NIST SPECIAL PUBLICATION 260-100

STANDARD REFERENCE MATERIALS

HANDBOOK FOR SRM USERS

John K. Taylor

(edited by N.M. Trahey)
Standard Reference Materials Program
National Institute of Standards and Technology
Gaithersburg, MD 20899

Supersedes NBS Special Publication 260-100, 1985 Edition

Issued February 1993

U.S. DEPARTMENT OF COMMERCE, *Ronald H. Brown, Secretary*
National Institute of Standards and Technology, *John W. Lyons, Director*

National Institute of Standards and Technology Special Publication 260-100
Natl. Inst. Stand. Technol. Spec. Publ. 260-100, 120 pages (Feb. 1993)
CODEN: NSPUE2

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1993

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402–9325

This edition of NIST SP 260-100
is dedicated to

JOHN K. TAYLOR

Chemist - Metrologist - Teacher

NOTE

This handbook has been edited to upgrade dated information and to reflect changes which have occurred at the National Institute of Standards and Technology (NIST) and in the NIST Standard Reference Materials Program since the handbook was first issued in 1985. References to "NBS" have been changed to "NIST" throughout the text, and more current information has been provided about available Standard Reference Materials and selected NIST publications. A general review of the handbook and its appendices (such as Appendix C. Statistical Tools) has also been undertaken to assure that the information remains consistent with NIST guidance on measurements. The editing process had not yet begun when Dr. Taylor passed away, and every effort has been made to preserve those parts of the text he prepared, essentially as he wrote them.

The contributions of James I. Shultz, ASTM Research Associate, NIST Standard Reference Materials Program, Susannah B. Schiller, Statistician, NIST Statistical Engineering Division, and Sharon K. Maxwell, NIST Standard Reference Materials Program, to the review and editing of this handbook are gratefully acknowledged.

NMT
February, 1993

PREFACE

Standard Reference Materials (SRMs), as defined by the National Institute of Standards and Technology (NIST), are well-characterized materials produced in quantity and certified for one or more physical or chemical properties. Their purpose is to assure the accuracy and compatibility of measurements throughout the United States. Both nationally and throughout the world, SRMs are widely used as primary standards in many different fields in science, industry and technology. They are also used extensively in the fields of environmental and clinical analysis. In many applications, traceability of measurement and quality control processes to the National Measurement System are carried out through the mechanism and use of SRMs. Therefore, for many of the Nation's scientists and technologists, it is of more than passing interest to know the details of the measurements made at NIST to arrive at the certified values of the SRMs produced. A NIST series of papers, referred to as the NIST Special Publication - 260 Series, is reserved for this purpose.

The 260 Series, of which this publication is a part, is dedicated to the dissemination of information on different phases of the preparation, measurement, certification, and use of NIST SRMs. In general, much more information will be found in these publications than is usually allowed (or desirable) in scientific journal articles. The papers enable SRM users 1) to assess the validity and accuracy of the measurement processes employed, 2) to judge the statistical analyses and, 3) to learn details of the techniques and methods utilized for work requiring the greatest care and accuracy. Often these papers also provide sufficient additional SRM information - not included in certificates - to allow their use for new applications in diverse fields unforeseen at the time the SRMs were originally issued.

A complete listing of the publications comprising the NIST 260 Series is provided in this document.

Stanley D. Rasberry
Director
Office of Measurement Services

William P. Reed
Chief
Standard Reference Materials Program

OTHER NIST PUBLICATIONS IN THIS SERIES

McKenzie, R. L., ed., NIST Standard Reference Materials Catalog 1990-91, NIST Spec. Publ. 260 (January 1990)

Michaelis, R. E., and Wyman, L. L., Standard Reference Materials: Preparation of White Cast Iron Spectrochemical Standards, NBS Misc. Publ. 260-1 (June 1964). COM74-11061**

Michaelis, R. E., Wyman, L. L., and Flitsch, R., Standard Reference Materials: Preparation of NBS Copper-Base Spectrochemical Standards, NBS Misc. Publ. 260-2 (October 1964). COM74-11063**

Michaelis, R. E., Yakowitz, H., and Moore, G. A., Standard Reference Materials: Metallographic Characterization of an NBS Spectrometric Low-Alloy Steel Standard, NBS Misc. Publ. 260-3 (October 1964). COM74-11060**

Alvarez, R., and Flitsch, R., Standard Reference Materials: Accuracy of Solution X-Ray Spectrometric Analysis of Copper-Base Alloys, NBS Misc. Publ. 260-5 (March 1965). PB168068**

Shultz, J. I., Standard Reference Materials: Methods for the Chemical Analysis of White Cast Iron Standards, NBS Misc. Publ. 260-6 (July 1965). COM74-11068**

Bell, R. K., Standard Reference Materials: Methods for the Chemical Analysis of NBS Copper-Base Spectrochemical Standards, NBS Misc. Publ. 260-7 (October 1965). COM74-11067**

Richmond, M. S., Standard Reference Materials: Analysis of Uranium Concentrates at the National Bureau of Standards, NBS Misc. Publ. 260-8 (December 1965). COM74-11066**

Anspach, S. C., Cavallo, L. M., Garfinkel, S. B., Hutchinson, J. M. R., and Smith, C. N., Standard Reference Materials: Half Lives of Materials Used in the Preparation of Standard Reference Materials of Nineteen Radioactive Nuclides Issued by the National Bureau of Standards, NBS Misc. Publ. 260-9 (November 1965). COM74-11065**

Yakowitz, H., Vieth, D. L., Heinrich, K. F. J., and Michaelis, R. E., Standard Reference Materials: Homogeneity Characterization of NBS Spectrometric Standards II: Cartridge Brass and Low-Alloy Steel, NBS Misc. Publ. 260-10 (December 1965). COM74-11064**

Napolitano, A., and Hawkins, E. G., Standard Reference Materials: Viscosity of Standard Lead-Silica Glass, NBS Misc. Publ. 260-11 (November 1966).

Yakowitz, H., Vieth, D. L., and Michaelis, R. E., Standard Reference Materials: Homogeneity Characterization of NBS Spectrometric Standards III: White Cast Iron and Stainless Steel Powder Compact, NBS Misc. Publ. 260-12 (September 1966).

Menis, O., and Sterling, J. T., Standard Reference Materials: Determination of Oxygen in Ferrous Materials"SRM 1090, 1091, and 1092, NBS Misc. Publ. 260-14 (September 1966).

Yakowitz, H., Michaelis, R. E., and Vieth, D. L., Standard Reference Materials: Homogeneity Characterization of NBS Spectrometric Standards IV: Preparation and Microprobe Characterization of W-20% Mo Alloy Fabricated by Powder Metallurgical Methods, NBS Spec. Publ. 260-16 (January 1969). COM74-11062**

Paule, R. C., and Mandel, J., Standard Reference Materials: Analysis of Interlaboratory Measurements on the Vapor Pressure of Gold (Certification of Standard Reference Material 745). NBS Spec. Publ. 260-19 (January 1970). PB190071**

Paule, R. C., and Mandel, J., Standard Reference Materials: Analysis of Interlaboratory Measurements on the Vapor Pressures of Cadmium and Silver, NBS Spec. Publ. 260-21 (January 1971). COM74-11359**

Yakowitz, H., Fiori, C. E., and Michaelis, R. E., Standard Reference Materials: Homogeneity Characterization of Fe-3 Si Alloy, NBS Spec. Publ. 260-22 (February 1971). COM74-11357**

Napolitano, A., and Hawkins, E. G., Standard Reference Materials: Viscosity of a Standard Borosilicate Glass, NBS Spec. Publ. 260-23 (December 1970). COM71-00157**

Sappenfield, K. M., Marinenko, G., and Hague, J. L., Standard Reference Materials: Comparison of Redox Standards, NBS Spec. Publ. 260-24 (January 1972). COM72-50058**

Hicho, G. E., Yakowitz, H., Rasberry, S. D., and Michaelis, R. E., Standard Reference Materials: A Standard Reference Material Containing Nominally Four Percent Austenite, NBS Spec. Publ. 260-25 (February 1971). COM74-11356**

Martin, J. F., Standard Reference Materials: National Bureau of Standards-US Steel Corporation Joint Program for Determining Oxygen and Nitrogen in Steel, NBS Spec. Publ. 260-26 (February 1971). PB 81176620**

Garner, E. L., Machlan, L. A., and Shields, W. R., Standard Reference Materials: Uranium Isotopic Standard Reference Materials, NBS Spec. Publ. 260-27 (April 1971). COM74-11358**

Heinrich, K. F. J., Myklebust, R. L., Rasberry, S. D., and Michaelis, R. E., Standard Reference Materials: Preparation and Evaluation of SRM's 481 and 482 Gold-Silver and Gold-Copper Alloys for Microanalysis, NBS Spec. Publ. 260-28 (August 1971). COM71-50365**

Geller, S. B., Standard Reference Materials: Calibration of NBS Secondary Standard Magnetic Tape (Computer Amplitude Reference) Using the Reference Tape Amplitude Measurement "Process A-Model 2," NBS Spec. Publ. 260-29 (June 1971). COM71-50282**

Gorozhanina, R. S., Freedman, A. Y., and Shaievitch, A. B. (translated by M. C. Selby), Standard Reference Materials: Standard Samples Issued in the USSR (A Translation from the Russian), NBS Spec. Publ. 260-30 (June 1971). COM71-50283**

Hust, J. G., and Sparks, L. L., Standard Reference Materials: Thermal Conductivity of Electrolytic Iron SRM 734 from 4 to 300 K, NBS Spec. Publ. 260-31 (November 1971). COM71-50563**

Mavrodineanu, R., and Lazar, J. W., Standard Reference Materials: Standard Quartz Cuvettes for High Accuracy Spectrophotometry, NBS Spec. Publ. 260-32 (December 1973). COM74-50018**

Wagner, H. L., Standard Reference Materials: Comparison of Original and Supplemental SRM 705, Narrow Molecular Weight Distribution Polystyrene, NBS Spec. Publ. 260-33 (May 1972). COM72-50526**

Sparks, L. L., and Hust, J. G., Standard Reference Materials: Thermal Conductivity of Austenitic Stainless Steel, SRM 735 from 5 to 280 K, NBS Spec. Publ. 260-35 (April 1972.) COM72-50368**

Cali, J. P., Mandel, J., Moore, L. J., and Young, D. S., Standard Reference Materials: A Referee Method for the Determination of Calcium in Serum NBS SRM 915, NBS Spec. Publ. 260-36 (May 1972). COM72-50527**

Shultz, J. I., Bell, R. K., Rains, T. C., and Menis, O., Standard Reference Materials: Methods of Analysis of NBS Clay Standards, NBS Spec. Publ. 260-37 (June 1972). COM72-50692**

Clark, A. F., Denson, V. A., Hust, J. G., and Powell, R. L., Standard Reference Materials: The Eddy Current Decay Method for Resistivity Characterization of High-Purity Metals, NBS Spec. Publ. 260-39 (May 1972). COM72-50529**

McAdie, H. G., Garn, P. D., and Menis, O., Standard Reference Materials: Selection of Thermal Analysis Temperature Standards Through a Cooperative Study (SRM 758, 759, 760), NBS Spec. Publ. 260-40 (August 1972) COM72-50776**

Wagner, H. L., and Verdier, P. H., eds., Standard Reference Materials: The Characterization of Linear Polyethylene, SRM 1475, NBS Spec. Publ. 260-42 (September 1972). COM72-50944**

Yakowitz, H., Ruff, A. W., and Michaelis, R. E., Standard Reference Materials: Preparation and Homogeneity Characterization of an Austenitic Iron-Chromium-Nickel Alloy, NBS Spec. Publ. 260-43 (November 1972). COM73-50760**

Schooley, J. F., Soulen, R. J., Jr., and Evans, G. A., Jr., Standard Reference Materials: Preparation and Use of Superconductive Fixed Point Devices, SRM 767, NBS Spec. Publ. 260-44 (December 1972). COM73-50037**

Greifer, B., Maienthal, E. J., Rains, T. C., and Rasberry, S. D., Standard Reference Materials: Powdered Lead-Based Paint, SRM 1579, NBS Spec. Publ. 260-45 (March 1973). COM73-50226**

Hust, J. G., and Giarratano, P. J., Standard Reference Materials: Thermal Conductivity and Electrical Resistivity Standard Reference Materials: Austenitic Stainless Steel, SRM's 735 and 798, from 4 to 1200 K, NBS Spec. Publ. 260-46 (March 1975). COM75-10339**

Hust, J. G., Standard Reference Materials: Electrical Resistivity of Electrolytic Iron, SRM 797, and Austenitic Stainless Steel, SRM 798, from 5 to 280 K, NBS Spec. Publ. 260-47 (February 1974). COM74-50176**

Mangum, B. W., and Wise, J. A., Standard Reference Materials: Description and Use of Precision Thermometers for the Clinical Laboratory, SRM 933 and SRM 934, NBS Spec. Publ. 260-48 (May 1974). Superseded by NIST Spec. Publ. 260-113.

Carpenter, B. S., and Reimer, G. M., Standard Reference Materials: Calibrated Glass Standards for Fission Track Use, NBS Spec. Publ. 260-49 (November 1974). COM74-51185**

Hust, J. G., and Giarratano, P. J., Standard Reference Materials: Thermal Conductivity and Electrical Resistivity Standard Reference Materials: Electrolytic Iron, SRM's 734 and 797 from 4 to 1000 K, NBS Spec. Publ. 260-50 (June 1975). COM75-10698**

Mavrodineanu, R., and Baldwin, J. R., Standard Reference Materials: Glass Filters As a Standard Reference Material for Spectrophotometry-Selection, Preparation, Certification, Use-SRM 930 NBS Spec. Publ. 260-51 (November 1975). COM75-10339**

Hust, J. G., and Giarratano, P. J., Standard Reference Materials: Thermal Conductivity and Electrical Resistivity Standard Reference Materials 730 and 799, from 4 to 3000 K, NBS Spec. Publ. 260-52 (September 1975). COM75-11193**

Durst, R. A., Standard Reference Materials: Standardization of pH Measurements, NBS Spec. Publ. 260-53 (February 1988, Revision of December 1975 version).

Burke, R. W., and Mavrodineanu, R., Standard Reference Materials: Certification and Use of Acidic Potassium Dichromate Solutions as an Ultraviolet Absorbance Standard, NBS Spec. Publ. 260-54 (August 1977). PB272168**

Ditmars, D. A., Cezairliyan, A., Ishihara, S., and Douglas, T. B., Standard Reference Materials: Enthalpy and Heat Capacity; Molybdenum SRM 781, from 273 to 2800 K, NBS Spec. Publ. 260-55 (September 1977). PB272127**

Powell, R. L., Sparks, L. L., and Hust, J. G., Standard Reference Materials: Standard Thermocouple Material, Pt-67: SRM 1967, NBS Spec. Publ. 260-56 (February 1978). PB277172**

Barnes, J. D., and Martin, G. M., Standard Reference Materials: Polyester Film for Oxygen Gas Transmission Measurements SRM 1470, NBS Spec. Publ. 260-58 (June 1979). PB297098**

Velapoldi, R. A., Paule, R. C., Schaffer, R., Mandel, J., and Moody, J. R., Standard Reference Materials: A Reference Method for the Determination of Sodium in Serum, NBS Spec. Publ. 260-60 (August 1978). PB286944**

Verdier, P H., and Wagner, H. L., Standard Reference Materials: The Characterization of Linear Polyethylene (SRM 1482, 1483, 1484), NBS Spec. Publ. 260-61 (December 1978). PB289899**

Soulen, R. J., and Dove, R. B., Standard Reference Materials: Temperature Reference Standard for Use Below 0.5 K (SRM 768), NBS Spec. Publ. 260-62 (April 1979). PB294245**

Velapoldi, R. A., Paule, R. C., Schaffer, R., Mandel, J., Machlan, L. A., and Gramlich, J. W., Standard Reference Materials: A Reference Method for the Determination of Potassium in Serum, NBS Spec. Publ. 260-63 (May 1979). PB297207**

Velapoldi, R. A., and Mielenz, K. D., Standard Reference Materials: A Fluorescence Standard Reference Material Quinine Sulfate Dihydrate (SRM 936), NBS Spec. Publ. 260-64 (January 1980). PB80-132046**

Marinenko, R. B., Heinrich, K. F. J., and Ruegg, F. C., Standard Reference Materials: Micro-Homogeneity Studies of NBS Standard Reference Materials, NBS Research Materials, and Other Related Samples, NBS Spec. Publ. 260-65 (September 1979). PB300461**

Venable, W. H., Jr., and Eckerle, K. L., Standard Reference Materials: Didymium Glass Filters for Calibrating the Wavelength Scale of Spectrophotometers-SRM 2009, 2010, 2013, and 2014, NBS Spec. Publ. 260-66 (October 1979). PB80-104961**

Velapoldi, R. A., Paule, R. C., Schaffer, R., Mandel, J., Murphy, T. J., and Gramlich, J. W., Standard Reference Materials: A Reference Method for the Determination of Chloride in Serum, NBS Spec. Publ. 260-67 (November 1979). PB80-110117**

Mavrodineanu, R., and Baldwin, J. R., Standard Reference Materials: Metal-On-Quartz Filters as a Standard Reference Material for Spectrophotometry SRM 2031, NBS Spec. Publ. 260-68 (April 1980). PB80-197486**

Velapoldi, R. A., Paule, R. C., Schaffer, R., Mandel, J., Machlan, L. A., Garner, E. L., and Rains, T. C., Standard Reference Materials: A Reference Method for the Determination of Lithium in Serum, NBS Spec. Publ. 260-69 (July 1980). PB80-209117**

Marinenko, R. B., Biancaniello, F., Boyer, P. A., Ruff, A. W., and DeRobertis, L., Standard Reference Materials: Preparation and Characterization of an Iron-Chromium-Nickel Alloy for Microanalysis, NBS Spec. Publ. 260-70 (May 1981). PB84-165349**

Seward, R. W., and Mavrodineanu, R., Standard Reference Materials: Summary of the Clinical Laboratory Standards Issued by the National Bureau of Standards, NBS Spec. Publ. 260-71 (November 1981). PB82-135161**

Reeder, D. J., Coxon, B., Enagonio, D., Christensen, R. G., Schaffer, R., Howell, B. F., Paule, R. C., and Mandel, J., Standard Reference Materials: SRM 900, Antiepilepsy Drug Level Assay Standard, NBS Spec. Publ. 260-72 (June 1981). PB81-220758

Interrante, C. G., and Hicho, G. E., Standard Reference Materials: A Standard Reference Material Containing Nominally Fifteen Percent Austenite (SRM 486), NBS Spec. Publ. 260-73 (January 1982). PB82-215559**

Marinenko, R. B., Standard Reference Materials: Preparation and Characterization of K-411 and K-414 Mineral Glasses for Microanalysis: SRM 470, NBS Spec. Publ. 260-74 (April 1982). PB82-221300**

Weidner, V. R., and Hsia, J. J., Standard Reference Materials: Preparation and Calibration of First Surface Aluminum Mirror Specular Reflectance Standards (SRM 2003a), NBS Spec. Publ. 260-75 (May 1982). PB82-221367**

Hicho, G. E., and Eaton, E. E., Standard Reference Materials: A Standard Reference Material Containing Nominally Five Percent Austenite (SRM 485a), NBS Spec. Publ. 260-76 (August 1982). PB83-115568**

Furukawa, G. T., Riddle, J. L., Bigge, W. G., and Pfieffer, E. R., Standard Reference Materials: Application of Some Metal SRM's as Thermometric Fixed Points, NBS Spec. Publ. 260-77 (August 1982). PB83-117325**

Hicho, G. E., and Eaton, E. E., Standard Reference Materials: Standard Reference Material Containing Nominally Thirty Percent Austenite (SRM 487), NBS Spec. Publ. 260-78 (September 1982). PB83-115576**

Richmond, J. C., Hsia, J. J., Weidner, V. R., and Wilmering, D. B., Standard Reference Materials: Second Surface Mirror Standards of Specular Spectral Reflectance (SRM's 2023, 2024, 2025), NBS Spec. Publ. 260-79 (October 1982). PB84-203447**

Schaffer, R., Mandel, J., Sun, T., Cohen, A., and Hertz, H. S., Standard Reference Materials: Evaluation by an ID/MS Method of the AACC Reference Method for Serum Glucose, NBS Spec. Publ. 260-80 (October 1982). PB84-216894**

Burke, R. W., and Mavrodineanu, R., Standard Reference Materials: Accuracy in Analytical Spectrophotometry, NBS Spec. Publ. 260-81 (April 1983). PB83-214536**

Weidner, V. R., Standard Reference Materials: White Opal Glass Diffuse Spectral Reflectance Standards for the Visible Spectrum (SRM's 2015 and 2016), NBS Spec. Publ. 260-82 (April 1983). PB83-220723**

Bowers, G. N., Jr., Alvarez, R., Cali, J. P., Eberhardt, K. R., Reeder, D. J., Schaffer, R., and Uriano, G. A., Standard Reference Materials: The Measurement of the Catalytic (Activity) Concentration of Seven Enzymes in NBS Human Serum SRM 909, NBS Spec. Publ. 260-83 (June 1983). PB83-239509**

Gills, T. E., Seward, R. W., Collins, R. J., and Webster, W. C., Standard Reference Materials: Sampling, Materials Handling, Processing, and Packaging of NBS Sulfur in Coal Standard Reference Materials 2682, 2683, 2684, and 2685, NBS Spec. Publ. 260-84 (August 1983). PB84-109552**

Swyt, D. A., Standard Reference Materials: A Look at Techniques for the Dimensional Calibration of Standard Microscopic Particles, NBS Spec. Publ. 260-85 (September 1983). PB84-112648**

Hicho, G. E., and Eaton, E. E., Standard Reference Materials: A Standard Reference Material Containing Two and One-Half Percent Austenite, SRM 488, NBS Spec. Publ. 260-86 (December 1983). PB84-143296**

Mangum, B. W., Standard Reference Materials: SRM 1969: Rubidium Triple-Point - A Temperature Reference Standard Near 39.30 C, NBS Spec. Publ. 260-87 (December 1983). PB84-149996**

Gladney, E. S., Burns, C. E., Perrin, D. R., Roelandts, I., and Gills, T. E., Standard Reference Materials: 1982 Compilation of Elemental Concentration Data for NBS Biological, Geological, and Environmental Standard Reference Materials, NBS Spec. Publ. 260-88 (March 1984). PB84-218338**

Hust, J. G., Standard Reference Materials: A Fine-Grained, Isotropic Graphite for Use as NBS Thermophysical Property RM's from 5 to 2500 K, NBS Spec. Publ. 260-89 (September 1984). PB85-112886**

Hust, J. G., and Lankford, A. B., Standard Reference Materials: Update of Thermal Conductivity and Electrical Resistivity of Eleetrolytic Iron, Tungsten, and Stainless Steel, NBS Spec. Publ. 260-90 (September 1984). PB85-115814**

Goodrich, L. F., Vecchia, D. F., Pittman, E. S., Ekin, J. W., and Clark, A. F., Standard Reference Materials: Critical Current Measurements on an NbTi Superconducting Wire Standard Reference Material, NBS Spec. Publ. 260-91 (September 1984). PB85-118594**

Carpenter, B. S., Standard Reference Materials: Calibrated Glass Standards for Fission Track Use (Supplement to NBS Spec. Publ. 260-49), NBS Spec. Publ. 260-92 (September 1984). PB85-113025**

Ehrstein, J.R., Standard Reference Materials: Preparation and Certification of Standard Reference Materials for Calibration of Spreading Resistance Probes, NBS Spec. Publ. 260-93 (January 1985). PB85-177921**

Gills, T. E., Koch, W. F., Stolz, J. W., Kelly, W. R., Paulsen, P. J., Colbert, J. C., Kirklin, D. R., Pei, P.T.S., Weeks, S., Lindstrom, R. M., Fleming, R. F., Greenberg, R. R., and Paule, R. C., Standard Reference Materials: Methods and Procedures Used at the National Bureau of Standards to Certify Sulfur in Coal SRM's for Sulfur Content, Calorific Value, Ash Content, NBS Spec. Publ. 260-94 (December 1984). PB85-165900**

Mulholland, G. W., Hartman, A. W., Hembree, G. G., Marx, E., and Lettieri, T. R., Standard Reference Materials: Development of a 1 μm Diameter Particle Size Standard, SRM 1690, NBS Spec. Publ. 260-95 (May 1985). SN003-003-02665-4*

Carpenter, B. S., Gramlich, J. W., Greenberg, R. R., Machlan, L. A., DeBievre, P., Eschbach, H. L., Meyer, H., Van Audenhove, J., Connolly, V. E., Trahey, N. M., and Zook, A. C., Standard Reference Materials: Uranium-235 Isotopic Abundance Standard Reference Materials for Gamma Spectrometry Measurements, NBS Spec. Publ. 260-96 (September 1986). PB87-108544**

Mavrodineanu, R., and Gills, T. E., Standard Reference Materials: Summary of the Coal, Ore, Mineral, Rock, and Refactory Standards Issued by the National Bureau of Standards, NBS Spec. Publ. 260-97 (September 1985). SN003-003-02688-3*

Hust, J. G., Standard Reference Materials: Glass Fiberboard SRM for Thermal Resistance, NBS Spec. Publ. 260-98 (August 1985). SN003-003-02674-3*

Callanan, J. E., Sullivan, S. A., and Vecchia, D. F., Standard Reference Materials: Feasibility Study for the Development of Standards Using Differential Scanning Calorimetry, NBS Spec. Publ. 260-99 (August 1985). SN003-003-02675-1*

Taylor, J. K., Standard Reference Materials: Handbook for SRM Users, NBS Spec. Publ. 260-100 (September 1985). PB86-110897**

Mangum, B. W., Standard Reference Materials: SRM 1970, Succinonitrile Triple-Point Standard: A Temperature Reference Standard Near 58.08 C, NBS Spec. Publ. 260-101 (March 1986). SN003-003-02722-7*

Weidner, V. R., Mavrodineanu, R., Mielenz, K. D., Velapoldi, R. A., Eckerle, K. L., and Adams, B., Standard Reference Materials: Holmium Oxide Solution Wavelength Standard from 240 to 640 nm - SRM 2034, NBS Spec. Publ. 260-102 (July 1986). PB86-245727**

Hust, J. G., Standard Reference Materials: Glass Fiberblanket SRM for Thermal Resistance, NBS Spec. Publ. 260-103 (September 1985). SN003-003-02687-5*

Mavrodineanu, R., and Alvarez, R., Standard Reference Materials: Summary of the Biological and Botanical Standards Issued by the National Bureau of Standards, NBS Spec. Publ. 260-104 (November 1985). SN003-003-02704-9*

Mavrodineanu, R., and Rasberry, S. D., Standard Reference Materials: Summary of the Environmental Research, Analysis, and Control Standards Issued by the National Bureau of Standards, NBS Spec. Publ. 260-105 (March 1986). SN003-003-02725-1*

Koch, W. F., ed., Standard Reference Materials: Methods and Procedures Used at the National Bureau of Standards to Prepare, Analyze, and Certify SRM 2694, Simulated Rainwater, and Recommendations for Use, NBS Spec. Publ. 260-106 (July 1986). PB86-247483**

Hartman, A. W., and McKenzie, R. L., Standard Reference Materials: SRM 1965, Microsphere Slide (10 μm Polystyrene Spheres), NBS Spec. Publ. 260-107 (November 1988).

Mavrodineanu, R., and Gills, T. E., Standard Reference Materials: Summary of Gas Cylinder and Permeation Tube Standard Reference Materials Issued by the National Bureau of Standards, NBS Spec. Publ. 260-108 (May 1987).

Candela, G. A., Chandler-Horowitz, D., Novotny, D. B., Marchiando, J. F., and Belzer, B. J., Standard Reference Materials: Preparation and Certification of an Ellipsometrically Derived Thickness and Refractive Index Standard of a Silicon Dioxide Film (SRM 2530), NIST Spec. Publ. 260-109 (October 1988).

Kirby, R. K., and Kanare, H. M., Standard Reference Materials: Portland Cement Chemical Composition Standards (Blending, Packaging, and Testing), NBS Spec. Publ. 260-110 (February 1988).

Gladney, E. S., O'Malley, B. T., Roelandts, I., and Gills, T. E., Standard Reference Materials: Compilation of Elemental Concentration Data for NBS Clinical, Biological, Geological, and Environmental Standard Reference Materials, NBS Spec. Publ. 260-111 (November 1987).

Marinenko, R. B., Blackburn, D. H., and Bodkin, J. B., Standard Reference Materials: Glasses for Microanalysis: SRM's 1871-1875, NIST Spec. Publ. 260-112 (February 1990).

Mangum, B. W., and Wise, J. A., Standard Reference Materials: Description and Use of a Precision Thermometer for the Clinical Laboratory, SRM934, NIST Spec. Publ. 260-113 (June 1990).

Vezzetti, C. F., Varner, R. N. and Potzick, J. E., Standard Reference Materials: Bright-Chromium Linewidth Standard, SRM 476, for Calibration of Optical Microscope Linewidth Measuring Systems, NIST Spec. Publ. 260-114 (January 1991).

Williamson, M. P., Willman, N. E., and Grubb, D. S., Standard Reference Materials: Calibration of NIST Standard Reference Material 3201 for 0.5 Inch (12.65 mm) Serial Serpentine Magnetic Tape Cartridge, NIST Spec. Publ. 260-115 (February 1991).

Mavrodineanu, R. , Burke, R. W., Baldwin, J. R., Smith, M. V., and Messman, J. D., Standard Reference Materials: Glass Filters as a Standard Reference Material for Spectrophotometry and Selection, Preparation, Certification and use of SRM 930 and SRM 1930, NIST Spec. Publ. 260-116 (in prep).

Vezzetti, C. F., Varner, R. N., and Potzick, J. E., Standard Reference Materials: Antireflecting-Chromium Linewidth Standard, SRM 475, for Calibration of Optical Microscope Linewidth Measuring Systems, NIST Spec. Publ. 260-117 (January 1992).

Williamson, M. P., Standard Reference Materials: Calibration of NIST Standard Reference Material 3202 for 18-Track, Parallel, and 36-Trace, Parallel Serpentine, 12.65 mm (0.5 in), 1491 cpmm (37871 cpi), Magnetic Tape Cartridge, NIST Spec. Publ. 260-118 (July 1992)

*Send order with remittance to Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20102. Remittance from foreign countries should include an additional one fourth of the purchase price for postage.

**May be ordered from: National Technical Information Services (NTIS), Springfield, VA 22161.

ABSTRACT

This handbook was prepared to provide guidance for the use of Standard Reference Materials (SRMs) to provide an accuracy base for chemical measurements. The general concepts of precision and accuracy, and their realization by quality assurance of the measurement process, are discussed. General characteristics of SRMs are described and guidance given on their selection for specific applications. Ways to effectively use SRMs are recommended, including the utilization of control charts to evaluate and monitor measurement accuracy. Appendices provide statistical guidance on the evaluation of measurement uncertainty.

Key words: accuracy; calibration; chemical analysis; control charts; measurement uncertainty; precision; quality assurance; reference materials; standard reference materials; statistical control.

TABLE OF CONTENTS

Page

PREFACE v
ABSTRACT xii
List of Tables xv
List of Figures xvi

1 Introduction 1

2 Precision and Accuracy 2
2.1 Concepts of Precision and Accuracy 2
2.1.1 Precision and Bias in a Measurement System 4
2.2 Dependence on Standards 5
2.3 Physical and Chemical Standards 6
2.4 Calibration, Standardization, and Analytical Response Function 6
2.5 National Measurement System for Analytical Chemistry 7

3 Quality Assurance 9
3.1 Quality Assurance of a Measurement Process 9
3.2 Statistical Control 10
3.3 Control Charts 10
3.4 Frequency of Use of Reference Materials 14

4 Reference Materials 16
4.1 Role of Reference Materials 16
4.2 Concept of Traceability 17

5 Standard Reference Materials 19
5.1 Philosophy of SRM Production 19
5.2 How a SRM is Produced 19
5.3 Differences Among Measurement Methods 22
5.4 Understanding Certificate Information 22
5.5 Uncertainty of Certified Values 23
5.6 Relation to Real-World Samples 26

6 Use of SRMs 26
6.1 Kinds of SRMs 26
6.2 Selecting a SRM 29
6.3 Use of SRMs 30
6.4 Interpretation of Reference Material Analyses 33
6.5 Evaluation of Measurement Error 36
6.6 Is a SRM Expensive? 38

7 Reporting Analytical Data 39
7.1 Limits of Uncertainty for Data 39
7.2 Significant Figures 39
7.3 Guidelines for Reporting Results of Measurements 40

8 NIST Services Related to SRMs 41

9 Bibliography 43
9.1 References 43
9.2 Publications Related to SRMs 44

Appendix A. Glossary of Terms 49

Appendix B. The SI System - Units and Conversion Factors 56
B.1 Units 56
B.2 Conversion Factors 57
B.3 Tables 58
B.3.1 Use of Range To Estimate Variability 58
B.3.2 Critical Values for the F Test 59
B.3.3 Factors for Computing Two-Sided Confidence Limits for σ 60
B.3.4 Percentiles of the t Distribution 61
B.3.5 Factors for Two-Sided Tolerance Limits for Normal Distribution 62
B.3.6 Short Table of Random Numbers 63
B.3.7 Z-Factors for Two-Sided Confidence Interval 64

Appendix C. Statistical Tools 65
C.1 Introduction 65
C.2 Estimating the Standard Deviation 67
C.2.1 Estimating the Standard Deviation from Replicate Measurements 68
C.2.2 Estimating the Measurement Standard Deviation from Duplicate Measurements 69
C.2.3 Pooling Estimates of Standard Deviation 70
C.3 Do Two Estimates of Precision Differ? 71
C.4 Confidence Limits for an Estimate of Standard Deviation 72
C.5 Confidence Interval for a Mean 72
C.6 Do Two Means Disagree Significantly? 73
C.6.1 Does a Measured Value Differ from a Certified (Standard) Value? 76
C.7 Statistical Tolerance Intervals 76
C.8 Pooling Means to Obtain a Grand Average, $\bar{\bar{x}}$ 77
C.9 Outliers 79
C.10 Use of Random Number Tables 81

Appendix D. Published Articles 82
D.1 Quality Assurance of Chemical Measurements 82
D.2 Sampling for Chemical Analysis 87
D.3 Reference Materials - What They Are and How They Should be Used 95
D.4 Validation of Analytical Methods 98

Guide for Requesting Development of Standard Reference Materials

LIST OF TABLES

		Page
Table 1.	Quality Assessment Using RMs	15
Table 2.	Quality Assessment Using Duplicates/Splits	16
Table 3.	SRM Category/Subcategory Listing	27
Table 4.	Some Recent New and Renewal SRMs/RMs	29
Table 5.	Uses of SRMs in Measurement Systems	30
Table B.3.1	Use of Range to Estimate Variability	58
Table B.3.2	Critical Values for the F Test	59
Table B.3.3	Factors for Computing Two-Sided Confidence Limits of σ	60
Table B.3.4	Percentiles of the t Distribution	61
Table B.3.5	Factors for Two-Sided Tolerance Limits for Normal Distribution	62
Table B.3.6	Short Table of Random Numbers	63
Table B.3.7	Z-Factors for Two-Sided Confidence Interval	64

LIST OF FIGURES

		Page
Figure 1.	Unbiased Measurement Processes	2
Figure 2.	Biased Measurement Processes	3
Figure 3.	National Measurement System for Analytical Chemistry	8
Figure 4.	Measurement Process Quality Assurance	10
Figure 5A.	Duplicate Measurements made on a SRM $\bar{\bar{x}}$ Chart	12
Figure 5B.	Duplicate Measurements made on a SRM; R Chart	12
Figure 6.	Measurement Compatibility by Intercalibration Using SRMs	18
Figure 7.	Systems Approach to Measurement Accuracy	32
Figure 8.	Interpreting SRM Measurements	34
Figure 9.	Identifying Measurement Bias	35
Figure 10.	Uncertainty of Measured Value	37
Figure C.1	Typical Distribution of the Means of Random Samples/Measurements	65
Figure C.2	Confidence Interval for σ	66
Figure C.3	Computed 50% Confidence Intervals for the Population Mean, m, from 100 Samples of 4, 40 Samples of 100, and 4 Samples of 1000 [19]	67

1. INTRODUCTION

Standard Reference Materials (SRMs) issued by the National Institute of Standards and Technology (NIST), formerly the National Bureau of Standards (NBS), have become well established as benchmarks for the quality assurance of measurements. Some analytical chemists find them indispensable for this purpose and use them systematically; others use them sporadically. Yet, there are others, and those especially in some areas of analytical chemistry, who rarely, if ever, use reference materials. From casual observation, the frequency of use of reference materials is closely coupled with one or more of the following factors:

- familiarity with the philosophy of SRM use
- degree of appreciation of the benefits of SRM use
- availability of directly applicable SRMs
- understanding of the role of directly and indirectly related SRMs in a measurement system
- degree of full appreciation of measurement as a system.

NIST has pioneered, and continues to be the leader in, the development of SRMs for quality assurance of measurements. The program to provide reference materials, originally known as standard samples, was initiated at NBS in 1906, largely in response to needs of the metals industry. It has since expanded to a multimaterial program of over 1300 items (up from 1000 items in 1985 when this publication was first issued) that serves most areas of modern analytical chemistry and metrology. A large number of materials useful in physical metrology and engineering are now also included. It should be noted that some areas of measurement are covered by SRMs more completely than others due to historical reasons, national issue priorities, and to some extent, the degree of industrial awareness of the quality assurance concept.

For the early users, SRMs were identical, in most respects, to the materials ordinarily analyzed. Thus the results of measurements of SRMs were easily interpreted. But as the NIST SRM program has expanded, it has become impossible to provide SRMs with a one-to-one correspondence to every conceivable application; as a result, generic standards which serve many purposes are commonly produced. A multipurpose SRM broadens the scope of application, conserves the technical effort and reduces the cost of SRM production.

This handbook was prepared with the objective of improving the understanding of the basis for the use of SRMs. While written from the viewpoint of a chemist, the basic concepts described are believed to be applicable to most areas of metrology. The handbook is arranged by section in a logical progression, starting with the basic concepts of precision and accuracy, followed by discussions of the calibration and quality assurance of the measurement process, the use of SRMs to evaluate various kinds of measurements, and the reporting of data with evaluated limits of uncertainty. The statistical considerations most frequently applicable for the evaluation and interpretation of measurement data are reviewed in the Appendices. Each section is written with some degree of independence so that it can be comprehended without frequent reference to the contents of other sections.

The treatment of each subject in a section is not claimed to be exhaustive, but is often an overview. Accordingly, a bibliography section containing 1) selected references which contain both background and other information supplemental to that in the text, 2) selected journal publications related to the preparation, analysis, and certification of specific SRMs and, 3) the NIST 260 Series Special Publications on SRMs are provided for additional information.

2. PRECISION AND ACCURACY

2.1 Concepts of Precision and Accuracy

Accuracy is an intuitively understandable and desirable requirement for most measurements. Data which are knowingly inaccurate or whose accuracy is unknown have little appeal to most users. Yet precision is sometimes confused with accuracy and the agreement of successive results can inspire a degree of confidence that the measurements may not merit.

Accuracy, the closeness of a measured value to the true value, includes the concepts of bias and precision (see Figures 1 and 2) and is judged with respect to the use to be made of the data. A measurement process must be unbiased to be capable of producing accurate values. In such a case, it must be sufficiently precise as well, or else the individual results will be inaccurate due to unacceptable variability. The following discussion is presented to clarify these concepts. The term "uncertainty" is used widely in describing the results of measurement and denotes an estimate of the bounds of inaccuracy. Strictly speaking, the actual error of a reported value is usually not known. However, limits of error ordinarily can be inferred - with some risk of being incorrect - from the precision so that reasonable limits for the possible bias of the measurement process can be developed.

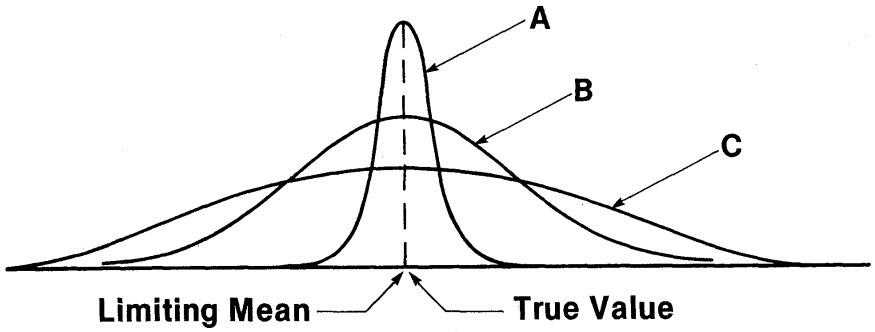


Figure 1. Unbiased Measurement Processes

The distributions of results from three unbiased processes are shown. The precision decreases in the order, A > B > C. Note that the limiting means of all approach the "true value," but process C is relatively inaccurate (compared with A) due to its imprecision.

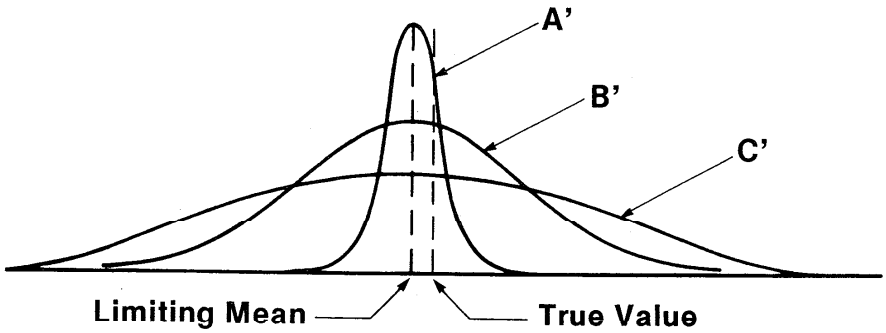


Figure 2. Biased Measurement Processes

As shown, all of the processes are biased and hence inaccurate since the limiting means do not coincide with the "true value." However, note that most of the results for process A will be more accurate than those of process C and even B (Figure 1) due to precision considerations.

The concept of precision is concerned with the variability of the individual results of replicate measurements. A process which shows a small scatter is said to be precise and vice versa. Obviously, such judgments are subjective and based on the intended use of the data. What might be considered as very precise for one purpose could be grossly imprecise for another. Random errors are responsible for the observed scatter of measured values. These may be reduced to the point at which they are negligible with respect to the tolerable error of the measured value, or are limited by inherent characteristics of the instrumentation of the methodology used. The averages of several series of measurements will show a smaller variability than the individual values and the grand average of the series is expected to approach a limiting value (limiting mean) as the number of measurements is increased.

The concept of bias is concerned with whether or not the limiting mean differs from the true (or accepted) value of the property measured. Here again, judgment is usually involved since it is impossible to eliminate all errors or even to know if this has been achieved. Such decisions are thus based on whether or not, for all practical purposes, bias exists.

In the case of individual measurements, each will exhibit some degree of inaccuracy; that is to say, it will deviate from the true value. This will occur because of random errors together with any bias of the measurement system. Indeed, it is highly improbable that any individual measurement made by an unbiased measurement system will be accurate, since the probability of zero random error is zero. Many individual values may appear to be the correct value, but this is due to truncation resulting from insensitivity of the measurement process or from rounding of the data.

A measurement process should be sufficiently precise to minimize the number of replicate measurements required for the intended use. A very precise system may need only a few measurements, even one, to provide data that would not be significantly improved by further replication. Also, a measurement system must be sufficiently

precise to identify whether or not biases of a comparable magnitude are present in the system. While possible in principle, an unbiased measurement process of low precision may be incapable of providing accurate data - from a practical point of view - because of the large number of measurements required to reduce the uncertainty of the random errors to reasonable limits.

Precision may be evaluated by the redundant process of replicate measurement. Results on a single material may be used for this purpose, or the information obtained on a number of samples (even duplicate measurements, see Appendix C.2.2), may be pooled. Accordingly, there is no reason why a laboratory cannot evaluate its own precision without external assistance [1]. While SRMs may be helpful in this regard, they are not necessary for this purpose.

In order to properly estimate precision, a number of measurements over an extended period of time may be required. Common experience shows that it is much easier to repeat a measurement than to reproduce it over a period of time, so a small number of measurements tend to underestimate the standard deviation. The repeatability, or short-term standard deviation, is needed to answer questions about the number of repetitive measurements that may be required while the long-term standard deviation, or reproducibility, is needed to answer such questions as the agreement of data obtained at different times, i.e., the statistical control of a measurement process.

Though precise measurements can serve useful purposes when limited comparisons are required, accuracy is more often the essential requirement. Whenever the true value of the measured quantity is needed or when data from different laboratories, different methodologies, or from the same laboratory using the same method over a period of time need to be interrelated, bias can be a serious problem. Working methods may be compared with reference methods to assess accuracy, but this is ordinarily a more difficult and time-consuming process (see Appendix D.4). The analysis of appropriate reference materials is the best and easiest way to ascertain bias.

2.1.1 Precision and Bias in a Measurement System

The precision of a measurement system may be influenced by a number of factors, each having its own precision. The precision of each factor, quantified in terms of the variance, contributes to the precision of the process. The variance is simply the square of the standard deviation, s. In a measurement process, the variances of the individual process steps, s^2_i, add up to define the variance of the process, i.e., $s^2 = s^2_1 + s^2_2 + s^2_3 + \cdots + s^2_n$. Some of the steps (or factors) can be easily identified and the individual variances estimated. Examples are weighing and extraction. As steps are identified, improvements conceivably can be made when there are 'assignable causes' for undesirable imprecision. Because of addition in quadrature, it is evident that one or a few sources of variance can be the major contributors to the total variance. Knowledge of the magnitude of the individual variances can indicate both directions for improvement and possible sources of trouble when 'out-of-control' measurements occur.

It is conceivable that variance can be reduced to very low levels with diligent effort. Laboratories commonly improve their precision as they gain experience with their methodology. Ordinarily, a laboratory will improve its quality control practices to the point when the precision attained is adequate for a particular application or when peer performance has been attained. Because analysis must be pragmatic, cost-benefit decisions will often dictate how far to go. For example, it is a matter of record that laboratories using the same methodology may differ in their precisions. This may be due to difference in skill levels, but also to different levels of tolerance for permissible errors.

Bias in measurement systems can result from several sources. The commonly recognized ones are 1) control of measurement variables, 2) interferences, 3) erroneous calibrations, 4) contamination, 5) losses, 6) deteriorations and, 7) inefficiencies in extractions or sample dissolution. Variability in some of these can contribute to random error as well, and often to a major degree. Inappropriate calibration techniques can also be a serious source of bias. Reliance on spiking which may not simulate a matrix-incorporated analyte, or the use of a pure matrix (e.g., pure

water) to simulate a natural matrix (waste water) are typical examples. A reference material that closely simulates the analytical samples is needed to identify and evaluate such biases.

Unlike random errors, systematic errors or biases from several sources are not necessarily randomly distributed; hence one must recognize that biases can add up algebraically, that is to say, the total bias, B, equals $B_1 + B_2 + \cdots + B_n$. Thus, a large number of small biases can equal or even exceed a large bias from a single source. While the effect of random error decreases as the number of measurements, n, is increased $s_{\bar{x}} = s_x/\sqrt{n}$, the effect of bias is independent of the number of measurements.

2.2 Dependence on Standards

All measurements depend on standards. Physical measurements depend almost entirely on physical standards with little or no dependence on chemical standards. Chemical measurements on the other hand depend on both - with greater dependence on the latter. The early recognition of the need for universally acceptable physical standards, and the chaos that could result from their unavailability, led to the development of the new internationally accepted physical standards for the primary units of length, mass, time, temperature, and radiant luminosity plus the units derived from them such as pressure, force, acceleration, power, and density. No corresponding chemical standards have ever been developed. There are, of course, the useful standards of atomic weights and a variety of physicochemical standards.

The early chemical analytical measurements were largely absolute in nature, which means they depended almost entirely on physical standards. Thus, classical analysis used gravimetry in which chemical constituents were removed quantitatively or isolated from their matrix, purified, and weighed. Relation of such masses to the chemical information desired was then calculated by stoichiometry. The critical sources of errors in such measurements were incomplete separations, mechanical losses, and contamination due to coprecipitation and analytical blanks. Physical standards were the primary standards and provided adequate and sufficient means to control the accuracy of such chemical measurements.

While classical methods, augmented by such physicochemical techniques as coulometry, still provide the basis for the most accurate measurement of major constituents or for the assay of pure materials, the bulk of modern chemical measurements are made by comparative techniques by which an instrument is used to compare an unknown sample with one of known composition. Some measurements require the removal of the substance of interest prior to analysis, or its isolation from the matrix using physical or chemical techniques. In others, the analytical process may combine the separation and measurement steps. Separation of a group of analytes, followed by selective detection, is yet another approach to analysis.

The trend toward comparative measurements has thus shifted the need in chemical analysis from heavy dependence on physical standards to heavy dependence on chemical standards. However, there is usually no problem in obtaining chemicals of requisite purity to serve as standards, so few national or international standard chemicals have been developed or exist today in the same sense as the physical standards. Where chemical reference materials do exist, they are ordinarily not chemical standards in the hierarchal sense of physical standards of measurement, but rather are quality assurance materials, as will be discussed later. Of course, some reference materials are high purity chemicals which may be used as primary standards in some areas of chemical analysis.

2.3 Physical and Chemical Standards

Seven basic units for physical measurements have been adopted by international agreement. From these, all other units of measurement may be derived [2]. The basic units are defined by appropriate artifacts or measurements. Transfer standards may be calibrated with respect to the basic standards maintained in national metrology laboratories (such as NIST). Such calibrations must be done with a sufficient degree of reliability, traceable to national standards. For most chemical measurements, uncertainties in the physical standards used do not contribute significantly to the analytical uncertainty.

Chemical standards differ from physical standards in several ways. They are chemical elements or compounds usually identical with, or related by, stoichiometry to the analytes measured. It is ordinarily possible to obtain such standards in sufficient purity or to purify them adequately so they need not be maintained in a national laboratory (such as NIST). Due to the complexity and variety of chemical measurements, it would be infeasible, if not impossible to do so.

The problem in the use of chemical standards is the degree to which they can be blended or incorporated into a sample matrix to produce a substance that can reliably be used to calibrate or define the response function of a chemical analyzer. Matrix match between standard and analytical sample is often critical, but difficult to achieve. When standards are carried through an entire analytical process, spikes, surrogates, and other artificially introduced constituents may not respond in the same manner as naturally occurring analytes, thus causing calibration problems. On the other hand, standards prepared to simulate the final analytical sample (e.g., an extract or a solution of the original sample) may not be able to calibrate the entire analytical process.

No matter what kind of standards are used, they must be prepared with care from reliable starting materials. The mode of preparation should be such that the uncertainty of the standards does not contribute significantly to the overall analytical uncertainty. Chemists usually assume that standards can be prepared with negligible error. Standards for very low concentration levels may, however, be exceptions. Furthermore, the stability of such standards and the degree of protection required to safeguard them from contamination, deterioration, or losses always needs consideration.

Standards should never be used in an extrapolative mode. They should always bracket the measurement range. No measurement should be reported at a value lower or higher than the lowest or highest standard used to calibrate the measurement process.

2.4 Calibration, Standardization, and Analytical Response Function

Calibration may be defined as the comparison of a measurement standard or instrument with another standard or instrument to identify or eliminate by adjustment, any variation (deviation) of the accuracy of the item being compared. Physical standards, such as mass standards, and instruments, such as thermometers, are calibrated. Physical standards or calibrated instruments traceable to national standards are required for calibration of other standards or instruments. The uncertainty of the calibrations will depend on the uncertainty of the values of the standards and the measurement processes used for the intercomparisons.

Chemical measurements require standards consisting of pure chemicals or liquid, solid, or gaseous mixtures prepared from them. For most applications, chemicals of sufficient purity for use as standards or their preparation, can be obtained from suppliers. But for critical applications, chemical standards are sometimes first assayed to determine their purity or analyzed for impurities in order to calculate their composition. The latter practice can be erroneous unless it is ascertained that all significant impurities have been determined. The concentrations of solutions used as analytical reagents or as calibrants are sometimes defined on the basis of their preparational data and knowledge (or assumptions) of purity. When concentrations are determined by comparison with other solutions of known concentrations, the process is called standardization.

In general, the calibration of a chemical analyzer consists of the evaluation of its response function relative to the chemical composition of the samples to be analyzed. The analyzer responds to some property of the analyte, the value of which needs to be quantified by use of known substances. Then it can be tacitly assumed that the instrument will respond analogously to the standard and test samples. The sources of uncertainty in this case are the uncertainty in composition of the known samples and the validity of the analogy.

It is a generally accepted principle of reliable analysis that chemical analyzers should be calibrated over the full range of measurement and that measurement data be restricted to the range calibrated. It is not good measurement practice to report extrapolated data, i.e., data outside of the range calibrated. The range of reliable calibration can be considered as the range of reliable measurement and conversely.

The necessary frequency of recalibration or reevaluation of a response function will depend on the stability of the measurement system and the accuracy requirements for the data. To ensure confidence in measured values, such reevaluations should be made before significant changes are to be expected.

The terms primary and secondary standards are used frequently and need some discussion. Strictly speaking, a primary standard is one whose value may be accepted without further verification by the user. In turn, it may be used to establish or ascertain a value for a secondary standard. Thus, a secondary standard provided by one laboratory (e.g., a mass standard calibrated at NIST) could serve as a primary standard for another laboratory. In any case, the uncertainty of the value of any standard must be known since the adjective designation (primary or secondary) does not define any limits of uncertainty for its value.

Analytical chemists have used the terms primary and secondary to indicate the relative purity of materials that may be used to prepare solutions with accurately known compositions [3]. The International Union of Pure and Applied Chemistry [4] has set minimum levels of purity for primary and secondary chemicals.

Specifications for certain classes of physical standards have been established to include design characteristics and permissible departures (tolerances) from nominal values. Thus, a 1-g weight of Class 1 will have a tolerance of 0.034 mg while the tolerance is 2 mg for a Class 6 weight of the same denomination [5]. When used, the nominal weight is assumed, but it should be recognized that the true value may lie anywhere within the tolerance range. If such an uncertainty is too large, the standard should be calibrated, but upgrading may be difficult due to design considerations.

2.5 National Measurement System for Analytical Chemistry

What might be called the National Measurement System for Analytical Chemistry is shown in Figure 3. Measurement is a complex process, involving the coordinated interaction of a number of areas of science and technology. The figure is admittedly complicated, yet only the major supporting activities are shown. It is presented here largely to emphasize the complexity of measurement and to stimulate the reader to consider the aspects that must be controlled in order to obtain reliable data.

The measurement of any specific sample requires a measurement system individually designed with consideration of the requirements for the data. This system must be calibrated, using physical and chemical standards. As already discussed, the physical standards may be traceable to national primary standards maintained by NIST and compatible with those of other nations. The chemical standards generally will be prepared by the measurement laboratory. SRMs can serve as calibrants in some cases, but ordinarily they serve as quality assurance materials to evaluate measurement accuracy, to intercalibrate laboratories in a measurement program, and to provide compatibility of measurement data.

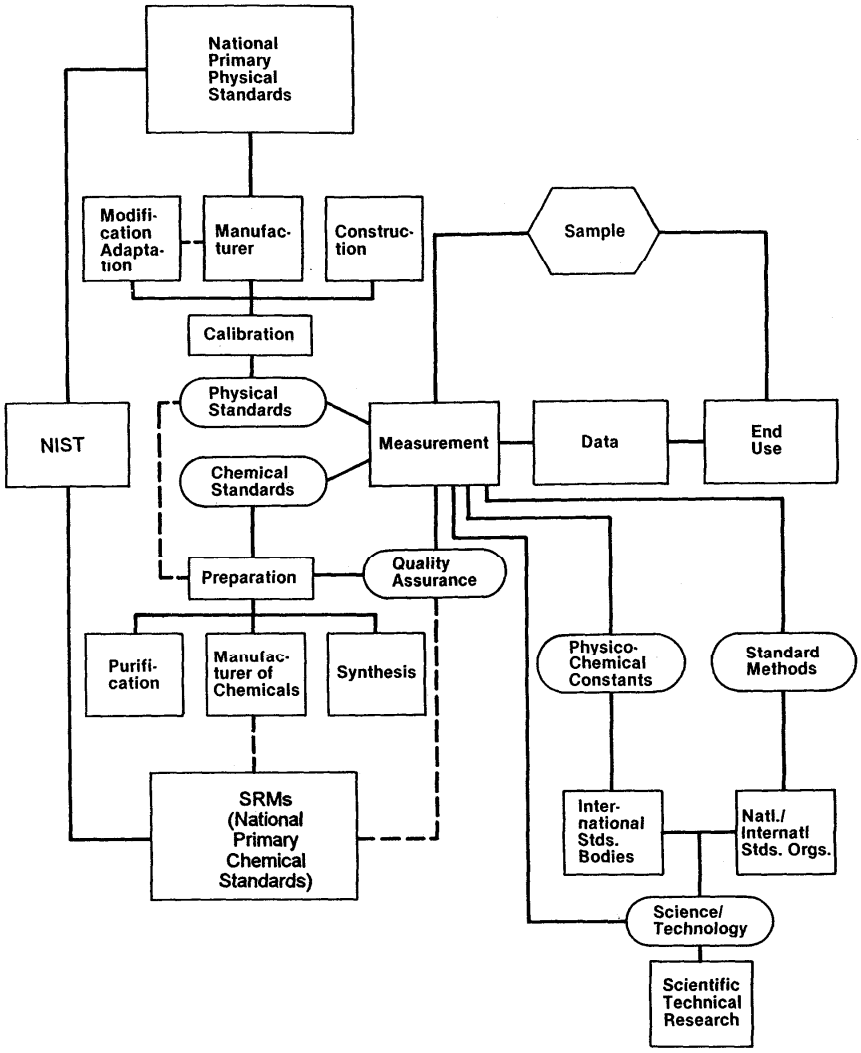


Figure 3. National Measurement System for Analytical Chemistry

The figure illustrates several points that have been discussed earlier. The critical dependence of modern analysis on both physical and chemical standards is indicated. Physical standards of requisite reliability usually are available from external sources, but as indicated, the chemist must often prepare all chemical standards used, starting with source materials. Questions about the matrix match of standards and test samples always must be considered. The measurement process is highly dependent on broad areas of science and technology as well.

Quality assurance of the measurement process is essential for reliable data. The important role of SRMs in controlling the calibration process and in assessing data quality is shown and will be elaborated on throughout this handbook.

3. QUALITY ASSURANCE

Quality assurance is the name given to the procedures used to ascertain that measurement data are good enough for their intended purpose [1,6]. It involves two distinct but related activities:

quality control - those procedures and activities developed and implemented to produce a measurement of requisite quality;

quality assessment - those procedures and activities utilized to verify that the quality control system is operating within acceptable limits and to evaluate the quality of the data.

The basic requirement for producing reliable data is appropriate methodology, adequately calibrated and properly used. This together with good laboratory and measurement practices, are the basic ingredients of a quality control program. The quality of the data may be assessed by use of reference materials to evaluate bias and the time-consuming process of redundancy to evaluate precision.

3.1 Quality Assurance of a Measurement Process

There is a growing awareness that analytical data for use in any decision process must be technically sound and defensible. Limits of uncertainty are required which need to be supported by suitable documentary evidence. Professional analytical chemists have always espoused this philosophy. Regulatory agencies and contracting parties increasingly are specifying it as a routine requirement. The formal and even informal procedures used to establish the limits of uncertainty of measurement data are generally referred to as "quality assurance", by which replicate measurements and independent procedures support claims for the accuracy of the data. When a measurement process is demonstrated to be in a state of statistical control, the accuracy of the process can be implied to characterize the accuracy of all data produced by it. Hence, the requirements for redundancy can be greatly reduced.

A measurement process of the type described above is illustrated in Figure 4. It utilizes methodology appropriate to the measurement program and appropriate quality control practices are followed. Statistical control is demonstrated by the measurement of replicate samples and internal reference materials, using control charts. This also permits the evaluation of the precision of the process.

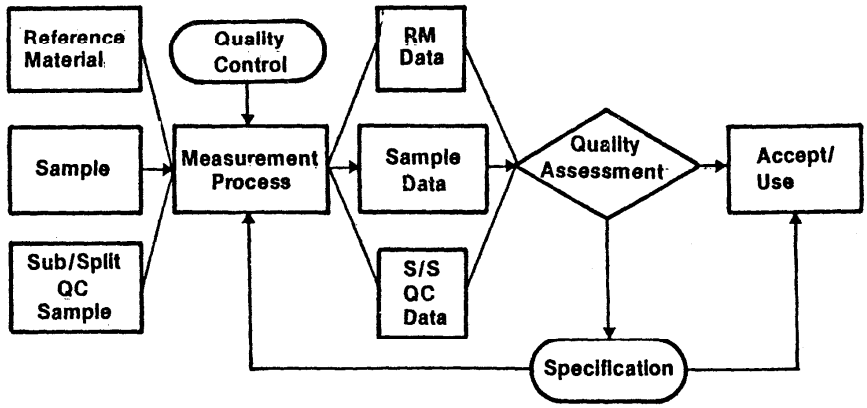


Figure 4. Measurement Process Quality Assurance

When the process is demonstrated to be in a state of statistical control, reference materials such as SRMs may be analyzed to assess measurement accuracy. The resulting judgment of precision and accuracy can be assigned to the sample data output of the process. The figure also shows how data quality assessment is used in a feedback mode to monitor the process or initiate corrective actions as required, and in a decision mode to release or use data.

3.2 Statistical Control

A stable measurement system is expected to produce reproducible data. Statistical control may be defined as the attainment of a state of predictability. Under such a condition, the mean of a large number of measurements will approach a limiting value (limiting mean) and the individual measurements should have a stable distribution as described by their standard deviation. Under such a condition, the limits within which any new measured value would be expected to lie can be predicted with a specified probability, the confidence limits for a measurement or mean of set of measurements can be calculated, and the number of measurements required to obtain a mean value with a given confidence may be estimated.

It is axiomatic that attainment of statistical control is the first objective of a measurement process. This is just another way of stating that it must achieve stability. Yet, it has the further connotation that the data produced are statistically describable. Eisenhart has stated -- "Until a measurement operation has been "debugged" to the extent that it has attained a state of statistical control, it cannot be regarded in any logical sense as measuring anything at all." [7]

When a measurement system is altered or disturbed, a new or modified measurement system may result with a limiting mean and/or a standard deviation different from the previous values. During normal use of a measurement system, changes can occur which are not known to the laboratory personnel. A well designed quality assurance program will monitor the system for such changes and indicate when corrective actions are required.

3.3 Control Charts

The philosophy of the use of control charts is based on the premise that analytical measurements may be systematized to provide a process simulating, in many respects, a manufacturing process. As the result of quality control procedures, a system may be debugged and attain a state of statistical control of its data output. Once this is achieved, the accuracy of the system can be evaluated for typical test samples and thus can be assigned to all similar measurement data generated by the system.

A control chart is simply a graphical way to interpret test data. In its simplest form, it consists of a graph, on which the periodic measurement results of a selected reference sample, are plotted sequentially (time-ordered). Limits for acceptable values are defined and the measurement system is assumed to be in control (variability is stable and due to chance alone) as long as the results stay within these limits. A second useful form of control chart is one in which the standard deviation or range (even differences between duplicates) of a series of measurements is plotted in a similar manner. The residence of the values within expected limits is accepted as evidence that the precision of measurement remains in control. The monitored precision of measurement and the accuracy of measurement of the reference sample may be transferred, by inference, to all other appropriate measurements made by the system while it is in a state of control.

Examples of each kind of control chart described above are given in Figure 5. In Figure 5A, the mean, $\bar{x}$, of two measurements is plotted sequentially. The central line is the most probable value for $\bar{x}$ (i.e., the grand average, $\bar{\bar{x}}$, of measurements of $\bar{x}$) and the limits, LWL to UWL (lower and upper warning limits), define the area in which 95% of the plotted points are expected to lie. The limits, LCL to UCL (lower and upper control limits), define the area in which almost all (99.7%) of the plotted points are expected to lie when the system is in a state of statistical control. It should be clear that, when more than 5% of the points lie outside of the warning limits or when values fall outside of the control limits, the system is behaving unexpectedly and corrective actions, even rejection of data, may be required.

A discussion of the strategy to follow in the use of control charts is beyond the scope of the present presentation, but laboratories using them need to develop such a strategy. Results are expected to scatter randomly within the limits. Systematic trends or patterns in the data plots could be early warning of incipient problems and are cause for concern. Hence, techniques to identify such problems should be practiced.

Control charts and the factors for calculating control limits are discussed more thoroughly elsewhere. ASTM Special Technical Publication MVL7 (formerly STP 15) is an excellent source of information [8]. Briefly described, the central line of a control chart is either the known value for the test sample (e.g., the certified value, if a SRM is used) or the mean of at least 15 sets of independent measurements. The standard deviation estimate, s, should be based on at least 15 such measurements. Control limits can then be calculated as follows.

<u>Control Limits</u>

Central Line	$\bar{x}$ (mean of at least 15 sets of measurements)
UCL	$\bar{x}+\frac{3s}{\sqrt{n}}$
UWL	$\bar{x}+\frac{2s}{\sqrt{n}}$
LWL	$\bar{x}-\frac{2s}{\sqrt{n}}$
LCL	$\bar{x}-\frac{3s}{\sqrt{n}}$

For the above limits, n represents the number of repetitive measurements of the reference sample, the mean of which will be plotted on the $\bar{x}$ chart. For an X chart (single measurement of the reference sample) n = 1.

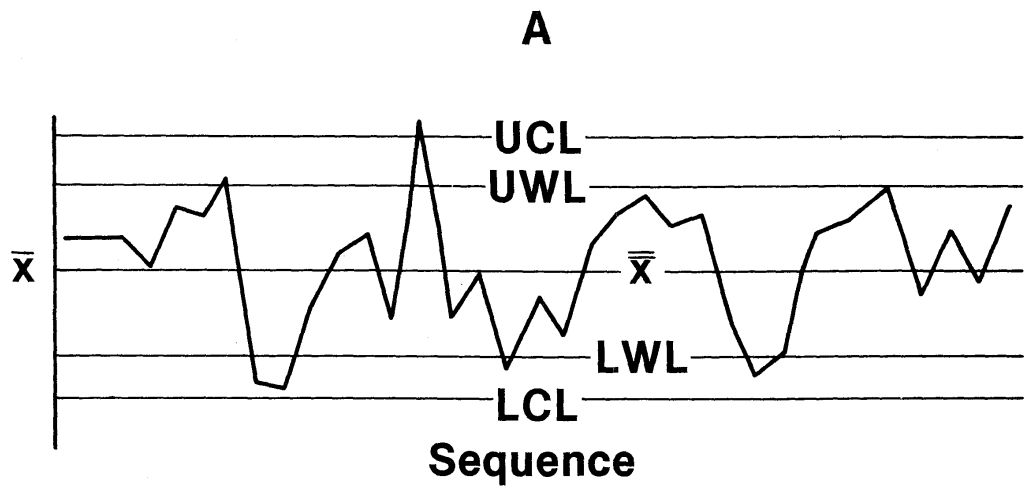


Figure 5A. Duplicate Measurements made on a SRM; $\bar{x}$ Chart

Figure 5B represents a range, R, control chart. In chemical measurements, the difference of duplicates is a good choice to plot on such a chart. The line, $\bar{R}$, represents the average range obtained as the result of a reasonably large number (e.g., >15) of sets of duplicate measurements. The warning and control limits are appropriate multiples of $\bar{R}$ and have the same significance as discussed in Figure 5A. The range is calculated without regard to sign (absolute value) so the lower limits are zero.

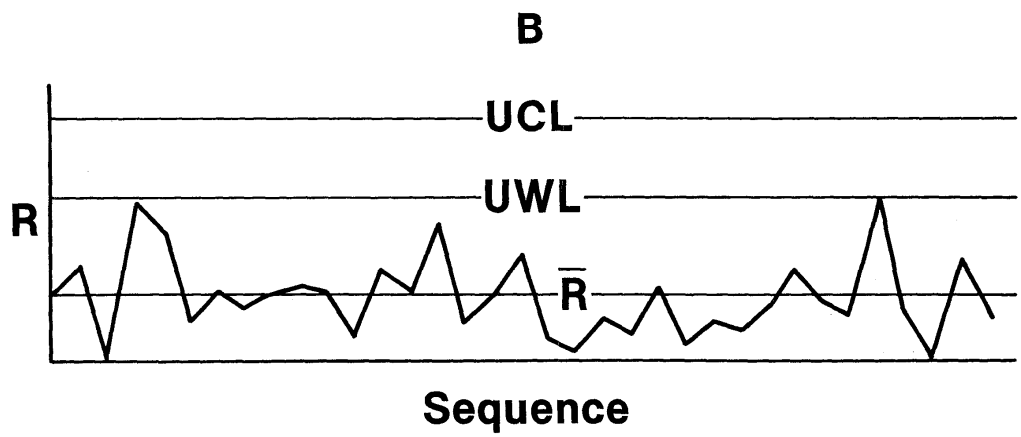


Figure 5B. Duplicate Measurements made on a SRM; R Chart

The factors for calculating limits are discussed in reference [8]. For duplicate measurements, the limits are

$\bar{R}$	mean of the differences of at least 15 sets of duplicate measurements
UWL	2.512 $\bar{R}$
UCL	3.267 $\bar{R}$
LWL	0
LCL	0

In practice, the differences of duplicate determinations are plotted on the control chart and statistical control is assumed as long as they fall within the expected limits. Again, they should not fall disproportionately outside of the warning limits and trends should not be observed.

The R chart is based on the known relation between the range and the standard deviation; hence, it is a form of standard deviation or precision chart. When used with an X or $\bar{x}$ chart, it is helpful in deciding whether an observed deviation is due to bias or to a change in precision. When used alone, an R chart will monitor precision (but not bias).

The test samples themselves, when measured in duplicate, may be plotted on a control chart to monitor precision. The ranges for duplicate measurements of a class of samples can be plotted on the same control chart, as long as they are expected to be measurable with comparable precision.

A s control chart is based on plotting the estimate of the standard deviation obtained from measurements of n replicates of the reference sample. Since a number of measurements are required (at least seven is recommended) to estimate the standard deviation, and since some calculations are required, such charts are seldomly used for chemical measurements. R charts can provide sufficient monitoring of precision with a reasonable amount of effort. They also offer the advantage of using a reasonable number of the actual test samples to monitor precision. The use of a R chart for test samples, and a X chart when utilizing a SRM, may be ideal choices for many laboratories.

A $\bar{x}$ control chart is more robust than an X chart. Since it is based on the mean of two or more measurements, an occasional outlier will have limited influence on the decision process. However, it requires additional work and this should be considered when using it in a quality assessment program. If the assessment strategy calls for confirmatory measurements of reference samples when out-of-control is indicated, the advantage of a $\bar{x}$ chart is lessened. Such a strategy is most effective when control charts are maintained and used in a real-time mode. This also provides the advantage of taking immediate corrective actions and thus minimizing the uncertainty of data.

The question of how to obtain the statistical information necessary to construct a control chart needs to be considered. Once the decision to develop a control chart is made, one might want to acquire the standard deviation and mean data as quickly as possible. But this could be misleading. It has been mentioned that measurements made over a short period of time show greater consistency than those obtained over a long period of time. Since the control chart will be used over a long period of time, the latter is more appropriate for judging performance.

To develop control limits based on long-term behavior, it is recommended that at least 15 data points be accumulated and that no two points be obtained on the same day. This recommendation applies only when obtaining the standard deviation data to establish control limits for a $\bar{x}$ or X chart and when preparing a s or R control chart.

If a SRM is used as the control chart reference sample, the value for the central line is known, namely the certified value. If a laboratory's own internal reference materials are used, much work may be required if the "true value" for the central line is to be used. Some laboratories use an analytical mean value for the central line, and assume that this is essentially the "true value". This may be true only if the measurement process has been demonstrated to have negligible bias. The use of an analytical mean as the value for the central line can be useful in indicating stability of a process, but bias can be evaluated only when true values are known.

3.4 Frequency of Use of Reference Materials

The optimum frequency of use of reference samples and also of replicates of actual test samples will depend on the stability of the measurement system and the risk involved when the system departs from statistical control [9]. Since all data obtained during the period, last-known-in-control to first-known-out-of-control are suspect, such intervals may need to be shortened. The real-time use of control charts and/or reference material data is a further consideration. While the following discussion is directed toward control chart maintenance, the same philosophy applies whether this is done or the reference samples' results are interpreted by other means.

There are several empirical approaches to deciding on the frequency of use of reference samples. The experience of the laboratory may indicate the expected frequency of occurrence of trouble, in which case reference sample measurements - at least three in number - should be equally spaced within such an interval. Another approach is the "length of run" concept. This is based on identifying recognizable breaks in the production (of data) process which could cause significant changes in precision or bias. Such breaks could include change of work shift; rest periods; the change, modification or adjustment of apparatus; use of new calibration standards; significantly long down-times in the process; and/or use of a new lot of reagents. At least three reference samples should be measured during any of these breaks when they are considered to be potentially significant.

In summary, the measurement of reference materials is a risk-reducing procedure. However, if it involves more than 10% of a laboratory's measurement effort, either the quality control process needs improvement or too much effort is being exerted in this direction. If less than 5% of effort is devoted to such measurements, the laboratory could be taking too high a risk of producing unacceptable data or may not even know the quality of the data it is producing. The two prior statements are made with a laboratory making a significant number of high-quality routine measurements in mind. However, if a laboratory's program involves occasional or one-of-a-kind measurements, the amount of quality assurance effort required, including the number of reference material measurements to be made, could be significantly more than that indicated above.

Suggested measurement schedules for efficient utilization of references materials are given in Tables 1 and 2 [9]. The sequence in Table 1 utilizes a combination of an Internal Reference Material (IRM) and a SRM. The sequence in Table 2 utilizes a limited number of duplicate or split samples together with reference materials. In either case, the use of control charts is recommended on a real-time basis. Recommended critical decision points in the measurement sequences are also indicated.

Table 1. Quality Assessment Using RMs

CALIBRATION - FULL EXPECTED RANGE

· IRM
TEST SAMPLES - GROUP 1

· IRM
TEST SAMPLES - GROUP 2

· IRM
·
·
· SRM
·
·
TEST SAMPLES - GROUP N-1

· IRM
TEST SAMPLES - GROUP N

· IRM
CALIBRATION - MIDRANGE POINT

NOTES · - DECISION POINT

1. MAINTAIN CONTROL
   X-CONTROL CHART, IRM
   R-CONTROL CHART, ΔIRM
2. SYSTEM MUST BE IN CONTROL AT DECISION POINTS
3. AT LEAST 2 GROUPS: MAXIMUM OF 10 SAMPLES IN EACH GROUP
4. AT LEAST ONE SRM MEASUREMENT SHOULD BE MADE DURING EACH SEQUENCE/DAY

Table 2. Quality Assessment Using Duplicates/Splits

CALIBRATION - FULL EXPECTED RANGE

- CALIBRATION CHECK - MIDRANGE POINT
  SAMPLE 1
- SAMPLE 1 D/S
  SAMPLE 2-9
  SAMPLE 10
- SAMPLE 10 D/S
- IRM or SRM
  SAMPLE 11-19
  SAMPLE 20
- SAMPLE 20 D/S
- CALIBRATION CHECK - MIDRANGE POINT
- CALIBRATION CHECK - MIDRANGE POINT/DUPLICATE

NOTES
- - DECISION POINT

1. MAINTAIN CONTROL CHARTS
   a. DUPLICATE MIDRANGE CALIBRATION
   b. DUPLICATE/SPLIT SAMPLE
   c. X-CONTROL CHARTS, SRM AND IRM
2. SYSTEM MUST BE IN CONTROL AT DECISION POINTS
3. IF MORE THAN 20 SAMPLES, REPEAT SEQUENCE
4. IF LESS THAN 20 SAMPLES, DIVIDE INTO TWO GROUPS AND FOLLOW SIMILAR PLAN
5. AT LEAST ONE SRM MEASUREMENT SHOULD BE MADE DURING EACH SEQUENCE/DAY

4. REFERENCE MATERIALS

4.1 Role of Reference Materials

In the most general terminology, a Reference Material (RM) is a substance for which one or more properties are established sufficiently well for use to calibrate a chemical analyzer or to validate a measurement process [10,11,12]. An Internal Reference Material (IRM) is such a material developed by a laboratory for its own internal use. An External Reference Material (ERM) is one provided by someone other than the end-user. A Certified Reference Material (CRM) is a RM issued and certified by an organization generally accepted to be technically competent to do so. A Standard Reference Material (SRM) is a CRM issued by NIST.

A reference material is for use in a decision process, hence the requirement for reliability of the value of the property measured must be consistent with the risk associated with a wrong decision. The appropriateness of the reference material in the decision process must also be considered. For some purposes, a simple substance, mixture, or solution will be adequate and the value of the property may be calculated from the data for its preparation. However, even this is best verified by suitable check measurements to avoid blunders. Many decision processes require a natural matrix reference material which may necessitate extensive blending and homogenization treatments and complex analytical measurements. In such cases, only a highly competent organization may have the resources and experience to do the necessary work.

The terms certificate and certification merely refer to the documentation that supports the reference material. Guidelines for the content of certificates for reference materials have been prepared by the International Organization

for Standardization (ISO) [13,14,15]. It recommends the kind of information the certificate should contain, but does not describe how the information should be obtained. Furthermore, there are no guidelines for judging the relative quality of reference materials.

The guiding principle in issuing a SRM is that it will be used for measurement quality assessment. Hence, the property certified must be accurately known. The uncertainty in the certified values takes into account that due to measurement and any variability (inhomogeneity) between and/or within samples of the material [16]. Thus, definitive methods are used for establishing the values of the certified properties or they are measured by two or more independent reliable methods. In the latter case, the results must agree to minimize the chance for measurement bias. All certification measurements are described in the certificate or are properly referenced. The certification measurements may be preceded by stability studies, as appropriate, to set limits on the life expectancy of the material.

4.2 Concept of Traceability

The concept of traceability to national standards has been advanced in recent years to facilitate intercalibration of laboratories and promote compatibility of measurements. Traceability, as related to a standard, may be likened to genealogy in that it may describe the chain of calibrations related to the establishment of its value, including the intermediate standards used and the various measurements involved. In the area of physical measurements, calibrations of standards or artifacts with respect to the national measurement standards can be made at NIST with high precision. These may be made for secondary calibration laboratories who, in turn, calibrate standards for others. Each time this is done, the uncertainty is increased due to uncertainties in a laboratory's own standards and propagation of the uncertainty of measurement. Measurement Assurance Programs (MAPs) are designed to minimize the latter and thus decrease the accumulation of uncertainty as measurements go lower down the measurement chain. The various measurements must be made with adequate quality assurance if reliable limits of uncertainty are to be assigned. The responsibility is that of the measurement laboratory.

While chemical measurements rely to some degree on physical measurements and require calibrated physical standards, very few chemical standards are disseminated in the same manner as are the physical ones. Hence, it is difficult, if not impossible, to establish the traceability of most chemical standards to other standards and especially to national standards. One exception, however, is when SRMs are used as calibrants for a measurement process or as primary standards for chemical analysis. All measurements using such SRMs have the capability of being traceable to a common set of standards and the intercalibration of laboratories is facilitated. Relatedly, certain commercial suppliers are producing reference materials, certified with respect to specific SRMs, using protocols developed to issue such materials as CRMs. [17].

While an SRM ordinarily does not provide traceability in its narrowest interpretation, it may serve a broader and more useful function; that being, to provide measurement assurance which ensures both proper calibration and acceptable utilization of methodology. When specific SRMs are commonly used in a systematic manner and tracked with control charts, intercalibration of all laboratories involved and compatibility of the data generated may be achieved. This is shown by Figure 6. Thus, measurement networks can specify the SRMs to be used and the quality assessment procedure to be followed in order to attain compatibility of monitoring data. But while acceptable SRM data indicate acceptable performance of the measurement system, discrepant results may not be simple to interpret since they could indicate calibration uncertainties, application problems, or both. Also, one cannot completely rule out the misapplication of methodology or the use of inappropriate methodology as a source of trouble. Nonetheless, a well-designed quality assurance program should facilitate the identification of the source(s) of the problem(s).

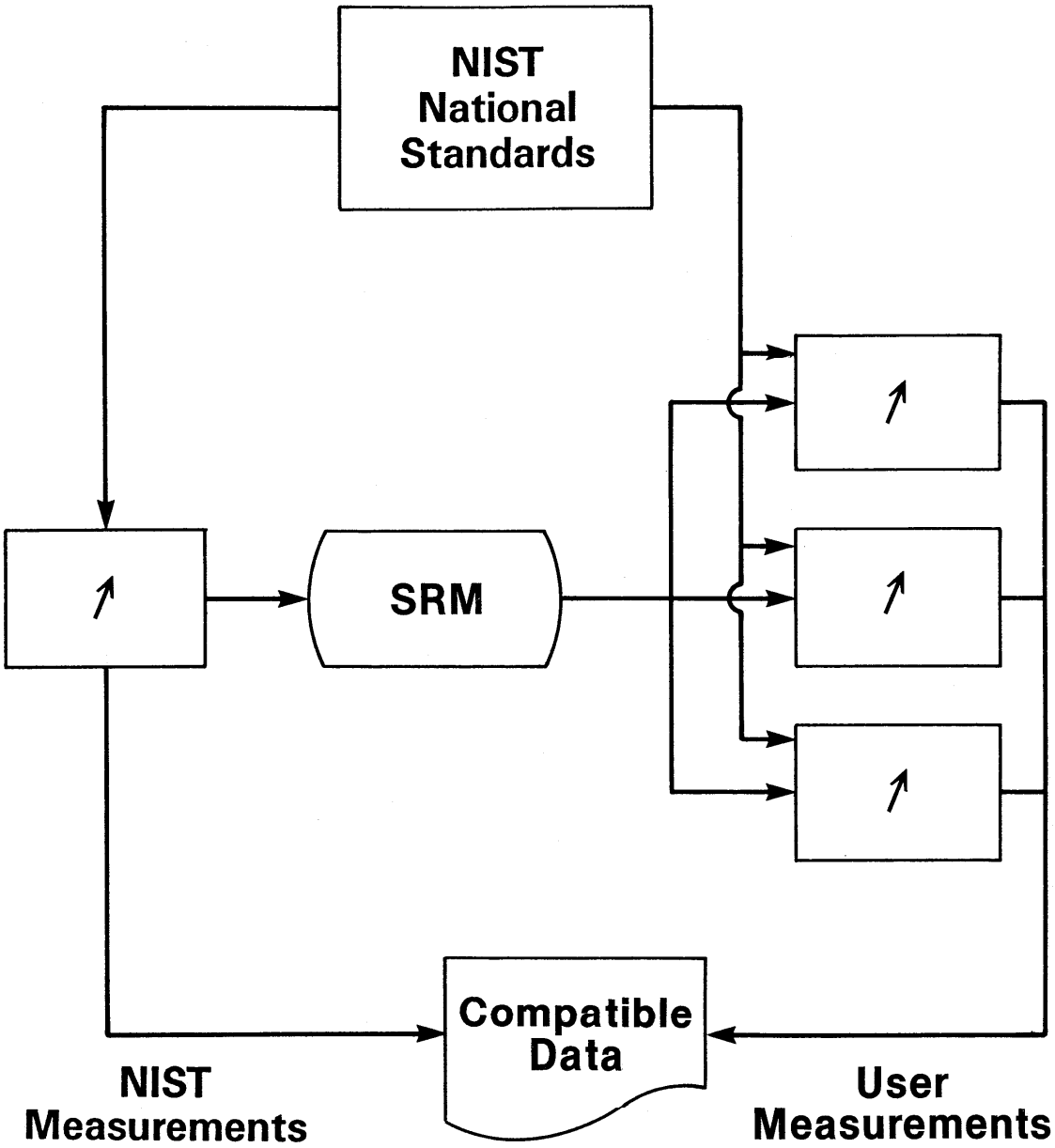


Figure 6. Measurement Compatibility by Intercalibration Using SRMs

5. STANDARD REFERENCE MATERIALS

5.1 Philosophy of SRM Production

SRMs are considered to be services to the individual user who must pay the full cost of the service provided. Costs of development, preparation, certification, and marketing are accumulated and pro-rated on the basis of the number of saleable units that are produced. Thus, the costs and benefits are prime considerations in authorizing and issuing a SRM. Accordingly, the production of low demand, high (unit) production cost SRMs is difficult, though not impossible.

SRM production is often preceded by a substantial research effort. Methodology may need to be developed or potential bias problems must be solved if accurate certification is to be done. Materials-related problems such as stability, homogenization techniques, and proper conditions for packaging and storage may need investigation. Occasionally, the results of such research are applicable to wider areas of science and technology, or at least to broader areas of SRM certification. In such cases, the costs of such work may be supported by general research funds and not charged to production of a SRM. Otherwise all costs, including research and development costs, must be recovered from sales. This increases the unit costs of SRMs and impacts on the development of new items for which substantial research and development funding are necessary.

5.2 How a SRM is Produced

Identification of Need

SRMs are developed to meet measurement needs. The need may be specific, e.g., the result of a regulatory issue, or it may be general, such as a wide-spread measurement problem. Often the need comes to the attention of NIST in the form of a specific request or as the result of interactions by NIST scientists with the measurement community. Because the Standard Reference Materials Program must be self-supporting, the magnitude of the need, cost of development and the prospect of cost recovery through sales, together with the technological chances of success, are important considerations in establishing the feasibility of issuing an SRM.

Determination of Characteristics/Properties/Specifications

The necessary properties of a useful reference material must be given careful consideration. The kind and level of parameters certified, the matrix and other physical characteristics, homogeneity requirements, and the maximum acceptable uncertainties for the certified values are key considerations. While a reference material is developed for a specific use, it is often possible to extend its usefulness by certifying additional parameters or by modifying the matrix. When doing this, however, it must be remembered that modification from a specific to a generic standard could possibly limit its usefulness for the initial purpose, while not significantly extending its areas of application. Moreover, the certification of additional parameters can increase costs and users do not ordinarily like to pay for information (in this case, additional certified values) which is not of direct use to them.

After considering factors such as those discussed above, and from discussions with the user community, minimum specifications for a candidate SRM can be drafted. These specifications may utilize materials available on the market or require that suitable materials be produced. In some cases, NIST must prepare the materials or prototypes for initial testing. Often, preliminary research and development efforts are necessary to evaluate the feasibility of producing SRMs and/or to develop specifications.

Preliminary Studies

After a material has been obtained, measurements are made to evaluate its compliance with the specifications. While the exact level of the analyte is seldom a controlling requirement, homogeneity, both within and between units

of issue, is always an important requirement. Ordinarily, it is desirable to certify the material as a lot rather than as individual items. If this is done, the homogeneity between units of issue must be acceptable.

Homogeneity

Homogeneity evaluation may be done in two phases. Preliminary measurements will need to be made to accept material for conformance with specifications and to decide on such questions as pre-mixing and subdivision into units of issue (e.g., bottling) prior to certification analysis. When a multicomponent/parameter SRM is involved, this can be a major undertaking - especially if homogeneity determinations for each constituent/property are to be undertaken at this time. When possible, a quick and precise method is sought to evaluate homogeneity. However, this may not be possible for multiparameter materials, in which case initial homogeneity may need to be judged on the basis of a limited number of constituents.

Final homogeneity evaluation is determined by interpretation of certification data on each individual constituent or property. This requires design and execution of the measurement program so that variance of measurement and sample composition can be individually evaluated.

Measurement

The certification measurements are conducted according to a quality assurance plan established before the work is actually begun. This requires development of a statistical plan for sampling and measurement, selection of methodology which is reliable, maintenance of statistical control of the measurement process, and quality assessment of the data by concurrent measurements of suitable reference materials.

The methodology is selected on the basis of the following considerations. When possible, the attainable accuracy of measurement should be better than that required for use of the data. The first choice for methodology is a method of known and demonstratable accuracy. The term "definitive method" has been coined for this and is finding considerable usage, especially in relation to reference material analysis. A definitive method is one based on sound theoretical principles and which has been experimentally demonstrated to have negligible systematic errors and a high level of precision. While a technique, that is a measurement principle, may be conceptually definitive, a method based on such a technique must be demonstrated to deserve such a status for each individual application.

An example of a definitive technique is isotope dilution mass spectrometry for trace analysis in which the concentration of unknown samples is related directly to the actual weights of spikes of isotopes or isotopically-labeled compounds. A mass spectrometer is used to measure isotopic ratios, obviating the need for instrumental corrections. The only theoretical uncertainty in such a process is the question of equilibration between a natural analyte and a spike. The accuracy attainable will depend on isotopic and chemical purity of the spike, the care used in preparation and measurement, and the degree of equilibration achieved.

Examples of other definitive techniques are gravimetry and coulometry. Both are based on fundamental measurements that can be made with high accuracy. As is the case with any methodology, it must be demonstrated that no significant systematic errors are relatable or present. When used in a specific application, possible biases can be minimized by utilizing multiple analysts/instruments to the greatest extent possible. Redundancy of measurements in random sequence is another technique to avoid application bias.

Because definitive methods are not always available, the multitechnique approach is one often used in certification of SRMs. Parameters are measured by at least two independent techniques, when possible, and such measurements must agree within reasonable limits to permit certification. If significant discrepancies occur, additional work must be carried out to reconcile them; otherwise the values cannot be certified. They may, however, be reported for informational purposes only.

Another mode of certification, which can be called the multilaboratory approach, is used for renewal of certain compositional SRMs. A group of laboratories of recognized competence are requested by NIST to analyze a renewal SRM, using methods of proven accuracy and the existing SRM as a control. NIST carefully scrutinizes the data for any significant discrepancies and, if necessary, requests additional analyses using the same or different methodologies.

In a few cases, SRMs are certified for the value of a constituent or property that is method dependent because existing technology requires such. An example of this is the Kjeldahl technique. In such a case, demonstration of statistical control of the measurement process and agreement of results by independent analysts are the requirements for certification.

Evaluation of Data

All SRM data undergo thorough statistical analysis to establish limits of uncertainty due to measurement and any variability among the units of issue. The advance cooperation of statisticians in planning the experimental program for a SRM is essential if the measurements are to be made properly, so that a thorough evaluation of the reported values for the SRM can be conducted. The interpretation of the certified values is discussed later in this handbook (see Sections 5.4 and 5.5).

Follow-up

SRM production is usually preceded by studies of the stability of candidate materials. When possible, only materials which have a long shelf life are selected. But if there is any limitation on stability, it is indicated on the certificate. In addition, NIST periodically performs "shelf life" analysis on certain SRMs when it is considered that some deterioration could be possible.

Most SRMs are prepared in sufficient quantity so that a several-year supply, based on anticipated demand, can be added to inventory at the time of issue. This also applies to renewal SRMs which are made available as existing stock is depleted. However, demand sometimes exceeds expectations, especially for SRMs which must be prepared in limited quantities, so delays in the availability of a particular SRM may occur. Users must also recognize that changes in technology and conflicting priorities of competitive items may also cause the cancellation of plans for continuous renewal of a SRM.

It is NIST's aim to make SRMs as useful as possible to purchasers. Standard Reference Materials Program (SRMP) and NIST technical staff are available to answer questions and inquiries about SRMs and every effort is made to provide satisfactory solutions to application problems.

A SRM is ordinarily certified using state-of-the-art methodology. This could be methodology already widely used in practical analysis, with special care given to calibration, to quality control, and to eliminating sources of bias. But in many cases, the methodology is suitable only for research laboratory use. For example, isotope dilution mass spectrometry is often used for certification; however, it is a time-comsuming and costly technique and would be inappropriate for routine analysis. In any case, certified values are usually independent of the measurement method. When certified values are method-dependent, the methodology used in certification, together with references where detailed information about the methodology can be found, are always provided in the certificate. For a number of SRMs, NIST 260 publications have been prepared which describe the measurement processes in detail. A current listing of these publications is provided in the front of this handbook.

5.3 Differences Among Measurement Methods

Agreement of measured values by two or more independent methods is one of the approved conditions for certification. Of course, measured values never agree perfectly, so the statistical significance of disagreement must be considered. Results may not agree within their respective uncertainty for one or more of the following reasons.

1. Matrix effects in a method have not been fully compensated for by the calibration procedure used.

2. Other systematic errors have not been fully compensated for or unsuspected ones may exist.

3. It is in the nature of things that the more precisely a measurement is made, the smaller the difference that can be detected.

4. The most common reason, perhaps, for two or more methods to disagree is because the standard deviation of the method has been underestimated.

In any event, systematic differences in measured values must be examined for their practical significance. Unless discrepancies can be reasonably resolved, questionable values should not be certified.

5.4 Understanding Certificate Information

SRM certificates provide a variety of information about the particular material. Compositional values with uncertainty limits are given for all certified analytes. Ordinarily the limits are 95% confidence intervals in homogeneous materials and include allowances for the uncertainties of known sources of systematic error as well as the random error of measurement. Many certificates also will include values for other parameters or analytes which are being reported for "informational purposes" only. The values are not certified because 1) they were measured by only one technique, 2) they are results from discrepant measurements by several techniques or, 3) there are homogeneity problems which detract from the values' analytical usefulness. Informational values may have uncertainty values assigned to them, but they simply represent the analyst's best judgment of the random error uncertainty.

The certificate will sometimes describe restrictions in the use of the sample, which must be adhered to for reliable results. One of these restrictions concerns drying. Whenever drying is critical, instructions for doing so are included and must be followed. In the case of some SRMs, some constituents must be determined on pre-dried samples while others (volatiles) are determined with subsequent correction to dry weight based on a moisture determination.

For heterogeneous materials, the minimum weight of an analytical sample is often specified. This requirement should be followed if certified values are to be realized.

In the case of some SRMs, segregation on standing is a potential problem. The certificate will instruct the user to shake, rotate, stir, or otherwise reconstitute the material. Failure to do so not only invalidates the present measurement, but jeopardizes further measurement from the same container, due to disproportionate withdrawal of the contents.

Storage of some SRMs may also need to be done under prescribed conditions. Refrigeration and/or freezing is often necessary as is protection from radiant energy or moisture (once the container is opened). Because of such problems, some SRMs are certified for first use only.

In other cases, certification is valid only for a finite lifetime (e.g., 1 yr, 5 yrs, etc.). Such time periods are established in order to reduce the need to notify users in case of a change in the material. The lifetime of a SRM is always calculated from the time of shipment -- it is neither related to any packaging date(s) found on the SRM container nor to the date the certificate was issued.

The aforementioned and other restrictions are necessary to protect the integrity of the sample or to ensure results that will be consistent with the certified values. NIST can accept no responsibility for integrity of the material if all certificate instructions are not carefully followed.

5.5 Uncertainty of Certified Values

(From a lecture by H.H. Ku, NBS, presented at the Precision and Accuracy Seminar, 27 March 1980. See also [10, p. 296] and [12].)

For the purpose of this discussion, SRM certification can be divided into two groups:

A. Each unit is measured and carries its own value (e.g., permeation tubes, microhardness blocks, etc.).

B. Samples chosen statistically from the SRM lot are measured and one certificate gives the value for all units (most chemical types).

The uncertainty of the certified values for Group A SRMs will depend entirely on the uncertainty of measurement. This will be based on the standard deviation and best estimates of the systematic error uncertainties which have been corrected.

For Group B SRMs, any differences in the certified property within the units or between units of the lot poses a problem. A sampling and measurement scheme must be devised to determine if inhomogeneity exists and to estimate its magnitude whenever it is important to the use of the SRM.

Homogeneity checks may be made using two different sampling schemes. In one, a batch of material is subsampled, using a statistically developed scheme, and measurements are made to detect significant differences in the compositions of the subsamples. This has the advantage that grossly heterogeneous material would be rejected, thus saving the time and cost of packaging unacceptable material. It has the disadvantage that further heterogeneity could be introduced into potentially acceptable material by segregation or contamination during packaging. For material believed to be homogeneous, bulk examination may be the method of choice. In some cases, this is an advantage because once homogeneity is confirmed, the material can be analyzed and packaged as required.

Material considered to have measurable heterogeneity is best checked after packaging into bottles, vials, etc. This way it is not only possible to detect the original heterogeneity, but also any that might have resulted from the packaging process. Three kinds of heterogeneity are generally possible:

a. Between units or within units

b. Trend or pattern - along a rod, within a sheet or block of material, in the order of preparation, etc.

c. Between blocks of units - processed on different days, between drums, between lots or batches, etc.

The sampling scheme used for each SRM depends to a large extent on knowledge of and experience with the types of heterogeneity most likely to occur. The sampling scheme will be designed predominately to check on variability due to that source. A knowledge of the details of the packaging process is also required. The filling sequence of bottles, or the order in which specimens were cut from a material mass, must also be factored into the design.

If the material is homogeneous, it is accepted; if the material shows large variability, it is rejected. Often the variability is at about the level that can be detected by a particular analytical method. In this case, both the analytical error and the heterogeneity of the material contribute to the uncertainty in the final product — SRM units.

Let σ_m be the standard deviation of the analytical method, and σ_c be the standard deviation of the value of the individual units about the mean value of the lot. Then the standard deviation, σ, of a single measurement on a unit drawn at random from the lot would be

$$\sigma = (\sigma^2_m + \sigma_c^2)^{1/2}$$

Real estimates of two of the three sigmas allow an estimate of the third.

When possible, σ_m is evaluated independently of the measurement of the SRM. In this case, the standard deviation, σ, of the measured values of individual samples, together with the value for σ_m, permits an estimation of σ_c.

Often, possible matrix effects preclude or raise questions about the independent estimation of σ_m. Even if σ_m can be assumed to have a certain value, it may be necessary to verify it for the SRM certification measurements.

By duplicate measurements of portions of n randomly selected samples, both σ_m and σ_c may be estimated. Let x_i and y_i be the first and second measurements on portions of the ith unit (bottle, disk, subsample, etc.). The results may be tabulated as:

$$x_1 - y_1 = d_1 \qquad \frac{1}{2}(x_1 + y_1) = z_1$$

$$x_2 - y_2 = d_2 \qquad \frac{1}{2}(x_2 + y_2) = z_2$$

$$\vdots \qquad \vdots$$

$$x_n - y_2 = d_n \qquad \frac{1}{2}(x_n + y_n) = z_n$$

. . . . . .

$$s_1^2 = \frac{1}{2n}\sum (d_i)^2 \qquad s_2^2 = \frac{1}{n-1}\sum (z_i - \bar{z})^2$$

s_1^2 estimates σ_m^2; $\qquad s_2^2$ estimates $\sigma_c^2 + \frac{\sigma_m^2}{2}$

From these results, the estimates of σ_m^2 and σ_c^2 can be calculated.

Another way in which σ_c may be estimated is by the use of an independent method of measurement. Occasionally a highly precise method, available for intercomparison of samples, may not be feasible for use for certification measurement due to calibration problems. However, such a procedure is especially useful for confirmation of the homogeneity status of a candidate SRM.

In any evaluation of homogeneity, it must be remembered that each analyte certified must be individually considered and examined. It is not justifiable to merely extend conclusions on the homogeneity for one analyte to another, even though they may be closely related in other respects.

Homogeneity statements always must be coupled with the size (mass) of sample to be used. Basically, heterogeneous materials, such as bulk solids, can exhibit gross heterogeneity as the sample size is diminished (see Appendix D.2). In crushed material, for example, individual particles could have widely different compositions. Apparent homogeneity is improved as larger sample sizes are considered, since individual differences will be averaged out. Accordingly, the minimum sample size necessary to realize the certified values often will be specified in the certificate. The NIST certified values cannot be extended to the composition of subsamples smaller than the recommended size.

Materials of Group B which are essentially homogeneous are certified on a lot or bulk basis. All subsamples are considered to have the same composition, which is certified together with a confidence interval based on estimates of uncertainties for both random and systematic errors of measurement.

Materials with significant, but usable levels of inhomogeneity, may be certified as a batch, in which case the statistical tolerance limits are given. This includes the average value of all samples in the batch and limits within which most individual samples are expected to lie, with a stated confidence. Because only a small number of samples have been analyzed, the limits for a given percentage of samples cannot be stated with certainty. Rather, there is only a probability that the limits are valid. Thus, it can be said with 95% confidence that these statistical tolerance limits cover the true values of 95% of the samples of the batch. This is often stated on a certificate for such a material.

In the case of most granular SRMs, the within unit-of-issue heterogeneity is essentially the same for all units-of-issue and no significant average difference is to be expected between individual units. In such a case, the average composition of subsamples within all units would not be expected to differ significantly from unit to unit. In using such SRMs, it should be remembered that the composition of any subsample is expected to be within the tolerance limits and the average composition of a number of subsamples is expected to approach the certified value.

5.6 Relation to Real-World Samples

A SRM is a homogeneous and stable material which is measured accurately and certified as a reference material for the evaluation of a measurement process. In the case of closely specified industrial materials, a SRM may be identical to the substances that are analyzed routinely by a measurement laboratory. In other cases, a SRM may be representative of a class of materials. In still other situations, a SRM may be developed to challenge the analytical process used in particular measurement programs.

Because a SRM must be prepared in relatively large quantities with stringent requirements for homogeneity, a natural material may be significantly modified. This process could result in the addition of small amounts of foreign matter or in the intentional removal of some component to improve homogeneity.

The collection of material can also change its composition in some cases. For example, some volatile components were undoubtedly lost during the 2-year period of collection of Urban Particulate Matter SRM 1648, and oxidation of some constituents may have resulted, as well.

From the above discussion, it is clear that a SRM should not be considered typical of a specific material, but as a measurement standard. Accordingly, it should not be a surprise to discover that the composition of a SRM may not be identical to any natural material. [This should not detract in any way from the SRM's usefulness since its analytical nature is not altered significantly by the way it is prepared]. The most important consideration is whether or not a SRM is appropriate to evaluate a specific measurement process. This subject is addressed further in Section 6.4.

6. USE OF SRMS

6.1 Kinds of SRMs

SRMs are defined as well-characterized and certified materials, produced in quantity to improve measurement science and technology. They fall into three general categories: SRMs for Chemical Composition; SRMs for Physical Properties; and, SRMs for Engineering Materials. All three categories are further subdivided as shown in Table 3.

NIST Special Publication 260, Catalog of Standard Reference Materials - which describes currently available SRMs and RMs and those scheduled to be available within a few months of the catalog issue date - has historically been constructed according to the above categories. However, the subcategories listed in Table 3 now better define each group for ease in locating specific materials. The catalog is revised and updated every other year so new subcategories are added at that time. In the interim, SRMP newsletters issued quarterly, are used to provide the most recent information on new materials issued or materials deleted from stock. A current copy of the catalog or the newsletter can be obtained by contacting the SRM Program sales office.

Table 3. SRM Category/Subcategory Listing

Category	Subcategory
SRMs for Chemical Composition	101. Ferrous Metals
	102. Nonferrous Metals
	103. Microanalysis
	104. High-Purity Materials
	105. Health and Industrial Hygiene
	106. Inorganics
	107. Analyzed Gases
	108. Fossil Fuels
	109. Organics
	110. Food and Agriculture
	111. Geological Materials and Ores
	112. Ceramics and Glasses
	113. Cement
	114. Engine Wear Materials
SRMs for Physical Properties	201. Ion Activity
	202. Polymeric Properties
	203. Thermodynamic Properties
	204. Optical Properties
	205. Radioactivity
	206. Electrical Properties
	207. Metrology
	208. Ceramics and Glasses
	209. Miscellaneous Properties
SRMs for Engineering Materials	301. Sizing
	302. Surface Finish
	303. Nondestructive Evaluation
	304. Automatic Data Processing
	305. Fire Research
	309. Miscellaneous Performance Engineering Materials

Industrial materials that must be analyzed frequently for quality control of production processes constitute a major fraction of all SRMs. Predominant in this group are metals in which essentially all major alloy types are specifically represented. Ores, minerals, cements, glasses, and ceramics are also included. Ordinarily, these SRMs are sampled directly from the container and analyzed by the user in the same manner as the day-to-day samples of the laboratory. The results of such SRM analyses are frequently plotted on control charts to monitor the measurement process.

High-purity chemicals for use as primary standards in a wide variety of chemical analyses also constitute an important group of SRMs. Typically, the user will prepare solutions from these materials to be used directly in the analytical process or for standardizing other analytical reagents.

The Health and Industrial Hygiene subcategory comprises a large and growing group of SRMs. Originally, such SRMs consisted largely of pure substances from which spikes or calibration solutions could be prepared. They are now being augmented by natural matrix materials containing analytes of interest that are measured directly without preliminary preparation.

The Inorganic, Analyzed Gases, Fossil Fuels, and Organic SRMs satisfy many routine environmental monitoring requirements and some special situations as well. SRMs addressing the criteria for air pollutant analyses were introduced early into the SRM Program. These were followed by SRMs related to the measurement of emissions from mobile and stationary sources, priority pollutants, and hazardous wastes. The SRMs for the gaseous pollutants of the atmosphere are particularly slanted toward the calibration of analyzers and provide traceability for industrially-produced working standards. Because of the wide variety of sample types and the number of constituents of interest, it is virtually impossible to provide matrix matches for most of the samples encountered by organic analysts. Accordingly, the SRMs in this area represent either high-priority sample types or generic materials that should be widely applicable. Some SRMs useful for spiking or other types of standards preparation are also available.

A number of natural matrix SRMs certified for a large number of the inorganic constituents of environmental interest and some organic substances have been produced. These include several biological matrix samples (orchard leaves, now replaced by apple and peach leaves, was the first SRM in this group) and also urban particulate matter and river and marine sediments. Industrial hygiene analysis materials are a small but important subcategory. More of these SRMs will be made available as the demand for additional natural matrix standards is identified.

The SRMs for Physical Properties reflect many of the kinds of measurements made in testing laboratories. These SRMs run the gamut from those useful for the conventional physical measurements of temperature, melting point, vapor pressure, calorimetry, conductivity, and thermal expansion to those for color, thickness of electrodeposits, and fineness of powders. Radioactivity standards are also classified in this category.

Engineering standards are a small but growing group which is rather diverse. Standard rubbers, SRMs for evaluating the performance of magnetic tapes, and flammability standards are just a few examples.

The SRM Program tries to keep abreast of and even anticipate changes in technologies, since it usually takes several years to develop and certify a new SRM. Input from the SRM user communities, from contacts with professional colleagues, and NIST's own measurement experience are crucial in guiding and establishing priorities for SRM development. Furthermore, SRM sales provide guidance on inventory maintenance and priorities when decisions about SRM renewals must be made. While SRM Program contacts are extensive, additional input to the decision processes for new and renewal SRMs is always sought, especially in those areas where the importance of SRMs has yet to be fully appreciated. Only with such input can SRM development continue to meet new and changing measurement requirements as technologies advance and as new ones appear.

SRMs constitute the largest grouping, but there is another type of reference material also produced or distributed by NIST. Reference Materials (RMs) are in addition to and distinct from the SRMs issued by NIST. The distinction between a RM and a SRM is in the information supplied with each and the purpose for which each should be used. Unlike a SRM, the RM is not issued with a Certificate or Certificate of Analysis; rather it is accompanied by a "Report of Investigation," the sole authority being the author of the report. A RM is intended primarily to further scientific or technical research on that particular material, so one of the principal considerations in issuing a RM is material homogeneity. An investigator using a RM in one laboratory should be assured the material under investigation is the same as that being investigated in a different laboratory.

Some of the RMs distributed by NIST have been standardized by other organizations. These include U.S. Government agencies other than NIST, foreign national measurement laboratories such as the National Physical Laboratory of the United Kingdom, or standards-writing bodies such as the American Society for Testing and Materials (ASTM), the American National Standards Institute (ANSI), and the Organization for International Standardization (ISO). For many of these RMs, NIST acts only as a distribution point and does not participate in their evaluation. Table 4 lists a few examples of SRMs and RMs recently issued (or about to be issued) by NIST.

Table 4. Some Recent New and Renewal SRMs/RMs

SRM	
345a	Chromium-Nickel Steel
659	Silicon Nitride Particle Size Distribution
909a	Human Serum
1480	Polyurethane
1976	XRD Instrument Sensitivity
2261	Chlorinated Pesticides in Hexane
1804a	Toxic Volatile Organic Compounds (~5 ug/kg)
1846	Infant Formula, Milk Based (Avail. 1993)
2390	DNA Profiling
2520	Optical Fiber Geometry
2579	Lead Paint Film for XRF Calibration
2709 - 2711	Soils - Baseline (2709), Highly Contaminated (2710) and Moderately Contaminated (2711)

RM	
8449	Drugs of Abuse in Powdered Human Hair
8535-8558	Light Stable Isotopic Materials
8600-8608	Hubei Province China Ores

6.2 Selecting a SRM

The large number of SRMs available may cause some confusion as to which to choose for a given purpose. Obviously, matrix match is a major consideration since there is little or no difficulty in interpreting test results of such materials. Close match is only possible when recurring analysis of well-defined materials is of concern, i.e., large volume industrial or natural products. However, all of the SRMs have been developed as the result of widespread needs and much consideration has been given to providing matrices that are either typical or that can satisfy generic purposes.

When a matrix match is possible, analysts are advised to use such SRMs. In consideration of this, many users stock a relatively large number of these SRMs, thus encompassing the variety of materials that they expect to analyze. But for many users, a perfect matrix match will not be possible, hence, professional judgment will be required to select the ones most useful for each situation. The NIST SRMP technical staff and NIST technical division scientists have special experience in most of the measurement areas represented by the SRM inventory. The assistance of a SRM Program manager or NIST scientist to provide advice in the selection and use of appropriate SRMs for a given purpose, can be obtained by contacting the SRM Program office.

In cases where matrix-matched reference materials are not available, benefits may still be derived from use of a different kind of SRM. An increasingly wider range of instrument performance SRMs now make it possible to

test, at the very least, for such parameters as alignment, wavelength, absorbance, resolution, sensitivity, or magnification.

Readers of this handbook wishing to request development of a SRM that NIST does not already supply, should reproduce, fill out and send in the request form provided as Appendix D.5.

6.3 Use of SRMs

Any use of a reference material depends on the ability to make valid inferences from the measurement results. This involves the tacit assumption or demonstrated evidence that the material is reliable and capable of challenging the measurement process. Furthermore, the measurement process must be known to be in a state of statistical control, since limited measurements of the reference material will be used for predictive or evaluation purposes.

Because of the high reliability of the certified values, SRMs find a wide variety of uses ranging from special occasions when a material of known properties is needed to test some aspect of measurement, to the continual quality assurance of measurement systems. Table 5 is a summary of the most common kinds of applications.

Table 5. Uses of SRMs in Measurement Systems

Method Development and Evaluation

- Verification/evaluation of the precision and accuracy of test methods
- Development of reference test methods
- Evaluation of field methods
- Validation of methods for a specific use

Establishment of Measurement Traceability

- Development of secondary reference materials
- Development of traceability protocols
- Direct field use

Assurance of Measurement Compatibility

- Direct calibration of methods and instrumentation
- Internal (intralaboratory) quality assurance
- External (interlaboratory) quality assurance

Examples of one-time uses of SRMs as known test materials are numerous. Whenever an analytical method is developed or modified, a well-characterized test material is needed to evaluate its performance characteristics. SRMs, as appropriate, are obvious choices in such cases and numerous examples are cited in the literature. A survey conducted over an 18-month period, identified 40 research articles citing the use of SRMs in the development or evaluation of a wide variety of different methods for chemical analysis, and in particular, for trace analysis.[10] Likewise, performance checks of instrumentation, such as linearity, stability, and sensitivity are highly dependent on reliable test materials. Here again, a SRM is a logical choice for such a purpose.

When new methodology is adopted, a laboratory should perform preliminary measurements in order to gain proficiency in its application. The use of SRMs in this process will eliminate questions of stability and homogeneity that might otherwise complicate interpretation of the test results obtained with less well-characterized materials.

Some laboratories also use SRMs to confirm an analyst's capability to perform the procedure or method before undertaking measurements in their test programs.

When a contract laboratory provides analytical services to an individual or in a monitoring program, evidence of the capability to do so is often a requirement. The analysis of test samples provided by the client is one approach to evaluate the competence of candidates. SRMs are virtually unexcelled for this purpose. The only questions in such usage are the selection of an appropriate SRM and the possibility of falsification of data due to recognition of the test sample as a SRM with a known composition.

The use of SRMs for educational purposes should not be overlooked. Understanding of analytical chemistry is best demonstrated by practical laboratory work in which students analyze real samples. No better ones are available than SRMs which provide the opportunity to test both the precision and the accuracy of the analytical results.

SRMs find use as calibrants in some cases. For example, the practical pH scale is defined by NIST SRMs, and SRMs for fixed temperature points are available. The use of SRMs as primary chemical standards has already been discussed. Some industrial matrix SRMs, and particularly metals, are used for calibrating chemical analyzers. Many of the standard methods developed by ASTM Committee E-1 on Analytical Chemistry for Metals, Ores, and related Materials, can be calibrated directly with appropriate SRMs. For example, when ASTM Standard Method E 322 is used to analyze low alloy steels by X-ray fluorescence, the 1200 and 1700 series of low-alloy steel SRMs are recommended as calibrants.

One of the major uses of SRMs, and the original driving force behind SRM development, is for the quality assurance of measurement processes. When various analysts use different methodologies (and even the same methods), unacceptable discrepancies may arise, usually attributable to calibration or procedural differences. The analysis of available reference materials can identify such problems and lead to their resolution.

Figure 7 depicts a hierarchical system of measurement methods and reference materials. The function of each component (I through VI) is to transfer accuracy to the level immediately below it and help provide traceability to the level immediately above it, thus helping to assure overall measurement compatibility. Proceeding from the bottom to the top of the measurement hierarchy, accuracy requirements increase at the expense of decreased measurement efficiency. At the top are the so-called definitive methods of analysis or test, which give the most accurate values obtainable. Unfortunately, most definitive methods (e.g., gravimetric techniques for preparing analyzed gas SRMs) are usually time-consuming and sophisticated. Thus, they are not economically acceptable for widespread and routine use. Definitive methods, however, are used whenever possible to certify NIST SRMs. With these materials, accuracy can be transferred throughout a measurement system.

SRMs are commonly used in developing reference methods and assuring their accuracy. Such methods may be suitable for direct use. Alternatively, they may serve as a basis for developing or evaluating other methods. Reference methods are also commonly used for producing secondary reference materials, which, in turn, are directly used in routine field measurement applications.

In principle, the accuracy of numerous field methods can be traced to a definitive method in a hierarchical accuracy-based measurement system. SRMs and other reference materials are essential in the transfer of accuracy. Also essential are good methods, good laboratory practices, well qualified personnel, and adequate quality assurance procedures.

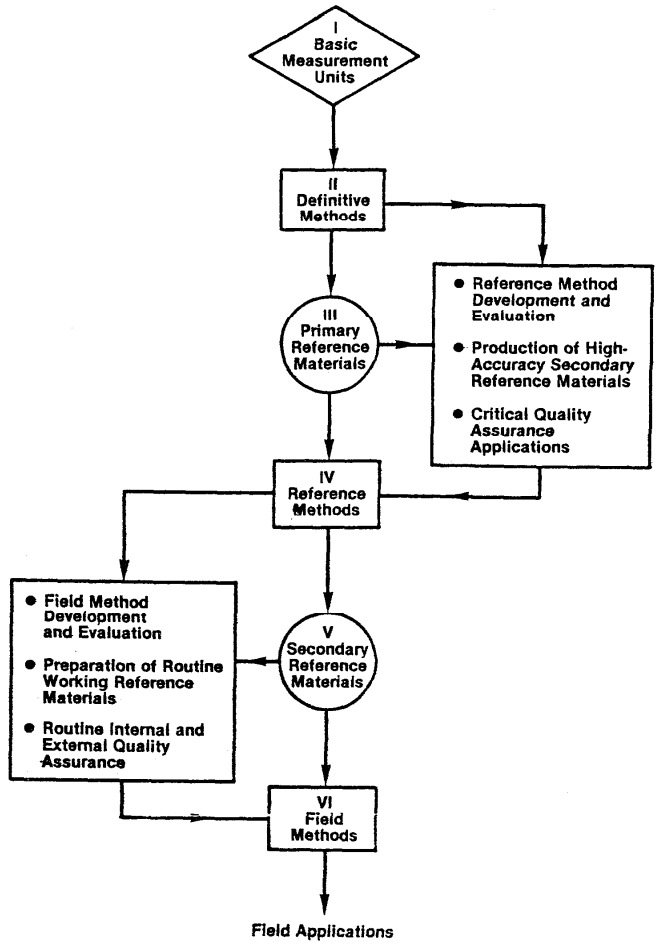


Figure 7. Systems Approach to Measurement Accuracy

The complexity of modern chemical analysis provides many sources of error and opportunities for introduction of bias and imprecision. Accordingly, such systems must be operated under a rigid quality assurance system if results are to be meaningful. It is not sufficient to check the calibration of instruments, although this is always necessary. Rather, the performance of the entire system needs to be monitored on a regular basis. SRMs are finding increasing use as test materials to monitor system performance.

SRMs are best used on a regular basis. The sporadic use of reference materials when trouble is suspected is a legitimate use, but systematic measurement in a control-chart mode of operation will generally be more informative and is highly recommended. SRMs may be used alone or they may be used with internal reference materials in a systematic manner, thus conserving the former and adding credence to the latter.

When using SRMs, one should carefully read all labels, certificates, and material safety data sheets - if included. These provide important information which bear on correct use of the material, such as storage conditions, lifetime of certification, useful life of standard solutions, minimum amount of material to be used and important precautions for the safety of workers handling the material.

The use of SRMs as quality assurance materials is discussed in more detail in the article contained in Appendix D.3.

6.4 Interpretation of Reference Material Analyses

Some SRMs have a matrix identity with test samples and can be used directly to establish the response function of chemical analyzers. Others may be used by a laboratory as their primary standards. However, the majority of the SRMs are quality assurance materials and should be analyzed regularly to monitor the performance of a measurement system.

The four general cases for use of SRMs as quality assurance materials are illustrated in Figure 8 (A-D). When a matrix match is possible (8A), the uncertainty in the sample measurements can be equated to that observed in measurement of the SRM. When such a match is not possible, but a SRM with a related matrix is available (8B), the test sample uncertainty may be related to those observed when measuring the SRM. Even when the above situations do not apply, the measurement of an appropriate SRM (8C) can monitor the measurement system and its performance when measuring test samples can be inferred in many cases. When a SRM is unavailable or not used, measurement uncertainty must be inferred from other evidence such as physical calibrations and the experience of others. Obviously, it is to the advantage of a laboratory to evaluate its own performance using SRMs, or other reliable reference materials, whenever possible.

The results obtained when analyzing reference materials should be interpreted with consideration given to the conditions of measurement. When measured consistently and utilizing control charts, reference materials can effectively monitor a measurement process. When measured in isolation, the results could be inconclusive or even misleading.

The inability to correctly analyze a reference material can cast serious doubts on the reliability of a measurement process but may not provide any diagnostic information. Also, the correct analysis of a SRM may not necessarily indicate the converse. Referral to Figure 9 will clarify this point. In this figure, measured values are plotted with respect to the expected or certified values. For an unbiased system, the data would be represented by line A.

Various kinds of linear measurement bias are illustrated in Figure 9. Line B corresponds to a constant bias (negative in this case, but it could be positive as well) while line C results from bias which is proportional to the concentration level of the sample. The proportionality factor could be less or greater than unity (shown). Line D results from a combination of constant and level-proportional bias. It is obvious that the measurement of one reference sample will not evaluate the performance of a measurement system throughout a concentration range unless supplemented by other information. One could even obtain a result, 2, and conclude a system was unbiased when the analysis of additional reference materials might indicate performance represented by line D. Occasionally, situations may occur where linear treatment of biases does not apply, but these are not described here.

When possible, the analysis of several reference samples, spanning the concentration range of interest, is the most useful way to investigate measurement bias. The three-sample approach, which involves the analysis of a low, middle, and upper range sample, is practical in most cases - as long as the reference samples are sufficiently homogeneous and the range of analytical interest is covered. Bias is even identifiable using relatively inhomogeneous samples, provided a sufficient number are analyzed, because it is highly unlikely that n randomly selected samples from a lot would all deviate in a systematic manner from a population mean value. However, n should be 5 or more in order to measure the kind of bias that may be present. In these cases, statistical advice may be necessary to plan the number of samples and the replicate measurements needed and to evaluate the test results.

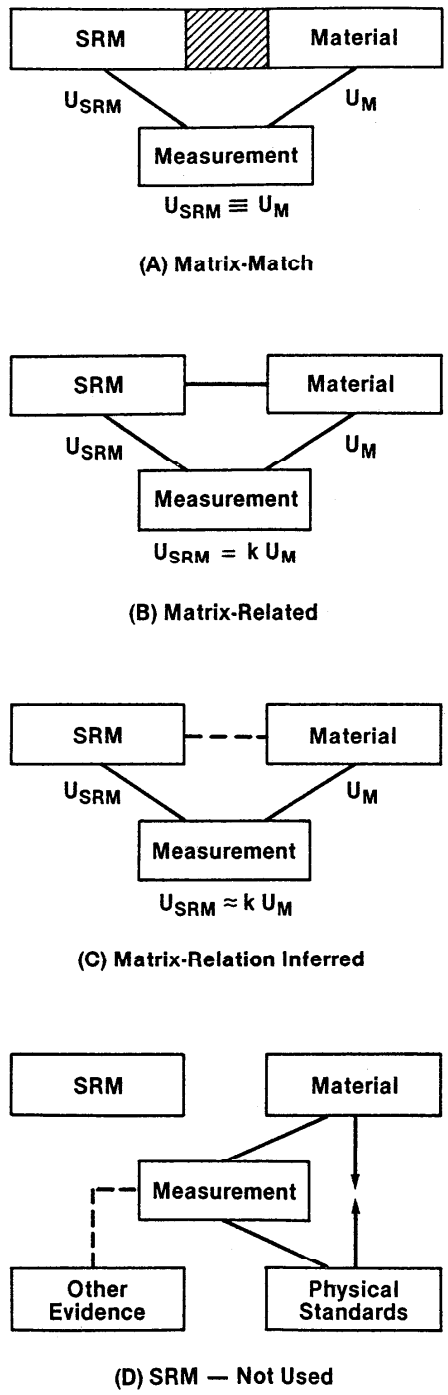


Figure 8. Interpreting SRM Measurements

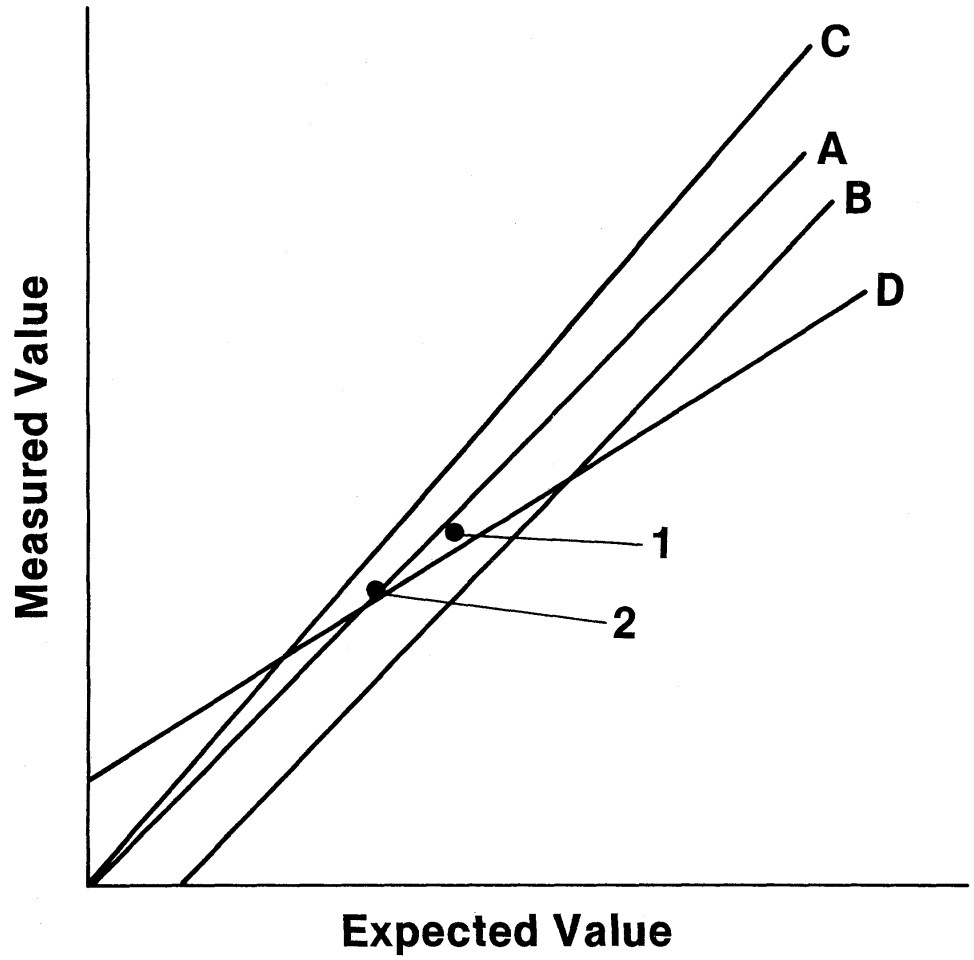


Figure 9. Identifying Measurement Bias

When supported by other data, the measurement of even a single reference sample can be meaningful. Thus, a knowledge of the standard deviation of measurement obtained from other data, would answer whether point (1) or point (2) in Figure 9, could be considered as represented by line A. Measurement of a series of nonreference material samples might provide some knowledge about the slope and hence assist in the interpretation of the SRM measurement data. The best diagnostic information would be obtained from the measurement of a series of SRMs containing graduated levels of analyte. When such are available, all the SRMs in the series should be used in order to maximize the information on the performance characteristics of an analytical system.

A SRM may have a reasonable matrix match with test samples, but differ from them in level of concentration. If the level of analyte in the SRM is higher than that of the test sample, it may be possible to quantitatively dilute the SRM. The best diluent is the matrix of the SRM, but a neutral matrix may be used in some cases. Also, two SRMs containing different levels of an analyte may be proportionally mixed to obtain a series of materials, ranging from the concentration level of the lower to that of the higher. These techniques are described in ASTM D-3975 Preparation of Samples for Collaborative Testing of Methods for Analysis of Sediments (Vol. 11.02). The expression used to calculate the composition of a blend of two samples, A and B, is as follows:

$$\alpha_{A+B} = \frac{\alpha_A W_A + \alpha_B W_B}{W_A + W_B}$$

where

α_A = Wt. % (or ppm) of constituent α in sample A
α_B = Wt. % (or ppm) of constituent α in sample B
α_{A+B} = Wt. % (or ppm) of constituent α in mixture
W_A = Wt. of sample A in mixture
W_B = Wt. of sample B in mixture

When a material is diluted with a second material containing an insignificant amount of the analyte of interest, the expression to be used is:

$$\alpha_{A+D} = \frac{\alpha_A W_A}{W_A + W_D}$$

where

W_D = Wt. of diluent sample mixed with W_A.

All dilutions must be made with care and, because uniform mixing could be difficult to achieve, the entire mixture prepared may have to be used at once rather than sub-sampled. But despite such problems, this technique is attractive since it can provide reference materials that simulate the test samples more closely, and evaluate a measurement process over a wider range of concentration levels.

6.5 Evaluation of Measurement Error

The values measured by a user for an analyte or a parameter will rarely agree fully with the certified value of a SRM due to uncertainties in each. The question naturally arises as to how large a difference is significant. This will depend on the uncertainty of the measurement by the user (see Figure 10) and the certification limits for the SRM. The former can be calculated, using the expression (see also Section C.5):

$$\bar{x} \pm \left(\frac{ts}{\sqrt{n}} + B\right)$$

where $\bar{x}$ is the mean of n measurements by the user whose estimated standard deviation of measurement is s. The student's t value will depend on the number of degrees of freedom in the estimation of s (n-1 if s is based on independent measurements) and the confidence level (a 95 % confidence level is the usual level for certified values of an SRM). The value, B, is the user's estimate of the magnitude of any uncorrected biases inherent in his measurement and is based on experience and professional judgment.

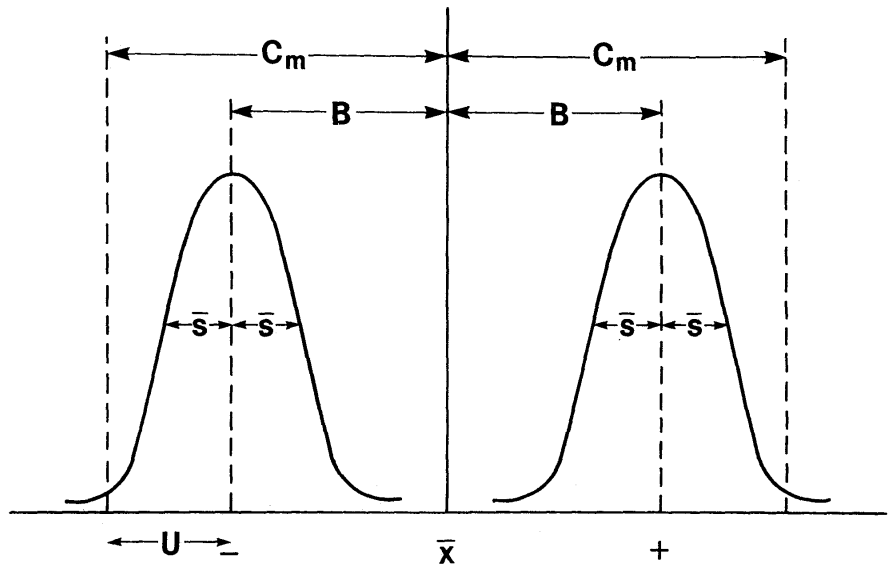


Figure 10. Uncertainty of Measured Value

C_m	=	Uncertainty of Measured Value
B	=	Biases, Errors Inherent in Measurement
s	=	Precision of Measurement
$s(\bar{x})$	=	Precision of Mean of n Measurements

$$C_m = \frac{ts}{\sqrt{n}} + B$$

If the confidence interval intersects the confidence or tolerance interval of the SRM, disagreement is not shown. If not, then a discrepancy exists which should be investigated. In the case of heterogeneous SRMs, several subsamples may need to be measured to evaluate measurement bias.

If an apparent discrepancy is found, it is advisable to look closer at the estimates of uncertainty. Rarely will a user's uncertainty, $(ts/\sqrt{n})$, be less than that of the NIST measurements, which are done with state-of-the-art techniques, so a discrepancy could signal unsuspected biases in the user's laboratory which the SRM has uncovered. If an explanation cannot be found, the user should communicate with the NIST SRM Program which will look into the matter and advise the user as to whether the sample may have deteriorated or become contaminated.

It cannot be emphasized too strongly that statistical control must be attained before any data can be believed and any errors identified or corrected. There is no easy way to identify assignable causes for either unacceptable bias or imprecision, but this must be the first step for corrective actions. The advice of experienced users of the methodology employed should be sought as they could suggest approaches to follow, if not the specific solutions.

Also, the magnitude of the errors encountered may rule out certain sources and indicate likely ones. However, the simultaneous existence of several sources of unacceptable error cannot be discounted.

In diagnosing error, it should be remembered that random errors add up in quadrature, which is to say the variance of random errors is additive, as discussed in 2.1.1. When the measurement system is well-understood, it may be possible to estimate the variance of the individual steps or operations of which it is composed and to compare them with the magnitude of the errors of concern. Obviously, the step or operation with the largest variance is the first one to be considered when other information is not available.

On the other hand, there is no reason to believe that biases are randomly distributed; rather, they add up algebraically. Furthermore, small systematic errors contribute differently than small random errors.

Whenever excessive bias or imprecision is found to be present, corrective action must be taken, or else the measurement process will have limited usefulness. In this regard, the first question that needs to be answered is whether the unsatisfactory situation is inherent to the methodology or is due to its applications in a given laboratory or even by specific persons. Collaborative test data and/or the research findings of others may supply partial answers to the methodology question. If the experience of a laboratory is inconsistent with the experience of others, misapplication of the measurement process should be suspected as the source of excessive imprecision or bias.

To reduce operational (nonstate-of-the-art) bias, the following improvements must be made: 1) better quality of calibrants, 2) improvement in calibration, 3) reduction of contamination, 4) reduction of mechanical losses, 5) reduction of solution/extraction inefficiencies and,6) removal of interferences.

To reduce random errors, the following improvements are necessary: a) improvement of technical skills, b) improvement of manipulative skills, c) better environmental control, d) closer tolerances in operational parameters, e) improved instrumentation and, f) reduction in variability of blanks.

6.6 Is a SRM Expensive?

During 1991, the average SRM selling price was $323. While this may seem expensive, the alternatives are even more costly. One alternative is to do without a SRM. The result, however, could be less accurate measurements, reduced quality and even greater exposure to liability costs. Another alternative is to prepare an "in-house" standard. Those who have tried know how expensive this alternative can be - especially when no connection to national standards is achieved.

Good measurements are costly. Bad measurements are even more costly. Fact or truism, this means the price of a SRM can be quite reasonable when the benefits derived from it are considered.

Considering the total cost of installing and operating a laboratory (including such costs as capital equipment, rent, labor, training, supplies, utilities, and SRMs) the SRM component is typically less than 1 % of the total. Yet, that small component provides the powerful service of relating the laboratory to benchmarks of national measurement.

Over the past 20 years, Federal Government policymakers have taken the position that SRM users capture benefits from the materials and, therefore, must pay all the costs of producing and distributing SRMs. Today 100% of such costs are paid by the users.

Producing each SRM type is a custom project, thus the market is very limited. In 1991 for example, the average number of chemical SRMs purchased by individual customers was about 50 per each SRM type. Even when producing 10-year supplies (if shelf-life considerations permit such an economy), the SRM batch sizes are usually in the range of 200 to 1,000 units - certainly not comparable to mass-produced goods.

7. REPORTING ANALYTICAL DATA

7.1 Limits of Uncertainty for Data

Data are of limited or no use unless limits of uncertainty can be estimated and assigned to them [1]. Included should be the limits of uncertainty of any systematic error (bias) and the random errors of measurements. Estimates of limits of systematic error are based on judgment and require a full understanding of the measurement system used. The experience of the analytical community is helpful in this respect. The random error component is based on the skill and experience of the laboratory making the measurements and is evaluated from the standard deviation of the measurement process.

When the measurement process is demonstrated to be in a state of statistical control, the process standard deviation may be used to evaluate the confidence interval for the mean of n measurements (see Appendix C.5). The use of control charts is the best way to demonstrate that a process is in a state of statistical control at the time the measurements are made and this can minimize the amount of work needed to establish confidence limits for the data.

In the absence of a control chart, a sufficient number of replicate measurements must be made to demonstrate statistical control and to estimate the standard deviation of measurement with a measurable degree of confidence. The minimum number of replicate measurements required is somewhat arbitrary and depends on the risk concerned with exceeding the limits of confidence that are stated. Metrologists often recommend 7 to 30 determinations as a reasonable number of replicates. The uncertainty of the standard deviation increases rapidly below 7 and little is to be gained by increasing the number above 30.

It is not considered good practice to correct for biases without understanding their origins. Biases such as analytical blanks should be measured as accurately as necessary and the results corrected directly for them [18]. This is proper because analytical chemists believe that blanks are additive errors. Biases such as those found when measuring SRMs, are investigated to identify their sources so that they can be eliminated, minimized, or corrected in a proper manner. As already pointed out (see Section 6.4), the method of correction will be dictated by the nature of the bias. The most reliable approach is to remove the bias rather than to correct for it.

The treatment of bias related to the question of recovery often troubles the trace analyst. Usually corrections are not made, but recoveries are reported as one of the qualifications for the data. The recovery, no matter what its value, should be shown to be in a state of statistical control, and this should be a requirement for reporting data of this kind. When a recovery determination is made and the value is variable or does not fall within control limits, corrective action is indicated and control should be reestablished before data can be reported.

7.2 Significant Figures

Numerical data are often obtained (or at least calculations can be made) with more digits than are justified by their accuracy or precision. In order not to be misleading, such data should be rounded to the number of figures consistent with the confidence that can be placed in them (see Section 7.3) when reported in final form. However, as many digits should be retained at intermediate stages. Since computers are so pervasive, this presents little burden to the scientist. Accordingly, metrologists have adopted the terminology of significant figures in describing the resulting data. The number of significant figures is said to be the number of digits remaining when the data are rounded; that is, the last digit or, at most, the last two digits, are expected to be the only ones that would be subject to change on further experimentation. Thus, for a measured value of 20.5, only the 5 and at most the 0.5, would be expected to be subject to change. Such a value would be described as having three significant figures.

In counting significant figures, any zeros used to locate a decimal point are not considered as significant. Thus, 0.0025 contains only two significant figures. Any zeros to the right of the digits are considered as significant, so only those that have significance should be retained. Thus 2500 and 2501 each have four significant figures. Zeros

should not be added to the right of significant digits to define their magnitude unless they (the zeros) are truly significant; otherwise, this will confuse the significance of the value. For example, it is not good practice to report a value as 2500 μg but rather 2.5 mg if the data are reliable to two significant figures. The use of exponential notation, e.g., 3.5×10^3, is an acceptable way to report data with two significant figures which would otherwise have to be reported as 3500, suggesting 4 significant figures.

In multiplication and division, the operation with the least number of significant figures determines the numbers to be reported in the result. For example, the product $1256 \times 12.2 = 15323.2$ is reported as 1.53×10^4. In addition and subtraction, the least number of figures to either the right or the left of the decimal point determines the number of significant figures to be reported. Thus, the sum of $120.05 + 10.1 + 56.323 = 156.473$ is reported as 156.5 because 10.1 defines the reporting level. In complex calculations involving multiplications and additions, for example, the operation is done serially, and the final result is rounded according to the least number of significant figures involved. Thus $(1256 \times 12.2) + 125 = 1.53 \times 10^4 + 125 = 1.54 \times 10^4$.

The following rules should be used in rounding data, consistent with their significance:

1. When the digit next beyond the one to be retained is less than five, the retained figure is kept unchanged. For example: 2.541 becomes 2.5 to two significant figures.

2. When the digit next beyond the one to be retained is greater than five, the retained figure is increased by one. For example: 2.453 becomes 2.5 to two significant figures.

3. When the digit next beyond the one to be retained is exactly five and the retained digit is even, it is left unchanged. But when the digit next beyond the one to be retained is exactly five and the retained digit is odd, it is increased by one. Thus, 3.450 becomes 3.4, but 3.550 becomes 3.6 to two significant figures.

4. When two or more figures are to the right of the last figure to be retained, they are to be considered as a group in rounding decisions. Thus in 2.4(501), the group (501) is considered to be >5 while for 2.5(499), (499) is considered to be <5.

7.3 Guidelines for Reporting Results of Measurements

The number of significant figures to be used in reporting results is often asked. This will depend on the number of figures in the original data and the confidence limits to support the results. Analysts sometimes feel that observed data have more digits than are meaningful and are tempted to round them to what is felt to be significant. This should be resisted and rounding operations should be deferred to the last. The following discussion provides guidelines for deciding what is significant.

The number of figures to retain in experimental data and even in preliminary calculations is unimportant, provided a certain minimum is exceeded. At the least, there should be variability between successive trials of the last figure and preferably, between the last two figures. If this is not the case, the data are probably being truncated by the operation (e.g., low attenuation), or rounded off by the observer (analyst), or imprecisely read. Through training, the precision of reading can often be improved. But when observers have preconceived ideas about the precision attainable (or that required for some application), they could be rounding off readings prematurely and thus actually throwing away data.

The average of several values should be calculated with at least one more significant figure than that of the data. This will then be rounded for reporting, consistent with the confidence limits estimated.

The standard deviation (necessary for computing confidence intervals) should be calculated to three significant figures and rounded to two when reported as data. The confidence interval also should be calculated, then rounded to two significant figures (use more figures in the calculations as available) before the result is reported. The 95 % level is the most commonly referenced, but any confidence level can be used and must be reported as part of the calculation.

As an example, the following data were observed:

15.2, 14.7, 15.1, 15.0, 15.3, 15.2, 14.9

where $\bar{x} = 15.057$ and $s = 0.207$

$$\frac{ts}{\sqrt{n}} = \frac{2.517 \times 0.207}{\sqrt{7}} = 0.1969$$

$$\bar{x} = 15.06 \pm 0.20$$

The uncertainty of the result represents the 95 % confidence interval for the mean of seven measurements.

8. NIST SERVICES RELATED TO SRMs

The NIST Standard Reference Materials Program and the various NIST scientists engaged in SRM development and certification believe that SRMs used in a systematic manner can provide a high level of confidence in analytical measurement data. To that end, the Program office offers its assistance to anyone needing information about SRMs and their applications. For most purposes, the advice given is fairly generic in nature, but if any questions regarding specialized measurement problems are received, they usually can be referred to an appropriate SRM project manager or NIST technical division for response. The following publications and notices are used to provide SRM information:

NIST SP 260 Standard Reference Materials Catalog
- issued every 2 years,

NIST SP 260 Standard Reference Materials Price List
- issued 1-2 times per year,

Brochures describing certain SRM types or applications
- issued every 2-4 years,

NIST SP 260-XXX technical publications on specific SRMs
- issued throughout the year (see list in this SP),

Notices of Availability announcing new and renewal SRMs
- issued each time a SRM is released to the public,

Newsletters updating/announcing SRMs and upcoming exhibits
- issued 3-4 times per year,

National technical meeting exhibit material (PITTCON, etc.)
- distributed at 10-14 meetings per year.

Requests for any of the above publications, for SRM price quotations and availability, or for placement of SRM orders should be directed to the following:

National Institute of Standards and Technology
Standard Reference Materials Program
Building 202, Room 204
Gaithersburg, MD 20899-0001
Telephone: (301) 975-6776
FAX: (301) 948-3730

Requests for replacement of lost certificates and answers to questions concerning SRM applications, experimental results when measuring specific SRMs, selection of suitable SRMs, details of certification and certified values, and related technical matters should be directed to the following:

National Institute of Standards and Technology
SRM Production and Certification Programs
Building 202, Room 215
Gaithersburg, MD 20899-0001
Telephone: (301) 975-2019
FAX: (301) 926-4751

NIST also provides physical measurement services in the areas of Calibration, Special Tests, and Measurement Assurance Programs. These services are described in Special Publication 250 NIST Calibration Services Users Guide. Requests for copies of the guide and current test fees schedule should be directed to the following:

National Institute of Standards and Technology
NIST Calibration Program
Building 411, Room A104
Gaithersburg, MD 20899-0001
Telephone: (301) 975-2002
FAX: (301) 926-2884

9. BIBLIOGRAPHY

9.1 References

1. Taylor, J.K., Principles of Quality Assurance, NBSIR 85-3105 (1985), National Bureau of Standards, Gaithersburg, MD 20899.

2. Brief History and Use of the English and Metric System of Measurement. NBS SP 304A (1968).

3. Kolthoff, I.M., et al., Quantitative Chemical Analysis, 4th Edition, p. 687, Macmillian (1969).

4. The Analyst, 90, 251 (1965).

5. ASTM E617, Specification for Laboratory Weights and Precision Mass Standards, 1981(1985), ASTM, Philadelphia, PA 19103.

6. Taylor, J.K., Quality Assurance of Chemical Measurements, Anal. Chem., 53:1588A-1596A (1981).

7. Eisenhart, C., Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems, Ref. 9, pp. 21-47.

8. ASTM Committee E-11, ASTM Manual on Presentation of Data and Control Chart Analysis, MVL7 (formerly STP 15), ASTM, Philadelphia, PA 19103, 6th edition.

9. Taylor, J.K., Reference Materials - What They Are and How They Should Be Used. J. Testing and Evaluation, 11:385-86 (1983).

10. Uriano, G.A. and Gravatt, C.C. The Role of Reference Materials and Reference Methods in Chemical Analysis. CRC Critical Reviews in Analytical Chemistry, October 1977.

11. Cali, J.P. et al., The Role of Standard Reference Materials in Measurement Systems, National Bureau of Standards (U.S.) Monogr. 148; 1975.

12. Cali, J.P. and Stanley, C.L. Measurement Compatibility and Standard Reference Materials, Annual Review of Materials Science, 5:329-342, (1975).

13. ISO Guide 31 - Contents of Certificates of Reference Materials; 1981. Available from ANSI, 11 West 42nd Street, New York, NY 10036.

14. ISO Guide 30 - Terms and Definitions Used in Connection with Reference Materials; 1981. Available from ANSI, 11 West 42nd Street, New York, NY 10036.

15. ISO Directory of Certified Reference Materials: 1982. Available from ANSI, 11 West 42nd Street, New York, NY 10036.

16. Ku, H.: Editor. Precision Measurement and Calibration Statistical Concepts and Procedures, National Bureau of Standards (U.S.) Spec. Publ. 300, Vol. 1; 1969.

17. Hughes, E.E. and Mandel, J., A Procedure for Establishing Traceability of Gas Mixtures to Certain NBS SRMs, NBSIR 81-2777; 1981.

18. Taylor, J.K., Guidelines for Evaluation of the Blank Correction, J. Testing and Evaluation 12, 54 (1984).

19. Natrella, M.G., Experimental Statistics, National Bureau of Standards (U.S.) Handbook 91; 1963.

9.2 Publications Related to SRMs

Alvarez, R., Rasberry, S.D., and Uriano, G.A, NBS Standard Reference Materials. Update 1982; Anal. Chem. 54:1226A (1982).

Alvarez, R., Seward, R.W., and Rasberry, S.D. Clinical Standard Reference Materials from NBS; Apr 19 (1983).

ASTM E305, Recommended Practices for Establishing and Controlling Spectrochemical Analytical Curves, ASTM, Philadelphia, PA 19103.

ASTM E826, Practice for Testing Homogeneity of Materials for the Development of Reference Materials, 1985 (1990), ASTM, Philadelphia, PA 19103.

Bowers, G.N., Jr., McComb, R.B., Christensen, R.G., and Schaffer, R., High Purity 4-nitrophenol: Purification, Characterization, and Specifications for Use as a Spectrophotometric Reference Material. Clin. Chem. 26(6), 724 (1980), SRM 938.

Chesler, S.N., Guenther, F.R., May, W.E., and Parris, R.M., SRMs for Accurate Analyses of the PCB's in Oil. ASTM Standardization News 10(6), 15 (1982), SRM 1581.

Christensen, R.G., White V.E., Meiselman, S., and Hertz, H.S., Quantitative Trace Analysis by Reversed-Phase Liquid Chromatography Mass Spectrometry, J. Chromatogr. 271, 61 (1983), SRM 1599.

Cohen, A., Hertz, H.S., Mandel, J., Paule, R.C., Schaffer, R., Sniegoski, L.T., Sun, T., Welch, M.J., and White, V.E., Total Serum Cholesterol by Isotope Dilution/Mass Spectrometry: A Candidate Definitive Method. Clin. Chem. 26(7), 854 (1980), SRM 911a.

Coxon, B. and Schaffer, R., Characterization and Quantitative Analysis of D-Glucose for Use in Clinical Analysis, Anal. Chem. 43, 1565 (1971), SRM 917.

Coxon, B. and Schaffer, R., Analysis of Cortisol Clinical Standard by Fourier Transform, Proton Magnetic Resonance Spectroscopy, Anal. Lett. 6(4), 291 (1973).

Doumas, B.T., Bayse, D.D., Carter, R.J., Peters, T. Jr., and Schaffer, R.A., Reference Method for the Determination of Total Serum Protein, I. Development and Validation: Clin. Chem. 27, 1642 (1981), SRM 927.

Doumas, B.T., Bayse, D.D., Bauer, K., Carter, R.J., Elevitch, F., Garber, C.C., Gradby, R.A., Hause, L.L., Mathes, A., Peters, Jr., T., Rand, R. N., Reeder, D.J., Russell, S.M., Schaffer, R., and Westgard, J.O., A Reference Method for the Determination of Total Protein, II. Test for Transferability. Clin. Chem. 27(10), 1651 (1981), SRM 927.

Duncan, P., Goshman, N., Bayse, D.D., and Schaffer, R., et al., A Candidate Reference Method for Uric Acid in Serum, II. Inter-laboratory Testing. Clin. Chem. 28, 291 (1982), SRM 913.

Epstein, M.S., et al., Application of Laser-Excited Atomic Fluorescence Spectrometry to the Determination of Iron, Spectrochem. Acta, 35B, 233 (1980), SRMs 394, 1633, 1643 (See also Applied Spectroscopy 34, 372 (1980)).

Fatiadi, A.J. and Schaffer, R., An Improved Procedure for Synthesis of DL-4-Hydroxy-3-Methoxymandelic Acid (DL-"Vanillyl"-Mandelic Acid, VMA), J. Res. Nat. Bur. Stand. (U.S.) 78A(3), 136 (1974), SRM 925.

Gladney, E.S., Compilation of Elemental Concentration Data for NBS Biological and Environmental SRMs, LA-8438-A-MS, UC-4, July 1980, Los Alamos National Laboratory.

Gramlich, J.W., Machlan, L.A., Brletic, K.A., and Kelly, W.R., Thermal Ionization Isotope Dilution Mass Spectrometry as a Definitive Method for Determination of Potassium in Serum, Clin. Chem. 28, 1309 (1982), SRMs 909, 985.

Greenberg, R.R. and Kingston, H.M., Trace Element Analysis of Natural Water Samples by Neutron Activation Analysis with Chelating Resin, Anal. Chem. 55(7), 1160 (1983), SRM 1643a.

Guenther, F.R., Parris, R.M., Chesler, S.N., and Hilpert, L.R., Determination of Phenolic Compounds in Alternate Fuel Matrices, J. Chromatogr. 207, 256 (1981), SRM 1580.

Howell, B.F., Margolis, S.A., and Schaffer, R., Detection of Inhibitors in Reduced Nicotinamide Adenine Dinucleotide by Kinetic Methods, Clin. Chem. 22, 1648 (1976), SRM for NADH.

Howell, B.F., Margolis, S.A., and Schaffer, R., Residual Fluorescence as an Index of Purity of Reduced Nicotinamide Adenine Dinucleotide, Clin. Chem. 19(11), 1280 (1973), SRM for NADH.

Hughes, E.E., Role of the National Bureau of Standards in Calibration Problems Associated with Air Pollution Measurements, in Calibration in Air Monitoring, ASTM STP, 598; 1976. ASTM, Philadelphia, PA 19103.

Kingston, H.M. and Pella, P.A., Preconcentration of Trace Metals in Environmental and Biological Samples by Cation Exchange Resin Filters for X-ray Spectrometry, Anal. Chem. 53, 223 (1981), SRM 1632.

Kline, W.R., Enagonio, D., Reeder, D.J., and May, W.E., Liquid Chromatographic Determination of Valproic Acid in Human Serum., J. Liq. Chromatogr. 5(9), 1697 (1982), SRM 1599.

Kingston, H.M. and Pella, P.A., Preconcentration of Trace Metals in Environmental and Biological Samples by Cation Exchange Resin Filters for X-ray Spectrometry, Anal. Chem. 53, 223 (1981), SRM 1632.

Kline, W.R., Enagonio, D., Reeder, D.J., and May, W.E., Liquid Chromatographic Determination of Valproic Acid in Human Serum, J. Liq. Chromatogr. 5(9), 1697 (1982), SRM 1599.

Kratochvil, B.G. and Taylor, J.K., Sampling for Chemical Analysis, Anal. Chem. 53, 924A (1981).

Kuehner, E.C. and Pella, P.A., Energy-Dispersive X-ray Spectrometric Analysis of NBS SRM 1571, Orchard Leaves After Oxidation and Borate Fusion. Applied Spectroscopy 33, 632 (1979).

Koch, W.F., Sample Preparation in Ion-Chromatography, J. Res. Nat. Bur. Stand. (U.S.). 84, 241 (1979).

Lindstrom, R.M., et al., Neutron Capture Prompt γ-Ray Activation Analysis for Multielement Determination in Complex Samples, Anal. Chem. 51(13), 2209 (1979).

Maienthal, E.J., Application of Linear Sweep Voltammetry to the Determination of Trace Elements in Biological and Environmental Materials, Trace Substances for Environmental Health VIII 243-6 (1974), SRM 1571.

Maienthal, E.J., Analysis of Botanical Standard Reference Materials by Cathode Ray Polarography, J.A.O.A.C. 55, 1109 (1972), SRM 1571.

Margolis, S.A., Howell, B.F., and Schaffer, R., The Purification and Analysis of the Purity of Reduced Nicotinamide Adenine Dinucleotide, Clin. Chem. 22, 1322 (1976), SRM for NADH.

Margolis, S.A., Howell, B.F., and Schaffer, R., Lactate Dehydrogenase Inhibitors in NADH Preparations, Clin. Chem. 23, 1581 (1977), SRM for NADH.

Margolis, S.A. and Konash, P.J., Non-Peptide Impurities in Angiotensin I and Other Commercial Peptides, J. High Res. Chromatogr. and Chromatogr. 3, 317 (1980), SRM 998.

Margolis, S.A. and Konash, P.L., The HPLC Analysis of Diasteromers and Structural Analogs of Angiotensins I and II, Anal. Biochem. In Press. SRM 998.

Margolis, S.A. and Schaffer, R., Reversed-Phase Liquid Chromatographic Analysis of the Impurities of Nicotinamide Adenine Dinucleotides, J. Liq. Chromatogr. 2, 837 (1979), SRM for NADH.

Margolis, S.A. and Schaffer, R., Development of a Standard Reference Material for Angiotensin I. Final Report on Agreement #1-YO1-HU-70042-06, 1979, SRM 998.

Margolis, S.A., Yap, W.T., Matthews, B., and Schaffer, R., Reversed-Phase Liquid Chromatographic Analysis of the Acid-Catalyzed Rearrangement of NADH: Separation of New Products and an Analysis of the First Two Reactions of the Rearrangements, J. Liq. Chromatogr., 1, 669 (1978), SRM for NADH.

Marinenko, G., Koch, W.F., and Etz, E.S., High Precision Coulometric Titration of Uranium, J. Res. NBS 88, 117 (1982), SRM 950.

Mavrodineanu, R., Editor, Procedures Used at the National Bureau of Standards to Determine Selected Trace Elements in Biological and Botanical Materials, National Bureau of Standards (U.S.) Spec. Publ. 492, 1977.

Mavrodineanu, R., Shultz, J.I., and Menis, O., Accuracy in Spectrophotometry and Luminescence Measurements, National Bureau of Standards (U.S.) Spec. Publ. 378; 1973.

May, W.E., Brown J.M., Chesler, S.N., Guenther, R.F., Hilpert, L.R., Hertz, H.S., and Wise, S.A., Development of an Aqueous Polynuclear Aromatic Hydrocarbon Standard Reference Material. Polynuclear Aromatic Hydrocarbons, P.W. Jones and P. Leber, Eds. Ann Arbor Science Publishers, Inc., Ann Arbor, MI, (1979), SRM 1644.

May, W.E., Brown-Thomas, J.M., Hilpert, L.R., and Wise, S.A., Certification of Selected Polynuclear Aromatic Hydrocarbons in SRM 1580, Organics in Shale Oil, Anal. Methods and Biol. Fate of Polynuclear Aromatic Hydrocarbons, M. Cooke, and A.J. Dennis, Eds. Battelle Press, Columbus, OH, 1-16; (1981), SRM 1580.

May, W.E., Chesler, S.N., Hertz, H.S., and Wise, S.A., Analytical Standards and Methods for the Determination of Polycyclic Aromatic Hydrocarbons in Environmental Samples, Intern. J. Environ. Anal. Chem. 12, 259 (1982), SRMs 1580, 1644, 1647, and 1649.

Orvini, E., Gills, T.E., and LaFleur, P.D., Method for the Determination of Selenium, Arsenic, Zinc, Cadmium, and Mercury in Environmental Matrices by Neutron Activation Analysis, Anal. Chem. 46, 1294 (1974), SRMs 1571, 1577, 1630, 1632, and 1633.

Pella, P.A., Kingston, H.M., Sieber, J.R., and Feng, L-Y., Effect of Sample Dissolution Procedures on X-ray Spectrometric Analysis of Biological Materials, Anal. Chem. 55, 1193 (1983), SRMs 1575 and 1577.

Pella, P.A., Lorber, K.E., and Heinrich, K.F.J., Energy-Dispersive X-ray Spectrometric Analysis of Environmental Samples after Borate Fusion, Anal. Chem. 50, 1268 (1978), SRMs 1633 and 1648.

Pella, P.A. and Sieber, J.R., Intercomparisons of Selected Semi-Empirical and Fundamental Parameter Interelement Correction Methods in X-ray Spectrometry, X-ray Spectrometry 11, 167 (1982), SRM 1633.

Perry, B.W. and Doumas, B.T. (Chairman), Bayse, D.D., Butler, T., Cohen, A., Fellows, W., Garber, C.C., Howell, B., Koch, T., Krishnamurthy, S., Louderback, A., McComb, R.B., Miller, D., Miller, R.R., Rand, R. N., and Schaffer, R., A Candidate Reference Method for Determination of Bilirubin in Serum. Test for Transferability, Clin. Chem. 29, 297 (1983), SRM 916.

Rains, T.C., Rush, T.A., and Butler, T.A., Innovations in Atomic Absorption Spectrophotometry with Electrothermal Atomization for Determinating Lead in Foods, J.A.O.A.C. 65, 994 (1982), SRM 1566.

Rasberry, S.D., The Role of NBS/SRMs in Quality Assurance, 1983 ASQC Quality Congress Transactions, pp. 343-8.

Rasberry, S.D. and Heinrich, K.F.J., Calibration for Interelement Effects in X-ray Fluorescence Analysis, Anal. Chem. 46, 81 (1974), SRM 1188.

Reeder, D.J., Enagonio, D.P., Christensen, R.G., and Schaffer, R., Development of a Standard Reference Material for Antiepilepsy Drugs in Serum, Trace Org. Anal.: A New Frontier in Anal. Chem. National Bureau of Standards (U.S.) Spec. Publ. 519, 1979, 447-453, SRM 900.

Rook, H.L., Gills, T.E., and LaFleur, P.D., Method for the Determination of Mercury in Biological Materials by Neutron Activation Analysis, Anal. Chem. 44, 1114 (1972), SRM 1571.

Schaffer, R., Editor, Organic Chemistry Section: Summary of Activities, July 1967 to June 1968, National Bureau of Standards (U.S.) Tech. Note 457; 1968, SRMs 911a, 912a, 913, and 914.

Schaffer, R., Editor, Organic Chemistry Section: Summary of Activities, July 1968 to June 1969, National Bureau of Standards (U.S.) Tech. Note 507, 1969, SRMs 911a, 912a, 913, 914, 916, and 917.

Schaffer, R., Editor, Organic Chemistry Section: Summary of Activities, July 1969 to June 1970, National Bureau of Standards (U.S.) Tech. Note 547, 1970, SRMs 911a, 912a, 916, 917, 920, 921, and 925.

Schaffer, R., Margolis, S.A., and Longenbach, P.J., Consideration in the Preparation and Certification of Pure Analyte Reference Materials, Proc. of the 1979 Tenovus Workshop on Quality Control in Clin. Endoc., D.W. Wilson, S.J. Gaskell, and K.W. Kemp, Eds., Alpha Omega Publishing, Cardiff, Wales 37-42 (1979).

Schaffer, R., Sniegoski, L.T., Welch, M.J., White, V.E., Cohen, A., Hertz, H.S., Mandel, J., Paule, R.C., Svensson, L., Bjorkhem, I., and Blomstrand, R., Comparison of Two Isotope Dilution/Mass Spectrometric Methods for the Comparison of Total Cholesterol, Clin. Chem. 28(1), 5 (1982), SRM 911a.

Schaffer, R., Velapoldi, R.A., Paule, R.C., Mandel, J., Bowers, G.N., Jr., Copeland, B.E., Rodgerson, D.O., and White, J.C., A Multilaboratory-Evaluated Reference Method for Determination of Serum Sodium, Clin. Chem. 27, 1842 (1981), SRM 919.

Schaffer, R., White, V.E., Sniegoski, L.T., and Cohen, A., A Candidate Definitive Method for the Determination of Total Cholesterol in Serum, Trace Org. Anal.: A New Frontier in Anal. Chem., Nat. Bur. Stand. (U.S.) Spec. Publ. 519, (1979) 405-410, SRM 911a.

Schwartz, M.K., Blatt, S., Blumenfeld, T.D., Fleinsker, M., Free, A.H., Goechman, N., Hainline, A., Jr., Laessig, R.H., Nino, H.V., Richards, A.H., Schaffer, R., and Vanderlinde, R.F., Clinical Chemistry in Quality Assurance Practices for Health Laboratories, S. Inhorn, M.D., Ed., American Public Health Assoc. Washington, DC, 451-508, 1978.

Seward, R.W., Editor, Standard Reference Materials and Meaningful Measurements, Nat. Bur. Stand. (U.S.) Spec. Publ. 408.

Small, J.A., Production and Characterization of Glass Fibers and Spheres for Microanalysis, Scanning Electron Microscopy 1, 445 (1978).

Small, J.A. and Steel, E.B., An Overview of the NBS-EPA Program for Development of Standards for Use in the Analysis of Environmental Samples for Asbestos by Electron Microscopy, Proc. Third Intern. Colloq. on Dust Meas. Techniques, Cannes, France 1980, p. 256.

Small, E.B., Newbury, D.E., and Pella, P.A., Preparation of Thin-Film Glass Standards for Analytical Electron Microscopy, Anal. Elect. Microscopy 1981, 65-70 (San Francisco Press, 1981).

Velapoldi, R.A., White, P.A., May, W.E., and Eberhardt, K., Spectrofluorometric Determination of Polycyclic Aromatic Hydrocarbons in Aqueous Effluents from Generator Columns, Anal. Chem. 55, 1896-1901, SRM 1644.

Welch, M.J., Cohen, A., Hertz, H.S., Ruegg, F.C., Schaffer, R., Sniegoski, L.T., and White, V.E., An Isotope Dilution Mass Spectrometric Method for the Accurate Determination of Serum Urea, Anal. Chem., Submitted for Publication, SRM 912a.

White, V.E., Welch, M.J., Schaffer, R., and Hertz, H.S., Definitive Measurement of Constituents in Human Serum, Dimensions; 1980, SRMs 911a, 912a, 913, and 917.

White, V.E., Welch, M.J., Sun, T., Sniegoski, L.T., Schaffer, R., Hertz, H.S., and Cohen, A., The Accurate Determination of Serum Glucose by Isotope Dilution/Mass Spectrometry --Two Methods, Biomed. Mass Spectrom. 9, 395 (1982), SRM 917.

Wise, S.A., Bowie, S.L., Chesler, S.N., Cuthrell, W.F., and Rebbert, R.E., Analytical Methods for the Determination of Polycyclic Aromatic Hydrocarbons on Air Particulate Matter, Polynuclear Aromatic Hydrocarbons: Phys. and Biol. Chem., M. Cooke, and A.J. Dennis, Eds., Battelle Press, Columbus, OH (1982), 919-929, SRMs 1648 and 1649.

APPENDIX A. GLOSSARY OF TERMS

Absolute Method -- a method in which characterization is based on physically defined (absolute) standards.

Accreditation -- (1) a formal process by which a laboratory is evaluated, with respect to established criteria, for its competence to perform a specified kind(s) of measurement(s); (2) the decision based upon such a process; (3) formal recognition that a testing laboratory is competent to carry out specific tests or specific types of tests. [(3) - ISO Guide 2: 1986 (E/F/R)]

Accuracy -- the degree of agreement of a measured value with the true or expected value of the quantity of interest.

Aliquant -- (1) a divisor that does not divide a sample into a number of equal parts without leaving a remainder; (2) a sample resulting from such a divisor.

Aliquot -- (1) a divisor that divides a sample into a number of equal parts leaving no remainder; (2) a sample resulting from such a divisor.

Analyte -- the specific component measured in a chemical analysis.

Assignable cause -- a cause believed to be responsible for an identifiable change in precision or accuracy of a measurement process.

Bias -- a systematic error inherent in a method or caused by some artifact or idiosyncrasy of the measurement system. Temperature effects and extraction inefficiencies are examples of errors inherent in the method. Blanks, contamination, mechanical losses and calibration errors are examples of artifact errors. Bias can be both positive and negative and several kinds of error can exist concurrently. Therefore, net bias is all that can be evaluated.

Blank -- the measured value obtained when a specified component of a sample is not present during the measurement. In such a case, the measured value (or signal) for the component is believed to be due to artifacts and should be deducted from a measured value to give a net value due solely to the component contained in the sample. The "blank" measurement must be made so that the correction process is valid.

Blind Sample -- a sample submitted for analysis whose composition is known to the submitter but unknown to the analyst. A blind sample is one way to test the proficiency of a measurement process.

Bulk Sampling -- sampling of a material that does not consist of discrete, identifiable, constant units but rather arbitrary, irregular units.

Calibrant -- a substance used to calibrate, or to establish the analytical response of, a measurement system.

Calibration -- comparison of a measurement standard or instrument with another standard or instrument to report or eliminate, by adjustment, any variation or deviation in the accuracy of the item being compared.

Cause-Effect Diagram -- a graphical representation of the causes that can produce a specified kind of error in measurement. A popular one is the so-called fish-bone diagram, first described by Ishikawa and given this name because of its suggestive shape.

Central Line -- the long-term expected value of a variable displayed on a control chart.

Certification - a written declaration that a particular product or service complies with stated criteria. [Compilation of ASTM Standard Definitions, 7th Edition, (1990)]

Certification of Conformity -- action by a third party, demonstrating that adequate confidence is provided that a duly identified product, process or service is in conformity with a specific standard or other normative document. [ISO Guide 2:1986 (E/F/R)]

Certified Reference Material (CRM) -- a reference material one or more of whose property values are certified by a technically valid procedure, accompanied by or traceable to a certificate or other documentation which is issued by a certifying body. [ISO Guide 30: 1981 (E)]

Certified Value -- the value that appears in a certificate as the best estimate of the value for a property of a certified reference material.

Chance Cause -- a cause for variability of a measurement process that occurs unpredictably, for unknown reasons, and believed to happen by chance alone.

Check Standard (in physical calibration) -- an artifact measured periodically, the results of which typically are plotted on a control chart to evaluate the measurement process.

Coefficient of Variation -- the standard deviation divided by the value of the parameter measured.

Common Cause -- a cause of variability of a measurement process inherent in and common to the process itself, as contrasted to a special defined cause.

Comparative Method -- a method which is based on the intercomparison of the sample with a chemical standard.

Composite Sample -- a sample composed of two or more increments selected to represent a population of interest.

Confidence Interval -- that range of values, calculated from an estimate of the mean and the standard deviation, which is expected to include the population mean with a stated level of confidence. In the same context, confidence intervals can also be calculated for standard deviations, lines, slopes, and points.

Control Chart -- a graphical plot of test results with respect to time or sequence of measurement together with limits in which they are expected to lie when the system is in a state of statistical control.

Control Limits - the limits shown on a control chart beyond which it is highly improbable that a point could lie while the system remains in a state of statistical control.

Control Sample -- a material of known composition that is analyzed concurrently with test samples to evaluate a measurement process. (See also Check Standard.)

Cross Sensitivity -- a quantitative measure of the response for an undesired constituent or interferent as compared to that for a constituent of interest.

Detection Limit -- the smallest concentration or amount of some component of interest that can be measured by a single measurement with a stated level of confidence.

Double Blind -- a sample, known by the submitter but supplied to an analyst in such a way that neither its composition nor its identification as a check sample or standard are known to the analyst.

Duplicate Measurement -- a second measurement made on the same or identical sample of material to assist in the evaluation of measurement variance.

Duplicate Sample -- a second sample randomly selected from a population of interest to assist in the evaluation of sample variance. (See also Split Sample.)

Education -- disciplining the mind through instruction or study. Education is general and prepares the mind to react to a variety of situations.

Error -- difference between the true or expected value and the measured value of a quantity or parameter.

Figure of Merit -- a performance characteristic of a method believed to be useful when deciding its applicability for a specific measurement situation. Typical figures of merit include: selectivity, sensitivity, detection limit, precision, bias.

Good Laboratory Practice (GLP) -- an acceptable way to perform some basic operation or activity in a laboratory that is known or believed to influence the quality of its output(s). GLPs are usually independent of the measurement techniques used.

Good Measurement Practice (GMP) -- an acceptable way to perform some operation associated with a specific measurement technique that is known or believed to influence the quality of the measurement.

Gross Sample -- one or more increments of material taken from a larger quantity or lot of material for assay or record purposes. Also called Bulk Sample or Lot Sample.

Homogeneity -- the degree to which a property or substance is randomly distributed throughout a material. Homogeneity depends on the size of the subsample under consideration; thus, a mixture of two minerals may be homogeneous at the particulate level and inhomogeneous at the molecular or atomic levels.

Increment -- an individual portion of material collected by a single operation of a sampling device from parts of a lot separated in time or space. Increments can be either tested individually or combined/composited and tested as a unit.

Individuals -- conceivable constituent parts of a population.

Informational Value -- an uncertifed value of a property, provided because it is believed to be reliable information important to the certified material.

Intercalibration - the process, procedures, and activities used to ensure the laboratories engaged in a monitoring program can produce compatible data. When compatible data outputs are achieved and maintained over time, the laboratories can be said to be intercalibrated.

Laboratory Sample -- a sample intended for testing or analysis, prepared from a gross sample or otherwise obtained. The laboratory sample must retain the composition of the gross sample even though a reduction in particle size is often necessary in the course of reducing the quantity.

Limit of Linearity (LOL) -- the upper limit of concentration or amount of substance for which incremental additions produce constant increments of response.

Limit of Quantitation (LOQ) -- the lower limit of concentration or amount of substance that must be present before a method is considered to provide quantitative results. By convention, LOQ = 10 s_o , where s_o is the estimate of the standard deviation at the lowest level of measurement.

Limiting Mean -- the value approached by the average as the number of measurements made by a stable measurement process, increases indefinitely.

Lot -- a quantity of bulk material of similar composition whose properties are under study.

Method -- an assemblage of measurement techniques and the order in which they are used.

Outlier -- a value which appears to deviate markedly from that for other members of the sample in which it occurs.

Pareto Analysis -- a statistical approach to ranking assignable causes according to the frequency of occurrence.

Performance Audit -- a process to evaluate the proficiency of an analyst or laboratory by evaluation of the results obtained on known test materials.

Population -- (1) a generic term denoting any finite or infinite collection of individual things, objects, or events; (2) in the broadest concept, an aggregate determined by some property that distinguishes things that do and do not belong.

Precision -- the degree of mutual agreement characteristic of independent measurements as the result of repeated application of the process under specified conditions. Precision is concerned with the closeness together of results.

Primary Standard -- a substance or artifact, the value of which can be accepted without question (within specific limits) when used to establish the value of the same or related property of another material. Note that the primary standards for one user may have been a secondary standard of another user.

Probability -- the likelihood of the occurrence of any particular form of an event, estimated as the ratio of the number of ways or times that the event may occur in that form, to the total number of ways that it could occur in any form.

Procedure -- a set of systematic instructions for using a method of measurement, of sampling, or of the steps/operations associated with such.

Protocol -- a procedure specified to be used when performing a measurement or related operation, as a condition to obtain results that could be acceptable to the specifier.

Protocol for a Specific Purpose (PSP) -- detailed instructions for the performance of all aspects of a measurement program. This is sometimes called a project quality assurance (QA) plan.

Quality -- (1) an estimation of acceptability or suitability for a given purpose of an object, item, tangible or intangible thing; (2) the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs. [ISO 8402:1986 (E/F/R)]

Quality Assessment -- the overall system of activities whose purpose is to provide assurance that the quality control activities are being done effectively. It involves a continuing evaluation of performance of the production system and the quality of the products being produced.

Quality Assurance -- (1) a system of two separate but related activities - quality control and quality assessment - whose purpose is to provide to the producer or user of a product or a service, the assurance that the product or service meets defined standards of quality; (2) all those planned and systematic actions necessary to provide adequate confidence that a product or service will satisfy given requirements for quality. [ISO 9000:1987 (E)]

Quality Circle -- a small group of individuals with related interests that meet at regular intervals to consider problems or other matters related to the quality of outputs of a process. The purpose of this group is to correct problems and improve quality.

Quality Control -- (1) the overall system of activities whose purpose is to control the quality of a product or service so that it meets the needs of users. The aim is to provide quality that is satisfactory, adequate, dependable, and economic; (2) the operational techniques and activites that are used to fulfill requirements for quality. [(2) - ISO 9000:1987 (E)].

Quality Management -- that aspect of the overall management function that determines and implements the quality policy. Note 2 - Quality management includes strategic planning, allocation of resources and other systematic activities for quality such as quality planning, operations, and evaluations. [ISO 9000:1987 (E)]

Quality System -- the organizational structure, responsibilities, procedures, processes, and resources for implementing quality management. [ISO 9000:1987 (E)]

Random Sample -- a sample selected from a population using a randomization process.

Reduction -- the process of preparing one or more subsamples from a sample.

Reference Material (RM) -- a material or substance one or more properties of which are sufficiently well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assignment of values to materials. [ISO Guide 30:1981 (E)]

Reference Method -- a method which has been specified as capable, by virtue of recognized accuracy, of providing primary reference data.

Relative Standard Deviation -- the coefficient of variation expressed as a percentage.

Replicate -- a counterpart of another, usually referring to an analytical sample or a measurement. It is the general case for which duplicate, consisting of two samples or measurements, is the special case.

Routine method -- a method used in recurring analytical problems.

Sample -- a portion of a population or lot. It may consist of an individual or groups of individuals; it may refer to objects, materials, or to measurements conceived to be part of a larger group that could have been considered.

Secondary Standard -- a standard whose value is based upon comparison with some primary standard. Note that a secondary standard once its value is established, can become a primary standard for some other user.

Segment -- a specifically demarcated portion of a lot, either actual or hypothetical.

Selectivity -- the capability of methodology or instrumentation to respond to a desired substance or constituent and not to others. It is sometimes quantified as cross sensitivity.

Sensitivity -- the capability of methodology or instrumentation to discriminate between samples having differing concentrations or containing differing amounts of an analyte.

Significant Figure -- a figure that remains to a number or decimal after the ciphers to the right or left are cancelled.

Special Cause -- a cause of variance or bias that is external (not inherent) to the measurement system.

Split Sample -- a replicate portion or subsample of a total sample obtained in such a manner that it is not believed to differ significantly from other portions of the same sample.

Standard -- a substance or material, the properties of which are believed to be known with sufficient accuracy to permit its use to evaluate the same property of another. In chemical measurements, it often describes a solution or substance, commonly prepared by the analyst, to establish a calibration curve or to determine the analytical response function of an instrument.

Standard -- (1) a physical reference used as a basis for comparison or calibration; (2) a concept that has been established by authority, custom, or agreement to serve as a model or rule in the meaasurement of quality or the establishment of a practice or procedure. [Compilation of ASTM Standard Definitions, 7th Edition (1990)]

Standard -- document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree or order in a given context. [ISO Guide 2:1986 (E/F/R)]

Standard Addition -- a method in which small increments of a substance under measurement are added to a sample under test to establsh a response function or, by extrapolation, to determine the amount of a constituent originally present in the test sample.

Standardization -- (in analytical chemistry) the assignment of a compositional value to one standard on the basis of another standard.

Standard Method -- a method or procedure of test developed by a standards-writing organization, based on consensus opinion or other criteria, and often evaluated for its reliability by a collaborative testing procedure.

Standard Operations Procedure (SOP) -- a procedure adopted for repetitive use when performing a specific measurement or sampling operation. It may be a standard method or one developed by the user.

Standard Reference Material -- a certified reference material produced by the National Institute of Standards and Technology. (See definition for Certified Reference Material.)

Strata -- segments of a lot that may vary with respect to the property under study.

Subsample -- a portion taken from a sample. A laboratory sample may be a subsample of a gross sample; similarly, a test portion may be a subsample of a laboratory sample.

Technique -- a physical or chemical principle utilized separately or in combination with other techniques to determine (analyze) the composition of materials.

Test Portion (also called a specimen, test unit, aliquot) -- that quantity of a material of proper size for measurement of the property of interest. Test portions may be taken from the gross sample directly, but often preliminary operations, such as mixing or further reduction in particle size, are necessary.

Tolerance Interval -- that range of values, calculated from an estimate of the mean and the standard derivation, within which a specified percentage of individual values of a population of measurements or samples, are expected to lie with a stated level of confidence.

Traceability -- (1) the ability to trace the source of uncertainty of a measurement or a measured value; (2) the property of a result of a measurement whereby it can be related to appropriate standards, generally international or national standards, through an unbroken chain of comparisons. [(2) - VIM:1984]

Training -- formal or informal instruction designed to provide competence of a specific nature.

Uncertainty -- the range of values within which the true value is estimated to lie. It is a best estimate of possible inaccuracy due to both random and systematic errors.

Validation -- the process by which a sample, measurement method, or a piece of data is deemed to be useful for a specified purpose.

Variance -- the value approached by the average of the sum of the squares of deviations of n individual measurements from the limiting mean, m. Mathematically it may be expressed as

$$\frac{\sum (x_i - m)^2}{n} \rightarrow \sigma^2 \text{ as } n \rightarrow \infty$$

Ordinarily it cannot be known, but only its estimate, s^2 , which is calculated by the expression

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Warning Limits -- the limits shown on a control chart within which most of the test results are expected to lie (within a 95% probability) while the system remains in a state of statistical control.

Youden Plot -- a graphical representation of data, recommended first by W.J. Youden, in which the result(s) obtained by a laboratory on one sample is plotted with respect to the result(s) it obtained on a similar sample. The plot helps in deciding whether discrepant results are due to random or systematic error.

APPENDIX B. THE SI SYSTEM - UNITS AND CONVERSION FACTORS

B.1 Units

The International System of Units (SI), more commonly known as the **modernized metric system**, is divided into three classes: base units, derived units and supplementary units. These units form what is considered to be a "coherent" set; that is, they are mutually related by rules of multiplication and division without any numerical factor. For each physical quantity there is only one SI unit (even if the name of the unit can be expressed in several forms), but the inverse is not true; thus, the same SI unit can correspond to several different quantities. The base and supplementary SI units are as follows:

BASE UNITS

Quantity	Unit	Symbol
length	meter	m
mass	kilogram	kg
time	second	s
electric current	ampere	A
thermodynamic temperature	kelvin	K
amount of substance	mole	mol
luminous intensity	candela	cd

SUPPLEMENTARY UNITS

Quantity	Unit	Symbol
plane angle	radian	rad
solid angle	steradian	sr

While there are only 7 base and 2 supplementary units, the number of derived units are many. Twenty have special names; the rest are described by the algebraic relations linking corresponding quantities. In every case, the symbols for derived units are represented by mathematical signs for multiplication, division, and exponents. A few derived units are as follows:

DERIVED UNITS

Quantity	Unit	Symbol	Formula
force	newton	N	$m \cdot kg \cdot s^{-2}$
pressure, stress	pascal	Pa	$m^{-1} \cdot N$
energy, work, quantity of heat	joule	J	$N \cdot m$
area	square meter	m^2	m^2
volume	cubic meter	m^3	m^3
density, mass density	kilogram per cubic meter	kg/m^3	$m^{-3} \cdot kg$

B.2 Conversion Factors

In the SI system, designations of multiples and submultiples of any unit can be arrived at by combining a prefix with the name of the unit. For example, the prefixes deca, hecto, and kilo mean 10, 100, and 1,000 respectively; deci, centi, and milli mean 0.1, 0.01, and 0.001 respectively. Together with the SI units, the prefixes and prefix symbols for the multiples and submultiples listed below, provide a logical and interconnected framework for scientific measurements.

SI PREFIXES

Factor	Prefix	Symbol
10^{24}	yotta	Y
10^{21}	zetta	Z
10^{18}	exa	E
10^{15}	peta	P
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{2}	hecto	h
10^{1}	deka	da
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a
10^{-21}	zepto	z
10^{-24}	yocto	y

The interrelationship of SI units is clearly demonstrated in the following examples.

Length - 1 meter (m)	=	10^3 millimeters (mm)	=	10^{-3} kilometer (km)
Area - 1 are (a)	=	10^2 meters^2 (m^2)	=	10^{-1} kilometer^2 (km^2)
Volume - 1 liter (L)	=	10^3 milliliters (mL)	=	10^{-3} meter (m^3)
Mass - 10^3 grams (g)	=	1 kilogram (kg)	=	10^{-6} microgram (μg)

The above information was obtained from NIST Special Publication 330, "The International System of Units (SI)," August 1991 and NIST Special Publication 811, "Guide for the Use of the International system of Units - The Modernized Metric System," September 1991. Copies of these publications can be obtained from: Superintendent of Documents, Government Printing office, Washington, DC 20402.

B.3 Tables

Table B.3.1 Use of Range to Estimate Variability

Number of Sets, k		Number of Measurements in a Set				
		2	3	4	5	6
3	d_2^*	1.23	1.77	2.12	2.38	2.58
	ν	2.83	5.86	8.44	11.1	13.6
5	d_2^*	1.19	1.74	2.10	2.36	2.56
	ν	4.59	9.31	13.9	18.4	22.6
10	d_2^*	1.16	1.72	2.08	2.34	2.55
	ν	8.99	18.4	27.6	36.5	44.9
20	d_2^*	1.14	1.70	2.07	2.35	2.54
	ν	17.8	36.5	55.0	72.7	89.6
∞	d_2^*	1.13	1.69	2.06	2.33	2.53

$$s = \frac{\bar{R}}{d_2^*}$$

$\bar{R}$ = mean of k sets of replicate measurements

ν = degrees of freedom in estimate of standard deviation

For $k > 20$, $\nu \approx 0.876\,k$

Adapted from Lloyd S. Nelson, J. Qual. Tech. 7 (1) January (1975).

Table B.3.2 Critical Values for the F Test

Critical values for a 2-tailed test of equality of standard deviation estimated at 5% level of significance.

$F_{.975}(n_1, n_2)$

n_1 = degrees of freedom for numerator

n_2 = degrees of freedom for denominator

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

Excerpted from "Experimental Statistics" [19] Table A.5 which may be consulted for more extensive listings.

Table B.3.3 Factors for Computing Two-Sided Confidence Limits for σ

Degrees of Freedom ν	$\alpha = .05$		$\alpha = .01$		$\alpha = .001$	
	B_L	B_U	B_L	B_U	B_L	B_U
1	.3576	17.79	.2969	86.31	.2480	844.4
2	.4581	4.859	.3879	10.70	.3291	33.29
3	.5178	3.183	.4453	5.449	.3824	11.65
4	.5590	2.567	.4865	3.892	.4218	6.938
5	.5899	2.248	.5182	3.175	.4529	5.085
6	.6143	2.052	.5437	2.764	.4784	4.128
7	.6344	1.918	.5650	2.498	.5000	3.551
8	.6513	1.820	.5830	2.311	.5186	3.167
9	.6657	1.746	.5987	2.173	.5348	2.894
10	.6784	1.686	.6125	2.065	.5492	2.689
11	.6896	1.638	.6248	1.980	.5621	2.530
12	.6995	1.598	.6358	1.909	.5738	2.402
13	.7084	1.564	.6458	1.851	.5845	2.298
14	.7166	1.534	.6549	1.801	.5942	2.210
15	.7240	1.509	.6632	1.758	.6032	2.136
16	.7308	1.486	.6710	1.721	.6116	2.073
17	.7372	1.466	.6781	1.688	.6193	2.017
18	.7430	1.448	.6848	1.658	.6266	1.968
19	.7484	1.432	.6909	1.632	.6333	1.925
20	.7535	1.417	.6968	1.609	.6397	1.886
21	.7582	1.404	.7022	1.587	.6457	1.851
22	.7627	1.391	.7074	1.568	.6514	1.820
23	.7669	1.380	.7122	1.550	.6568	1.791
24	.7709	1.370	.7169	1.533	.6619	1.765
25	.7747	1.360	.7212	1.518	.6668	1.741
26	.7783	1.351	.7253	1.504	.6713	1.719
27	.7817	1.343	.7293	1.491	.6758	1.698
28	.7849	1.335	.7331	1.479	.6800	1.679
29	.7880	1.327	.7367	1.467	.6841	1.661
30	.7909	1.321	.7401	1.457	.6880	1.645
31	.7937	1.314	.7434	1.447	.6917	1.629
32	.7964	1.308	.7467	1.437	.6953	1.615
33	.7990	1.302	.7497	1.428	.6987	1.601
34	.8015	1.296	.7526	1.420	.7020	1.588
35	.8039	1.291	.7554	1.412	.7052	1.576
36	.8062	1.286	.7582	1.404	.7083	1.564
37	.8085	1.281	.7608	1.397	.7113	1.553
38	.8106	1.277	.7633	1.390	.7141	1.543
39	.8126	1.272	.7658	1.383	.7169	1.533
40	.8146	1.268	.7681	1.377	.7197	1.523
41	.8166	1.264	.7705	1.371	.7223	1.515
42	.8184	1.260	.7727	1.365	.7248	1.506
43	.8202	1.257	.7748	1.360	.7273	1.498
44	.8220	1.253	.7769	1.355	.7297	1.490
45	.8237	1.249	.7789	1.349	.7320	1.482
46	.8253	1.246	.7809	1.345	.7342	1.475
47	.8269	1.243	.7828	1.340	.7364	1.468
48	.8285	1.240	.7847	1.335	.7386	1.462
49	.8300	1.237	.7864	1.331	.7407	1.455
50	.8314	1.234	.7882	1.327	.7427	1.449

Excerpted from "Experimental Statistics" [19] Table A.20 which may be consulted for more extensive listings.

Table B.3.4 Percentiles of the t Distribution

Confidence level of 2-sided interval	20	40	60	80	90	95	98	99
df	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	.325	.727	1.376	3.078	6.314	12.706	31.821	63.657
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925
3	.277	.584	.978	1.638	2.353	3.182	4.541	5.841
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032
6	.265	.553	.906	1.440	1.943	2.447	3.143	3.707
7	.263	.549	.896	1.415	1.895	2.365	2.998	3.499
8	.262	.546	.889	1.397	1.860	2.306	2.896	3.355
9	.261	.543	.883	1.383	1.833	2.262	2.821	3.250
10	.260	.542	.879	1.372	1.812	2.228	2.764	3.169
11	.260	.540	.876	1.363	1.796	2.201	2.718	3.106
12	.259	.539	.873	1.356	1.782	2.179	2.681	3.055
13	.259	.538	.870	1.350	1.771	2.160	2.650	3.012
14	.258	.537	.868	1.345	1.761	2.145	2.624	2.977
15	.258	.536	.866	1.341	1.753	2.131	2.602	2.947
16	.258	.535	.865	1.337	1.746	2.120	2.583	2.921
17	.257	.534	.863	1.333	1.740	2.110	2.567	2.898
18	.257	.534	.862	1.330	1.734	2.101	2.552	2.878
19	.257	.533	.861	1.328	1.729	2.093	2.539	2.861
20	.257	.533	.860	1.325	1.725	2.086	2.528	2.845
21	.257	.532	.859	1.323	1.721	2.080	2.518	2.831
22	.256	.532	.858	1.321	1.717	2.074	2.508	2.819
23	.256	.532	.858	1.319	1.714	2.069	2.500	2.807
24	.256	.531	.857	1.318	1.711	2.064	2.492	2.797
25	.256	.531	.856	1.316	1.708	2.060	2.485	2.787
26	.256	.531	.856	1.315	1.706	2.056	2.479	2.779
27	.256	.531	.855	1.314	1.703	2.052	2.473	2.771
28	.256	.530	.855	1.313	1.701	2.048	2.467	2.763
29	.256	.530	.854	1.311	1.699	2.045	2.462	2.756
30	.256	.530	.854	1.310	1.697	2.042	2.457	2.750
40	.255	.529	.851	1.303	1.684	2.021	2.423	2.704
60	.254	.527	.848	1.296	1.671	2.000	2.390	2.660
120	.254	.526	.845	1.289	1.658	1.980	2.358	2.617
∞	.253	.524	.842	1.282	1.645	1.960	2.326	2.576

Excerpted from "Experimental Statistics" [19] Table A-4, which may be consulted for more extensive listings.

Table B.3.5 Factors for Two-Sided Tolerance Limits for Normal Distribution

	$\gamma = 0.95$					$\gamma = 0.99$				
n \ P	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
2	22.858	32.019	37.674	48.430	60.573	114.363	160.193	188.491	242.300	303.054
3	5.922	8.380	9.916	12.861	16.208	13.378	18.930	22.401	29.055	36.616
4	3.779	5.369	6.370	8.299	10.502	6.614	9.398	11.150	14.527	18.383
5	3.002	4.275	5.079	6.634	8.415	4.643	6.612	7.855	10.260	13.015
6	2.604	3.712	4.414	5.775	7.337	3.743	5.337	6.345	8.301	10.548
7	2.361	3.369	4.007	5.248	6.676	3.233	4.613	5.488	7.187	9.142
8	2.197	3.136	3.732	4.891	6.226	2.905	4.147	4.936	6.468	8.234
9	2.078	2.967	3.532	4.631	5.899	2.677	3.822	4.550	5.966	7.600
10	1.987	2.839	3.379	4.433	5.649	2.508	3.582	4.265	5.594	7.129
11	1.916	2.737	3.259	4.277	5.452	2.378	3.397	4.045	5.308	6.766
12	1.858	2.655	3.162	4.150	5.291	2.274	3.250	3.870	5.079	6.477
13	1.810	2.587	3.081	4.044	5.158	2.190	3.130	3.727	4.893	6.240
14	1.770	2.529	3.012	3.955	5.045	2.120	3.029	3.608	4.737	6.043
15	1.735	2.480	2.954	3.878	4.949	2.060	2.945	3.507	4.605	5.876
16	1.705	2.437	2.903	3.812	4.865	2.009	2.872	3.421	4.492	5.732
17	1.679	2.400	2.858	3.754	4.791	1.965	2.808	3.345	4.393	5.607
18	1.655	2.366	2.819	3.702	4.725	1.926	2.753	3.279	4.307	5.497
19	1.635	2.337	2.784	3.656	4.667	1.891	2.703	3.221	4.230	5.399
20	1.616	2.310	2.752	3.615	4.614	1.860	2.659	3.168	4.161	5.312
21	1.599	2.286	2.723	3.577	4.567	1.833	2.620	3.121	4.100	5.234
22	1.584	2.264	2.697	3.543	4.523	1.808	2.584	3.078	4.044	5.163
23	1.570	2.244	2.673	3.512	4.484	1.785	2.551	3.040	3.993	5.098
24	1.557	2.225	2.651	3.483	4.447	1.764	2.522	3.004	3.947	5.039
25	1.545	2.208	2.631	3.457	4.413	1.745	2.494	2.972	3.904	4.985
26	1.534	2.193	2.612	3.432	4.382	1.727	2.469	2.941	3.865	4.935
27	1.523	2.178	2.595	3.409	4.353	1.711	2.446	2.914	3.828	4.888

Excerpted from "Experimental Statistics" [19] Table A.6 which may be consulted for more extensive listings.

Table B.3.6 Short Table of Random Numbers

46	96	85	77	27	92	86	26	45	21	89	91	71	42	64	64	58	22	75	81	74	91	48	46	18
44	19	15	32	63	55	87	77	33	29	45	00	31	34	84	05	72	90	44	27	78	22	07	62	17
34	39	80	62	24	33	81	67	28	11	34	79	26	35	34	23	09	94	00	80	55	31	63	27	91
74	97	80	30	65	07	71	30	01	84	47	45	89	70	74	13	04	90	51	27	61	34	63	87	44
22	14	61	60	86	38	33	71	13	33	72	08	16	13	50	56	48	51	29	48	30	93	45	66	29
40	03	96	40	03	47	24	60	09	21	21	18	00	05	86	52	85	40	73	73	57	68	36	33	91
52	33	76	44	56	15	47	75	78	73	78	19	87	06	98	47	48	02	62	03	42	05	32	55	02
37	59	20	40	93	17	82	24	19	90	80	87	32	74	59	84	24	49	79	17	23	75	83	42	00
11	02	55	57	48	84	74	36	22	67	19	20	15	92	53	37	13	75	54	89	56	73	23	39	07
10	33	79	26	34	54	71	33	89	74	68	48	23	17	49	18	81	05	52	85	70	05	73	11	17
67	59	28	25	47	89	11	65	65	20	42	23	96	41	64	20	30	89	87	64	37	93	36	96	35
93	50	75	20	09	18	54	34	68	02	54	87	23	05	43	36	98	29	97	93	87	08	30	92	98
24	43	23	72	80	64	34	27	23	46	15	36	10	63	21	59	69	76	02	62	31	62	47	60	34
39	91	63	18	38	27	10	78	88	84	42	32	00	97	92	00	04	94	50	05	75	82	70	80	35
74	62	19	67	54	18	28	92	33	69	98	96	74	35	72	11	68	25	08	95	31	79	11	79	54
91	03	35	60	81	16	61	97	25	14	78	21	22	05	25	47	26	37	80	39	19	06	41	02	00
42	57	66	76	72	91	03	63	48	46	44	01	33	53	62	28	80	59	55	05	02	16	13	17	54
06	36	63	06	15	03	72	38	01	58	25	37	66	48	56	19	56	41	29	28	76	49	74	39	50
92	70	96	70	89	80	87	14	25	49	25	94	62	78	26	15	41	39	48	75	64	69	61	06	38
91	08	88	53	52	13	04	82	23	00	26	36	47	44	04	08	84	80	07	44	76	51	52	41	59
68	85	97	74	47	53	90	05	90	84	87	48	25	01	11	05	45	11	43	15	60	40	31	84	59
59	54	13	09	13	80	42	29	63	03	24	64	12	43	28	10	01	65	62	07	79	83	05	59	61
39	18	32	69	33	46	58	19	34	03	59	28	97	31	02	65	47	47	70	39	74	17	30	22	65
67	43	31	09	12	60	19	57	63	78	11	80	10	97	15	70	04	89	81	78	54	84	87	83	42
61	75	37	19	56	90	75	39	03	56	49	92	72	95	27	52	87	47	12	52	54	62	43	23	13
78	10	91	11	00	63	19	63	74	58	69	03	51	38	60	36	53	56	77	06	69	03	89	91	24
93	23	71	58	09	78	08	03	07	71	79	32	25	19	61	04	40	33	12	06	78	91	97	88	95
37	55	48	82	63	89	92	59	14	72	19	17	22	51	90	20	03	64	96	60	48	01	95	44	84
62	13	11	71	17	23	29	25	13	85	33	35	07	69	25	68	57	92	57	11	84	44	01	33	66
29	89	97	47	03	13	20	86	22	45	59	98	64	53	89	64	94	81	55	87	73	81	58	46	42
16	94	85	82	89	07	17	30	29	89	89	80	98	36	25	36	53	02	49	14	34	03	52	09	20
04	93	10	59	75	12	98	84	60	93	68	16	87	60	11	50	46	56	58	45	88	72	50	46	11
95	71	43	68	97	18	85	17	13	08	00	50	77	50	46	92	45	26	97	21	48	22	23	08	32
86	05	39	14	35	48	68	18	36	57	09	62	40	28	87	08	74	79	91	08	27	12	43	32	03
59	30	60	10	41	31	00	69	63	77	01	89	94	60	19	02	70	88	72	33	38	88	20	60	86
05	45	35	40	54	03	98	96	76	27	77	84	80	08	64	60	44	34	54	24	85	20	85	77	32
71	85	17	74	66	27	85	19	55	56	51	36	48	92	32	44	40	47	10	38	22	52	42	29	96
80	20	32	80	98	00	40	92	57	51	52	83	14	55	31	99	73	23	40	07	64	54	44	99	21
13	50	78	02	73	39	66	82	01	28	67	51	75	66	33	97	47	58	42	44	88	09	28	58	06
67	92	65	41	45	36	77	96	46	21	14	39	56	36	70	15	74	43	62	69	82	30	77	28	77
72	56	73	44	26	04	62	81	15	35	79	26	99	57	28	22	25	94	80	62	95	48	98	23	86
28	86	85	64	94	11	58	78	45	36	34	45	91	38	51	10	68	36	87	81	16	77	30	19	36
69	57	40	80	44	94	60	82	94	93	98	01	48	50	57	69	60	77	69	60	74	22	05	77	17
71	20	03	30	79	25	74	17	78	34	54	45	04	77	42	59	75	78	64	99	37	03	18	03	36
89	98	55	98	22	45	12	49	82	71	57	33	28	69	50	59	15	09	25	79	39	42	84	18	70
58	74	82	81	14	02	01	05	77	94	65	57	70	39	42	48	56	84	31	59	18	70	41	74	60
50	54	73	81	91	07	81	26	25	45	49	61	22	88	41	20	00	15	59	93	51	60	65	65	63
49	33	72	90	10	20	65	28	44	63	95	86	75	78	69	24	41	65	86	10	34	10	32	00	93
11	85	01	43	65	02	85	69	56	88	34	29	64	35	48	15	70	11	77	83	01	34	82	91	04
34	22	46	41	84	74	27	02	57	77	47	93	72	02	95	63	75	74	69	69	61	34	31	92	13

Excerpted from "Experimental Statistics" [19] Table A.36 which may be consulted for more extensive listings.

Table B.3.7 Z-Factors for Two-Sided Confidence Interval

Confidence Level	Z-Factor
50 %	0.67
68	1.00
75	1.15
90	1.645
95	1.960
95.45	2.000
99.00	2.575
99.74	3
99.9934	4

APPENDIX C. STATISTICAL TOOLS

C.1 Introduction

The following pages contain a brief description, with examples, of statistical calculations related to some of the questions that arise when evaluating chemical measurement data. The reader is referred to the many excellent books that are available which discuss these matters in more detail and are the basis for the relationships used. General information on precision of measurement is contained in reference [16], and NBS Handbook 91 [19] is especially recommended for a detailed discussion of statistical concepts. It contains many numerical examples as well as extensive tables, from which several of the ones included in Appendix B were taken.

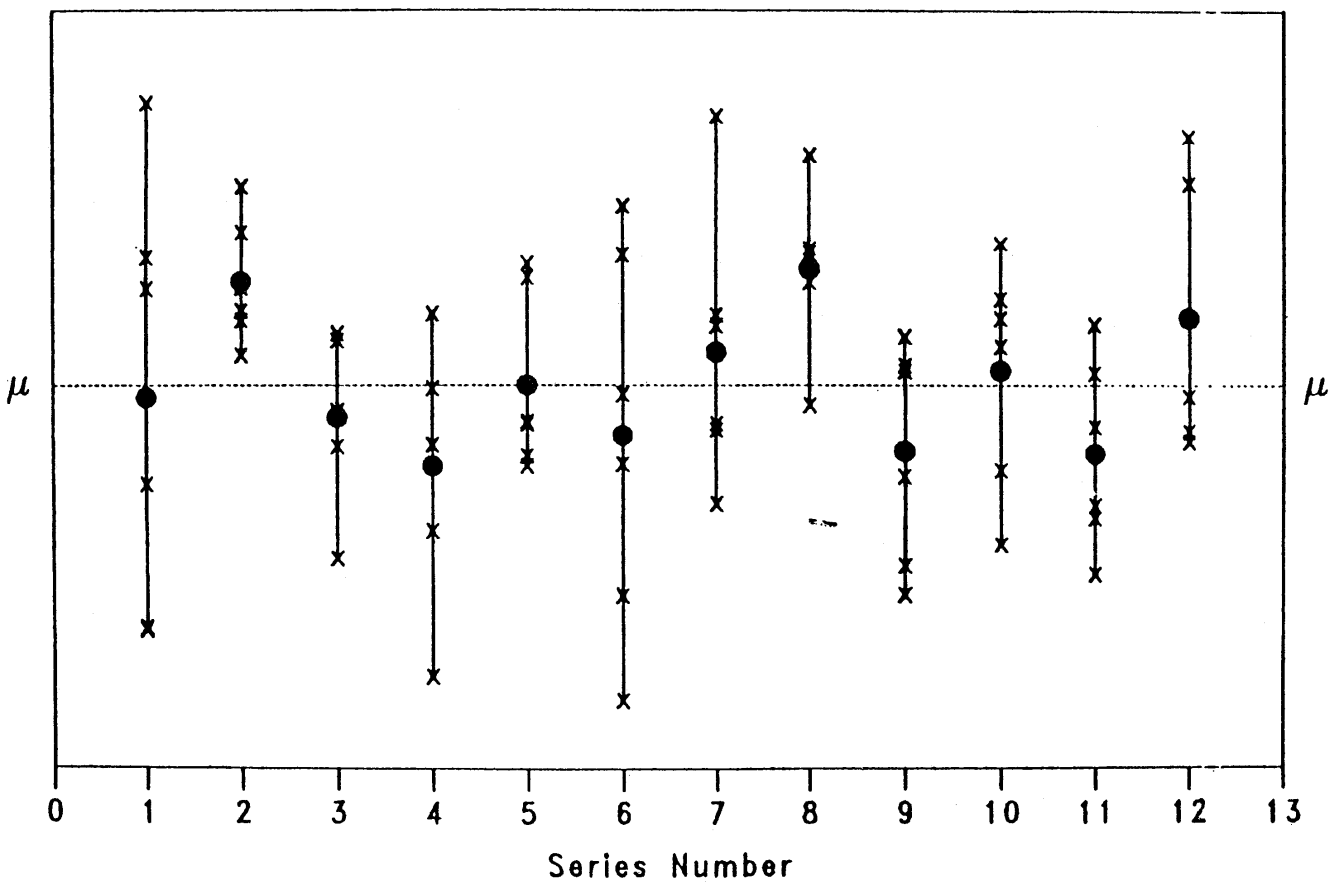


Figure C.1 Typical Distribution of the Means of Random Samples/Measurements. The individual measurements are indicated by x and the means, $\bar{x}$, by dots

The results of repetitive measurements are usually expected to be normally distributed and representable by a bell-shaped curve. If a series of n measurements were made many times, one would expect to obtain data sets represented by distributions such as those in Figure C.1. The means of the sets will differ from each other. If the sample standard deviation, s, is calculated for each set of such measurements, a different result would be expected each time. Figure C.2 shows the two-sided (asymmetric) confidence limits for the true standard deviation, σ, based on such estimates, for several confidence levels.

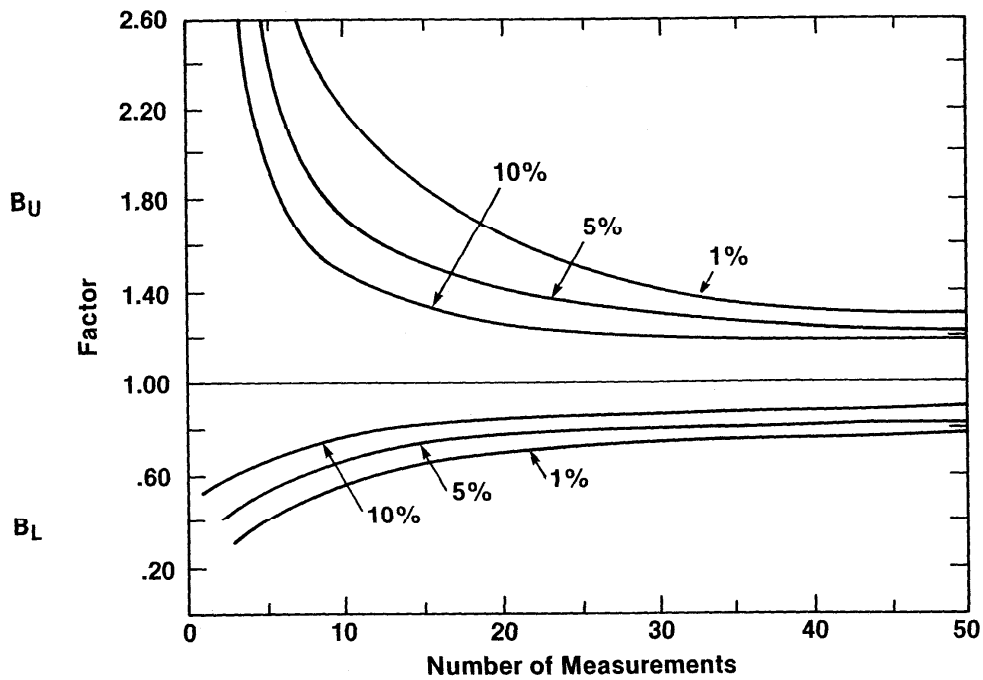


Figure C.2 Confidence Interval for σ

This figure illustrates the expected variability of estimates of standard deviations made on various occasions, as a function of the number of measurements involved. The factor, when multiplied by the estimate of the standard deviation, gives the interval that is expected to include the population standard deviation for a given percentage of occasions. The labels, e.g., 10 %, indicate the percentage of time that such an interval would not be expected to include σ. See section C.4 for a discussion and Table B.4 for the factors to be used in such calculations.

When several series of unbiased measurements are made, both the sample means and the sample standard deviations will vary from series to series as illustrated in Figure C.3. For example, as n increases from 4 to 1000, the variation of the means decreases, but never disappears.

Considering this, it should be obvious that even the best of measurements will differ among themselves, whether made by the same or different laboratories or scientists. One often needs to answer questions such as the confidence that can be placed in measurement data and the significance of apparent differences resulting from measurements. The various equations given in this appendix take into account both the expected variability within populations and the uncertainties in the estimates of the population parameters that must be considered when answering such questions.

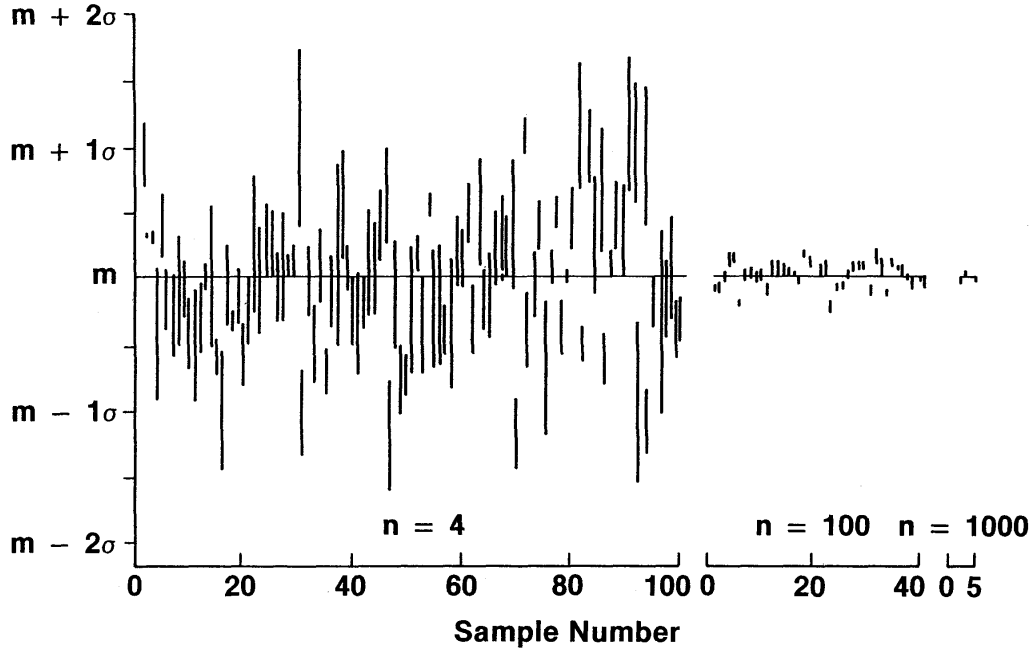


Figure C.3 Computed 50 % Confidence Intervals for the Population Mean, m, from 100 Samples of 4, 40 Samples of 100, and 4 Samples of 1000 [19]

The vertical lines are essentially error bars. The sample means, located at the center of each, are not indicated. The sample means and standard deviations (proportional to the error bars) vary with each set of measurements. The error bars decrease inversely as the square root of the sample size is increased and the means show correspondingly smaller deviations from the population mean.

C.2 Estimating the Standard Deviation

The basic parameters which characterize a population (universe) of samples or measurements on a given sample are the mean, μ, and the standard deviation, σ. Unless the entire population is examined, μ and σ cannot be known but are estimated from sample(s) randomly selected from it. The result is a sample mean, $\bar{x}$, and an estimate of the standard deviation, s, which must be used if such things as confidence intervals, tolerance intervals, comparison of precision, and the significance of apparent discrepancies in measured values are to be evaluated.

Two ways by which the standard deviation may be estimated are given in the following sections.

C.2.1 Estimating the Standard Deviation from Replicate Measurements

For a series of n independent measurements

$$x_1, x_2, x_3 \ldots\ldots x_n$$

$$\bar{x} = \frac{x_1 + x_2 + x_3 \ldots + x_n}{n}$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

s is estimated with ν = n-1 degrees of freedom.

- Example: C.2.1 - Series of Measurements -

x_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$
15.2	0.143	0.0204
14.7	-0.357	0.1257
15.1	0.043	0.0018
15.0	-0.057	0.0033
15.3	0.243	0.0590
15.2	0.143	0.0204
14.9	-0.157	0.0247

$\bar{x}$ = 15.057 $\qquad$ Σ = 0.2572

n = 7 $\qquad$ ν = 6

$$s = \sqrt{\frac{0.2572}{6}} = 0.207$$

C.2.2 Estimating the Measurement Standard Deviation from Duplicate Measurements

$$s = \sqrt{\frac{\sum d^2}{2k}}$$

where k = number of sets of duplicate measurements

d = difference of a duplicate measurement

ν = k degrees of freedom

Note: It is not necessary that the duplicate measurements be made on the same materials. It is only necessary that the materials measured are expected to have the same measurement standard deviation.

- Example A: C.2.2 - Duplicates, Same Material -

x_f	x_s	$\lvert d \rvert$	d^2
14.7	15.0	0.3	0.09
15.1	14.9	0.2	0.04
15.0	15.1	0.1	0.01
14.9	14.9	0.0	0.0
15.3	14.8	0.5	0.25
14.9	15.1	0.2	0.04
14.9	15.0	0.1	0.01
			$\Sigma = 0.44$

$$s = \sqrt{\frac{0.44}{14}} = 0.18$$

ν = 7 degrees of freedom

- Example B: C.2.2 - Estimation from Duplicates, Different Materials -

x_f	x_s	$\lvert d \rvert$	d^2
14.7	15.0	0.3	0.09
20.1	19.8	0.3	0.09
12.5	13.0	0.5	0.25
23.6	23.3	0.3	0.09
15.1	14.9	0.2	0.04
18.2	18.0	0.2	0.04
20.7	20.9	0.2	0.04
			$\Sigma = 0.64$

$$s = \sqrt{\frac{0.64}{14}} = 0.21$$

$\nu = 7$ degrees of freedom

C.2.3 Pooling Estimates of Standard Deviations

Several estimates of the standard deviation may be pooled to obtain a better estimate. Given several estimates of the standard deviation obtained on several occasions with the corresponding measurements:

s_1	n_1	$\nu_1 = n_1-1$
s_2	n_2	$\nu_2 = n_3-1$
s_3	n_3	$\nu_3 = n_3-1$
•	•	• •
•	•	• •
s_k	n_k	$\nu_k = n_k-1$

$$s_p = \sqrt{\frac{\nu_1 s_1^2 + \nu_2 s_2^2 + \cdots + \nu_k S_k^2}{\nu_1 + \nu_2 + \cdots + \nu_k}}$$

s_p will be based on $(\nu_1 + \nu_2 + \bullet\bullet\bullet + \nu_k)$ degrees of freedom.

Note: If s is not pooled, $\nu = n-1$ for n independent measurements.

- Example C.2.3 Pooling Standard Deviations -

The standard deviation of a measurement process was estimated on five occasions. These are to be pooled to improve the estimate of σ.

Trial	s	n	ν
1	0.171	7	6
2	0.205	5	4
3	0.185	7	6
4	0.222	4	3
5	0.180	5	4

$$s_p = \sqrt{\frac{6(0.171)^2 + 4(0.205)^2 + 6(0.185)^2 + 3(0.222)^2 + 4(0.180)^2}{6 + 4 + 6 + 3 + 4}}$$

$$s_p = \sqrt{\frac{0.1755 + 0.1681 + 0.2054 + 0.1479 + 0.1296}{23}}$$

$s_p = 0.190$ with 23 degrees of freedom.

C.3 Do Two Estimates of Precision Differ?

In order to determine whether there is a statistically significant difference between two estimates of precision, conduct an F test, as follows:

Let s_1 = estimate of standard deviation (larger value) based on n_1 independent measurements, the degrees of freedom are n_1-1.

Let s_2 = estimate of standard deviation (smaller value) based on n_2 independent measurements, the degrees of freedom are n_2-1.

$$\text{Calculate } F = \frac{s_1^2}{s_2^2}$$

Look up the critical value of F_c in Table B.3, based on the respective degrees of freedom for the estimates of s_1 and s_2 and the chosen level of confidence.

If $F \leq F_c$, there is no reason to believe that the true precisions are different, at the chosen level of confidence.

If $F > F_c$, reject the assumption of no difference in precision, and accept the alternative that the standard deviation estimated by s_1 is larger than that estimated by s_2.

- Example C.3 - Comparison of Precision Estimates -

$s_1 = 2.00$ $\quad n_1 = 6 \quad \nu_1 = 5$

$s_2 = 1.00$ $\quad n_2 = 6 \quad \nu_2 = 5$

$F = 4.00/1.00 = 4.0$

$F_c = 7.15$ at 5% level of significance

Conclusion: s_1 is not larger than s_2 at the 5% level of significance.

C.4 Confidence Limits for an Estimate of Standard Deviation

The width of the confidence interval for an estimated standard deviation will depend on the number of degrees of freedom, ν, upon which the estimate is based. The interval is not symmetric (see Figure C.2), as in the case for a mean, since a small number of measurements tend to underestimate the standard deviation. To calculate the bounds of the interval, one may use a table such as Table B.4 and find the factors B_L and B_U corresponding to the number of degrees of freedom involved and the confidence level sought. In the table, $\alpha = 0.05$ corresponds to a confidence of 95% for the interval so calculated. The confidence interval is then sB_L to sB_U.

- Example C.4 - Confidence Limits for a Standard Deviation -

$s = 0.15 \qquad n = 10 \qquad \nu = 9$

For $\alpha = .05$. $\nu = 9$ (see Table B.4) one finds $B_L = 0.6657$; $B_U = 1.746$

The confidence interval for s is 0.15 x 0.6657 to 0.15 x 1.746 or 0.10 to 0.26.

C.5 Confidence Interval for a Mean

The confidence interval for the mean will depend on the number of independent measurements, n, the standard deviation, s, and the level of confidence desired. The confidence interval is calculated using the expression

$$\bar{x} \pm \frac{ts}{\sqrt{n}}$$

The value for t (see Table B.5) will depend on the level of confidence desired and the number of degrees of freedom, ν, associated with the estimation of s. If s is based on the set of measurements used to calculate the mean, $\bar{x}$, then $\nu = n-1$. If the measurements are made by a system under statistical control, as demonstrated by a control chart, ν will depend on the number of measurements made to establish the control limits.

- Example A: C.5 - Confidence Interval Based on s Estimated from Data Set of Seven Measurements -

$\bar{x} = 10.05$
$s = 0.11$
$n = 7$
$\nu = 6$

For a 95% level of confidence, t = 2.447, hence

$$10.05 \pm \frac{2.447 \times 0.11}{\sqrt{7}} = 10.05 \pm 0.10, \text{ or } 9.95 \text{ to } 10.15$$

- Example B: C.5 - Confidence Interval Based on s Obtained from Control Chart Limits, One Measurement of x -

$$x = 10.05$$
$$s = 0.11$$
$$n = 1$$
$$\nu = 45 \text{ (control chart)}$$

For a 95 percent level of confidence, t = 2.014*, hence

$$10.05 \pm \frac{2.014 \times 0.11}{\sqrt{1}} = 10.05 \pm 0.22, \text{ or } 9.83 \text{ to } 10.27$$

*t-value obtained from Table B.5 by interpolation.

Note: There is no statistical basis for a confidence level statement for one measurement unless supported by a control chart or other evidence of statistical control, which also supplies an estimate of the standard deviation of that measurement.

C.6 Do Two Means Disagree Significantly?

The decision on disagreement is based on whether the difference, Δ, of the two values exceeds its statistical uncertainty, U. The method used for calculation of the uncertainty depends on whether or not the respective standard deviation estimates may be considered to estimate the same value.

Case I No reason to believe that the standard deviations differ (e.g., same method, analyst, experimental conditions, etc.)

Step 1. Choose the significance level of the test.

Step 2. Calculate a pooled standard deviation from the two estimates to obtain a better estimate of the standard deviation.

$$s_p = \sqrt{\frac{\nu_A s_A^2 + \nu_B s_B^2}{\nu_A + \nu_B}}$$

s_p will be based on $\nu_A + \nu_B$ degrees of freedom

Step 3. Calculate the uncertainty, U, of the difference.

$$U = t\, s_p \sqrt{\frac{n_A + n_B}{n_A n_B}}$$

Step 4. Compare $\Delta = | \bar{x}_A - \bar{x}_B |$ with U

If $\Delta \leq U$, there is no reason to believe that the true means disagree.

- Example, Case I -

$\bar{x}_A = 4.25$ $\quad$ $\bar{x}_B = 4.39$
$s_A = 0.13$ $\quad$ $s_B = 0.17$
$n_A = 7$ $\quad$ $n_B = 10$
$\nu_A = 6$ $\quad$ $\nu_B = 9$

$$\Delta = | \bar{x}_A - \bar{x}_B | = | 4.25 - 4.39 | = 0.14$$

Step 1. $\alpha = 0.05$ (95% confidence)

Step 2.

$$s_p = \sqrt{\frac{6\ (0.13)^2 + 9\ (0.17)^2}{15}}$$

$$s_p = 0.155$$

Step 3.

$$U = 2.131 \times 0.155 \sqrt{\frac{7 + 10}{70}}$$

$$U = 0.080$$

Step 4. $0.14 > .080$

Conclude that the two means differ at the 95% level of confidence.

<u>Case II</u> Reason to believe that the standard deviations differ (e.g., different experimental conditions, different laboratories, etc.)

Note: Even if there is a reason to believe that the standard deviations differ, if the sample sizes are the same (i.e., $n_A = n_B$) then Case I should be followed.

Step 1. Choose α, the significance level of the test.

Step 2. Compute the estimated variance of each mean.

$$V_A = \frac{s_A^2}{n_A} \quad , \quad V_B = \frac{s_B^2}{n_B}$$

Step 3. Compute the effective number of degrees of freedom, f.

$$f = \frac{(V_A + V_B)^2}{\frac{V_A^2}{n_A + 1} + \frac{V_B^2}{n_B + 1}} - 2$$

Step 4. Compute the uncertainty, U, of the difference

$$U = t_f \sqrt{V_A + V_B}$$

where t_f is chosen from Table B.5 for f degrees of freedom and two-sided confidence 100 $(1-\alpha)$ %.

Step 5. Compare Δ with U

If $\Delta \leq U$ there is no reason to believe that the means disagree.

- Example, Case II. -

$\bar{x}_A = 4.25$ $\quad$ $\bar{x}_B = 4.39$
$s_A = 0.13$ $\quad$ $s_B = 0.17$
$n_A = 7$ $\quad$ $n_B = 10$

Step 1. $\alpha = 0.05$ (95% confidence)

Step 2.

$$V_A = \frac{(.13)^2}{7} = 2.414 \times 10^{-3}$$

$$V_B = \frac{(.17)^2}{10} = 2.89 \times 10^{-3}$$

Step 3.

$$f = \frac{(2.414 \times 10^{-3} + 2.89 \times 10^{-3})^2}{\frac{2.414 \times 10^{-3}}{8} + \frac{2.89 \times 10^{-3}}{11}} - 2$$

$$f = 17$$

Step 4.

$$U = 2.11 \sqrt{2.414 \times 10^{-3} + 2.89 \times 10^{-3}}$$

$$U = 0.153$$

Step 5. $\Delta = 0.14$; $U = 0.153$

Conclude there is no reason to believe that the true means differ at the 95% level of confidence.

C.6.1 Does a Measured Value Differ from a Certified (Standard) Value?

This is a frequently occurring special case of the above, in which it is assumed that the standard value does not have significant error. One chooses the level of significance of the test, say 5%, then calculates the uncertainty of the measured value, using the procedures of C.5. If the observed difference is less than the uncertainty of the measurement, there is no reason to believe that the difference is significant.

- Example C.6.1 Measurement of SRM 122h -

Certified Value $C = 3.52$
$\bar{x} = 3.55$
$s = 0.020$
$n = 3$

$$U = \frac{4.303 \times 0.020}{\sqrt{3}}$$

$$U = 0.050$$

Conclusion: There is no reason to believe that the measured value differs from the certified value at the 95% level confidence level.

C.7 Statistical Tolerance Intervals

A tolerance interval represents the limits within which a specified percentage of the population is expected to lie with a given probability. A tolerance interval is frequently used to specify the variability in composition of samples. If the mean and standard deviation of the population of samples were known, the limits for a given percentage of the population could be calculated with certainty. Because only estimates are usually available based on a limited sampling of the population, a tolerance interval is all that can be calculated.

The calculation is made as follows:

$$\text{Tolerance Interval} = \bar{x} \pm ks$$

where k = a factor (obtained from Table B.6, for example) based on the percentage, p, of the population to be included, the probability, γ, of inclusion, and the number, n, of measurements used to calculate $\bar{x}$ and s.

- Example C.7 - Statistical Tolerance Interval -

For measurements of ten samples of a shipment of coal, the sulfur content was found to be

$\bar{x} = 1.62$ $s = 0.10$ $n = 10$

From Table B.6, $k = 3.379$ for $\gamma = 0.95$, $p = 0.95$, $n = 10$.
The tolerance interval is thus 1.62 ± 0.34 or 1.28 to 1.96.

C.8 Pooling Means to Obtain a Grand Average, $\bar{\bar{x}}$

<u>Case I</u> All means based on the same number of measurements of equal precision

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 \cdots + \bar{x}_k}{k}$$

$$S^2_{\bar{\bar{x}}} = \frac{1}{k^2}\left(S^2_{\bar{x}_1} + S^2_{\bar{x}_2} + \cdots + S^2_{\bar{x}_k}\right)$$

- Example C.8 - Calculation of Grand Average, Case I -

Calculate the grand average, $\bar{\bar{x}}$, of the following means, all believed to be equally precise.

$$\bar{x}_1 = 10.50$$
$$\bar{x}_2 = 10.37$$
$$\bar{x}_3 = 10.49$$
$$\bar{x}_4 = 10.45$$
$$\bar{x}_5 = 10.47$$

$$\bar{\bar{x}} = \frac{10.50 + 10.37 + 10.49 + 10.45 + 10.47}{5}$$

$$\bar{\bar{x}} = 10.456$$

<u>Case II</u> Means based on different number of measurements, but no reason to believe the precisions differ.

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n_1 + n_2 + \cdots + n_k}$$

$$S^2_{\bar{\bar{x}}} = \frac{\left(n_1^2 S^2_{\bar{x}_1} + n_2^2 S^2_{\bar{x}_2} + \cdots + n_k^2 S^2_{\bar{x}_k}\right)}{(n_1 + n_2 + \cdots + n_k)^2}$$

- Example C.8 - Calculation of Grand Average, Case II -

$$\bar{x}_1 = 10.50 \quad n_1 = 10$$
$$\bar{x}_2 = 10.37 \quad n_2 = 5$$
$$\bar{x}_3 = 10.49 \quad n_3 = 20$$
$$\bar{x}_4 = 10.45 \quad n_4 = 5$$
$$\bar{x}_5 = 10.47 \quad n_5 = 7$$

$$\bar{\bar{x}} = \frac{10.50 \times 10 + 10.37 \times 5 + 10.49 \times 20 + 10.45 \times 5 + 10.47 \times 7}{47}$$

$$\bar{\bar{x}} = 10.472$$

<u>Case III</u> Means based on different number of measurements with differing precisions.

Step 1. Compute weight to be used for each mean,

$$w_i = \frac{n_i}{s_i^2} \text{ e.g., } w_1 = \frac{n_1}{s_1^2}$$

$$\bar{\bar{x}} = \frac{w_1\bar{x}_1 + w_2\bar{x}_2 + \cdots + w_k\bar{x}_k}{w_1 + w_2 + \cdots + w_k}$$

$$s^2_{\bar{\bar{x}}} = \frac{1}{w_1 + w_2 + \cdots w_k}$$

- Example C.8 - Calculation of Grand Average, Case III -

i	$\bar{x}_i$	n_i	s_i	w_i
1	10.50	10	.10	1000
2	10.37	5	.15	222
3	10.49	20	.11	1652
4	10.45	5	.10	500
5	10.47	7	.16	273

$$\bar{\bar{x}} = \frac{10.50 \times 1000 + 10.37 \times 222 + 10.49 \times 1652 + 10.45 \times 500 + 10.47 \times 273}{1000 + 222 + 1652 + 500 + 273}$$

$$\bar{\bar{x}} = 10.478$$

$$S^2_{\bar{\bar{x}}} = 2.74 \times 10^{-4}$$

$$S_{\bar{\bar{x}}} = 0.0166$$

C.9 Outliers

Outliers are data values that do not belong, or have a very low probability of belonging, to the data set in which they occur. They can result from such causes as analyst blunders or malfunctions of the methodology, or from unusual losses or contamination. If outliers occur too often, there may be deficiencies in the quality control program which need to be corrected to improve the measurements. One should always look for causes when data are rejected.

Outliers may be suspected when data are plotted, when results are ranked, and when control limits are exceeded. Only when a measurement system is well understood and the variance is well established, or when a large body of data are available, it is possible to distinguish between extreme values and true outliers with any degree of confidence.

The following rules for rejection of data should be used with caution since an outlier in a well-behaved measurement system should be a rare occurrence.

A. Rejection for Assignable Cause

System malfunction, misidentification of a sample, suspected transcription error, known contamination, etc.

B. Rule of Huge Error

If the questioned value differs from the mean by some multiple, M, of the standard deviation, it may be considered to be an outlier. The size of the multiple depends on the confidence required for rejection. One evaluates

$$M = \frac{|x - \bar{x}|}{s}$$

A practical rule might be to use $M \geq 4$ as a criterion for rejection. This corresponds to a significance level of $< 2\%$ when the standard deviation is well established, such as based on a data set of 15 or larger.

If s is not well established but depends on the data set in question, the odds for rejection are much larger. For example, if $\bar{x}$ and s are based on 6 measurements, $M > 6$ would be the criterion for rejection for a 2% level of significance.

- Example C.9 - Outliers, Huge Error Concept -

x (original order)	x (ranked order)
10.50	10.45
10.47	10.47
10.49	10.47
10.45	10.48
10.47	10.49
10.57	10.50
10.52	10.50
10.50	10.52
10.48	10.53
10.53	10.57

In the above list, 10.57 appears to be an outlier.

Therefore, calculating mean and sample standard deviation, s, ignoring 10.57

$$\bar{x} = 10.490 \qquad s = 0.0255$$

$$\frac{10.57 - 10.49}{0.0255} = 3.13$$

Since $3.13 < 4$, conclude that 10.57 should be retained in the data and calculate mean and sample standard deviation including 10.57

$$\bar{x} = 10.498 \qquad s = 0.0349$$

C. Statistical Tests

Several statistical tests are available for identifying outliers based on ranking data and testing extreme values for credibility. The Dixon criterion is described on page 17-3 of NBS Handbook 91 [19] and the critical values for decision on rejection using this criterion are given in Table A-14 of the same reference.

A method for identifying outlier laboratories in a collaborative test or proficiency testing program is described by Youden on page 148 of NBS Special Publication 300 [16] where a table of the test score values necessary to use his criterion also will be found.

ASTM E178-80 (1989), Recommended Practice for Dealing with Outlying Observations, contains extensive statistical tables for use in deciding on the significance of apparent outliers.

C.10 Use of Random Number Tables

It is often desirable to randomize the sequence in which measurements are made, samples are chosen, and other variables of an analytical program are set. Randomization is a cheap insurance policy against confounding measurement process problems with material problems.

For example, if samples are analyzed in the order in which they were packaged (or in any way systematic with the material itself) and the measurements show a trend, there is no way (from the immediate data) to determine whether there is a trend in the samples, or whether the trend was in the measurement process. Obviously, the ramifications are very different depending on which was the case. If the run sequence had been randomized, the source of the trend could be determined much more easily.

Tables of random numbers, such as Table B.7, are a convenient and simple way to accomplish this. The following procedure may be used.

1. Number the samples (or measurements) serially, say 00 to xy. For example, 00 to 15 for 16 items.

2. Start at any randomly selected place in the table and proceed from that point in any systematic path. The order in which the item numbers are located becomes the random sequence number to be assigned to them.

- Example C.10 - Using a Random Number Table -

Start at Row 7, Column 3 (chosen by chance) of Table B.7 and proceed from left to right as in reading. The first number is 76 which is not usable for the above series of items. The first usable number is 15. Proceeding as above, the items are located in the following order: 15, 06, 02, 03, 05, 00, 11, 13, 07, 10, 09, 08, 04, 14, 01, 12. If a number already chosen is encountered, pass over it to the next usable number.

Appendix D Published Articles

Appendix D.1

John K. Taylor
Center for Analytical Chemistry
National Bureau of Standards
Washington, D.C. 20234

Quality Assurance of

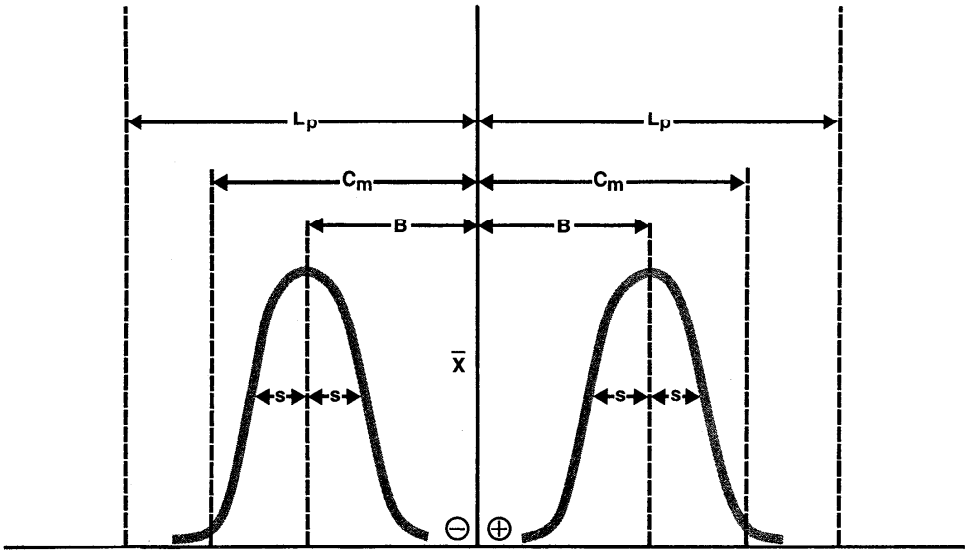


Figure 1. Measurement tolerances and errors

The objective of quality assurance programs for analytical measurements is to reduce measurement errors to tolerable limits and to provide a means of ensuring that the measurements generated have a high probability of being of acceptable quality. Two concepts are involved. *Quality control* is the mechanism established to control errors, while *quality assessment* is the mechanism to verify that the system is operating within acceptable limits. General handbooks that discuss quality assurance in more detail are given in References 1–3.

Quality is a subjective term. What is high quality in one situation may be low or unacceptable quality in another case. Clearly the tolerable limits of error must be established for each. Along with this there must be a clear understanding of the measurement process and its capability to provide the results desired.

The tolerance limits for the property to be measured are the first conditions to be determined. These are based upon the considered judgment of the end user of the data and represent the best estimate of the limits within which the measured property must be known, to be useful for its intended purpose. The limits must be realistic and defined on the basis of cost–benefit considerations. It is better to err on the side of too-narrow limits. Yet, measurement costs normally increase as tolerances are decreased, so that the number of measurements possible for a fixed budget may be inadequate when coupled with material-variability considerations.

Once one has determined the tolerance limits for the measured property, the permissible tolerances in measurement error may be established. The basis for this is shown in Figure 1. The tolerance limits for the measured property are indicated by L_p. Uncertainties in the measurement, based on the experience and judgment of the analyst, are indicated by C_m. These include estimates of the bounds for the biases (systematic errors), B, and the random errors as indicated by s, the estimate of the standard deviation. Obviously, C_m must be less than L_p if the data are to be useful. The confidence limits for $\bar{x}$, the mean of n replicate measurements, are:

$$C_m = \pm \left[B + \frac{ts}{\sqrt{n}} \right]$$

in which t is the so-called student factor. While the effect of random error is minimized by replication of measurements, there are practical limitations, and any measurement process that requires a large number of replicates has a serious disadvantage.

Well-designed and well-implemented quality assurance programs provide the means to operate a measurement system in a state of statistical control, thereby providing the basis for establishing reliable confidence limits for the data output.

Until a measurement operation . . . has attained a state of statistical control, it cannot be regarded in any logical sense as measuring anything at all.

C. E. Eisenhart

The Analytical System

Analytical measurements are made because it is believed that compositional information is needed for some end use in problem solving. Explicitly or implicitly, a measurement system such as that depicted in Figure 2 is involved. One must have full understanding of the measurement system for each specific situation in order to generate quality data.

The conceptualization of the problem, including the data requirements and their application, constitutes the model. The plan, based on the model, includes details of sampling, measurement, calibration, and quality assurance. Various constraints such as time, resources, and the availability of samples may necessitate compromises in the plan. Adequate planning will require the collaboration of the analyst, the statistician, and the end user of the data in all but the most routine

Published in Analytical Chemistry, December 1981, pp. 1588A–1596A, by the American Chemical Society

Chemical Measurements

cases. In complex situations, planning may be an iterative process in which the actual data output may require reconsideration of the model and revision of the plan.

Sampling has been discussed in a recent paper (*4*). Obviously, the sample is one of the critical elements of the measurement process. Closely related is the measurement methodology to be used. The method used must be adequate for the intended purpose and it must be properly utilized. The necessary characteristics of a suitable method include: adequate sensitivity, selectivity, accuracy, and precision. It is desirable that it also have the following characteristics: large dynamic measurement range; ease of operation; multiconstituent applicability; low cost; ruggedness; portability. To judge its suitability, the following information must be known about it: type of sample; forms determined; range of applicability; limit of detection; biases; interferences; calibration requirements; operational skills required; precision; and accuracy. Obviously all of the above characteristics must match the measurement requirements. In case of doubt, trial measurements must be made to demonstrate applicability to a given problem. A cost–benefit analysis may be needed to determine which of several candidate methods is to be selected. A method, once adopted, must be used in a reliable and consistent manner, in order to provide reproducible data. This is best accomplished by following detailed written procedures called Standard Operating Procedures (SOPs) in quality assurance terminology. Standard methods developed by voluntary standardization organizations are often good candidates for SOPs, when they are available.

Two kinds of calibrations are required in most cases. Physical calibrations may be needed for the measurement equipment itself and for ancillary measurements such as time, temperature, volume, and mass. The measurement apparatus may include built-in or auxiliary tests such as voltage checks, which may need periodic verification of their stability if not of their absolute values. But especially, most analytical equipment requires some kind of chemical calibration, often called standardization, to establish the analytical function (i.e., the relation of instrument response to chemical quantification). Obviously, the analyst must thoroughly understand each of the calibrations required for a particular measurement. This includes a knowledge of the standards needed and their relation to the measurement process, the frequency of calibration, the effect on a measurement system due to lack of calibration, and even the shock to the system resulting from recalibration.

Quality Control

Quality control encompasses all of the techniques used to encourage reproducibility of the output of the measurement system. It consists of the use of a series of protocols developed in advance and based on an intimate understanding of the measurement process and the definite requirements of the specific measurement situation. Protocols, i.e., procedures that must be rigorously followed, should be established for sampling, measurement, calibration, and data handling. Some of these, or at least selected portions, may be applicable to most or all of the measurements of a particular laboratory and become the basis of a good laboratory practices manual (GLPM).

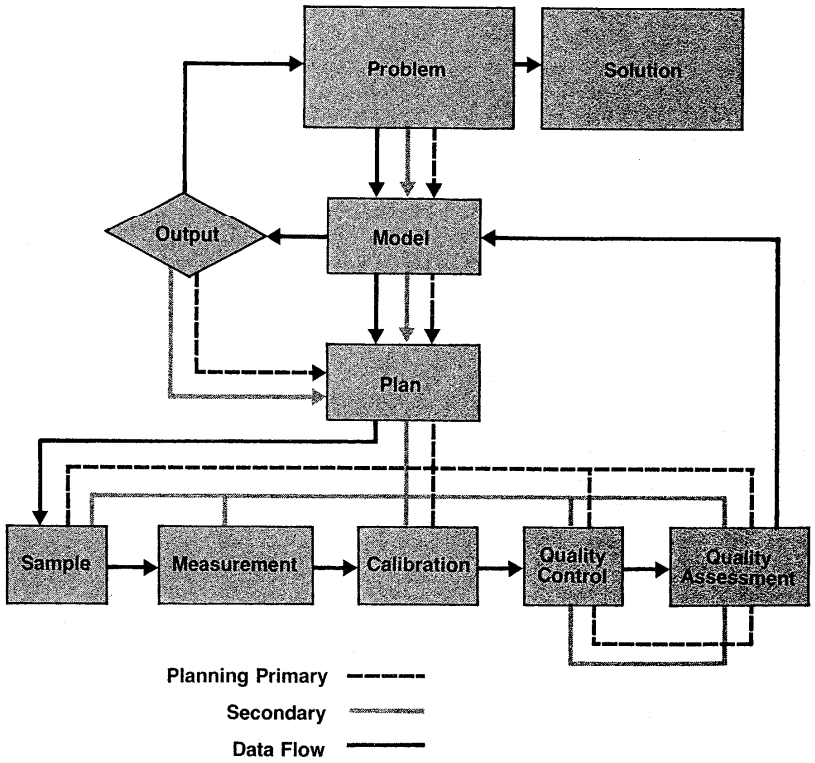


Figure 2. Analytical measurement system

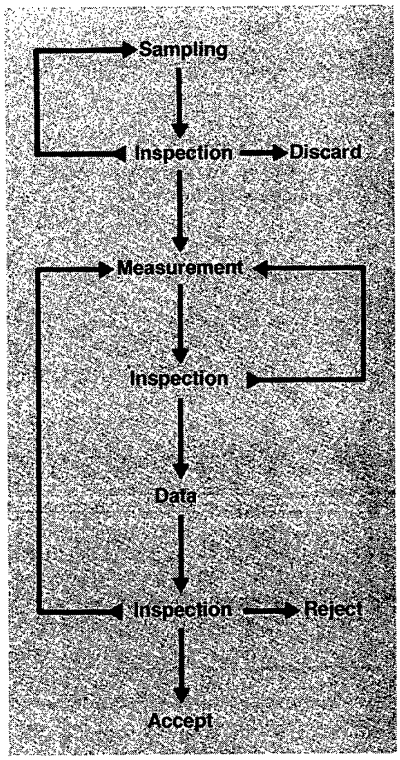


Figure 3. Quality control by inspection

In fact, the GLPM should cover the generalities, if not the specifics, of all measurement practices of the laboratory. The protocols for a specific measurement process include the GLPs together with any requirements of the specific situation.

The GLPM and protocols should be developed collaboratively by all of those involved in the measurements, and this development process may be the most important aspect of their function. It encourages a keen consideration of the measurement process and creates an awareness of potential problems that GLPs attempt to avert.

Protocols are of little use unless they are followed rigorously, and the attitudes of laboratory personnel are certainly key factors in this regard. Analysts must aspire to produce high quality data and must be their own most severe critics. Notwithstanding, good quality control systems should include provisions for inspection, both periodically and aperiodically (unannounced) to ascertain how well they are functioning. Large laboratories may have a quality control officer or group, independent of the laboratory management, that oversees the operation of the quality control system.

Quality Control by Inspection

An informal kind of quality control involves the frequent if not constant inspection of certain aspects of the measurement system for real or apparent problems (*5*). The essential features of such a system are depicted in Figure 3. Based on an intimate knowledge of the measurement process, samples may be casually inspected for their adequacy. The rejection and possible replacement of obviously unsuitable ones can eliminate not only extra work but also erroneous data that might be difficult to identify later. Difficulties in the actual measurement may often be identified as they occur and remedial measures, including remeasurement, may be taken either to save data that might otherwise be lost or at least to provide valid reasons for any rejections. Likewise, data inspection can identify problems and initiate remedial actions, including new measurements, while it is still possible to do so.

Control Charts

The performance of a measurement system can be demonstrated by the measurement of homogeneous and stable control samples in a planned repetitive process. The data so generated may be plotted as a control chart in a manner to indicate whether the measurement system is in a state of statistical control. Either the result of a single measurement on the control sample, the difference between duplicate measurements, or both may be plotted sequentially. The first mode may be an indicator of both precision and bias, while the second monitors precision only.

To effectively use such a chart, the standard deviation of a single measurement of the control sample must be known. This may be obtained by a series of measurements of the control sample, or it may be obtained from the experience of the laboratory on measuring similar samples. Control limits, i.e., the extreme values believed to be credible, are computed from the standard deviation. For example, the 2σ limit represents those within which the values are expected to lie 95% of the time. The 3σ limit represents the 99.7% confidence level. Departures from the former are warnings of possible trouble, while exceeding the latter usually means corrective action is needed. In the event that the standard deviation cannot be estimated with sufficient confidence initially, the control chart may be drawn using the best estimate, and the limits may be modified on the basis of increasing measurement experience.

The development of a control chart must include the rationale for its use. There must be a definite relation between the control measurements and the process they are designed to control. While the control chart only signifies the degree of replication of measurements of the control sample, its purpose is to provide confidence in the measurement process. To do this, the control measurements must simulate the measurements normally made. In chemical measurements, this means simulation of matrix, simulation of concentration levels, and simulation of sampling. The latter objective may be difficult if not impossible to achieve. It must be further emphasized that the control measurements should be random members of the measurement routine, or at least they should not occupy biased positions in any measurement sequence.

To the extent that control samples are representative of the test samples, and to the extent that measurements of them are representative of the measurement process, the existence of statistical control for these samples can imply such control of the measurement process and likewise of the results obtained for the test samples.

No specific statements can be made about the frequency of use of control samples. Until a measurement process is well understood, control samples may need to be measured frequently. As it is demonstrated to be in control, the need may become less and the incentive to do "extra" work may diminish. Along with the decision on how much effort should be devoted to quality control the risks and consequences of undetected loss of control must be weighed. Many laboratories consider that the 5–15% extra effort ordinarily required for all aspects of quality control is a small price to pay for the quality assurance it provides. When measurements are made on a frequently recurring schedule, internal controls, such as duplicate measurements of test samples, can provide evidence of reproducibility so that control samples may be used largely to identify systematic errors, drifts, or other types of problems.

When laboratories are engaged in a variety of measurements, the use of representative control samples may be difficult if not impossible. In such cases, often only the measurement methodology can be tested, and evaluation of the quality of the measurement output requires considerable judgment. In such cases, the experience of the lab becomes a key factor.

In some complex measurement systems, certain steps or subsystems are more critical than others, and hence it may be more important to develop control charts for them than for the entire system. The control of such steps may indeed prevent propagation of error into the end result. An example is the sampling step, which may be very critical with respect to the end result. In such a case, the records of periodic inspections may be adaptable to the control chart technique of quality control.

Quality Assessment

Procedures used to evaluate the effectiveness of the quality control system may be classified according to whether the evidence arises from internal or external sources. Internal procedures, useful largely for estimating precision, include the use of internal reference samples and control charts to monitor the overall performance of the measurement system as described in an earlier section. Replicate measurements on replicate or split samples can provide valuable insight into the reproducibility of both the measurement and sampling processes. Comparison of the results obtained as a consequence of interchange of analysts, equipment, or combinations of these can attest to operational stability as well as identify malfunctions. Measurements made on split samples using a completely independent method can lend confidence to the method normally in use or indicate the presence of measurement bias.

External quality assessment is always needed since it can detect problems of bias that are difficult to identify by internal procedures. Participation in collaborative tests, exchange of samples with other laboratories, and the use of certified reference materials are time-honored assessment devices. NBS Standard Reference Materials (SRMs) (*6*) are especially useful for quality assessment in cases where they are available and applicable. The information that can be obtained or inferred by their use is described in a later section. Operators of monitoring networks may provide proficiency testing or audit samples to assess laboratory performance. Ordinary practices should be used here, so that normal rather than optimum performance is measured.

A laboratory should diligently use the information obtained in the quality assessment process. Adverse data should not be treated in a defensive manner but the reason for it should be investigated objectively and thoroughly. When laboratory records are reliably and faithfully kept, the task of identifying causes of problems is made easier. This is an important reason for developing data handling protocols and ensuring that all protocols are strictly followed.

Systematic Errors

Systematic errors or biases are of two kinds—concentration-level independent (constant), and concentation-level related. The former are sometimes called additive while the latter are called multiplicative. Both kinds may be present simultaneously in a given measurement system. An example of the first kind is the reagent blank often present in measurements involving chemical processing steps. The second kind can result from, for example, use of an inaccurately certified calibrant.

Systematic errors may arise from such sources as faulty calibrations, the use of erroneous physical constants, incorrect computational procedures, improper units for measurement or reporting data, and matrix effects on the measurement. Some of these can be eliminated or minimized by applying corrections or by modification of the measurement technique. Others may be related to fundamental aspects of the measurement process. The most insidious sources of error are those unknown or unsuspected of being present.

One of the most important sources of error in modern instrumental measurements concerns uncertainties in the calibrants used to define the analytical function of the instrument. The measurement step essentially consists of the comparison of an unknown with a known (calibrant) so that any error in the latter results in a proportional error in the former. The need to use calibrants of the highest reliability is obvious.

The measurement protocol should include a detailed analysis of the sources of error and correction for them to the extent possible. The uncertainties, *B*, referred to earlier, represent the uncertainties in the corrections for the systematic errors. In making such an estimate, the 95% confidence limits should be assigned to the extent possible. The magnitudes of these uncertainties can be estimated from those assigned by others in the case of such factors as calibration standards and physical constants. Other constant sources of error may be more subtle both to identify and to evaluate, and the judgment and even intuition of the experimenter may be the only sources of information.

The effectiveness of elimination of, or correction for, systematic errors is best evaluated from external quality assessment procedures. Differences found between known and measured values of test samples, such as SRMs, need to be reconciled with the laboratory's own estimates of bounds for its random and systematic errors. When the random error is well established, as by the quality control process, significant discrepancies can be attributed to unsuspected or incorrectly estimated systematic errors.

The Use of SRMs for Quality Assessment

An SRM is a material for which the properties and composition are certified by the National Bureau of Standards (*6, 7*). To the extent that its compositional properties simulate those of the sample ordinarily measured, its "correct" measurement can imply "correct" measurement of the usual samples. Such a conclusion requires that the protocol of measurement was the same in each case. Hence it is necessary that no special care be exercised in measuring the SRM, other than that ordinarily used.

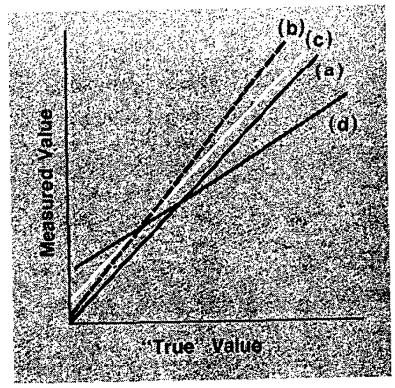


Figure 4. Typical analytical systematic errors (bias). (a) = unbiased; (b) = measurement-level related; (c) = constant error; and (d) = combination of b and c

Analysis of SRMs has been recommended as a means of providing "traceability" to national measurement standards. However, a word of caution is appropriate on this point. Measurement processes are seldom identical, so that traceability is most often based on inference. Also, the fact that an acceptable result is or is not obtained for an SRM provides no unique explanation for such a result.

The use of an SRM should never be attempted until the analytical system has been demonstrated to be in a state of statistical control. An SRM is not needed for such a purpose and such use is discouraged. Ordinarily, the SRM will be available in limited amount so that the statistics of the measurement process should be demonstrated by measurements on other materials. Only under such a situation can the results of an SRM measurement be considered as representative of the measurement system.

A consideration of the nature of analytical errors, shown in Figure 4, will clarify why the measurement of a single SRM may not be fully informative. It will be noted that errors may be constant, measurement-level related, or a combination of these, and a single right or wrong result will not indicate on which of several possible curves it might lie. Measurement of a series of SRMs may clarify the nature of the measurement process and this should be done whenever possible. An intimate understanding of the operation of a particular measurement system may also make it possible to eliminate some of the possible sources of error and to better interpret the data from measurement of SRMs.

Record Keeping

Adequate record keeping in an easily retrievable manner is an essential part of the quality assurance program. Records needed include the description of test samples, experimental procedures, and data on calibration and testing. Quality control charts should be diligently prepared and stored. A chain of custody of test materials should be operative and such materials should be retained and safeguarded until there is no doubt about their future use or need.

Data Control

The evaluation, review, and release of analytical data is an important part of the quality assurance process. No data should be released for external use until it has been carefully evaluated. Guidelines for data evaluation, applicable to almost every analytical situation, have been developed by the ACS Committee on Environmental Improvement (*8*). A prerequisite for release of any data should be the assignment of uncertainty limits, which requires the operation of some kind of a quality assurance program. Formal release should be made by a professional analytical chemist who certifies that the work was done with reasonable care and that assigned limits of uncertainty are applicable.

Laboratory Accreditation

Laboratory accreditation is one form of quality assurance for the data output of certified laboratories. Accreditation is based on criteria that are considered essential to generate valid data and is a formal recognition that the laboratory is competent to carry out a specific test or specific type of test (*9*, *10*). The certification is as meaningful as the care exercised in developing certification criteria and evaluating laboratory compliance. Generic criteria developed by national and international standardization organizations have been influential in this respect (*11*). These criteria are well conceived and provide general guidance for the sound operation of analytical laboratories, whether or not certification is involved.

Implementation

Detailed quality assurance plans are ineffective unless there is commitment to quality by all concerned. This commitment must be total, from management to technical staff. The former must provide the resources, training, facilities, equipment, and encouragement required to do quality work. The latter must have the technical ability and motivation to produce quality data. Some may argue that if there is such commitment, there is no need for a formal quality assurance program. However, the experience of many laboratories has demonstrated that a formal quality assurance program provides constant guidance for the attainment of the quality goals desired.

References

(1) "Quality Assurance Handbook for Air Pollution Measurement Systems: Principles"; Vol. 1, E.P.A. Publication No. 600/9-76-005.

(2) "Handbook for Analytical Quality Control in Water and Wastewater Laboratories"; EPA Publication No. 600/4-79-019.

(3) Juran, J. M. "Quality Control Handbook," 3rd ed.; McGraw-Hill: New York, 1974.

(4) Kratochvil, B. G.; Taylor, J. K. *Anal. Chem.* **1981**, *53*, 924 A.

(5) Brewers, J. M., et al. "Data Are for Looking At, or Quality Control by Interpretation." In "Water Quality Paramters"; American Society for Testing and Materials: Philadelphia, 1975, ASTM STP573.

(6) "Catalogue of NBS Standard Reference Materials"; National Bureau of Standards: Washington, NBS Special Publication 260.

(7) Cali, J. P., et al. "The Role of Standard Reference Materials in Measurement Systems"; National Bureau of Standards: Washington, January 1975, NBS Monograph 148.

(8) ACS Subcommittee on Environmental Analytical Chemistry. "Guidelines for Data Acquisition and Data Quality Evaluation in Analytical Chemistry"; *Anal. Chem.* **1980**, *52*, 2242.

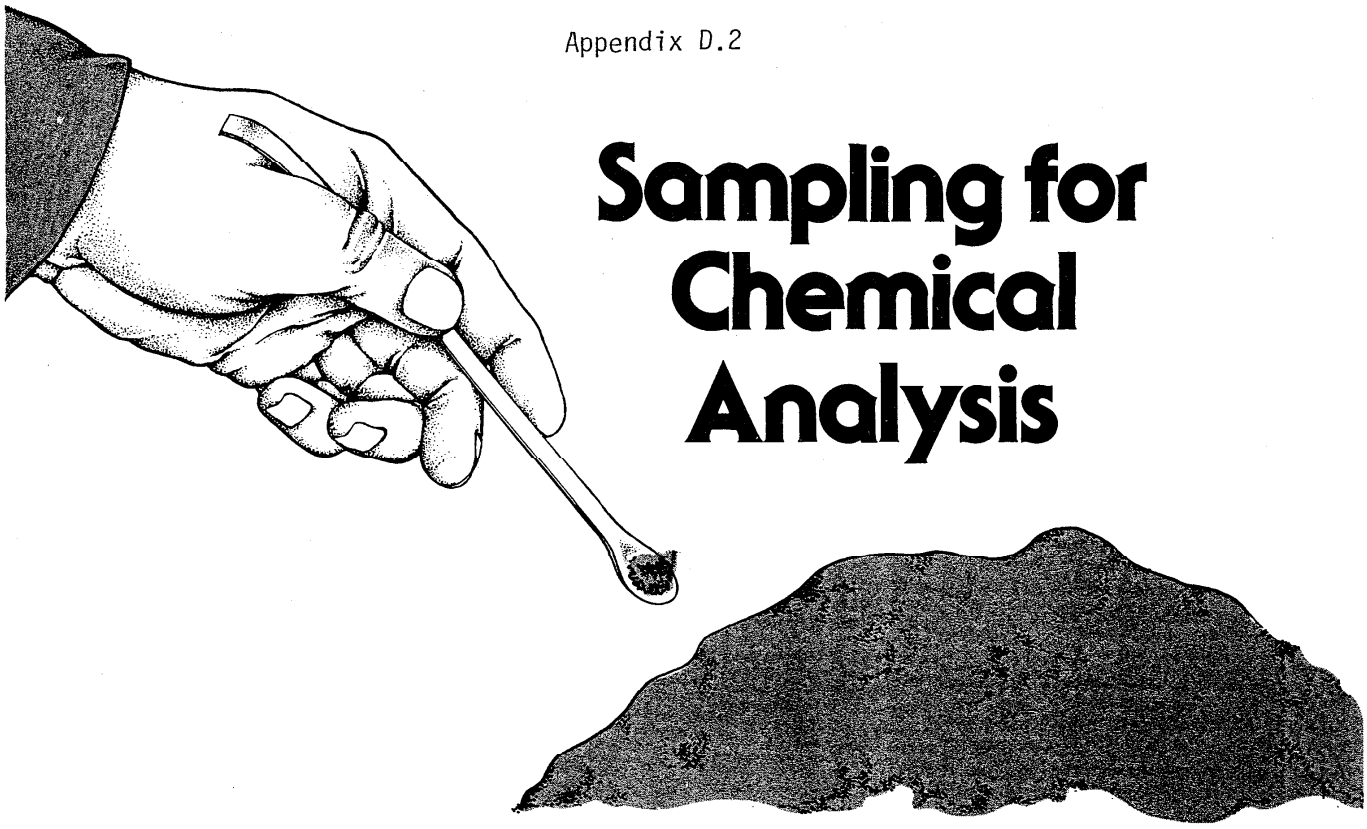
(9) "Quality Control System-Requirements for a Testing and Inspections Laboratory"; American Council of Independent Laboratories: Washington.

(10) "Testing Laboratory Performance: Evaluation and Accreditation"; Berman, G. A., Ed.; National Bureau of Standards: Washington, NBS Special Publication 591.

(11) "Standard Recommended Practice for Generic Criteria for Use in Evaluation of Testing and/or Inspection Agencies"; American Society for Testing and Materials: Philadelphia, ASTM Publication No. E-548.

John K. Taylor, coordinator for quality assurance and voluntary standardization activities at the National Bureau of Standards Center for Analytical Chemistry, received his BS from George Washington University, and his MS and PhD degrees from the University of Maryland. His research interests include electrochemical analysis, refractometry, isotope separations, standard reference materials, and the application of physical methods to chemical analysis.

Appendix D.2



Sampling for Chemical Analysis

A major consideration in the reliability of any analytical measurement is that of sample quality. Too little attention is directed to this matter. The analyst often can only report results obtained on the particular test specimen at the moment of analysis, which may not provide the information desired or needed. This may be because of uncertainties in the sampling process, or in sample storage, preservation, or pretreatment prior to analysis. The sampling plan itself is often so poorly considered as to make relation of the analytical results to the population from which the sample was drawn uncertain, or even impossible to interpret.

All of the above aspects of sampling merit full consideration and should be addressed in every analytical determination. Because the scope is so broad, we will limit the present discussion to a small segment of the total problem, that of sampling bulk materials. For such materials the major steps in sampling are:

- identification of the population from which the sample is to be obtained,
- selection and withdrawal of valid gross samples of this population, and
- reduction of each gross sample to a laboratory sample suitable for the analytical techniques to be used.

The analysis of bulk materials is one of the major areas of analytical activity. Included are such problems as the analysis of minerals, foodstuffs, environmentally important substances, and many industrial products. We shall discuss the major considerations in designing sampling programs for such materials. While our discussion is specifically directed toward solid materials, extension to other materials will often be obvious.

A brief list of definitions commonly used in bulk sampling is provided in the glossary.

Preliminary Considerations in Sampling

Poor analytical results may be caused in many ways—contaminated reagents, biased methods, operator errors in procedure or data handling, and so on. Most of these sources of error can be controlled by proper use of blanks, standards, and reference samples. The problem of an invalid sample, however, is special; neither control nor blank will avail. Accordingly, sampling uncertainty is often treated separately from other uncertainties in an analysis. For random errors the overall standard deviation, s_o, is related to the standard deviation for the sampling operation, s_s, and to that for the remaining analytical operations, s_a, by the expression: $s_o^2 = s_a^2 + s_s^2$. Whenever possible, measurements should be conducted in such a way that the components of variance arising from sample variability and measurement variability can be separately evaluated. If the measurement process is demonstrated to be in a state of statistical control so that s_a is already known, s_s can be evaluated from s_o, found by analysis of the samples. Otherwise, an appropriate series of replicate measurements or replicate samples can be devised to permit evaluation of both standard deviations.

Youden has pointed out that once the analytical uncertainty is reduced to a third or less of the sampling uncertainty, further reduction in the analytical uncertainty is of little importance (*1*). Therefore, if the sampling uncertainty is large and cannot be reduced, a rapid, approximate analytical method may be sufficient, and further refinements in the measurement step may be of negligible aid in improving the overall results. In fact, in such cases a rapid method of low precision that permits more samples to be examined may be the best route to reducing the uncertainty in the average value of the bulk material under test.

An excellent example of the importance of sampling is given in the deter-

Published in Analytical Chemistry, July 1981, pp. 924A–938A, by the American Chemical Society

Report

Byron Kratochvil
Department of Chemistry
University of Alberta
Edmonton, Alberta, Canada T6G 2G2

John K. Taylor
National Bureau of Standards
Washington, D.C. 20234

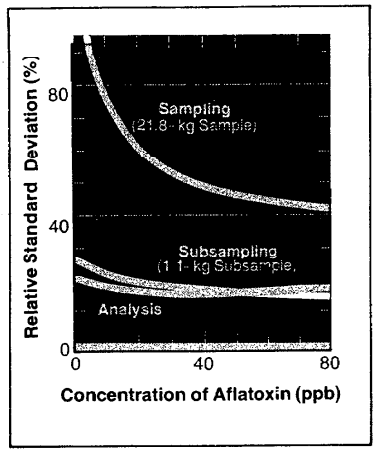


Figure 1. Relative standard deviation associated with the sampling and analysis operations in testing peanuts for aflatoxins (after T. B. Whittaker, *Pure and Appl. Chem.*, **49**, 1709 (1977))

mination of aflatoxins in peanuts (2). The aflatoxins are highly toxic compounds produced by molds that grow best under warm, moist conditions. Such conditions may be localized in a warehouse, resulting in a patchy distribution of highly contaminated kernels. One badly infected peanut can contaminate a relatively large lot with unacceptable levels (above about 25 ppb for human consumption) of aflatoxins after grinding and mixing. The standard deviations of the three operations of sampling, subsampling, and analysis are shown in Figure 1. The analytical procedure consists of solvent extraction followed by thin-layer chromatography and measurement of the fluorescence of the aflatoxin spots. Clearly, sampling is the major source of the analytical uncertainty.

Types of Samples

Random Samples. In common with the statistician, the analytical chemist ordinarily wishes to generalize from a small body of data to a larger body of data. While the specimen/sample actually examined is sometimes the only matter of interest, the characteristics of the population of specimens are frequently desired. Obviously, the samples under examination must not be biased, or any inferences made from them will likewise be biased.

Statisticians carefully define several terms that are applied to statistical inference. The *target population* denotes the population to which we would like our conclusions to be applicable, while the *parent population* designates that from which samples were actually drawn. In practice these two populations are rarely identical, although the difference may be small. This difference may be minimized when the selection of portions for examination is done by a random process. In such a process each part of the population has an equal chance of being selected. Thus, random samples are those obtained by a random sampling process and form a foundation from which generalizations based on mathematical probability can be made.

Random sampling is difficult. A sample selected haphazardly is not a random sample. On the other hand, samples selected by a defined protocol are likely to reflect the biases of the protocol. Even under the most favorable circumstances, unconscious selection and biases can occur. Also, it can be difficult to convince untrained individuals assigned the task of obtaining samples that an apparently unsystematic collection pattern must be followed closely for it to be valid.

Whenever possible, the use of a table of random numbers is recommended as an aid to sample selection. The bulk material is divided into a number of real or imaginary segments. For example, a body of water can be conceptually subdivided into cells, both horizontally and vertically, and the cells to be sampled selected randomly. To do this each segment is assigned a number, and selection of segments from which sample increments are to be taken is made by starting in an arbitrary place in a random number table and choosing numbers according to a predecided pattern. For example, one could choose adjacent, alternate, or *n*th entries and sample those segments whose numbers occur until all of the samples decided upon have been obtained.

The results obtained for these and other random samples can be analyzed by some model or plan to identify whether systematic relations exist. This is important because of the possible introduction of apparent correlations due to systematic trends or biases in the measurement process. Accordingly, measurement plans should always be designed to identify and minimize such problems.

Despite the disadvantages, sampling at evenly spaced intervals over the bulk is still often used in place of random sampling owing to its simplic-

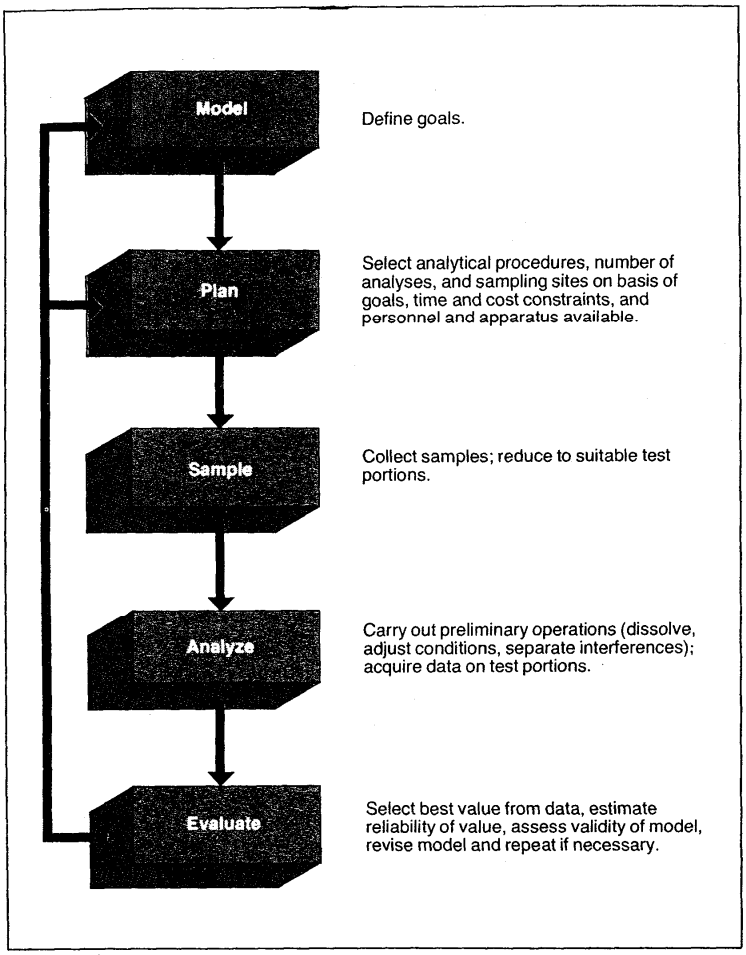


Figure.2. The place of sampling in the overall analytical process

ity. Because this procedure is more subject to bias than random sampling, it is not recommended. If it is used, the results must be closely monitored to ensure that errors from periodicity in the material are not introduced.

Systematic Samples. Frequently, samples are obtained and analyzed to reflect or test some systematic hypothesis, such as changes in composition with time, temperature, or spatial location. Such samples, if collected in a systematic manner, may each be considered to represent a separate discrete population under the existing conditions. However, the results may still be statistically tested for the significance of any apparent differences.

In a carefully designed sampling plan, consideration should be given to the possible concurrence of unanticipated events or phenomena that could prejudice the information on the sample measured. For example, measurements to be taken at time intervals are sometimes made with a random start or other superimposed random time element. Needless to say, the less known about a given process, the more randomness is merited. Conversely, as a process is more fully understood, systematic approaches can provide maximum efficiency of data acquisition.

Representative Samples. The term "representative sample" is frequently used in analytical discussions to connote a single sample of a universe or population (e.g., waste pile, lagoon, ground water) that can be expected to exhibit average properties of the population (see glossary). Obviously, such a sample cannot be selected by a random process. And even if it could, to ascertain the validity of its representativeness would require considerable effort.

The concept of a truly representative sample would appear to be valid in only two cases. The first case involves samples defined *a priori* as representative for a specific purpose. For example, the Hazardous Waste Management System prescribes seven protocols for sampling wastes—ranging from viscous liquids, solids, or containerized liquids to reservoirs—to provide samples that "will be considered by the Agency (EPA) to be representative of the waste" (*3*). The second case involves the sampling of truly homogeneous materials.

While the measurement of samples defined as representative may reduce analytical costs, the information so obtained ordinarily does not enjoy the status of that obtained from valid random samples of the population. An exception is when effort has been vigorously exerted to homogenize the population prior to sampling. Such processes are difficult and are ordinarily only justified when the objective is to produce a number of subsamples of essentially similar properties.

Because of the difficulties of selecting or producing a "representative sample" it is recommended that this concept be discouraged for general purposes and reserved only for cases where the effort required to prepare such a sample is justified. An appreciation of the compositional information that is lost as a result is a further reason to discourage the practice. With a properly designed and executed random sampling plan, the valuable characteristics of sample mean and variation between members can be ascertained, neither of which can be obtained by measurement of one "representative sample."

Composite Samples. A composite sample (see glossary) may be considered as a special way of attempting to produce a representative sample. Many sampling procedures are based on the assumption that average composition is the only information desired. Such averages may be bulk averages, time-weighted averages, and flow-proportional averages, for example, and may be obtained by measurement of a composite, suitably prepared or collected. Elaborate procedures involving crushing, grinding, mixing, and blending have been developed and even standardized for the preparation of solid composites, while sampling systems for liquids (especially water) have been developed to obtain various composite samples.

Analysis of a number of individual samples permits determination of the average (at the expense of extra analytical effort) and the distribution of samples within the population (between-sample variability). In some cases, it may be of interest to isolate the within-sample variability as well. All this information is necessary for collaborative test samples and in reference material usage, especially when apparent differences in analytical results within and between laboratories need to be evaluated.

Because of the limited information provided by a composite sample, full consideration should be given to the consequences before deciding between this approach and the analysis of individual samples.

Subsampling. Usually, the sample received by the analytical laboratory will be larger than that required for a single measurement, so some subsampling (see glossary) will be required. Often, test portions (see glossary) must be taken for replicate measurements or for measurement of different constituents by several techniques. Obviously, such test portions must be sufficiently alike that the results are compatible. Frequently it is necessary to reduce particle size, mix, or otherwise process the laboratory sample (see glossary) before withdrawing portions (subsamples) for analysis. The effort necessary at this stage depends on the degree of homogeneity of the original sample. In general, the subsampling standard deviation should not exceed one-third of the sampling standard deviation. Although this may sound appreciable, it is wasteful of time and effort to decrease it below this level. But this does not mean care is unnecessary in subsampling. If a sample is already homogeneous, care may be needed to avoid introducing segregation during subsampling. Even though analysts may not be involved with sample collection, they should have sufficient knowledge of sampling theory to subsample properly. They should also be provided with any available information on the homogeneity of the samples received so that they can subsample adequately and efficiently.

Table I. Confidence Intervals and Statistical Tolerance Limits[a]

n	t[b]	$\frac{ts}{\sqrt{n}}$[c]	K[c]	Ks
2	12.70	±18	37.67	±75
4	3.18	±3.2	6.37	±12.9
8	2.36	±1.7	3.73	±7.4
16	2.13	±1.1	2.90	±5.8
32	2.04	±0.7	2.50	±5.0
100	1.98	±0.4	2.23	±4.4
∞	1.96	0	1.96	±4.0

[a] Calculated for $s = 2$, based on measurement of n samples
[b] 95% confidence limits for the mean of n samples
[c] Based on a 95% confidence that the interval will contain 95% of the samples

Model of the Sampling Operation

Before sampling is begun, a model of the overall operation should be established (Figure 2). The model should consider the population to be studied, the substance(s) to be measured, the extent to which speciation is to be determined, the precision required, and the extent to which the distribution of the substance within the population is to be obtained.

The model should identify all assumptions made about the population under study. Once the model is complete, a sampling plan can be established.

The Sampling Plan

The plan should include the size, number, and location of the sample increments and, if applicable, the extent of compositing to be done. Procedures for reduction of the gross sample (see glossary) to a laboratory sample, and to the test portions, should be specified. All of this should be written as a detailed protocol before work is begun. The protocol should include procedures for all steps, from sampling through sample treatment, measurement, and data evaluation; it should be revised as necessary during execution as new information is obtained. The guidelines for data acquisition and quality evaluation in environmental chemistry set out by the ACS Subcommittee on Environmental Analytical Chemistry are sufficiently general to be recommended reading for workers in all fields (*4*).

The sampling protocol should include details of when, where, and how the sample increments are to be taken. On-site criteria for collection of a valid sample should be established beforehand. Frequently, decisions must be made at the time of sampling as to components likely to appear in the sample that may be considered foreign, that is, not part of the population. For example, a portion of dredged sediment in which the mercury content is to be determined might contain cans, discarded shoes, rocks or other extraneous material. For the information sought these items might be considered foreign and therefore legitimately rejected. Decisions as to rejection become less clear with smaller items. Should smaller stones be rejected? How small? And what about bits of metal, glass, leather, and so on? Criteria for such decisions should be made logically and systematically, if possible before sampling is initiated.

The type of container, cleaning procedure, and protection from contamination before and after sampling must be specified. The question of sample preservation, including possible addition of preservatives and refrigeration, should be addressed. Some sampling plans call for field blanks and/or field-spiked samples. The critical nature of the latter and the difficulties possible under field conditions require the utmost care in planning and execution of the sampling operation if the results are to be meaningful.

Whenever possible, the analyst should perform or directly supervise the sampling operation. If this is not feasible, a written protocol should be provided and the analyst should ensure that those collecting the samples are well-trained in the procedures and in use of the sampling equipment, so that bias and contamination are minimized. No less important is careful labeling and recording of samples. A chain of custody should be established such that the integrity of the samples from source to measurement is ensured. Often auxiliary data must be recorded at the time the sample is taken: temperature, position of the collecting probe in the sample stream, flow velocity of the stream, and so on. Omission or loss of such information may greatly decrease the value of a sample, or even render it worthless.

Sampling Bulk Materials. Once the substances to be determined, together with the precision desired, have been specified, the sampling plan can be designed. In designing the plan, one must consider:

- How many samples should be taken?
- How large should each be?
- From where in the bulk material (population) should they be taken?
- Should individual samples be analyzed, or should a composite be prepared?

These questions cannot be answered accurately without some knowledge of the relative homogeneity of the system. Gross samples should be unbiased with respect to the different sizes and types of particles present in the bulk material. The size of the gross sample is often a compromise based on the heterogeneity of the bulk material on the one hand, and the cost of the sampling operation on the other.

When the properties of a material

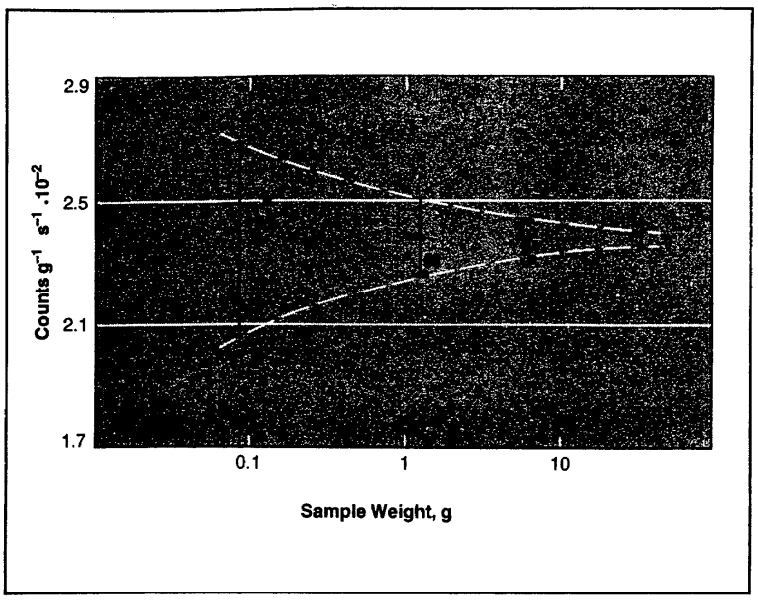


Figure 3. Sampling diagram of sodium-24 in human liver homogenate (from Reference 7)

to be sampled are unknown, a good approach is to collect a small number of samples, using experience and intuition as a guide to making them as representative of the population as possible, and analyze for the component of interest. From these preliminary analyses, the standard deviation s_s of the individual samples can be calculated, and confidence limits for the average composition can be established using the relation

$$\mu = \bar{x} \pm ts_s/\sqrt{n} \qquad (1)$$

where μ is the true mean value of the population, $\bar{x}$ is the average of the analytical measurements, and t is obtained from statistical tables for n measurements (often given as $n - 1$ degrees of freedom) at the desired level of confidence, usually 95%. Table I lists some t values; more extensive tables are provided in books on quantitative analysis and statistics (5).

On the basis of this preliminary information, a more refined sampling plan can be devised, as described in the following sections. After one or two cycles the parameters should be known with sufficient confidence that the optimum size and number of the samples can be estimated with a high level of confidence. The savings in sampling and analytical time and costs by optimizing the sampling program can be considerable.

Minimum Size of Individual Increments. Several methods have been developed for estimation of the amount of sample that should be taken in a given increment so as not to exceed a predetermined level of sampling uncertainty. One approach is through use of Ingamells's sampling constant (6). Based on the knowledge that the between-sample standard deviation s_s (Equation 1), decreases as the sample size is increased, Ingamells has shown that the relation

$$WR^2 = K_s \qquad (2)$$

is valid in many situations. In Equation 2, W represents the weight of sample analyzed, R is the relative standard deviation (in percent) of sample composition, and K_s is the sampling constant, corresponding to the weight of sample required to limit the sampling uncertainty to 1% with 68% confidence. The magnitude of K_s may be determined by estimating s_s from a series of measurements of samples of weight W.

Once K_s is evaluated for a given sample, the minimum weight W required for a maximum relative standard deviation of R percent can be readily calculated.

An example of an Ingamells sampling constant diagram is shown in Figure 3 for a human liver sample under study in the National Environmental Specimen Bank Pilot Program at the National Bureau of Standards (NBS) in conjunction with the Environmental Protection Agency (7). A major goal of the program is to evaluate specimen storage under different conditions. This requires analysis of small test portions of individual liver specimens. The material must be sufficiently homogeneous that variability between test portions does not mask small variations in composition owing to changes during storage. The homogeneity of a liver sample for sodium was assessed by a radiotracer study in which a portion was irradiated, added to the remainder of the specimen, and the material homogenized. Several test portions were then taken and the activity of ^{24}Na measured as an indicator of the distribution of sodium in the samples. From Figure 3 it can be seen that the weight of sample required to yield an inhomogeneity of 1% (±2.4 counts g^{-1}s^{-1}) is about 35 g. For a subsample of one gram, a sampling uncertainty of about 5% can be expected.

Minimum Number of Individual Increments. Unless the population is known to be homogeneous, or unless a representative sample is mandated by some analytical problem, sufficient replicate samples (increments) must be analyzed. To determine the minimum number of sample increments, a sampling variance is first obtained, either from previous information on the bulk material or from measurements made on the samples. The number of samples necessary to achieve a given level of confidence can be estimated from the relation

$$n = \frac{t^2 s_s^2}{R^2 \bar{x}^2} \qquad (3)$$

where t is the student's t-table value for the level of confidence desired, s_s^2 and $\bar{x}$ are estimated from preliminary measurements on or from previous knowledge of the bulk material, and R is the percent relative standard deviation acceptable in the average. Initially t can be set at 1.96 for 95% confidence limits and a preliminary value of n calculated. The t value for this n can then be substituted and the system iterated to constant n. This expression is applicable if the sought-for component is distributed in a positive binomial, or a Gaussian, distribution. Such distributions are characterized by having an average, μ, that is larger than the variance, σ_s^2. Remember that values of σ_s (and s_s) may depend greatly on the size of the individual samples.

Two other distributions that may be encountered, particularly in biological materials, should be mentioned. One is the Poisson distribution, in which the sought-for substance is distributed randomly in the bulk material such that σ_s^2 is approximately equal to μ. In this case

$$n = \frac{t^2}{R^2 \bar{x}} \qquad (4)$$

The other is the negative binominal distribution, in which the sought-for substance occurs in clumps or patches, and σ_s^2 is larger than μ. This pattern often occurs in the spread of contamination or contagion from single

sources, and is characterized by two factors, the average, $\bar{x}$, and a term, k, called the index of clumping. For this system

$$n = \frac{t^2}{R^2}\left[\frac{1}{\bar{x}} + \frac{1}{k}\right] \quad (5)$$

Here k must be estimated, along with $\bar{x}$, from preliminary measurements on the system.

Sometimes, what is wanted is not an estimate of the mean but instead the two outer values or limits that contain nearly all of the population values. If we know the mean and standard deviation, then the intervals $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ contain 95% and 99.7%, respectively, of all samples in the population. Ordinarily, the standard deviation σ is not known but only its estimate s, based on n observations. In this case we may calculate statistical tolerance limits of the form $\bar{x} + Ks$ and $\bar{x} - Ks$, with the factor K chosen so that we may expect the limits to include at least a fraction P of the samples with a stated degree of confidence. Values for the factor K (*8*) depend upon the probability γ of including the proportion P of the population, and the sample size, n. Some values of K are given in Table I. For example, when $\gamma = 0.95$ and $P = 0.95$, then $K = 3.38$ when $n = 10$, and $K = 37.67$ for duplicates ($n = 2$).

Sampling a Segregated (Stratified) Material. Special care must be taken when assessing the average amount of a substance distributed throughout a bulk material in a nonrandom way. Such materials are said to be segregated. Segregation may be found, for example, in ore bodies, in different production batches in a plant, or in samples where settling is caused by differences in particle size or density.

The procedure for obtaining a valid sample of a stratified material is as follows (*9*):

- Based on the known or suspected pattern of segregation, divide the material to be sampled into real or imaginary segments (strata).
- Further divide the major strata into real or imaginary subsections and select the required number of samples by chance (preferably with the aid of a table of random numbers).
- If the major strata are not equal in size, the number of samples taken from each stratum should be proportional to the size of the stratum.

In general, it is better to use stratified random sampling rather than unrestricted random sampling, provided the number of strata selected is not so large that only one or two samples can be analyzed from each stratum. By keeping the number of strata sufficiently small that several samples can be taken from each, possible variations within the parent population can be detected and assessed without increasing the standard deviation of the sampling step.

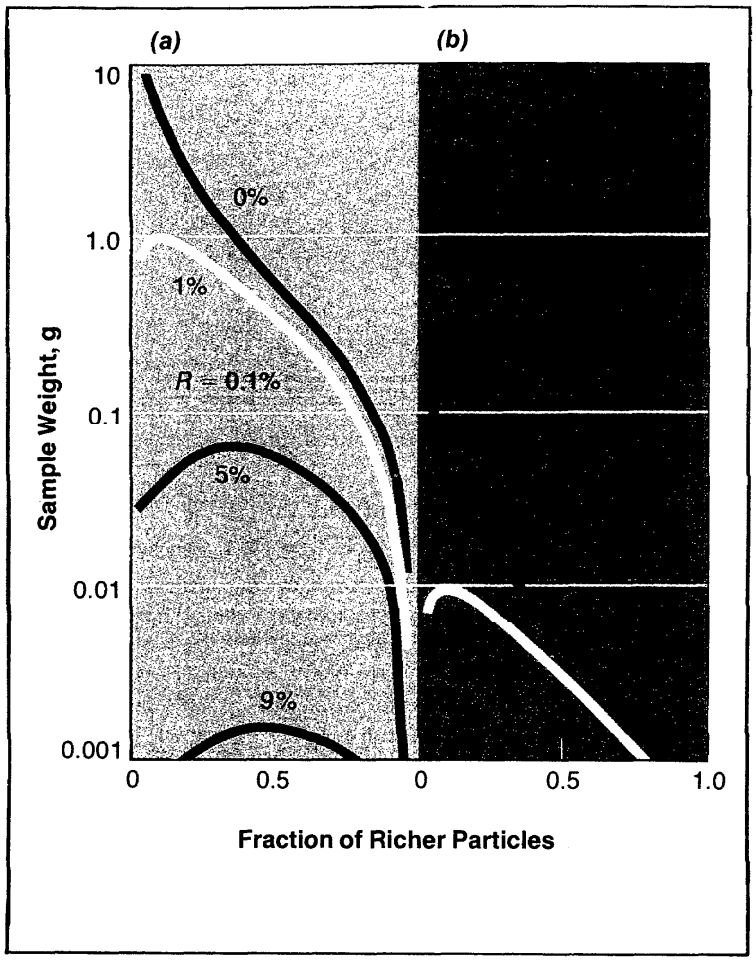


Figure 4. Relation between minimum sample size and fraction of the richer particles in a mixture of two types of spherical particles (diameter 0.1 mm and density 1) for a sampling standard deviation (R) of (a) 0.1% and (b) 1%. Richer particles contain 10% of substance of interest, and leaner ones contain 0, 1, 5, or 9% (after Reference 12, p 554)

Minimum Number of Individual Increments. When a bulk material is highly segregated, a large number of samples must be taken from different segments. A useful guide to estimating the number of samples to be collected is given by Visman (*10*), who proposed that the variance in sample composition depends on the degree of homogeneity within a given sample increment and the degree of segregation between sample increments according to the relation

$$s_s^2 = A/W + B/n \quad (6)$$

where s_s^2 is the variance of the average of n samples using a total weight W of sample, and A and B are constants for a given bulk material. A is called a homogeneity constant, and can be calculated from Ingamells's sampling constant and the average composition by

$$A = 10^4 \bar{x}^2 K_s \quad (7)$$

Sampling Materials in Discrete Units. If the lot of material under study occurs in discrete units, such as truckloads, drums, bottles, tank cars, or the like, the variance of the analytical result is the sum of three contributions: (1) that from the variance between units in the lot, (2) that from the average variance of sets of samples taken from within one unit, and (3) that from the variance of the analytical operations. The contribution from each depends upon the number of units in the lot and the number of samples taken according to the following relation (*9*):

$$\sigma_x^2 = \frac{\sigma_b^2}{n_b}\frac{(N - n_b)}{N} + \frac{\sigma_w^2}{n_b n_w} + \frac{\sigma_t^2}{n_t} \quad (8)$$

Glossary

Bulk sampling—sampling of a material that does not consist of discrete, identifiable, constant units, but rather of arbitrary, irregular units.

Composite—a sample composed of two or more increments.

Gross sample (also called bulk sample, lot sample)—one or more increments of material taken from a larger quantity (lot) of material for assay or record purposes.

Homogeneity—the degree to which a property or substance is randomly distributed throughout a material. Homogeneity depends on the size of the units under consideration. Thus a mixture of two minerals may be inhomogeneous at the molecular or atomic level, but homogeneous at the particulate level.

Increment—an individual portion of material collected by a single operation of a sampling device, from parts of a lot separated in time or space. Increments may be either tested individually or combined (composited) and tested as a unit.

Individuals—conceivable constituent parts of the population.

Laboratory sample—a sample, intended for testing or analysis, prepared from a gross sample or otherwise obtained. The laboratory sample must retain the composition of the gross sample. Often reduction in particle size is necessary in the course of reducing the quantity.

Lot—a quantity of bulk material of similar composition whose properties are under study.

Population—a generic term denoting any finite or infinite collection of individual things, objects, or events in the broadest concept; an aggregate determined by some property that distinguishes things that do and do not belong.

Reduction—the process of preparing one or more subsamples from a sample.

Sample—a portion of a population or lot. It may consist of an individual or groups of individuals.

Segment—a specifically demarked portion of a lot, either actual or hypothetical.

Strata—segments of a lot that may vary with respect to the property under study.

Subsample—a portion taken from a sample. A laboratory sample may be a subsample of a gross sample; similarly, a test portion may be a subsample of a laboratory sample.

Test portion (also called specimen, test specimen, test unit, aliquot)—That quantity of a material of proper size for measurement of the property of interest. Test portions may be taken from the gross sample directly, but often preliminary operations, such as mixing or further reduction in particle size, are necessary.

where

σ_x^2 = variance of the mean,
σ_b^2 = variance of the units in the lot,
σ_w^2 = average variance of the samples taken from a segment,
σ_t^2 = variance of the analytical operations,
N = number of units in the lot,
n_b = number of randomly selected units sampled,
n_w = number of randomly drawn samples from each unit selected for sampling, and
n_t = total number of analyses, including replicates, run on all samples.

If stratification is known to be absent, then much measurement time and effort can be saved by combining all the samples and mixing thoroughly to produce a composite sample for analysis. Equation 8 is applicable to this situation also. If the units vary significantly in weight or volume, the results for those units should be weighted accordingly.

For homogeneous materials σ_w^2 is zero, and the second term on the right-hand side of Equation 8 drops out. This is the case with many liquids or gases. Also, if all units are sampled, then $n_b = N$ and the first term on the right-hand side of Equation 8 also drops out.

Particle Size in Sampling Particulate Mixtures

Random sampling error may occur even in well-mixed particulate mixtures if the particles differ appreciably in composition and the test portion contains too few of them. The problem is particularly important in trace analysis, where sampling standard deviations may quickly become unacceptably large. The sampling constant diagram of Ingamells and the Visman expression are useful aids for estimating sample size when preliminary information is available. Another approach that can often provide insight is to consider the bulk material as a two-component particulate mixture, with each component containing a different percentage of the analyte of interest (*11*). To determine the weight of sample required to hold the sampling standard deviation to a preselected level, the first step is to determine the number of particles n. The value of n may be calculated from the relation

$$n = \left[\frac{d_1 d_2}{\bar{d}^2}\right]^2 \left[\frac{100(P_1 - P_2)}{R\bar{P}}\right]^2 p(1-p) \tag{9}$$

where d_1 and d_2 are the densities of the two kinds of particles, $\bar{d}$ is the density of the sample, P_1 and P_2 are the percentage compositions of the component of interest in the two kinds of particles, $\bar{P}$ is the overall average composition in percent of the component of interest in the sample, R is the percent relative standard deviation (sampling error) of the sampling operation, and p and $1 - p$ are the fractions of the two kinds of particles in the bulk material. With knowledge of the density, particle diameter, and n, the weight of sample required for a given level of sampling uncertainty can be obtained through the expression, weight = $(4/3)\pi r^3 dn$ (assuming spherical particles).

Figure 4 shows the relation between the minimum weight of sample that should be taken and the composition of mixtures containing two kinds of particles, one containing 10% of the sought-for substance and the other 9, 5, 1, or 0%. A density of 1, applicable in the case of many biological materials, is used, along with a particle diameter of 0.1 mm. If half the particles in a mixture contain 10% and the other half 9% of the substance of interest, then a sample of 0.0015 g is required if the sampling standard deviation is to be held to a part per thousand. If the second half contains 5%, a sample of 0.06 g is necessary; if 1%, 0.35 g would be needed. In such mixtures it is the relative difference in composition that is important. The same sample weights would be required if the compositions were 100% and 90, 50, or 10%, or if they were 0.1% and 0.09, 0.05, or 0.01%. The same curves can be used for any relative composition by substitution of x for 10%, and 0.1 x,

0.5 x, and 0.9 x for the curves corresponding to 1, 5, and 9% in Figure 4. If a standard deviation of 1% is acceptable, the samples can be 100 times smaller than for 0.1%.

An important point illustrated by the figure is that if the fraction of richer particles is small, and the leaner ones contain little or none of the substance of interest, large test portions are required. If a sample of gold ore containing 0.01% gold when ground to 140 mesh (0.1 mm in diameter) consists, say, of only particles of gangue and of pure gold, test portions of 30 g would be required to hold the sampling standard deviation to 1%. (An ore density of 3 is assumed.)

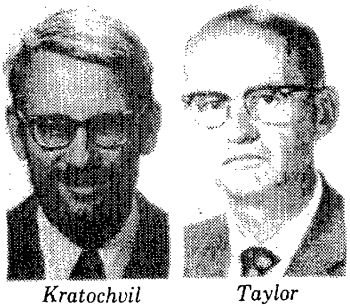
Concluding Comments

Sampling is not simple. It is most important in the worst situations. If the quantities $\bar{x}$, s, K_s, A, and B are known exactly, then calculation of the statistical sampling uncertainty is easy, and the number and size of the samples that should be collected to provide a given precision can be readily determined. But if, as is more usual, these quantities are known only approximately, or perhaps not at all, then preliminary samples and measurements must be taken and on the basis of the results more precise sampling procedures developed. These procedures will ultimately yield a sampling plan that optimizes the quality of the results while holding down time and costs.

Sampling theory cannot replace experience and common sense. Used in concert with these qualities, however, it can yield the most information about the population being sampled with the least cost and effort. All analytical chemists should know enough sampling theory to be able to ask intelligent questions about the samples provided, to take subsamples without introducing additional uncertainty in the results and, if necessary, to plan and perform uncomplicated sampling operations. It is the capability of understanding and executing all phases of analysis that ultimately characterizes the true analytical chemist, even though he or she may possess special expertise in a particular separation or measurement technique.

References

(1) W. J. Youden, *J. Assoc. Off. Anal. Chem.*, **50**, 1007 (1967).
(2) T. B. Whitaker, J. W. Dickens, and R. J. Monroe, *J. Am. Oil Chem. Soc.*, **51**, 214 (1974); T. B. Whitaker, *Pure Appl. Chem.*, **49**, 1709 (1977).
(3) Hazardous Waste Monitoring System, General, *Fed. Regist.*, Vol. 45, No. 98, pp 33075–33127 (May 19, 1980).
(4) Staff, ACS Subcommittee on Environmental Analytical Chemistry, *Anal. Chem.*, **52**, 2242 (1980).
(5) See, for example, W. J. Dixon and F. J. Massey, Jr., "Introduction to Statistical Analysis," 3rd ed., McGraw-Hill, New York, 1969; M. G. Natrella, "Experimental Statistics," National Bureau of Standards Handbook 91, August 1963, U.S. Government Printing Office.
(6) C. O. Ingamells and P. Switzer, *Talanta*, **20**, 547 (1973); C. O. Ingamells, *Talanta*, **21**, 141 (1974); **23**, 263 (1976).
(7) S. H. Harrison and R. Zeisler, NBS Internal Report 80-2164, C. W. Reimann, R. A. Velapoldi, L. B. Hagan, and J. K. Taylor, Eds., U.S. National Bureau of Standards, Washington, D.C., 1980, p 66.
(8) M. G. Natrella, "Experimental Statistics," National Bureau of Standards Handbook 91, August 1963, U.S. Government Printing Office, pp 2–13 and Table A6.
(9) ASTM E-300 Standard Recommended Practice for Sampling Industrial Chemicals, American Society for Testing and Materials, Philadelphia, 1973 (reapproved 1979).
(10) J. Visman, *Materials Research and Standards*, November, p 8 (1969).
(11) A. Benedetti-Pichler, in "Physical Methods of Chemical Analysis," W. M. Berl, Ed., Academic Press, New York, 1956, Vol. 3, p 183; W. E. Harris and B. Kratochvil, *Anal. Chem.*, **46**, 313 (1974).
(12) W. E. Harris and B. Kratochvil, "Introduction to Chemical Analysis," Saunders, Philadelphia, 1981, Chapter 21.



Kratochvil *Taylor*

Byron Kratochvil, professor of chemistry at the University of Alberta, received his BS, MS, and PhD degrees from Iowa State University. His research interests include solvent effects on solute properties and reactions, applications of nonaqueous systems to chemical analysis, and methods for determining ionic solutes.

John K. Taylor, coordinator for quality assurance and voluntary standardization activities at the National Bureau of Standards Center for Analytical Chemistry, received his BS from George Washington University, and his MS and PhD degrees from the University of Maryland. His research interests include electrochemical analysis, refractometry, isotope separations, standard reference materials, and the application of physical methods to chemical analysis.

Appendix D.3

SYMPOSIUM ON REFERENCE MATERIALS AND THEIR USE IN NUCLEAR FUEL CYCLE

John K. Taylor[1]

Reference Materials—What They Are and How They Should Be Used

REFERENCE: Taylor, J. K., **"Reference Materials—What They Are and How They Should be Used,"** *Journal of Testing and Evaluation*, JTEVA, Vol. 11, No. 6, Nov. 1983, pp. 385-387.

ABSTRACT: The role of reference materials in monitoring the chemical measurement process is considered. Requirements for reliable reference materials are reviewed. The use of reference material data in estimating the uncertainties of the results of measurements on test samples is discussed.

KEYWORDS: accuracy, chemical analysis, precision, quality assurance, quality control, reference materials, standard reference materials

The increasing requirements for accuracy in chemical analysis and the necessity to interrelate and combine data sets from several laboratories or from the same laboratory over intervals of time have created a demand for well-characterized reference materials (RMs) for quality assurance purposes. When properly used, such materials can provide a high degree of confidence in analytical data.

On the basis of inquiries received and discussions with various analysts, it is clear that RMs are not fully used in some situations and are poorly used in others. It is the purpose of this paper to clarify the role of RMs in the chemical measurement process and to suggest ways in which they may be used to the best advantage.

Role of Reference Materials

An RM is any substance that may be measured simultaneously or sequentially in a measurement process to provide information about the process or the measurements arising therefrom. RMs may be internally developed to monitor a specific measurement process, or they may be provided by an external source. Externally developed reference materials (ERMs) are usually certified by some organization and frequently called certified reference materials (CRMs) [*1*]. The National Bureau of Standards (NBS) Standard Reference Materials (SRMs) are a special class of CRMs that have been carefully analyzed (sometimes with the aid of cooperators) and certified by NBS [*2-4*]. While the RMs described above may differ in status, each can provide useful information to the analyst for the assessment of data quality.

The only rational basis for use of a RM is as a monitor of a measurement system that is in a state of statistical control [*5*]. This means that a valid measurement principle has been identified and put into practice using quality control procedures that assure a requisite degree of reproducibility. Indeed, such a measurement process should, conceptually, be capable of producing an infinite population of measurements, some of which at any moment may be considered as a sample. The measurements of the RM may then be considered as random samplings of the output of the measurement system, which would permit their interpretation for evaluation of the measurement process.

Requirements for RMs

RMs of any type must be appropriate in matrix and composition and of stable composition over the intended period of use. They must be sufficiently uniform in composition when subsampled (homogeneous) and available in sufficient quantity to be useful over a reasonable period of time.

CRMs have the further requirement that they must be issued with a certificate in which their measured parameters and assigned uncertainties are fully documented [*6*]. Internal reference materials (IRMs) must have an equal degree of reliability with respect to stability and homogeneity. The requirement for accuracy of the assigned values of specified parameters of the latter will depend upon the end use of the materials, but this may be of lesser importance than homogeneity in some cases.

Laboratories are well advised to upgrade the accuracy of their IRMs to the highest extent possible. Intercomparisons with high quality CRMs, such as SRMs, can be used to accomplish this purpose.

Interpretation of RM Data

The primary function of IRMs (often called control samples) is to evaluate the attainment of statistical control of the measurement system. As long as such samples are stable and homogeneous in

Presented at the Symposium on Reference Materials and Their Use in Nuclear Fuel Cycle, sponsored by ASTM Committee C-26 on Nuclear Fuel Cycle on 9 Aug. 1982, Knoxville, TN.

[1]Coordinator for quality assurance and voluntary standardization, Center for Analytical Chemistry, National Bureau of Standards, Washington, DC 20234.

0090-3973/83/0011-0385$00.00

composition, the precision observed in the analysis of the IRM may be inferred as the precision of the measurement system for that particular measurement. Accordingly, it is clear that stability and uniformity are the prime requirements of such materials and that accurate knowledge of compositions is only a secondary consideration. While such information is highly desirable, the proof of absence of bias using such samples may be difficult to establish based on internally generated evidence alone.

ERMs and CRMs in particular are best used to demonstrate accuracy, that is, the freedom from bias, of measurement systems that are demonstrated to be in a state of statistical control. Because the CRM may be costly or available in limited amounts or both, this latter demonstration is often best left to the use of IRMs. Because of their high quality and the care used in their certification, SRMs often stand at the top of the RM hierachy and, hence, are especially useful to evaluate the accuracy of a measurement process. Needless to say, SRMs should be used in carefully designed test sequences together with IRMs if maximum information is to be provided.

Figure 1 contains a typical sequence in which IRMs and SRMs may be measured together with the test samples to monitor a measurement process. Figure 1 assumes that control charts are maintained [7], the kinds of which are indicated in the "Notes." The control limits of the charts are determined from the results of previous measurements of the IRM. Points on the control charts are plotted immediately after the data are obtained, and they must fall within the control limits, at the decision points, in order to continue the measurements sequence. Otherwise any measurements obtained since the last time the system was known to be in control are suspect and discarded or held in abeyance. Furthermore, the system must be demonstrated to have regained control before data may again be accepted.

Figure 2 is similar to Fig. 1 but uses duplicate or split samples to monitor statistical control. Control charts based on the differences between the results obtained for the duplicate/split samples are the basis for monitoring statistical control. Such charts are described in several publications [7]. The rationale to be followed at the decision points is essentially the same as described in the discussion of Fig. 1.

No matter what kinds of RMs are used, their ability to monitor the measurement process and especially the measurements in progress must be demonstrated, and this is often a matter of inference. Though statistical control of the measurement system and freedom from bias may be readily demonstrated for the case of measurement of RMs, the performance of the system on measurements of other test samples is the matter of concern. To the extent that the RMs simulate the test samples, the inference drawn from measurement of the former may be transferred to the latter. Conversely, the confidence may diminish as the degree of simulation is decreased. In every case, the experience and professional judgment of the analyst must be used to infer how well an RM monitors the actual measurement process.

RMs may also be used to evaluate the suitability of a proposed method for a special purpose or to determine the performance characteristics of methods under development. Such use of RMs has much the same limitations as in the evaluation of monitoring a process. As stable test samples, RM analyses can provide data on the

Daily/Event Schedule

CALIBRATION - FULL EXPECTED RANGE
* IRM
TEST SAMPLES - GROUP 1
* IRM
TEST SAMPLES - GROUP 2
* IRM
.
.
* SRM
.
.
TEST SAMPLES - GROUP N-1
* IRM
TEST SAMPLES - GROUP N
* IRM
* CALIBRATION - MIDRANGE POINT

NOTES

* - DECISION POINT
1. MAINTAIN CONTROL CHARTS
X-CONTROL CHART, IRM
R-CONTROL CHART, ΔIRM
2. SYSTEM MUST BE IN CONTROL AT DECISION POINTS
3. AT LEAST 2 GROUPS: MAXIMUM OF 10 SAMPLES IN EACH GROUP
4. AT LEAST ONE SRM MEASUREMENT SHOULD BE MADE DURING EACH SEQUENCE/DAY

FIG. 1—*Quality assessment using IRM samples.*

Sequence Schedule

CALIBRATION - FULL EXPECTED RANGE
* CALIBRATION CHECK - MIDRANGE POINT
SAMPLE 1
* SAMPLE 1 D/S
SAMPLE 2-9
SAMPLE 10
* SAMPLE 10 D/S
* IRM or SRM
SAMPLE 11-19
SAMPLE 20
* SAMPLE 20 D/S
* CALIBRATION CHECK - MIDRANGE POINT
* CALIBRATION CHECK - MIDRANGE POINT/DUPLICATE

NOTES

* = DECISION POINT
1. MAINTAIN CONTROL CHARTS
a. DUPLICATE MIDRANGE CALIBRATION
b. DUPLICATE/SPLIT SAMPLE
c. X-CONTROL CHARTS, SRM AND IRM
2. SYSTEM MUST BE IN CONTROL AT DECISION POINTS
3. IF MORE THAN 20 SAMPLES, REPEAT SEQUENCE
4. IF LESS THAN 20 SAMPLES, DIVIDE INTO TWO GROUPS AND FOLLOW SIMILAR PLAN
5. AT LEAST ONE SRM MEASUREMENT SHOULD BE MADE DURING EACH SEQUENCE DAY

FIG. 2—*Quality assessment using duplicates/splits.*

precision of a method of measurement. How well this may be transferred to a practical measurement situation, and how well potential biases are evaluated is again a matter of judgment, which may need to be supported by additional information.

Conclusion

Laboratories should refrain from reporting data unless they are in a position to assign uncertainties to the reported values [8]. Such an assignment requires the attainment of statistical control of the measurement system and estimation of the bounds of systematic error. The analysis of reliable RMs in a planned measurement sequence can provide the basis of estimating both the random and systematic components of the measurement uncertainty. However, the assignment of such uncertainties to the test results must be done with due consideration of any matrix differences between the RMs and the test samples. The analysis of high quality ERMs, such as SRMs, together with the laboratory's IRMs (control samples) is the best approach to monitoring a chemical measurement system for quality assurance of the data output.

References

[1] "Terms and Definitions Used in Connection with Reference Materials," ISO Guide 30-1981, International Standards Organization, Geneva, Switzerland, 1981.

[2] Cali, J. P. and Stanley, C. L., "Measurement Compatibility and Standard Reference Materials," *Annual Review of Materials Science*, Vol. 5, 1975, pp. 329-342.

[3] Cali, J. P., Mears, T. W., Michaelis, R. E., Read, W. P., Seward, R. W., Stanley, C. L., and Yolken, H. T., "The Role of Standard Reference Materials in Measurement Systems," NBS Monograph 148, National Bureau of Standards, Washington, DC 20234, Jan. 1975.

[4] *Catalogue of NBS Standard Reference Materials*, SP 260, National Bureau of Standards, Washington, DC 20234.

[5] Taylor, J. K., "Quality Assurance of Chemical Measurements," *Analytical Chemistry*, Vol. 53, 1981, pp. 1588A-1596A.

[6] "Contents of Certificates of Reference Materials," ISO Guide 31-1981, International Standards Organization, Geneva, Switzerland, 1981.

[7] ASTM Committee E-11 on Statistical Methods, *Manual on Presentation of Data and Control Chart Analysis, STP 15D*, American Society for Testing and Materials, Philadelphia, 1976.

[8] American Chemical Society Subcommittee on Environmental Chemistry, "Guidelines for Data Acquisition and Data Quality Evaluation in Analytical Chemistry," *Analytical Chemistry*, Vol. 52, 1980, p. 2242.

Report

John K. Taylor
Center for Analytical Chemistry
National Bureau of Standards
Washington, D.C. 20234

Validation of Analytical Methods

Validation of analytical methods is a subject of considerable interest. Documents such as the "ACS Guidelines for Data Acquisition and Data Quality Evaluation" (*1*) recommend the use of validated methods. The promulgation of federal environmental regulations requires the inclusion of validated reference methods. Standards-writing organizations spend considerable time in collaborative testing of methods they prepare, validating them in typical applications and determining their performance characteristics. Nevertheless, questions about the appropriateness of methods and the validity of their use in specific situations often arise. Some of these questions may be due to differences in understanding both what a method really is and what the significance of the validation process is. This paper attempts to clarify the nomenclature of analytical methodology and to define the process of validating methods for use in specific situations.

Hierarchy of Methodology

The hierarchy of methodology, proceeding from the general to the specific, may be considered as follows: *technique → method → procedure → protocol.*

A ***technique*** is a scientific principle that has been found to be useful for providing compositional information; spectrophotometry is an example. Analytical chemists historically have investigated new measurement techniques for their ability to provide novel measurement capability, or to replace or supplement existing methodology. As a result of innovative applications, analysts can now analyze for myriad substances in exceedingly complex mixtures at ever lower trace levels, with precision and accuracy undreamed of only a few years ago (*2*).

A ***method*** is a distinct adaptation of a technique for a selected measurement purpose. The pararosaniline method for measurement of sulfur dioxide is an example. It involves measuring the intensity of a specific dye, the color of which is "bleached" by the gas. Several procedures for carrying out this method may be found in the literature. Modern methodology is complex and may incorporate several measurement techniques; a method may thus be interdisciplinary.

A ***procedure*** consists of the written directions necessary to utilize a method. The "standard methods" developed by ASTM and AOAC are, in reality, standardized procedures. ASTM D2914—Standard Test Method for the Sulfur Dioxide Content of the Atmosphere (West–Gaeke Method)—is an example (*3*). While a precise description is the aim, it is difficult, if not impossible, to describe every de-

Hierarchy of Analytical Methodology

	Definition	Example
Technique	Scientific principle useful for providing compositional information	Spectrophotometry
Method	Distinct adaptation of a technique for a selected measurement purpose	Pararosaniline method for measurement of sulfur dioxide
Procedure	Written directions necessary to use a method	ASTM D2914—Standard Test Method for the Sulfur Dioxide Content of the Atmosphere (West–Gaeke Method)
Protocol	Set of definitive directions that must be followed, without exception, if the analytical results are to be accepted for a given purpose	EPA Reference Method for the Determination of Sulfur Dioxide in the Atmosphere (Pararosaniline Method)

This REPORT is based on a talk given at the 184th ACS National Meeting, Sept. 12–17, 1982, Kansas City, Mo.

Published in Analytical Chemistry, May 1983, 600A-608A, by the American Chemical Society

tail of every operation in a procedure. Accordingly, some level of sophistication is presumed for the user of every published procedure; if very sophisticated users are contemplated, only a minimum of detail will be provided and vice versa. However, it should be noted that any omission in the description of critical steps is a potential source of variance or bias, even in the hands of knowledgeable analysts. Because of the flexibility intentionally or unintentionally provided to the analyst, or because of differences in interpretation, it is fair to say that minor-to-major differences of application occur in the use of even the most precisely defined procedures. Such differences often account for the interlaboratory variability observed in many collaborative tests. Further, at some point of departure from a published procedure, a new method results that may need its own validation.

The term ***protocol*** is the most specific name for a method. A protocol is a set of definitive directions that must be followed, without exception, if the analytical results are to be accepted for a given purpose. Protocols may consist of existing methods or procedures, modifications of such, or they may be developed especially for specific purposes. Typically, they are prescribed by an official body for use in a given situation such as a regulatory process. The EPA Reference Method for the Determination of Sulfur Dioxide in the Atmosphere (Pararosaniline Method) is an example of a protocol (*4*). The test method specified as part of a contractual arrangement for the acceptance of data or a product or material is another example of a protocol, although it may not be called that in the contract.

A plethora of methods, procedures, and protocols based on the same measurement principle can arise for a given analytical determination. Usually, they are worded differently, and they may contain subtle or major differences in technical details. The extent to which each needs to be individually validated is a matter of professional judgment. It is evident that some validation tests could be merely a matter of experimentally testing the clarity of the written word.

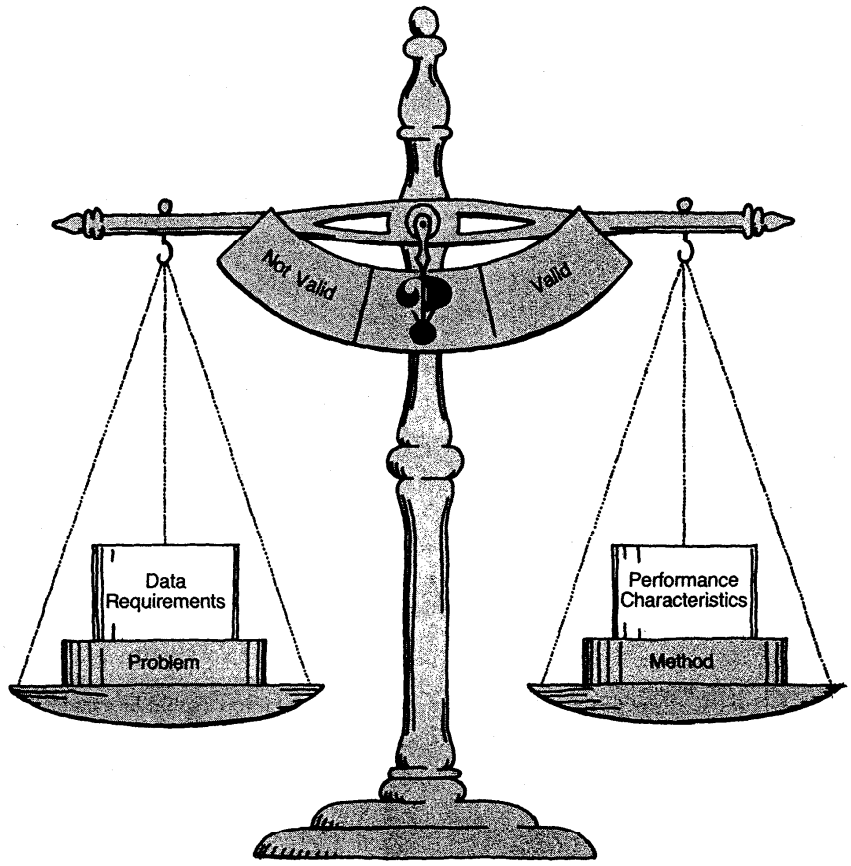


Figure 1. Basic concept of the validation process

Goals for Validation

Validation is the process of determining the suitability of methodology for providing useful analytical data. This is a value judgment in which the performance parameters of the method are compared with the requirements for the analytical data, as illustrated in Figure 1. Obviously a method that is valid in one situation could be invalid in another. Accordingly, the establishment of firm requirements for the data is a prerequisite for method selection and validation. When data requirements are ill-considered, analytical measurement can be unnecessarily expensive if the method chosen is more accurate than required, inadequate if the method is less accurate than required, or utterly futile if the accuracy of the method is unknown.

Fortunately, typical and even standard measurement problems often exist. Examples include a wide variety of clinical analyses, environmental determinations, and recurring measurements for the characterization of industrial products. The kinds of samples for which methods have been val-

idated should be clearly described, and users should be aware of the need to demonstrate their own abilities to use the method in their own laboratories.

Statements of precision and accuracy are often a result of a validation process, especially in the case of a collaborative test exercise. Such statements are often misinterpreted; they merely describe the results of the exercise and are, at best, estimates of typical performance expectations for the method. They should not be construed to be performance parameters nor should they be used to estimate the uncertainty of any future data obtained by using the method. However, information on precision and accuracy should be obtained to the extent possible since it provides a quantitative basis for judging general performance capability.

Other information useful for characterizing methodology or for judging its suitability for a given use includes: sensitivity to interferences, limits of detection, and useful range of measurement. The specific details for evaluating methodology in these respects are beyond the scope of the present paper. Ordinarily, such information is best obtained as a result of applied research during the method development stage. Because the limit of detection is closely related to the attainable precision at the lower limit of measurement, both the limit of detection and the lowest concentration range measurable (often called limit of quantitation) should be evaluated, as pertinent, in every laboratory (*1*, *5*).

Validation Process

The validation process verifies that the methodology is based on sound technical principles and that it has been reduced to practice for practical measurement purposes. Both the need to validate methodology and the procedure to be followed are matters for professional judgment. The validation can be either general or specific.

General Validation. Validation of measurement techniques depends on the elucidation of the scientific principles upon which they are based. Such validation results from the research of the scientific community, and its soundness is evaluated by peer review. Better understanding of measurement principles can extend their scope and improve the quality of their use. To confirm the above statement, one need only think about the varied research that has contributed to the understanding of the principles of gas chromatography and that has led to development of its status as a prime measurement technique.

Methods arise as the result of applied research, typically by individuals, that often involves both a comprehensive understanding of measurement techniques and a high degree of ingenuity and innovation in their application. Testing of the methods in typical practical situations plays a key role in both the development process and in validation. While ordinarily limited in scope, validation at the research stage can be comprehensive and can apply to a wide variety of end uses.

Procedures are developed for the end use of methods in practical analytical situations. The user laboratory ordinarily needs more experimental details than are contained in a published research report of a method to use it in practical measurements. Frequently, as a method gains widespread use, procedures evolve that the users may decide need to be standardized. This is often done by consensus in a standards organization forum. During this process, the resulting standard procedure is examined both technically and editorially. A thorough review process includes collaborative testing in which typical stable test materials are analyzed to verify the procedure's usefulness and to identify both technical and editorial weaknesses. The process is illustrated in Figure 2. If the composition of the reference samples is known, precision and bias, both intra- and interlaboratory, can be evaluated; otherwise, only precision can be evaluated. If a method of known accuracy is available, the collaborative test may consist of its comparison with the candidate method, in which case both precision and bias can be evaluated. The performance parameters of the procedure so evaluated are for the conditions of the collaborative test that are considered typical. Any extension of them to other kinds of samples is by inference only, and may need to be justified. Although it can be time-consuming, the development of a standard method is one of the best ways to validate a procedure because of the breadth of examination that is involved.

A protocol is prescribed by fiat of an organization requiring a specific kind of measurement. Presumably it results from an intelligent decision based on the organization's validation process or that of others. This may consist of an extensive collaborative test or publication of a proposed protocol for public comment. Unfortunately, expediency has overruled sound scientific judgment in some cases, resulting in the promulgation of unvalidated and scientifically defective protocols (*6*). Protocols that are specified in a contractual arrangement may be selected arbitrarily or through a well-conceived selection process. Verification of their validity for the specific use should be a prime consideration.

Validation for Specific Use. The ultimate use of analytical methodology is to produce compositional information about specific samples necessary for the solution of particular problems ranging from exotic research investigations to the very mundane. The selection of appropriate measurement methodology is often a major consideration. Methods or procedures, even if previously validated in general

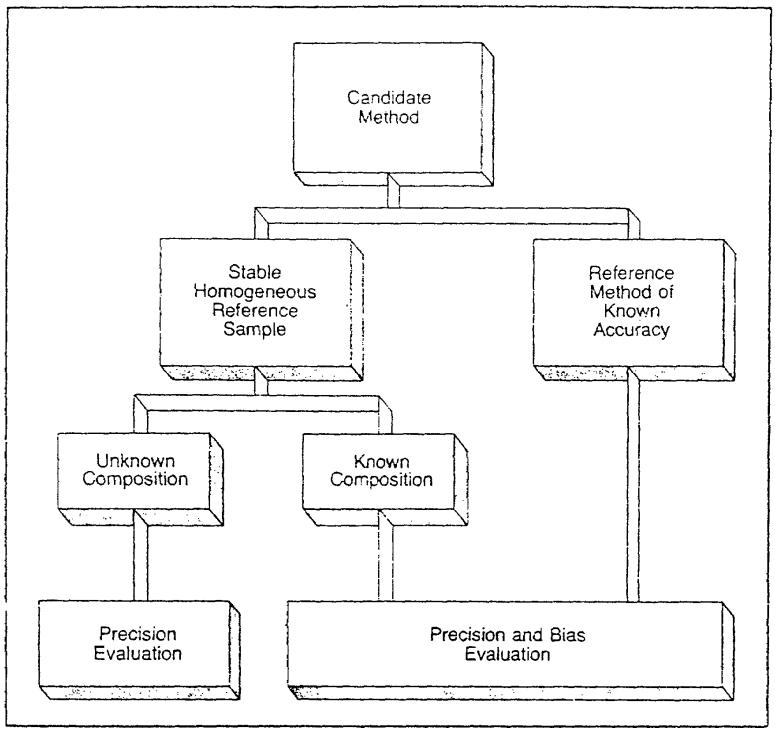


Figure 2. Collaborative test process

terms, cannot unequivocally be assumed to be valid for the situation in hand, because of possible differences in sample matrix and other considerations. Professional analytical chemists traditionally have recognized this and their responsibility to confirm or prove (if necessary) both the validity of the methodology used for specific application (*2*) and their own ability to reduce it to practice.

The classical validation process is illustrated in Figure 3. When reference samples are available that are similar in all respects to the test samples, the process is very simple: It consists of analyzing a sufficient number of reference samples and comparing the results to the expected or certified values (*7*). Before or during such an exercise, the analyst must demonstrate the attainment of a state of statistical control of the measurement system (*8*) so that the results can be relied upon as representative of those expected when using the methodology–measurement system.

When a suitable reference material is not available, several other approaches are possible. One consists of comparing the results of the candidate method with those of another method known to be applicable and reliable, but not useful in the present situation because of cost, unavailability of personnel or equipment, or other reasons. Even the agreement of results with those obtained using any additional independent method can provide some useful information.

Spiked samples and surrogates may be used as reference samples. This approach is less desirable and less satisfactory because of the difficulty in the reliable preparation of such samples and because artificially added materials such as spikes and surrogates may exhibit matrix effects differing from those of natural samples. Split samples of the actual test samples may be used to evaluate the precision of a method or procedure, but they provide no information about the presence or magnitude of any measurement bias.

Another approach is to infer the appropriateness of methodology from measurements on analogous but dissimilar reference materials. The critical professional judgment of the analyst is necessary to decide the validity of the inference.

In all cases, sufficient tests must be made to evaluate the methodology for

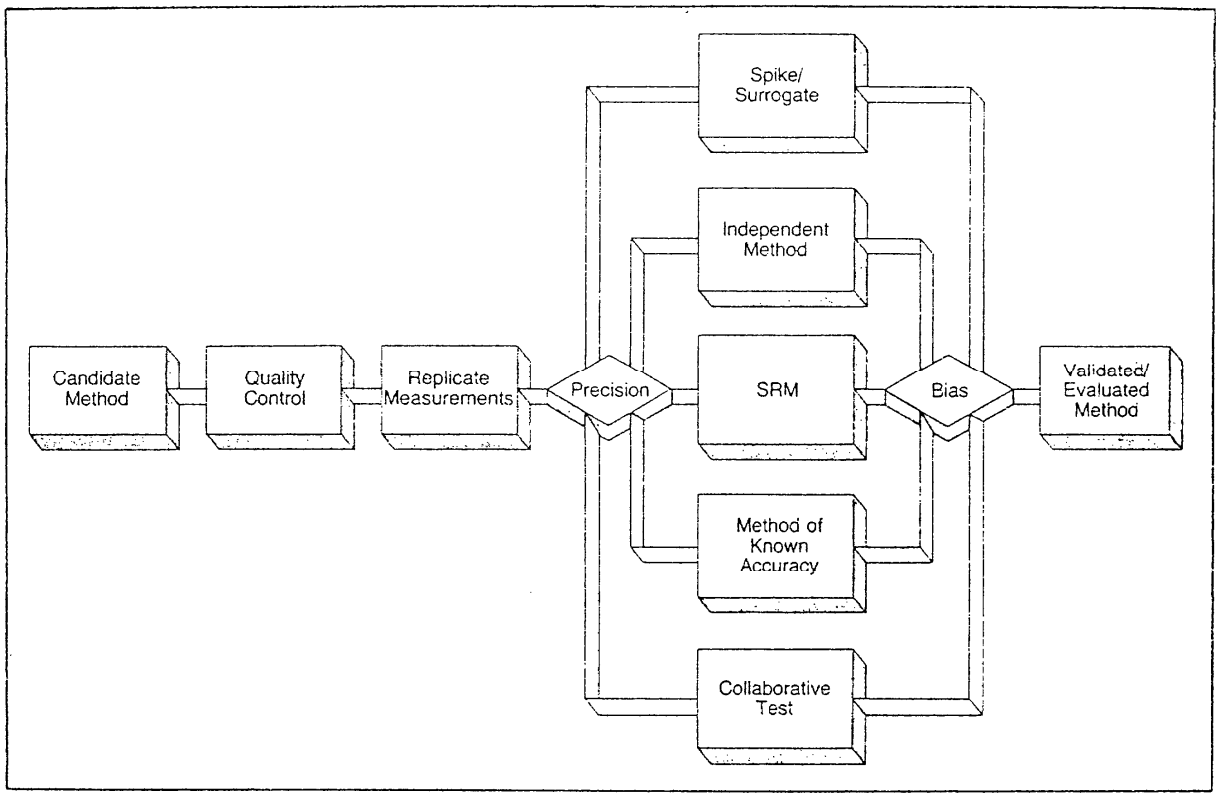


Figure 3. General process for evaluation/validation of methodology

the variety of matrices and ranges of composition expected during the measurement process. Ordinarily, the latter should include three levels of concentration, namely, the extremes and the mid-range of compositions expected. Statistical considerations suggest that at least six degrees of freedom (ordinarily seven measurements) should be involved at each decision point.

Conclusion

A valid method is necessary but not sufficient for the production of valid data. Most methods require a degree of skill on the part of the analyst; this skill constitutes a critical factor in the measurement process. It is common knowledge that data obtained by several laboratories on the same test sample using the same methodology may show a high degree of variability. The alleviation of such a problem is in the area of quality assurance of the measurements (*8*). Data obtained by a valid method used in a well-designed quality assurance program should allow the assignment of limits of uncertainty that can be used to judge the data's validity.

It should be remembered that the validity of any data will also depend upon the validity of the model and of the sample (*8, 9*). The model represents the conceptualization of the problem to be solved, describes the samples that should be analyzed, the data base required, and the way the model will be utilized. Obviously, even flawless measurement data will be of little value if the basic concepts are faulty. Likewise the samples analyzed must be valid if the results obtained for them are to be intelligently interpreted.

The key role of reliable reference materials in the validation of analytical measurements cannot be overemphasized. Their use in validating the methodology has already been discussed. A planned sequential analysis of reference materials in a quality assurance program can assess the quality of the data output and thus validate the overall aspects of the analytical measurement system (*7*).

References

(1) ACS Subcommittee on Environmental Analytical Chemistry. *Anal. Chem.* **1980,** *52*, 2242.
(2) Taylor, John K. CHEMTECH **1982,** *12*, 225.
(3) Annual Book of ASTM Standards, Part 26. American Society for Testing and Materials, Philadelphia, Pa. 19103, 1981.
(4) National Primary and Secondary Ambient Air Quality Standards, *Fed. Regist.* April 30, 1971, Vol. 36, Part 84, p. 8187.
(5) Glaser, J. A.; Foerst, D. L.; McKee, G. D.; Quave, S. A.; Budde, W. L. *Environ. Sci. Technol.* **1981,** *15*, 1426–35.
(6) Purdue, L. T. et al. *Environ. Sci. Technol.* **1972,** *6*, 152–54.
(7) Taylor, John K. "Reference Materials—How They Are or How They Should Be Used," *ASTM Journal of Testing and Technology*, in press.
(8) Taylor, John K. *Anal. Chem.* **1981,** *53*, 1588–96 A.
(9) Kratochvil, Byron; Taylor, John K. *Anal. Chem.* **1981,** *53*, 924–38 A.

John K. Taylor is coordinator for chemical measurement assurance and voluntary standardization at the National Bureau of Standards' Center for Analytical Chemistry. He received his BS degree from George Washington University and his MS and PhD degrees from the University of Maryland. His research interests include electrochemical analysis, refractometry, isotope separations, standard reference materials, and the application of physical methods to chemical analysis.

Guide for Requesting Development of Standard Reference Materials

The National Institute of Standards and Technology develops Standard Reference Materials (SRM's) to provide a basis for comparison of measurements on materials and to aid in the control of production processes. The Office of Standard Reference Materials evaluates the requirements of science, industry, and government for carefully characterized reference materials, then directs the production and distribution of these materials.

NIST currently has over 1000 SRM's available, about 100 new ones in preparation, and requests for the development of many more. The demand for new SRM's greatly exceeds NIST's capacity to produce and certify these materials. Consequently, requests for new SRM's of limited use are deferred in favor of those that serve a substantial area of interest. In determining which requests receive top priority, NIST relies heavily upon information supplied by industry and interested organizations.

NIST welcomes all requests for SRM's. Both NIST and potential users would be helped if these requests included as much of the information below as possible.

1. *Short title of the proposed Standard Reference Material.*
2. *Purpose for which this SRM is intended.*
3. *Reason why the SRM would be useful.*
4. *Special characteristics and/or requirements of the material. Include necessary additional information, if more than one SRM is needed for standardization in an area.*
5. *An estimate of the possible present and future (6-10 years) demand for such an SRM in your operations and elsewhere. (National and international estimates are very useful.)*
6. *Facts about whether such an SRM (or a similar one) could be produced by, or obtained from a source other than NIST. If so, justify preparation by NIST.*
7. *Other pertinent information to justify the SRM, such as: (a) an estimate of the range of application, economic significance of the measurement affected, and scientific and/or technological significance, including estimates of the impact upon industrial productivity or growth, and (b) supporting letters from industry leaders, trade organization, interested committees, and others.*

In developing an NIST-SRM, the candidate material must meet one or more of the criteria listed below:

1. *The SRM must permit users to attain more accurate measurements.*
2. *The production of the SRM must not be economically or technically feasible elsewhere.*
3. *The SRM must serve as an industry-wide standard for commerce, provided by a unique neutral source.*
4. *NIST production of the SRM would provide readily available, highly characterized material useful to science, industry, or government.*

NIST has recognized the need to enlarge the scope of the SRM program to include all types of well-characterized materials that can be used to calibrate a measurement system, or to produce scientific data that can be widely used. Input from science, industry, and government assists NIST in continuing to provide Standard Reference Materials that will be valuable in many areas.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1994 - 386-627 - 814/22789

1. Title of proposed SRM:

2. Purpose for which SRM is intended:

3. Reason SRM would be useful:

4. Special characteristics and requirements:

5. Estimate of present and future demand for proposed SRM:

6. Why Proposed SRM should be produced by NIST:

7. Other pertinent information:

8. Requested by: Telephone ( )

9. Company:

U.S. DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
GAITHERSBURG, MD 20899

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

NO POSTAGE NECESSARY IF MAILED IN THE UNITED STATES

BUSINESS REPLY MAIL
FIRST CLASS MAIL PERMIT NO. 12856 GAITHERSBURG, MD

POSTAGE WILL BE PAID BY
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

STANDARD REFERENCE MATERIALS PROGRAM
BUILDING 202, ROOM 204
NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY
GAITHERSBURG, MD 20878-9950

NIST

QUALITY MEASUREMENTS = STANDARD REFERENCE MATERIALS

UNITED STATES
DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY

MEASUREMENTS and STANDARDS are important to everyone who needs quality. The National Institute of Standards and Technology has over 1,000 different Standard Reference Materials that can help you calibrate instruments and check on measurements accuracy. When you use Standard Reference Materials you know your measurements are right. Listed below are some areas of available SRM's:

Atomic Absorption
Calorimetry
Clinical Analysis
Electron Probes
Emission Spectroscopy
GC/MS
HPLC
Inductively Coupled Plasma
Mass Spectrometry
Neutron Activation Analysis
Optical Emission
Optical Microscopes
Particle Counters
pH Meters
Scanning Electron Microscopes
Spectrophotometers
X-ray Diffraction
X-ray Fluorescence

NIST will continue to bring you more and even better Standard Reference Materials to meet your measurement requirements. If we can assist you, please contact us:

STANDARD REFERENCE MATERIALS PROGRAM
Building 202, Room 204
National Institute of Standards and Technology
Gaithersburg, MD 20899
Telephone: (301) 975-0SRM (6776)
FAX: (301) 948-3730

SRM