibrations is to use CRMs as the calibration standards. Many of them also believe the valid range of a calibration extends just from the lowest value from a CRM used as a calibrant to the highest value from a CRM used as a calibrant, and no farther in either direction. Of course, these beliefs ignore basic analytical chemistry and the concepts of limit of quantification at the low end and acceptable levels of uncertainty at the high end of a calibration curve.

One important part of calibration is evaluation of matrix influence and spectral interferences. RMs developed for this purpose may be the only viable tools for evaluating the magnitudes of such effects and calculating correction factors, in the absence of fundamental parameters modeling, such as with XRF methods. It may be difficult, if not impossible, to find CRMs that have the levels and variety of compositions necessary for empirical estimation of corrections for any given material matrix. Perhaps the most expedient way to obtain such materials is the development of in-house RMs.

It may also be possible to commission a private sector CRM producer to develop a set of RMs.

ASTM Committee E01 published ASTM E2972 Standard Guide for Production, Testing, and Value Assignment of In-House Reference Materials for Metals, Ores, and Other Related Materials (2015). This standard guide was developed to provide explanations of the types and amounts of information necessary to create an in-house RM.

IV. WHAT TO LOOK FOR IN A CERTIFICATE OF ANALYSIS

This discussion below covers just five categories of information required in a certificate of analysis. A complete list is found in ISO Guide 31. The first category is a description of the material. It is up to the CRM user to decide which materials are the best for their intended purpose. The description explains the composition of the material and gives the unit form and size. The form may be a bottle of chips of metal or powder, or it may be a disk of metal or glass. For bottled material, the mass of the bottle contents is given. For solid forms, the shape and dimensions are given.

The next category is a description of the intended uses of the material. For NIST SRMs designed for chemical metrology, the intended uses are summarized by the following sentence. This SRM is intended primarily for evaluation of methods of analysis for similar materials and for validation of value assignment of in-house RMs. There may be specialized uses, possibly related to specific test methods, validation of results for regulatory purposes, or calibration of special equipment.

Next, and perhaps most important is an explanation of the certified values. There are three topics which must be addressed: (1) definition of measurands, (2) explanation of metrological traceability, and (3) definitions of uncertainty estimates. For the definition of measurands, a typical statement is as follows. The measurands are the mass fractions of the total amounts of the elements in a steel matrix. Other possibilities include measurands that are chemical compounds or chemical states of elements. The matrix may be any type of material from steel and other alloys to plastics, foodstuffs, water, fuel oils, and many more. Metrological traceability states the units system for the certified values, which is discussed further below. Uncertainty estimates are typically defined as expanded uncertainty estimates expressed at a coverage level of approximately 95 %, and a more detailed discussion of uncertainty follows the details on traceability.

The period of validity of certification is relatively simple in that it can be an expiration date or indefinitely, which means the assigned values and uncertainty estimates for the material are expected to remain valid for a very long period of time. This option is used when a material is known to be highly stable, e.g., most metals and alloys, when stored correctly. Because CRM producers must perform stability testing and respond to customer inquiries, all CRMs are monitored for their stability.

The last category of information on this top five list is instructions for storage, handling, and use. Users need to know how much of a unit is certified. Perhaps the material is a disk of chill-cast metal, and it may be certified for only the first 10 mm deep from the original test surface. The user will also need to know any instructions for sample

preparation. If the assigned values are given in a dry basis, instructions for drying are given. Another example is a warning about potential contamination of a disk of metal when a fresh surface is prepared by grinding. Bottles of powder and liquids must be carefully mixed prior to sampling. Every certificate of analysis should provide a minimum recommended sample quantity for an analysis. This will be explained later under the heading of heterogeneity. Finally, the instructions may explain how to use the uncertainty estimates for comparisons of values and for propagation of uncertainty, for which there is more information later in this paper.

V. METROLOGICAL TRACEABILITY

It is the responsibility of the CRM issuer to establish metrological traceability (traceability for short) of assigned values to a higher-order reference system such as the SI. Traceability is the association of units with an assigned value. In Figure 1, the measured value at the bottom has units from the SI at the top, when the two are connected by calibrations with uncertainty estimates. In the figure, the traceability chain is the assemblage of actions in the boxes and the arrows connecting those boxes. Here is an example SRM statement: “The certified values are metrologically traceable to the SI derived unit of mass fraction expressed as percent”. For NIST SRMs, this is usually sufficient wording to ensure SRM customers that NIST has done the work needed to establish traceability. Many CRMs from commercial producers have no statement of this kind. Instead, they list CRMs and chemicals used in the certification process, without an explicit statement that traceability was established to a specific SI unit. It is an indication that they need more training in this area.

For a testing lab to discuss traceability, they must have first dealt with uncertainty estimates for their results, and then, they may wish to state something like the following: “The result value is traceable to the SI unit as realized by NIST through the value for element XX in Standard Reference Material YYY”. While there may be value in being able to show a link to NIST, the traceability link can be through values provided by any CRM producer as long as that producer can demonstrate traceability to the SI units system.

It is also possible for a lab to achieve metrological traceability to an SI unit without a CRM. The simplest way is to use a high-purity material having an assay with stated uncertainty and a balance calibrated to establish traceability to the kilogram. There are discussions and demonstrations of this concept in the published papers of Staats (1988, 1990) and Staats and Strieder (1993).

Traceability can be established by using a CRM as a quality assurance material. A comparison between measurement results may be viewed as a calibration, if the comparison to a RM is used to check and, if necessary, correct the quantity value and measurement uncertainty attributed to the measured material (JCGM 200:2012).

VI. HETEROGENEITY

Composition variance is a part of the overall uncertainty of the value of a measurand. Heterogeneity of a material is evaluated based on the purposes for which the material is intended. With that knowledge, the decision can be made as

to whether the material is fit for the intended purpose. To test heterogeneity, choose a method or methods based on considerations of minimal sample preparation, small quantity tested, low counting uncertainty, instrument stability, and testing *within-unit* variance versus *among-units* variance. In most cases, this testing must be done after the candidate material has been prepared and packaged as units for sale.

One or more quantitative analysis methods may adequately account for material heterogeneity as an effect contributing to one of its components of variance. The CRM developer should plan quantitative analyses with that in mind and balance it against the amount of work requested of analysts. In the final assessment of overall uncertainty, it may or may not be necessary to have a component of uncertainty explicitly stated to inform a user of the material heterogeneity. The decision must be made for every CRM development project. Some CRM producers believe that a standard uncertainty component for heterogeneity must be published for every measurand in every certificate of analysis. This is not true; however, the COA should contain a statement about heterogeneity, especially when sampling is strongly affected. For example, the heterogeneity and structure of a material may be such that certain portions should not be sampled or that for a particular test method, multiple portions should be tested, and the results averaged.

On the concept of minimum sample quantity, think of choosing small aliquots like using a more powerful microscope. At some magnification, it becomes apparent the material is very heterogeneous. Users of CRMs need to know how small a sample can be taken that still allows the analyst to expect to obtain a single result with a value and an uncertainty estimate comparable to the certified value and its uncertainty estimate. If smaller specimens than the recommended minimum must be used, the user must take multiple samples and calculate the mean for comparison to the certified value.

When heterogeneity among units of a CRM is large, the difference between the true value for the amount of measurand in one unit and the certified value may be large, or the difference between any two units may be large. Comparisons among units and of a unit to the certified value become more difficult. Figure 2 represents this concept, where the top, black curve shows the range of values for individual

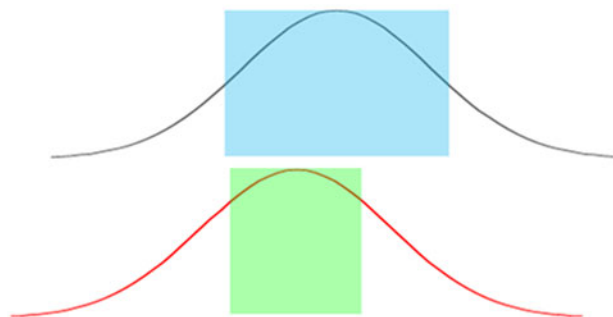


Figure 2. (Color online) Conceptualization of uncertainty with the top curve representing the certified value for a CRM as the central point of the Gauss curve, the width of the Gauss curve representing 95% coverage interval, and the top (blue) rectangle representing an exaggerated contribution by overall material heterogeneity. The bottom curve represents the true value for a single CRM unit as the central point of the (red) curve with the curve and (green) rectangle representing the overall and within-unit heterogeneity for that unit, respectively.

bottles as the width of the superimposed (blue) rectangle. This is an exaggerated case of high unit-to-unit variance. The central point of the upper curve represents the overall certified value with the width of the curve representing the uncertainty interval estimated for the certified value. The bottom (red) curve represents the composition of a single unit of the material with the central point being the true value of the amount of substance in that one unit. The (green) rectangle superimposed on the lower curve represents composition variance within the single unit as being smaller than the upper (blue) *among-units* variance. Greater heterogeneity among CRM units causes more units to be shifted farther from the overall certified value. Some units may have their (red) curve shifted way to the left, and some may be shifted way to the right. Therefore, comparisons between any two units must allow for the probability of larger differences, which makes the CRM *less* useful for high precision comparisons. The overall composition variance must be minimized with the contribution of *among-units* variability being less significant than *within-unit* variability. That is, the lower (green) rectangle for a single unit should be wider than the upper (blue) rectangle but not so wide as to cause the problem of excessive heterogeneity within each unit. When heterogeneity is sufficiently low overall, a material is described as having homogeneity fit for purpose.

VII. UNCERTAINTY

NIST consensus values and uncertainty estimates for measurands are calculated using statistical methods with many different approaches available, and they can be tailored to each CRM project. Examples of the tools can be seen in the NIST Consensus Builder at <https://consensus.nist.gov/>, which provides three different approaches for exploration. For those interested in general purpose, statistical tools for evaluation of uncertainty, NIST offers the Uncertainty Machine at <https://uncertainty.nist.gov/>. Both web tools have online manuals with examples.

In most cases, certified values are accompanied by a symmetric uncertainty interval, expressed as a half-interval, U , with approximately 95% coverage. That means the full interval, $2U$, is a range within which the true value is expected to be with 95% confidence. Propagate uncertainty using the combined standard uncertainty, $u_c = U/k$, where k is a coverage factor chosen for the effective degrees of freedom in the evaluation of the certified value. Values for k are given in the certificate of analysis. If for some reason, there is no k -value given, the user is advised to set $k = 2$ and to assume the effective degrees of freedom are high enough for that approximation. Older NIST COAs may give just u_c based on variance among collaborator mean results and expert judgement of other components of uncertainty.

In some cases, the uncertainty estimate may be given as an asymmetric interval. Then, the COA will show the certified value accompanied by the range of the interval having approximately 95% coverage. The decision to provide an asymmetric interval may result from the distributions of values from individual test methods having markedly different widths on either side of the consensus value. In some cases, the consensus value may be so close to either zero or 100% that the statistics indicate values <0 or $>100\%$ are possible. Then, the uncertainty interval will be truncated at 0 or 100, making

the interval asymmetric. Users can approximate u_c by dividing the range of the coverage interval by four. A conservative alternative is division of the wider side of the range by two. An approach recommended by statisticians is to use a Monte Carlo method for propagation of error, using the actual error distributions representing the uncertainties of the certified values. This method requires fewer assumptions and approximations. To make it possible, the CRM producer must supply files containing the error distributions for the assigned values. NIST SRM 2780a Hard Rock Mine Waste provides that information through a link to the NIST SRM Online Request System: https://www-s.nist.gov/srmors/view_datafiles.cfm?srm=2780a.

CRM users must be aware of the great variety of uncertainty definitions to be found in commercial CRM certificates. When in doubt, ask the issuing body to explain their approach and the components of uncertainty included in the estimate. At a bare minimum, the components must include an estimate of repeatability standard deviation from analyses of the material and a standard uncertainty estimate based on the uncertainty of calibration. There may be additional uncertainty components, but without these two things, an uncertainty budget is incomplete.

VIII. COMPARING YOUR RESULT TO A CERTIFIED VALUE

The primary purpose of a comparison between a found result and a certified value is to test for bias in the analytical method. ISO Guide 33 (2015) provides an equation for a bias detection limit as shown in Eq. (1), where x is the value from each source, u is the combined standard uncertainty for each value, and k is a coverage factor. Most experts agree that $k = 2$ is the most reasonable and convenient choice to approximate 95% confidence. When the difference between the found result and the certified value is less than this detection limit, there is no evidence for bias in the found result:

$$|x_{\text{found}} - x_{\text{CRM}}| \leq k\sqrt{u_{\text{found}}^2 + u_{\text{CRM}}^2} \quad (1)$$

The first thing needed for this calculation is an estimate of the uncertainty of the found result. Again, a complete uncertainty budget providing an overall estimate similar in definition to the certificate estimate would be ideal. However, one can begin with repeatability standard deviation and add components for other sources of uncertainty, including calibration, to improve the coverage of the uncertainty of the found result. The more inclusive the uncertainty estimate, the better the information to be gained from the comparison.

The concepts explained above can be illustrated as in Figure 3 in which the four cases are designated: (1) Found result (indicated by X) with no uncertainty either falls outside the certified value uncertainty interval (vertical error bar) (1a), or if the analyst is lucky, it falls inside the interval (1b); (2) Certified value falls within the found value uncertainty interval; (3) Found and Certified intervals overlap and either both values fall inside each other's interval (3a) or both values fall outside each other's interval (3b); and (4) the two intervals do not overlap. For case 1a, there is no evidence that the found and certified values agree. For case 4, there is evidence of a bias between the values. For all other cases, there is evidence providing some level of confidence there is agreement. Note

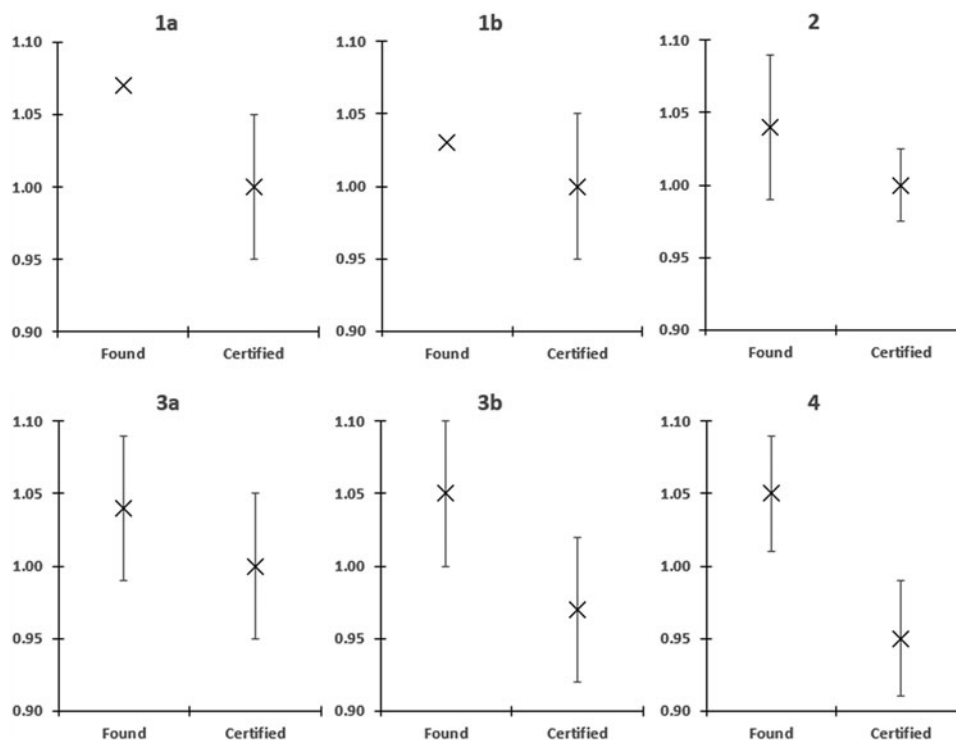


Figure 3. Example cases for comparisons of found results to a certified value and accompanying uncertainty estimate. Case 1a: bias indicated, found value without uncertainty outside certified value coverage interval; Case 1b: no evidence of bias, found value without uncertainty falls inside certified value interval; Case 2: no evidence of bias, the uncertainty of found value overlaps certified value; Case 3a: no evidence of bias, found coverage interval overlaps certified value and certified value interval overlaps found value; Case 3b: may or may not be biased, two coverage intervals overlap but do not cover the values; and Case 4: bias indicated, no overlap of coverage intervals.

that Eq. (1) may indicate a detected bias in contradiction to Figure 3b because the standard uncertainty estimates are added in quadrature. It should also be understood that case 1b was referred to as luck because at NIST the goal is to make the uncertainty estimate as small as possible to show the level of confidence that the certified value is a good estimate of the true value. This goal is in opposition to the goal of including the material heterogeneity variance in the uncertainty estimate, which tends to broaden the coverage interval. In other words, a compromise must be struck to improve confidence in the certified value without hindering the practical utility of the CRM.

IX. EQUIVALENCE OF CRMs AND RMs

The last topic for discussion addresses an example of a faulty accreditation requirement. One well-known requirement in the metals industry is that all CRMs must have been developed under an ISO/IEC 17034 accredited system. While such a situation would be ideal, it causes significant problems in practice. It may be that this requirement resulted from an attempt at universal application of the relatively new ISO/IEC 17034 to all accreditations. Because ISO/IEC 17034 (and its predecessor ISO Guide 34) is a relatively recent phenomenon, there are relatively few CRMs that are compliant out of the hundreds of extant CRMs. The implementation of Guide 34 did not automatically invalidate all existing CRMs at that time. To do that would be to try to argue that none of the CRM producers and national metrology institutes knew how to develop CRMs prior to sometime in the 1990s. That simply is not true. In fact, Guide 34 grew out of efforts by

qualified organizations to codify the processes in CRM development for the general benefit of all.

A good way to demonstrate that many CRMs in existence prior to the advent of ISO Guide 34 are still good even after the conversion to ISO/IEC 17034 is through the concept of equivalence. Equivalence means the CRMs, test methods, or test results from different sources are all fit for the same purpose. The example shown in Figure 4 is the calibration of copper in low alloy steel, using CRMs developed from 1965 through 1993. There are 36 CRMs with 33 developed by NIST. The other three were issued by Bureau of Analysed Samples Ltd. (Middlesbrough, England). In the graph, the calibration curve passes through every set of horizontal error bars. That fact and the relatively narrow sizes of the error bars indicate that all 36 CRMs are fit for the purpose of calibration of Cu in steel. In other words, they are all equivalent for that application. Consequently, a subset of the CRMs would be sufficient for the calibration task, and no biases would be expected among results from different subsets of these CRMs.

The concept of demonstration of equivalence has a practical application in testing laboratories. Perhaps a lab has a small collection of CRMs and RMs with two or three each from multiple CRM producers and a number of RMs included. When it can be demonstrated that all of them fit on the same calibration curve within their respective uncertainty estimates, the lab has validation evidence to show to auditors, who may challenge the lab's practices, their choices of CRMs and their inclusion of RMs as calibration standards. Of course, the RMs must have defensible uncertainty estimates that can be used as part of the equivalence comparisons. However, they need not all be certified under an ISO/IEC 17034 compliant quality system.

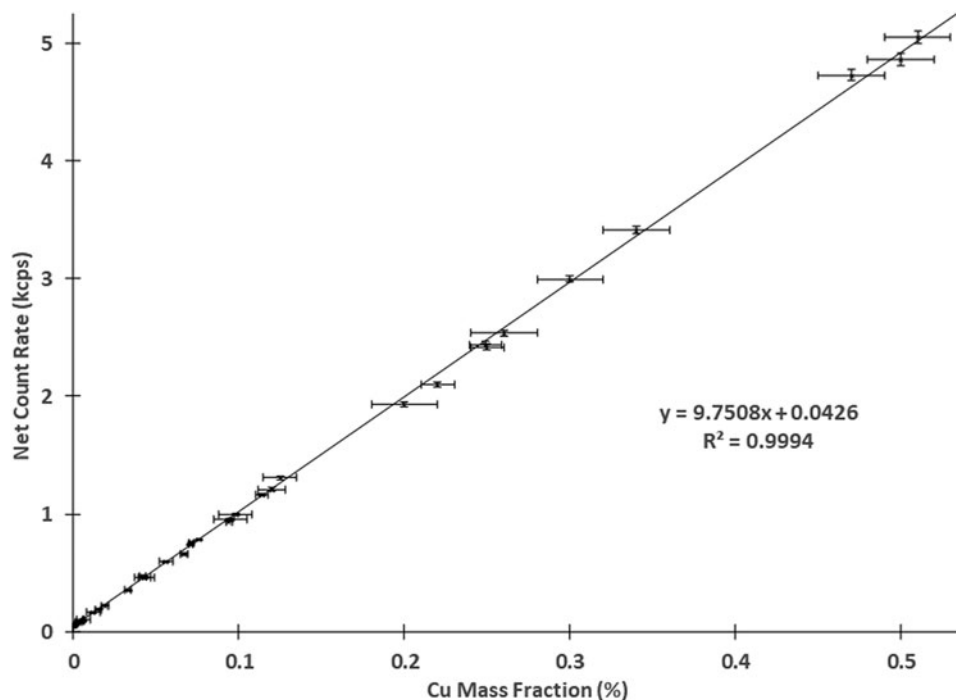


Figure 4. Calibration of a wavelength-dispersive XRF spectrometer for copper in low alloy steel, using 33 NIST SRMs and three materials from Bureau of Analysed Samples. Horizontal error bars equal the published uncertainty estimate from each certificate of analysis. Vertical error bars are X-ray counting standard uncertainty.

X. CONCLUSION

This manuscript is intended to teach a broad audience some fundamentals of CRM development, documentation, and use. It began by setting the stage with a description of the state of affairs in industrial chemical metrology in which there is a strong need for education in basic analytical chemistry and in method validation. The discussion covered the primary purposes for CRMs and contrasted them to uses better suited to non-certified RMs. Then, the fundamentals of CRM documentation were described to explain what a CRM user needs to ascertain from a certificate of analysis. Of those fundamentals, metrological traceability, heterogeneity, and uncertainty were explained in greater detail from the NIST perspective. The paper finished with two examples of how to interpret measured results for CRMs, including testing for bias between found results and certified values and demonstration of equivalence of RMs. Following the references list below, the reader will find a list of recommended reading on some of the topics discussed herein.

SRM development and other support provided by NIST are done with user needs in mind. NIST welcomes feedback and requests for support. The most direct way to provide feedback is by contacting the NIST Technical Contact for the individual SRM of interest, or one that is similar to the composition needed. For example, see SRM 8k (NIST, 2017) for which this author is Technical Contact. Questions of a more general nature can be sent to srms@nist.gov, or one may visit the Standard Reference Materials website at www.nist.gov/srm/.

XI. DISCLAIMER

Certain commercial equipment, instrumentation, or materials are identified in this document to adequately specify the

experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

- ASTM Workshop on the Impact of Accreditation on Reference Materials and Method Development in Analytical Laboratories Supporting the Metals and Ores Industries, Gaithersburg, MD, USA, 19 May 2014. Available at: <https://www.nist.gov/news-events/events/2014/05/astm-workshop-impact-accreditation-reference-materials-and-method> (accessed September 2019)
- ASTM E2857-11 (Reapproved 2016). "Standard guide for validating analytical methods," in *Annual Book of ASTM Standards*, Vol. 03.05 (ASTM International, West Conshohocken, PA). doi:10.1520/E2857-11R16.
- ASTM E2972-15 (2015). "Standard guide for production, testing, and value assignment of in-house reference materials for metals, ores, and other related materials," in *Annual Book of ASTM Standards*, Vol. 03.05 (ASTM International, West Conshohocken, PA). doi:10.1520/E2972-15.
- Duewer D. L., Lippa K. A., Rukhin A. L., and Sharpless K. E. (2014). "The ABCs of Using Standard Reference Materials in the Analysis of Foods and Dietary Supplements: A Practical Guide," NIST Spec. Pub. 260-181. Available at: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.260-181.pdf> (accessed August 2019).
- ISO Guide 31 (2015). *Reference Materials – Contents of Certificates and Labels* (International Organization for Standardization, Geneva).
- ISO Guide 33 (2015). *Uses of Reference Materials – Good Practice in Using Reference Materials* (International Organization for Standardization, Geneva).
- ISO/IEC 17025 (2017). *General Requirements for the Competence of Calibration and Testing Laboratories* (International Organization for Standardization, Geneva).
- ISO/IEC 17034 (2016). *General Requirements for the Competence of Reference Material Producers* (International Organization for Standardization, Geneva).
- JCGM 100 (2008). *Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement (GUM 1995 with Minor Corrections)* (International Bureau of Weights and Measures, Paris).

- Available at: https://www.bipm.org/utis/common/documents/jcgm/JCGM_100_2008_E.pdf (accessed September 2019).
- JCGM 101 (2008). *Evaluation of Measurement Data – Supplement 1 to the “Guide to the Expression of Uncertainty in Measurement” - Propagation of Distributions Using a Monte Carlo Method* (International Bureau of Weights and Measures, Paris). Available at: https://www.bipm.org/utis/common/documents/jcgm/JCGM_101_2008_E.pdf (accessed September 2019).
- JCGM 200 (2012). *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)*, 3rd ed. 2008 version with minor corrections (International Bureau of Weights and Measures, Paris). Available at: https://www.bipm.org/utis/common/documents/jcgm/JCGM_200_2012.pdf (accessed September 2019).
- May, W., Parris, R., Beck, C., Fassett, J., Greenberg, R., Guenther, F., Kramer, G., Wise, S., Gills, T., Colbert, J., Gettings, R., and MacDonald, B. (2000). “Definitions of Terms and Modes Used at NIST for Value-Assignment of Reference Materials for Chemical Measurements,” NIST Spec. Pub. 260-136.
- NIST (2017). Certificate of Analysis “SRM 8k - Bessemer Steel (Simulated) 0.1% Carbon (Chip Form),” 19 July 2017. Available at: https://www-s.nist.gov/srmors/view_detail.cfm?srm=8K (accessed September 2019).
- Staats, G. (1988). “Computer aided conversion of repeatability into trueness as applied to inorganic analytical chemistry,” *Fresenius J. Anal. Chem.* **330**, 469.
- Staats, G. (1990). “On the role of pure oxide materials for primary calibration and validation in inorganic bulk analysis,” *Fresenius J. Anal. Chem.* **336**, 132.
- Staats, G. and Strieder, S. (1993). “Validation of sulphide-bearing certified reference samples by primary calibration as applied to X-ray fluorescence,” *X-Ray Spectrom.* **22**(3), 132.
- Taylor, B. N. and Kuyatt, C. E. “Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results,” NIST Technical Note 1297. Available at: <https://physics.nist.gov/cuu/Uncertainty/index.html> (accessed August 2019).