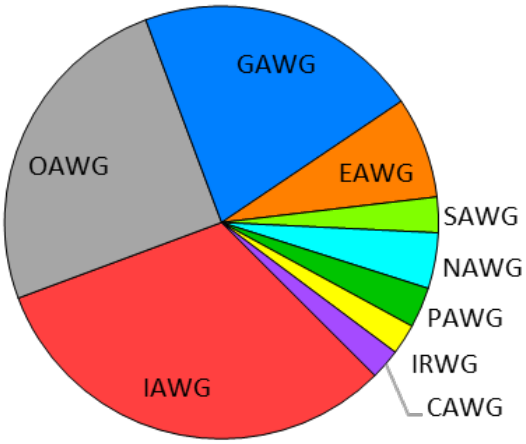
NIST Internal Report
NIST IR 8524

NIST's Engagement with CCQM Studies from 1992 to 2023: History and Performance



David L. Duewer
Christina E. Cecelski
Megan H. Cleveland
Nancy J. Lin
Katrice A. Lippa
Jacqueline L. Mann
John L. Molloy
Michael A. Nelson
James E. Norris
Kenneth W. Pratt
George C. (Jerry) Rhoderick
Sumona Sarkar
Michele M. Schantz
Hratch G. Semerjian
Robert L. Watters, Jr.
Donald A. Windover

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8524

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

(This page intentionally blank.)

NIST Internal Report
NIST IR 8524

NIST's Engagement with CCQM Studies from 1992 to 2023: History and Performance

David L. Duewer, Christina E. Cecelski, Megan H. Cleveland, Nancy J. Lin, Katrice A. Lippa, Jacqueline L. Mann, John L. Molloy, Michael A. Nelson, James E. Norris, Kenneth W. Pratt*, George C. (Jerry) Rhoderick*, Michele M. Schantz*, Hratch G. Semerjian*, Sumona Sarkar, Robert L. Watters*, Donald A. Windover
Material Measurement Laboratory

** Retired*

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8524

May 2024

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

NIST IR 8524
May 2024

Publication History
Approved by the NIST Editorial Review Board on 2024-05-17

How to Cite this NIST Technical Series Publication
Duewer DL, Cecelski CE, Cleveland MH, Lin NJ, Lippa KA, Mann JL, Molloy JL, Nelson MA, Norris JE, Pratt KW, Rhoderick GC, Sarkar S, Schantz MM, Semerjian HG, Watters RL Jr, Windover DA (2024) NIST's Involvement with CCQM Studies from 1992 to 2023: History and Performance. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Internal Report (IR) NIST IR 8524. https://doi.org/10.6028/NIST.IR.8524

NIST Author ORCID iDs

CE Cecelski	0000-0001-6782-3106
MH Cleveland	0000-0002-5584-9314
DL Duewer	0000-0002-3924-3064
NJ Lin	0000-0002-6980-3288
KA Lippa	0000-0001-8651-8326
JL Mann	0000-0001-6210-097X
JL Molloy	0009-0003-8901-5494
MA Nelson	0000-0003-0503-4501
JE Norris	0000-0002-5274-3274
KW Pratt	0009-0008-2509-1670
GC Rhoderick	0000-0002-7946-8217
S Sarkar	0000-0002-3192-843X
MM Schantz	0009-0009-4732-0005
HG Semerjian	0000-0003-2126-5887
DA Windover	0000-0002-8365-2092

Contact Information
david.duewer@nist.gov

Abstract

This report documents the engagement and measurement performance of the National Institute of Standards and Technology (NIST) in studies coordinated by the Consultative Committee for Amount of Substance: Metrology in Chemistry and Biology (CCQM) from 1992 to 2023. The data and graphical tools used to document performance are part of the *CCQM_Retrospectoscope* analysis system, https://doi.org/10.18434/mds2-2952.

Keywords

Cell Analysis Working Group (CAWG); Consultative Committee for the Amount of Substance: Metrology in Chemistry and Biology (CCQM); Electroanalytical Working Group (EAWG); Gas Analysis Working Group (GAWG); Inorganic Analysis Working Group (IAWG); Isotope Ratio Working Group (IRWG); Key Comparison (KC); Nucleic Acid Analysis Working Group (NAWG), Protein Analysis Working Group (PAWG); Organic Analysis Working Group (OAWG); Surface Analysis Working Group (SAWG); pilot studies; Supplementary Comparison (SC).

Table of Contents

List of Tables

List of Figures

Acknowledgments

We thank our colleagues David M. Bunk, John Kucklick, Willie E. May, Nathanael D. Olson, Karen W. Phinney, Katherine E. Sharpless, Abneesh Srivastava, Michael R. Winchester, and Stephen A. Wise for their comments and suggestions.

We recognize all NIST staff, Guest Researchers, and students who have contributed to one or more CCQM-sponsored studies over the past decades: Dan Bearden, Lyn Beary, Charles Beck, Mary Bedner, Joe Bennett, Arezue Boroujerdi, David Bunk, Carolyn Burdette, Therese Butler, George Caceres, Johanna Camara, Jennifer Carney, Christina Cecelski, Steven Christopher, Pamela Chu, Don Cleveland, Megan Cleveland, Bruce Coxon, Clay Davis, Russell Day, Thiago de Oliveira Araujo, David Duewer, Joe Dura, Regina Easley, Keith Eberhardt, James Ehrstein, Michael Ellisor, John Fassett, Lorena Francini, Adolfas Gaigalas, Lyn Gameson, Hugo Gasca-Aragon, Cassie Goodman, Justin Gorham, Robert Greenberg, Frank Guenther, Mike Halter, Diane Hancock, Kim Harris, Ross Haynes, Hua-Jun He, Andrew Herzing, Joseph Hodges, Marcia Holden, Monique Johnson, Brittany Kassim, Michael Kelley, Robert Kelly, Paul Kienzle, Eric Kilpatrick, Margaret Kline, Rebecca Kraft, John Kucklick, Brian Lang, Yunqiao Li, Nancy Lin, Katrice Lippa, Stephen Long, Mark Lowenthal, Steve Lund, Elizabeth Mackey, Jacqueline Mann, Ryna Marinenko, Anthony Marlow, Willie May, Jenny McDaniel, Curt Meuse, Walter Miller, Gerald Mitchell, John Molloy, John Moody, Sarah Munro, Karen Murphy, Jackie Murray, Rick Myers, Bryant Nelson, Michael Nelson, James Norris, Yoana Nuevo-Ordonez, Rabia Oflaz, Donna O'Kelly, William Osborn, Rick Paul, Curtis Phinney, Karen Phinney, Scott Pine, Anne Plant, Barbara Porter, Antonio Possolo, Cedric Powell, Kenneth Pratt, Jeanita Pritchett, Savelas Rabb, Michael Rearick, Jan Redman, Catherine Rimmer, Jerry Rhoderick, Marc Salit, Lane Sander, Sumona Sarkar, Michele Schantz, Nick Sharp, Katherine Sharpless, Rubina Shreiner, John Sieber, David Simons, John Small, Jackie Smeller, Lorna Sniegoski, Abneesh Srivastava, Eric Steel, Christopher Szakal, Susan Tai, William Thorn, Blaza Toman, Bryan Tomlin, Peter Trask, Gregory Turk, Peter Vallone, Robert Vocke, Lili Wang, Jason Waters, Robert Watters, Michael Welch, Edward White, Michael Winchester, Donald Windover, Stephen Wise, Laura Wood, Bei Xu, Lee Yu, and Rolf Zeisler.

1. Introduction

The International Committee for Weights and Measures (CIPM) created what is now the Consultative Committee for the Amount of Substance: Metrology in Chemistry and Biology (CCQM) in 1993 [1]. The CCQM coordinates comparison studies of the chemical and biological measurements made by National Metrology Institutes (NMIs) and Designated Institutes (DIs) that are of global interest. The CCQM also liaises with the Regional Metrology Organizations (RMOs) responsible for developing the chemical and biological measurement capabilities that are of more regional interest.

Many of the studies sponsored by CCQM and the RMOs are designed to assess the degrees of agreement among results generated by study participants using well-developed measurement systems. Fully attributed results from these studies are (sometimes after considerable delay) made public via the Key Comparison Database (KCDB) [2] maintained by the Bureau International des Poids et Mesures (BIPM). The information from each study is static, representing measurement and data analytic capabilities at a given time. Evaluating longer-term performance of individual NMI and DIs or measurement communities requires assessment of results from multiple studies. The *CCQM_Retrospectoscope*, a spreadsheet-based graphical analysis system and integrated database [3], has been developed to enable such studies.

This document uses some of the *CCQM_Retrospectoscope* system's graphical analysis tools to document the National Institute of Standards and Technology's (NIST's) engagement and performance in CCQM studies conducted from 1992 to 2023. The sections following this Introduction describe the data used in the analysis, present an overview of NIST's engagements with the CCQM studies finalized by this report's publication date, and detailed analyses for each of the of the CCQM's measurement focus areas. Appendix A defines the acronyms and symbols used In the document. Appendix B translates the codenames used for the NMI and DI participants in CCQM studies, and Appendix C identifies the *CCQM_Retrospectoscope* subsystems used to prepare this document.

But first, a little pre-CCQM history and NIST's roles therein.

1.1. 1990: CIPM *ad hoc* Working Group on Chemical and Physico-Chemical Measurements

While "Many members of the Comité [the CIPM] expressed the view that the Comité should be very cautious about becoming involved in this field," in September 1990 the 79th Meeting of the CIPM established an *ad hoc* Working Group on Chemical and Physico-Chemical Measurements (AHWG) "to advise the CIPM on whether or not the [Bureau International des Poids et Mesures] BIPM should have a significant role in addressing the problem of providing uniformity and traceability in chemical and physico-chemical measurements" [4 §6]. The members of the AHWG were the leaders of metrology institutes in France, Germany, Japan, Poland, the Union of Soviet Socialist Republics, the United Kingdom – and the newly confirmed Director of NIST, John W.

Lyons. Dr. Lyons was appointed Chairman of the AHWG. Notably, Dr. Lyons was a Ph.D. physical chemist.

The first meeting of the AHWG was held at NIST on 4-5 June 1991. In addition to Dr. Lyons and metrological leaders from France, Germany, Japan, and the United Kingdom, attendees included the head of the Thermodynamics Research Center at Texas A & M University, the Director of the BIPM (Terry Quinn), the Director of NIST's Chemical Science and Technology Laboratory (Harry S. Hertz), and the Director of the NIST Office of International and Academic Affairs (George A. Sinnott).

The outcome of the meeting was the recommendation that "CIPM form a Working Group on Metrology in Chemical Analysis to address the specific problem of international traceability in quantitative chemical analysis. The central hypothesis of this proposal is that coordinated international activities on only one or two reference methods of wide application on a few key reference materials will allow participating laboratories to extend the international traceability thereby gained to a wider range of methods and reference materials. The terms of reference of the proposed Working Group on chemical measurements are to: (1) draw up and conduct an exploratory programme of cooperative work among the national metrology laboratories in the area of quantitative chemical analysis to test the above hypothesis; (2) propose a future programme of work for the BIPM and the national metrology laboratories; (3) consider the possible formation of a Consultative Committee on Chemical Metrology" [5 §7].

The AHWG's recommendations were presented at the 80th Meeting of the CIPM in September 1991. After long discussion, "the CIPM approved the first term of reference proposed for [the] Working Group, but did not approve the second and the third. It was decided that the Working Group will be called the 'Working Group on Metrology in Chemistry' and Dr. Lyons was asked to chair it. The membership of the Working Group was to remain the same, but CIPM members were invited to designate additional experts and other specialists they considered necessary. The Working Group should initiate some international comparisons among national laboratories, in collaboration with specialists from other competent laboratories and relevant international bodies. This programme of international comparisons would be limited to a few reference methods of wide application applied to a few key reference materials. The intention is to discover if such a limited programme can lead to improved traceability for a much wider range of methods and materials" [5 §6].

1.2. 1992: CIPM Working Group on Metrology in Chemistry

Responding to the charge to explore the need for improved metrological traceability, in 1992 the CIPM Working Group on Metrology in Chemistry (WGMC) initiated two international comparisons. "Study I involved isotope dilution mass spectrometry (IDMS) for the analysis of simple aqueous solutions of inorganic elements made up gravimetrically at the NIST, and Study II the analysis of simple gas mixtures using IDMS, high accuracy mass spectrometry (HAMAS) using external calibration, gas chromatography (GC),

chemiluminescence or non-dispersive infrared (NDIR) spectrometry according to a protocol established at the [Netherlands NMI]" [6 §6].

At the CIPM meeting of 1993 it was reported that for Study I "… only a small number of laboratories were within this [± 1 %] error band. The results seem to be better for Study II, but the number of reporting laboratories is too small to draw general conclusions. The Study I results underscore the need for traceability systems to be established for chemical measurements" [6 §6].

Partly based upon the available results from the two comparisons but largely spurred by the then rapidly increasing efforts at establishing metrological networks for chemical measurements among NMIs and other organizations, "… the CIPM decided to establish a Comité Consultatif pour la Quantite de Matiere (consultative committee for the amount of substance) with the following provisional terms of reference: (1) to advise the CIPM on matters relating to the traceability to the SI base units of quantitative chemical measurements; (2) to coordinate the activities of the national metrology laboratories in establishing this traceability at the highest level; and (3) to keep under review the question of whether there is a need for a programme of work at the BIPM to support this activity. The CIPM requested the Chairman of its Working Group on Metrology in Chemistry, as acting President of the new Comité Consultatif, to undertake the necessary consultations prior to drawing up final terms of reference and a list of members of the Comité Consultatif to be proposed to the CIPM at its meeting in 1994 [6 §6].

1.2.1. Study I: Metals in Water

Study I was "an interlaboratory comparison of a relatively simple matrix using isotope dilution mass spectrometry (IDMS). … IDMS was chosen as the reference method because its inaccuracies and imprecisions have been proven to be relatively small and it is used within some national traceability networks to evaluate the accuracy of other chemical methods. … The Inorganic Analytical Research Division [a Division of NIST's Chemical Science and Technology Laboratory (CSTL), a precursor to the Material Measurement Laboratory] took the lead in designing the IDMS interlaboratory project, coordinating the preparation and distribution of the unknown samples, and providing the separated stable isotopes required for IDMS. In addition, the samples were analyzed (blind) in the Mass Spectrometry Group using inductively coupled plasma (ICP-MS) and thermal ionization (TIMS) mass spectrometry. The unknown samples were prepared from primary reference materials, in a manner similar to a class of Solution Standard [Standard Reference Materials®, SRMs®]. The gravimetrically calculated concentrations were taken as target values and the stated goal of the exercise is to be within ± 1 % of these target values" [7 §III-B-13].

Study I was designed by NIST and conducted (and completed) in 1993. Six metal measurands were evaluated, three each in two aqueous solutions: Mn, Pb, and Cd in one and Li, Mo, and Fe in the other. In addition to results from nine NMIs or DIs using IDMS, ten results from non-IDMS techniques were provided by NMI, DIs, and expert laboratories.

The results from Study I were presented in detail to the participants, other members of WGMC, and to the CIPM. However, no public report of the results was issued. Fortunately, the leader of Study I, Dr. Robert L. Watters, Jr., preserved a copy of the workbook used to collect and analyze the results. While the complete results for Study I are available (as pilot study "CCQM-P0") in the master copy of the *CCQM_Retrospectoscope* held at NIST, since there was no publication agreement among the participants in Study I and several of the participating organizations are no longer active, results cannot be distributed nor can the participants (other than NIST) be identified. To help document the lessons learned from Study I, all results for the six measurands are displayed in Fig. 1.

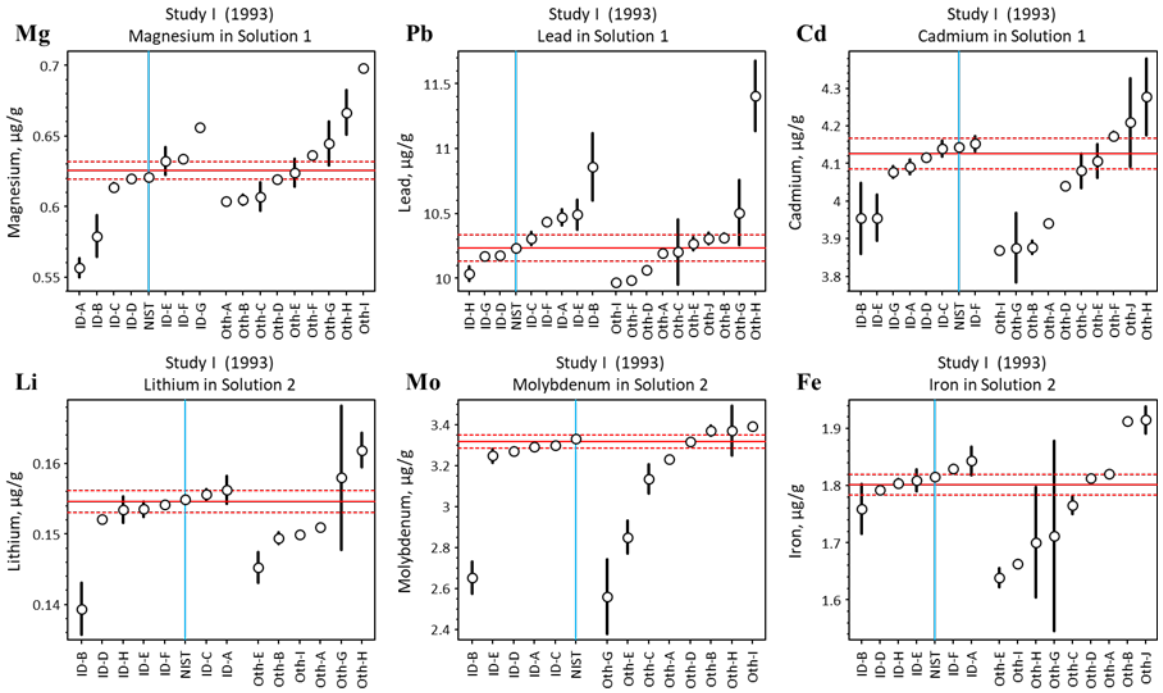


Fig. 1. Results of Study I, CIPM Working Group on Metrology in Chemistry.

Each panel shows all reported results for one Study I measurand. Results are sorted in order of increasing magnitude within IDMS and non-IDMS groupings. Open circles represent results; bars span ± standard uncertainties. Horizontal solid red lines denote gravimetric preparative reference values; dashed red lines bound ± 1 % intervals about the reference value. The blue vertical lines mark NIST results (analyzed "blind" by staff not involved in preparing the solutions.)

1.2.2. Study II: Primary Standard Gas Mixtures

The first international comparison of primary standard gas mixtures, initially called "Study II" was also initiated in 1992 but ran in stages until 1998. It was designed and conducted by the Netherlands NMI. Like Study I, the target agreement was initially set at 1 % relative although IDMS was only one of the five preferred methods of measurement. NIST was one of the ten NMI or DI participants with experience and expertise with these methods. The study evaluated four two-component mixtures (CO, CO_2, NO, and SO_2 in N_2) and three six-component "natural gas" mixtures of (N_2, CO_2, ethane, propane, and butane in methane).

Unlike Study I, most of the results of even the initial mixtures (CO and CO_2 in N_2) were within the initial ± 1 % target range. In 1999, the Study II results provided by the experienced NMIs and DIs were designated as the first CCQM Key Comparison, CCQM-K1.a to CCQM-K1.g [8]. These fully attributed results are available in the *CCQM_Retrospectoscope*.

1.3. 1994: Consultative Committee for the Quantity of Matter (CCQM)

Although Dr. Lyons resigned from the CIPM at the end of the 1993 meeting and would leave NIST later in the year, at the request of the CIPM he and Dr. Quinn undertook "the necessary consultations prior to proposing definitive terms of reference and a list of members to the Comité in September 1994" [9 §4.3]. Presented to the CIPM in 1994 by Dr. Quinn, "The working group on metrology in chemistry recommended to the CIPM the following terms of reference: (1) to advise the CIPM on matters relating to the accuracy of quantitative chemical measurements and traceability to the SI; (2) to coordinate the activities of national metrology laboratories in establishing this traceability at the highest level; (3) to stimulate the understanding of the concept of uncertainty and the assignment of uncertainty statements in chemical measurements, thereby encouraging the establishment of traceability, taking into account other initiatives at regional and international levels; and (4) to keep under review the question of whether or not there is a need for a programme of work at the BIPM to support this activity" [9 §4.3].

The report also observed that "The results [of Study I] show, as might have been expected, that the agreement between laboratories was not as good as the uncertainties given by each laboratory imply. Examination of the detailed information given by the participating laboratories shows that significant differences existed in the practice of IDMS among the laboratories. Many of these could be removed quite easily by following procedures developed by the small number of laboratories highly expert in the field. The working group proposed, therefore, that a small task group be established to draw up guidelines for the high-accuracy use of IDMS and at the same time to study the limitations to the accuracy of the method. It appears that no internationally agreed document of this sort exists. The participants at the meeting were of the opinion that such a document would be of great value not only to the metrological laboratories that participated in the study, but also to the much wider community of analytical chemists. It was emphasized that a key component of the document would be the treatment of the uncertainties inherent in the method. The working group also proposed that a "second inter-laboratory study be carried out when the guidelines have been prepared" [9 §4.3].

Dr. Robert Kaarls of the Netherlands NMI was elected as President of the CCQM.

1.3.1. Study III: Lead in Water

In response to the charge from the CIPM and now under the auspices of the CCQM, NIST staff prepared a detailed protocol for the IDMS analysis of inorganic elements using inductively coupled plasma ionization [10] and designed what was then called Study III to

test its utility for improving metrological traceability. Study III only characterized one measurand (Pb) at a concentration and in an acidic aqueous solution very similar to that used in Study I.

Study III is now known as the pilot study CCQM-P1, but no public report was issued. However, again thanks to Dr. Watters, the data were preserved and are available in the master copy of the *CCQM_Retrospectoscope* held at NIST. Since there was no publication agreement among the participants in Study III, results cannot be distributed nor can the participating organizations (other than NIST) be identified. But to document the dramatic improvement possible when participants use the same materials and methods, the anonymized results of Study III are displayed in Fig. 2.

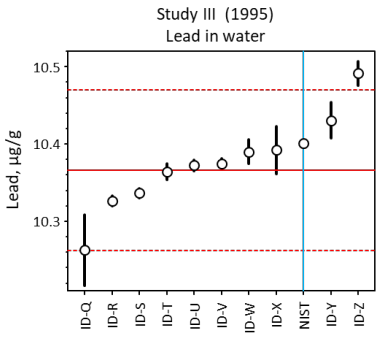


Fig. 2. Results of CIPM Study III/CCQM-P1 Lead in Water.

Results are sorted in order of increasing magnitude. Open circles represent results; bars span ± one standard uncertainties. The horizontal solid red line denotes the gravimetric preparative reference value; dashed red lines bound the ± 1 % interval about the reference value. The blue vertical line marks the NIST results.

1.4. 1995: CCQM Studies

The first meeting of the CCQM was held at the BIPM on 19-20 April 1995. This meeting and those for the following two years were mostly focused on organization and planning. With the exception of the later stages of the "Study II" (now CCQM-K1.c to CCQM-K1.g), the first CCQM-sponsored comparison providing attributed results was completed in 1998. This study, CCQM-K2 cadmium and lead in natural water, is acknowledged to be a continuation of CCQM-P1 [11], i.e., CIPM Study III. Like Study III, all participants used IDMS although they used their own choice of materials, procedures, and data analysis methods. While not prescriptive like Study III and addressing measurands in a more complicated matrix and at levels 1000-fold lower than in Studies I and III, as can be seen in Fig. 3 most of the CCQM-K2 results are within ± 1 % of the consensus-estimated reference values. The results of chemical measurements can in practice, as well as principle, be made metrologically traceable to established references.

Note: CCQM-K2 was designed and conducted by the European Commission's Joint Research Centre. NIST staff (John Moody) prepared the natural water samples used in the study [11].

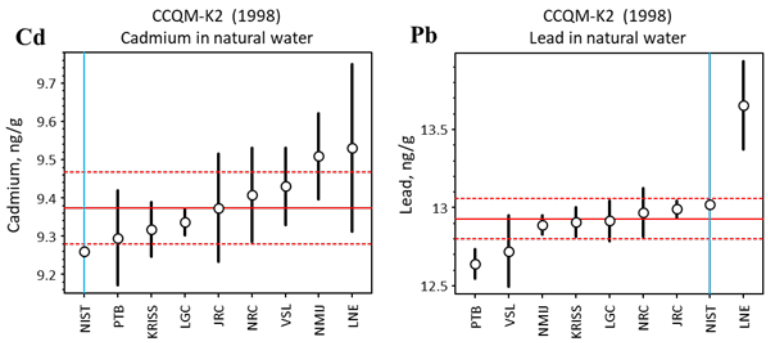


Fig. 3. Results of CCQM-K2, Cadmium and Lead in Natural Water.

Each panel shows all reported results for one CCQM-K2 measurand. Results are sorted in order of increasing magnitude. Open circles represent results; bars span ± standard uncertainties. Horizontal solid red lines denote consensus reference values; dashed red lines bound ± 1 % intervals about the reference value. The blue vertical lines mark NIST results.

1.5. BIPM and the Terms of Reference

The CIPM's original (1990) sole term of reference for Dr. Lyons and the AHWG was whether or not the BIPM should have a "significant role" in chemical metrology. By 1992, the terms of reference had expanded to focus on metrological traceability within the chemical measurement community but continued the charge "to keep under review" the BIPM's role in chemical metrology. The 1994 terms were further expanded to include a focus on the chemical measurement community's evaluation of the uncertainty in their measurements but once again continued the charge to review the BIPM's potential role.

In 1998 (and without CCQM review) the BIPM initiated a four-person effort in chemical metrology with an initial focus on gas measurements. In May 2000 Robert Wielgosz became Head of the BIPM Chemistry Section. In cooperation with the GAWG, a program in the analysis of gas mixtures was proposed in 2001. Also in 2001, a collaborative program between BIPM and NIST in ozone photometry (see Section 7.4.1) was identified as a major focus for the Section. In cooperation with the OAWG, by 2003 a program in purity assessment of relatively pure organic compounds was identified; this program was realized in 2005 with the hiring of additional Section staff. Additional programs continue to be proposed by the Chemistry Section and reviewed by the CCQM.

2. Data

All versions of the *CCQM_Retrospectoscope* database contain results from the CCQM comparisons published in the Bureau International des Poids et Mesures (BIPM) Key Comparison Database (KCDB) [2], available on the BIPM's "Final reports – QM: Pilot studies" website [12] or have otherwise been made publicly available. The NIST version of the *CCQM_Retrospectoscope* also contains results from many confidential exploratory or training studies.

2.1. CCQM Studies

CIPM-authorized studies are intended to support the "mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes" [13]. In the absence of national standards for most chemical and biological measurands, it is (tacitly) understood that CCQM studies support the mutual recognition of measurement capabilities. Unofficially (but compellingly), they are also tools for helping the participant communities improve measurement comparability.

There are three formally recognized types of CCQM study:

- Key Comparisons (KCs) are carried out by the various Working Groups (WGs) of the CCQM, the BIPM, and by RMO Technical Committees (TCs). Participation in KCs is limited to NMIs and DIs. Fully attributed results from these studies are made public via the KCDB along with the estimated degrees of equivalence (DoE) among the participants [14].
- Supplementary Comparisons (SCs) are carried out by RMOs in support of measurement issues of concern to member organizations within their geographical region of responsibility that are not addressed by KCs. Like KCs, fully attributed results from SCs are made public via the KCDB.

  Note: The term "Subsequent Comparison" refers to a (typically bilateral) study to allow participants who did not perform well in a KC to demonstrate their improved measurement performance. Subsequent comparisons are KCs, not SCs.
- Pilot studies (PSs) are carried out chiefly by WGs and occasionally by RMOs to assess newly established measurement capabilities, explore new measurement challenges, or provide participants with experience without the necessity of public disclosure. Results from PSs need not be communicated to any except the participants themselves but are generally disclosed to interested members of the sponsoring organization. With the agreement of all participants and the sponsor, results may be publicly disclosed in summary form, with anonymized results, or fully attributed. PSs are often conducted in parallel with a KC, sometimes involving more challenging measurands or offering the ability to judge level of competence against KC participants.

Results from KCs and SCs are added to the public version of the *CCQM_Retrospectoscope* as they become available in the KCDB. Results from published pilot studies are added when they are posted on the BIPM's "Final reports – QM: Pilot studies" webpage (https://www.bipm.org/en/committees/cc/ccqm/pilot-studies) or discovered by web-search. Results from confidential studies are **not** available in the public version of the *CCQM_Retrospectoscope* but are added to NIST's version as reports become available.

2.1.1. Study Codes

All CCQM studies are provided by the BIPM with a multi-part code of the form XQM-YZ where X identifies the organizing body (see Section 2.4); Y identifies the study type ("K" for Key Comparisons, "S" for Supplementary Comparisons, and "P" for pilot studies), and Z is a one-to-three digit study number followed by a variety of version designators as needed. While unambiguous, the official codes do not distinguish between published and confidential pilot studies, and the variability in the number of study digits complicates sorting.

The *CCQM_Retrospectoscope* database uses a standardized version of the official code. The variable length X (organizing body) is reduced to two letters, "Q" is used as the study type for published PSs, and the variable number of digits in Z (the study number) is always three digits with leading zeros as needed. That is:

- CCQM-K001.a in the *CCQM_Retrospectoscope* is officially CCQM-K1.a, the first Key Comparison, with sub-studies labeled a through g, organized by the consultative committee
- CCQM-K010 is officially CCQM-K10,
- CCQM-P001 is officially CCQM-P1, the first Pilot Study organized by the consultative committee
- CCQM-Q055.2.2018 is officially CCQM-P55.2.2018
- BIQM-K001 is officially BIPM.QM-K1, and
- EUQM-S003 is officially EUROMET.QM S3.

2.2. Datasets

Each of the *CCQM_Retrospectoscope*'s dataset contains the reported results for one measurand from one study, an assigned reference value (*RV*), an expanded uncertainty on the RV at a 95 % level of confidence ($U_{95}(RV)$), and descriptive metadata. Each dataset has a unique title that identifies the comparison and the measurand.

A CCQM study may address multiple measurands, thus generating multiple datasets.

Note: Only KCs and SCs are required to provide RVs, although many PSs also provide them. In the absence of a study-provided {RV, U95(RV)}, the CCQM_Retrospectoscope uses robust consensus estimates (see Section 2.10.5).

2.3. Measurement Results

Each measurement result reported by a study participant consists of a value (x_i), an expanded uncertainty at a 95 % level of confidence ($U_{95}(x_i)$), and a unit. To be included in the study, results must be reported to the study coordinator in a timely manner. In KCs and SCs, each participant may report only one "official" result per measurand. Participants may (and often do) report multiple values in PSs; however, the *CCQM_Retrospectoscope* only uses the average of these values in its analyses.

2.4. Organizing Bodies

Not all the world's NMIs and DIs participate directly in CCQM studies, which are reserved for the NMIs and DIs of typically large economies. RMOs are responsible for supporting the NMIs and DIs of the smaller economies within their regions. The geographic regions covered by the six current RMOs are pictured in Fig. 4. GULFMET is still developing its capabilities in the areas of chemistry and biology and has not yet sponsored CCQM studies.

Fig. 4. Geographic Areas of Metrological Responsibility.

In addition to the CCQM and the RMOs, the BIPM currently sponsors one chemically related KC – an ongoing comparison of standard reference photometer ozone measurements.

Table 1 lists the number of datasets and studies in the *CCQM_Retrospectoscope* database as of 1-April-2024, itemized by organizing body.

Table 1. Number of Datasets and Studies Attributable to Different Sponsoring Bodies.

		Number of Datasets					Number of Studies				
Body	Code	KC	SC	PPS	PS	Total	KC	SC	PPS	PS	Total
AFRIMETS	AFQM	2				2	1				1
APMP	APQM	21	57	6	11	95	13	22	1	3	39
BIPM	BIQM	36				36	17				17
CCQM	CCQM	784		238	576	1598	230		73	182	485
COOMET	CoQM	18	14			32	8	6			14
EURAMET	EUQM	17	88			105	9	13			22
SIM	SIQM	11	44			55	6	11			17
	Total	889	203	244	587	1923	284	52	74	185	595

Note: “EURAMET” (European Association of National Metrology Institutes) was known as “EUROMET” (European Collaboration in Measurement Standards) until 2007 [15]. The KCDB regards them as separate organizations, but they are combined in the CCQM_Retrospectoscope.

2.5. Working Groups

Within the CCQM, responsibilities for coordinating studies of the diverse types of chemical and biological measurands are spread among eight Working Groups (WGs). RMO Technical Committees (TCs) coordinate studies using the same division of responsibilities. Unlike the CCQM and the RMOs, the *CCQM_Retrospectoscope* system regards Technical Committees as extensions of the Working Groups. For example, the BIPM’s ozone studies are treated as an RMO’s extension of activities within the Gas Analysis Working Group (GAWG).

Several KCs and PSs have been conducted by more than one WG. The *CCQM_Retrospectoscope* assigns the results from such joint studies to the most appropriate WG on the basis of the units used to report results. For example, results from joint Inorganic Analysis Working Group (IAWG) and Electrochemical Analysis Working Group (EAWG) studies that were reported in units of mass fraction (g/g) are assigned to the IAWG.

The number and identity of the WGs has evolved over time. The Bioanalysis Working Group (BAWG) was the original WG devoted to biological measurements. In 2015 it was divided into the Cell Analysis (CAWG), Nucleic Acid analysis (NAWG), and Protein Analysis (PAWG) Working Groups. Within the *CCQM_Retrospectoscope*, the studies sponsored by the BAWG have been (retrospectively) assigned to the most appropriate current WG based on the nature of the measurand.

The Isotopic Ratio Working Group (IRWG) was split from the IAWG in 2018. However, the IAWG carried out several earlier studies that produced datasets that are now within the remit of the IRWG. Within the *CCQM_Retrospectoscope*, these have been (retrospectively) assigned to the IRWG.

Table 2 lists the number of datasets and studies in the *CCQM_Retrospectoscope* database as of 1-April-2024, itemized by WG.

Table 2. Number of Datasets and Studies Attributable to Different CCQM Working Groups.

Working Group		Number of Datasets					Number of Studies				
Responsibility	Code	KC	SC	PPS	PS	Total	KC	SC	PPS	PS	Total
Cell Analysis	CAWG			4	4	8			3	4	7
Electrochemical Analysis	EAWG	100	8	7	41	156	27	2	2	15	46
Gas Analysis	GAWG	365	132	74	7	578	101	31	23	4	159
Inorganic Analysis	IAWG	200	52	28	251	531	63	13	9	76	161
Isotopic Ratio	IRWG	15		39	10	64	5		6	4	15
Nucleic Acid Analysis	NAWG	8		10	58	76	4		2	19	25
Organic Analysis	OAWG	139	11	16	181	347	65	6	8	58	137
Protein Analysis	PAWG	9		23	19	51	5		11	4	20
Surface Analysis	SAWG	53		43	16	112	14		10	1	25
	Total	889	203	244	587	1923	284	52	74	185	595

In addition to these permanent WGs, the CCQM has created a number of *ad hoc* Working, Steering, and Task Groups to address specific issues. One such, the *ad hoc* Steering Group on Microbial Measurements (MBSG), conducted "pre-pilot" investagative studies adressing microbiological quantification and identification technologies. Since they were designed to assess metrological interest in relevant technologies rather than participants' capabilities, results from these studies are not included in the *CCQM_Retrospectoscope*. See Section 14 for further information.

2.6. Participating Organizations

The *CCQM_Retrospectoscope* maintains a list of every (identified) participant in a CCQM study that reported quantitative results. The list is updated as necessary when new studies are added to the datasheets.

The code names used to designate NMIs and DIs are (generally) the "official" codes listed on the BIPM's https://www.bipm.org/en/cipm-mra/participation webpage. However, in the interest of brevity some codes are modified; e.g., Turkiye's TÜBITAK Ulusal Metroloji Enstitüsü (TÜBITAK UME) is coded UME. Results from institutions that once participated independently but have been incorporated into a larger entity have been recoded. For example, results reported by Japan's National Institute of Advanced Industrial Science and Technology (AIST) are now assigned to the National Metrology Institute of Japan (NMIJ) and results from Russia's Ural'skiy Nauchno-Issledovatel'skiy Institut Metrologii (UNIIM) are assigned to D. I. Mendeleev Institute for Metrology (VNIIM).

Measurement laboratories associated with several international organizations are official participants in selected studies and for the purposes of the *CCQM_Retrospectoscope* are considered to be NMIs or DIs. These laboratories include the International Atomic Energy Agency (IAEA) for selected inorganic measurands, World Meteorology Organization (WMO) designees for atmospheric gases, and the BIPM for organic purity.

While participation in CCQM KCs and SCs is restricted to NMIs and DIs, a number of university and commercial laboratories participate in pilot studies. Some of these pilot studies have been published with full attribution of all participants. Very occasionally a non-

NMI or DI has been an “unofficial” participant in a KC or SC. While results from such participants are not used to assign reference values, their results have been included in the *CCQM_Retrospectoscope*’s datasets.

2.7. Year of Measurement

The *CCQM_Retrospectoscope* maintains a complete list of CCQM studies, their completion status, and the (approximate) year that the measurement results had to be reported to the study coordinator. When possible, this measurement year is derived from information provide in the study reports and may differ from the information provided in the KCDB. However, many reports for early CCQM studies did not provide the study timetable. When dates were not explicitly provided or could not be inferred from the reports, they have been assigned based on information provided in the KCDB.

2.8. Base units

The GAWG, with responsibilities for measuring the components of gas mixtures, describes the results of its studies in terms of mole analyte per mole mixture (amount-of-substance fraction, henceforth referred to as mole fraction, mol/mol). Likewise, the IAWG, with responsibilities for measuring elemental composition in solid and liquid matrices describes results in terms that either are or can be readily converted to mass analyte per mass matrix (mass fraction, g/g). However, other WGs have responsibilities in several measurement domains. For example the EAWG supports measurements of pH (pH) and conductivity (S/m, S/S) and the Surface Analysis Working Group (SAWG) supports measurements related to the thickness, composition, and sorption properties of surfaces and films. Since these measurement areas differ qualitatively, datasets within a WG are grouped by the “base unit” of the measurand.

A base unit is the unit of measure stripped of prefixes: M, d, c, m, µ, n, p, and f. With apologies to the General Conference on Weights and Measures (CGPM), the keepers of the Système international d'unités (SI), this Report accordingly uses “g” rather than “kg” as the base unit of mass. There are also measurement results that cannot be expressed in SI units: effective fluorescence (EFF), the isotopic delta scale (‰), and practical salinity (PSU). There are a few PS and PPS qualitative datasets that provide participant identity without associated quantitative results that use “arbitrary unit” (a.u.) as a placeholder base unit.

Table 3 lists the base units used in KC and SC datasets by the various WGs; Table 4 lists the base units used in PPS and PS datasets. Both tables are sorted in order of decreasing number of useful datasets per base unit, where “useful” is defined as containing data from just one study and holding at least one attributed quantitative result. Most (118) of the 150 “not useful” datasets contain results from multiple related studies (typically a CCQM KC with one or more derivative RMO KCs). All but one of the remainder are PPSs or PSs that do not provide attributed results; the lone exception is a dataset in which the mass fraction values are reported as at-or-below detectable limits.

Table 3. Base Units Used in Key and Subsequent Datasets by Working Group.

Base Unit		CAWG	EAWG	GAWG	IAWG	IRWG	NAWG	OAWG	PAWG	SAWG	Total
mole fraction	mol/mol			409	5	4				4	422
mass fraction	g/g				234			146	9	1	390
acidity	pH		72								72
conductivity	S/m		20								20
length	m									17	17
number fraction	n/n					8	6				14
amount content	mol/g									11	11
number concentration	n/L			10							10
charge concentration	C/L			10							10
surface area	m^2/g									3	3
isotopic δ-scale	‰					2					2
mass concentration	g/L									2	2
pore volume	cm^3/g						2				2
molar mass	g/mol					1					1
	Total	0	92	429	239	15	8	146	9	38	976

Table 4. Base Units Used in Pilot Study Datasets by Working Group.

Base Unit		CAWG	EAWG	GAWG	IAWG	IRWG	NAWG	OAWG	PAWG	SAWG	Total
mass fraction	g/g				268			186	37	1	492
mole fraction	mol/mol			79				10			89
number fraction	n/n					24	28				52
length	m									34	34
number concentration	n/L	2					21				23
acidity	pH		22								22
isotopic δ-scale	‰					21					21
arbitrary unit	a.u.									16	16
conductivity ratio	S/S		14								14
base pair count	bp						9				9
conductivity	S/m		9								9
mass concentration	g/L						3		1		4
amount content	mol/g									3	3
molar mass	g/mol					2					2
pore volume	cm^3/g									1	1
fluorescence	EFF	1									1
surface area	m^2/g									1	1
salinity	PSU		1								1
	Total	3	46	79	268	47	61	196	38	56	794

2.9. Degrees of Equivalence

"The degrees of equivalence of national measurement standards are understood as the degrees to which those standards are consistent with Key Comparison reference values and hence consistent with other national standards. A degree of equivalence is expressed quantitatively by two terms: a deviation from the Key Comparison reference value and an expanded uncertainty in that deviation, evaluated at a 95 % level of confidence (in practice, this is often approximated by using a coverage factor *k* equal to 2)" [14].

In the context of CCQM comparisons, the degrees of measurement capability equivalence (DoE) for a given measurement challenge is the deviation of a measurement result, $x_i \pm u(x_i)$, from a reference value, $x_{\text{ref}} \pm u(x_{\text{ref}})$, relative to the expanded uncertainty of the deviation at a 95 % level of confidence:

$$d_i = x_i - x_{\text{ref}} \tag{1}$$

$$u(d_i) = \sqrt{u^2(x_i) + u^2(x_{\text{ref}}) - 2 \times \text{cov}(x_i, x_{\text{ref}})} \tag{2}$$

$$U_{95}(d_i) = 2 \times u(d_i) \tag{3}$$

where "cov(x_i,x_{ref})" is the covariance between result x_i and the reference value [16]. DoE are formally specified as the pair {d_i, $U_{95}(d_i)$}.

A result is considered equivalent to the *RV* if the interval d_i - $U_{95}(d_i)$ to d_i + $U_{95}(d_i)$ includes zero. Assuming the d_i are (approximately) normally distributed, about 95 % of the $x_i \pm u(x_i)$ that are truly equivalent to the $x_{\text{ref}} \pm u(x_{\text{ref}})$ will have $|d_i/u(d_i)| \leq 2$.

The covariance term is zero when x_{ref} is independent of x_i. When x_{ref} is estimated using the x_i in some way, the magnitude of the covariance depends on the number of results used and the statistical methods employed [16]. To avoid confusion with published DoE, the *CCQM_Retrospectoscope* provides the related ζ (zeta)-score [17] bias metric (Section 3.2.1)

$$\zeta_i = \frac{x_i - x_{\text{ref}}}{\sqrt{u^2(x_i) + u^2(x_{\text{ref}})}} \tag{4}$$

where x_{ref} and $u(x_{\text{ref}}) = U_{95}(x_{\text{ref}})/2$ are provided in the study report and the covariances are all assumed to be zero.

Assuming the ζ_i are (approximately) normally distributed, about 95 % of the $x_i \pm u(x_i)$ that are truly equivalent to the $x_{\text{ref}} \pm u(x_{\text{ref}})$ should have $|\zeta_i| \leq 2$.

2.10. Transformations

Comparing results across datasets is facilitated when expressed in a standardized manner. To this end, results from some studies have been transformed in various ways.

2.10.1. Units

For the most part, studies conducted by the IAWG and OAWG report results in units of mass fraction (g/g). However, some studies have used units of amount content (mol/g), amount-of-substance concentration (mol/L), or mass concentration (g/L). Where practical, such results are transformed to have units of mass fraction:

$$z_i \pm u(z_i) = K\big(x_i \pm u(x_i)\big) \tag{5}$$

where z_i is the result for the i^{th} participant expressed as mass fraction, $u(z_i)$ is its standard uncertainty, x_i is the result as reported, $u(x_i)$ is its standard uncertainty, and K is a units conversion constant. For amount content, K is the relative molecular mass of the analyte in grams per mole; for mass concentration, K is the reciprocal of the mass in grams per liter of solvent; the amount-of-substance concentration transformation combines the two factors. Since the conversion constant scales the reference value as well as the individual results, any uncertainty in K is ignored as being of little consequence.

2.10.2. Uncertainties

CIPM DoE are required to be "evaluated at a 95 % level of confidence" with the caveat that "in practice" the expanded uncertainty is often estimated from the standard uncertainty, $U_{95}(x) = k \cdot u(x)$, "using a coverage factor k equal to 2" [14]. This explicitly assumes that individual results, $x_i \pm u(x_i)$, and reference values, $x_{\text{ref}} \pm u(x_{\text{ref}})$, represent well-defined "normal" $\mathrm{N}(\mu, \sigma)$ distributions.

Most CCQM studies report $U_{95}(x_i)$. Regardless of whether the participant's $u(x_i)$ are also provided or how the $U_{95}(x_i)$ were determined, the *CCQM_Retrospectoscope* estimates participant standard uncertainties as

$$u(x_i) = U_{95}(x_i)/2\,. \tag{6}$$

A very few CCQM results have been reported as asymmetric intervals: $x^{+U_{95}(x_{\text{hi}})}_{-U_{95}(x_{\text{lo}})}$. Lacking the infrastructure to make full use of this information the uncertainty recorded in the *CCQM_Retrospectoscope* datasheets is the average:

$$U_{95}(x) = \frac{U_{95}(x_{\text{lo}}) + U_{95}(x_{\text{hi}})}{2}\,. \tag{7}$$

Most CCQM studies report $U_{95}(x_{\text{ref}})$. When only $u(x_{\text{ref}})$ is provided the *CCQM_Retrospectoscope* assumes that

$$U_{95}(x_{\text{ref}}) = 2 \times u(x_{\text{ref}})\,. \tag{8}$$

2.10.3. Reference Values

A frequent CCQM study design has a coordinator send a subsample of a well-characterized material to each participant. To the extent that the material is homogeneous and stable, a single reference value is appropriate for each measurand determined for that material. However, many GAWG studies involve separate gas cylinders for each participant, each cylinder containing a similar-but-different gas mixture and thus requiring individual reference values. The cylinders may be prepared and value-assigned by the coordinator and sent to the participants for analysis or they may be prepared and value-assigned by each participant and sent to the coordinator for analysis under repeatability conditions. In either case if the between-cylinder differences in the mixture are small the results are readily converted to a single representative reference value:

$$z_i \pm u(z_i) = y_{\text{ref}} \left(\frac{x_i \pm u(x_i)}{y_i \pm u(y_i)} \right) \tag{9}$$

where: $y_i \pm u(y_i)$ are the coordinator's measurement results and y_{ref} is their mean. The 95 % level of confidence uncertainty for y_{ref} is twice the pooled $u(y_i)$:

$$U_{95}(y_{\text{ref}}) = 2 \times \sqrt{\frac{\sum_{i=1}^{n} u^2(y_i)}{n}} \tag{10}$$

where n is the number of the coordinator's measurements.

2.10.4. Reference Functions

When a study involves materials of quite different measurand level (e.g., certified reference materials for the same measurand, gas mixtures of very different analyte concentration, or measurands that degrade in a predictable manner over time), the DoE are estimated by comparing the reported results to a reference function rather than a single reference value. How the reference function is established depends on the design of the study but always involves material-specific measurements made by a coordinator. Rather than attempting to reproduce the (often fairly complex) calculations used to convert participant-reported values to a common scale, the published DoE are transformed to have the units of original measurement.

DoE expressed as values with units of the measurement, $d_i \pm U_{95}(d_i)$, can be transformed:

$$z_i = x_{\text{ref}} + d_i;\ u(z_i) = U_{95}(d_i)/2\ . \tag{11}$$

DoE expressed as percent relative values, $\%d_i \pm U_{95}(\%d_i)$, can be transformed:

$$z_i = x_{\text{ref}} \times \left(1 + \frac{\%d_i}{100}\right);\ u(z_i) = z_i \times \left(\frac{U_{95}(\%d_i)/2}{100}\right). \tag{12}$$

Since the materials used may have very different measurand levels, there is no simple transformation that will completely represent measurement reproducibility as a function of measurand level. However, in both cases x_{ref} estimated as the median of the participant results, $x_i \pm u(x_i)$, for all the materials in the study (probably) minimizes the distortion. The uncertainty assigned to this x_{ref} can be estimated from the pooled relative uncertainties:

$$U_{95}(x_{\text{ref}}) = 2 \times x_{\text{ref}} \times \sqrt{\sum_{i=1}^{n} \left(\frac{u(x_i)}{x_i}\right)^2 / n} \quad (13)$$

where *n* is the number of participant measurements.

2.10.5. Pilot Study Reference Values

KC and SC datasets from successfully completed studies are published with what (at the time the study was finalized was considered) an appropriately estimated reference value and associated uncertainty. These estimates provide the reference values for "parallel pilot" studies that address the same measurands in the same sample materials at the same time as the related KC or SC. Many other PSs provide statistical summaries of the reported results that can be used as reference values although they are not referred to as such. However, some PSs do not provide summary values in the explicit (if futile) intention to hinder use of the study results in quantitative comparisons: the *CCQM_Retrospectoscope* uses the median of the reported results as x_{ref} (see Section 3.1.1) and a robust standard deviation of the mean as $U_{95}(x_{\text{ref}})$:

$$U_{95}(x_{\text{ref}}) = 2 \times Q_n / \sqrt{n} \quad (14)$$

where Q_n is a robust standard deviation of the reported results (see Section 3.1.2) and *n* is the number of reported results.

2.10.6. Multiple Results

Since "Participating institutes shall be listed only with one degree of equivalence per measurand" [14], KC and SC datasets contain only one value per measurand. However, many PS participants report multiple values, typically from different measurement procedures but also from different laboratories or analysts. To facilitate analysis, every value is assumed to be equally important, and the multiple results replaced with their arithmetic average:

$$\bar{x}_i = \left(\sum_{j=1}^{n_i} x_{i,j}\right) / n_i \quad (15)$$

where n_i is the number of values reported by the i^{th} participant. The uncertainty associated with $\bar{x}_i$ combines the standard deviation of the values with the pooled uncertainty of the individual value:

$$u(\bar{x}_i) = \sqrt{\frac{\sum_{j=1}^{n_i} \frac{\left(x_{i,j} - \bar{x}_i\right)^2}{n_i - 1} + \sum_{j=1}^{n_i} u^2(x_{i,j}) / n_i}{n_i}}. \quad (16)$$

3. Summary Statistics

3.1. Distribution Characterization

Given the typically small number of results in CCQM datasets, the *CCQM_Retrospectoscope* summarizes sets of values assuming that at least the majority of the values can be usefully described as following a "normal" $\mathrm{N}(\hat{\mu}, \hat{\sigma})$ distribution, where $\hat{\mu}$ and $\hat{\sigma}$ are robust estimates of the "true" location (mean) and scale (standard deviation) of the distribution.

3.1.1. Median: a Robust Estimate of Location

The median is a widely used robust (i.e., fairly insensitive to atypical values) estimator of the central location of unimodal distributions that is reasonably statistically efficient (i.e., provides values close to the true value when applied to truly normally distributed values) [18]. It is calculated as the middle value of the set of values and is denoted in this document as Median{·}, where the {·} denotes a set of values.

The median has a breakdown point (that is, it ceases to provide a useful estimate) when the proportion of "atypical" values exceeds 50 %.

3.1.2. Q_n: a Robust Estimate of Scale

The Q_n is a robust and efficient estimator of scale for unimodal distributions [19]. Q_n is the name assigned by its developer. It is calculated from the first quartile (smallest 25 %) of the absolute pairwise differences between values, scaled by a function of the number of values being summarized. The value of the Q_n estimator for a given set of values is denoted in this document as $Q_n\{\cdot\}$.

While the standard deviation is extremely sensitive to atypical values (it has a breakdown point of 0 %, potentially ceasing to be a meaningful summary when there is even one atypical value), the Q_n has the same 50 % breakdown point as the median. The Q_n shares this robust breakdown with the more commonly encountered "adjusted median absolute deviation from the median" ($\mathrm{MAD_E}$) estimator but is considerably more efficient, 88 % compared to the $\mathrm{MAD_E}$'s 37 % [20]. The Q_n has been proposed as the most generally useful scale estimator for characterizing interlaboratory precision studies [21].

3.2. Trends

Participating in a given CCQM study illuminates measurement performance characteristics attributable to a particular set of analysts for a particular set of measurands at a particular time. Examining performance characteristics for sets of measurands that present qualitatively similar measurement challenges using metrics that facilitate comparisons among studies may provide more generalizable illumination.

3.2.1. Measurement Bias

Assuming that the dataset *RV*s, $x_{ref} \pm U_{95}(x_{ref})$, provide good estimates of true value, the ζ-score between a NIST result, $x_{NIST} \pm U_{95}(x_{NIST})$, and the *RV* is a nominally measurand-independent measure of measurement bias:

$$\zeta_{NIST} = \frac{x_{NIST} - x_{ref}}{\sqrt{(U_{95}(x_{NIST})/2)^2 + (U_{95}(x_{ref})/2)^2}}. \quad (17)$$

The ζ_{NIST} for qualitatively similar measurands can be charted as functions of time (measurement year) and measurand level (x_{ref}), potentially revealing systematic trends.

3.2.1.1. Degrees of Equivalence Ratio

As noted in Section 2.9, zeta-scores, ζ_i, are related to the DoE ratio, $d_i/u(d_i)$, used to judge whether a result is metrologically equivalent to the *RV* [13]. However, the two metrics are not identical. It is quite possible for $|\zeta_{NIST}| > 2$ while $d_{NIST}/u(d_{NIST}) < 2$ and *vice versa*.

Before the mid-2010s, few studies that assigned consensus-based *RV*s considered the covariance between the *RV* and the results used to estimate the *RV.* Including the covariance makes $u(d_i)$ smaller, thus the ratio $d_i/u(d_i)$ larger. However, at about the same time the presence of "dark uncertainty" [22] – differences among results greater than can be attributed to the stated measurement uncertainties – began to be recognized and some studies enlarged $u(d_i)$, thus making the ratio $d_i/u(d_i)$ smaller. Published DoEs are not well suited for cross-study analysis since they incorporate study-specific decisions beyond those embedded in the assigned *RV* and $U_{95}(RV)$.

3.2.1.2. "Outsider" Results

The analyses reported in this document use $|\zeta_{NIST}|$ as a consistent (albeit approximate) estimate of whether results overlap the reference interval, with $|\zeta_{NIST}| \leq 2$ as the criterion for being sufficiently "inside" for establishing metrological equivalence. Results where $|\zeta_{NIST}| > 2$ are therefore considered as "outside" equivalence. These "outsider" results are of interest in that understanding why $|\zeta_{NIST}| > 2$ may lead to improved measurement and/or data analytical performance.

Note: Since perhaps about 80 % of IAWG and OAWG datasets have significant dark uncertainty [23], it is likely that many "outsider" results are metrologically compatible with a properly evaluated reference interval.

3.2.2. Uncertainty Over Time

If the distribution of the $U_{95}(x_i)$ within each dataset is reasonably consistent across measurands and participants, then the ratio between NIST's uncertainty assessment and the median of all the uncertainties in the dataset

$$R = \frac{U_{95}(x_{NIST})}{\text{Median}\{U_{95}(x_i)\}} = \frac{u(x_{NIST})}{\text{Median}\{u(x_i)\}} \quad (18)$$

provides insight into NIST's uncertainty assessment capability relative to that of the study participants. Charting *R* as a function of time potentially reveals changes in the uncertainty-assessing capabilities of NIST and/or the community of participants.

Note: The $U_{95}(x_{NIST})/U_{95}(x_{ref})$ ratio does not provide a measurand-independent metric since $U_{95}(x_{ref})$ reflects the distribution of the measurement values used to assign the value (which need not be participant results) rather than the distribution of participant uncertainty estimates.

3.2.3. Linear Trend Characterization

For fairly consistent sets of $\{x_i, y_i\}$ data and assuming that (1) the uncertainty in all the x_i are small relative to those of the y_i and (2) the uncertainties in the y_i are all about the same magnitude, linear trends of *y* as a function of *x*

$$y = \beta_0 + \beta_1 x \tag{19}$$

can be parameterized using classical least squares regression. The values of the β coefficients are those that minimize the root-mean-squared-error (RMSE) between the observed and predicted *y* values. The "standard error" of the parameters, $u(\beta_0)$ and $u(\beta_1)$, express the uncertainty in the estimated value of their coefficient at about a 68 % level of confidence.

However, classical regression is sensitive to the presence of atypical $\{x_i, y_i\}$. For "noisy" but plausibly linear trends, the *CCQM_Retrospectoscope* uses robust non-parametric Theil-Sen estimation [24,25]. The slope, β_1, is estimated as the Median$\{(y_j - y_i)/(x_j - x_i)\}$ for all data pairs where x_j differs from x_i. The intercept, β_0, is then estimated as the Median$\{y_i - \beta_1 x_i\}$. The $u(\beta_1)$ is estimated as the MAD_E of the $(y_j - y_i)/(x_j - x_i)$; $u(\beta_0)$ is estimated as the MAD_E of the $y_i - \beta_1 x_i$.

While all of the trends evaluated in this report are calculated using the Theil-Sen estimator, the form of the prediction equation depends on whether the *x* and/or the *y* have been logarithmically base-10 (log) transformed. Expressing *y* always in the native (untransformed) form, the four equations are as follows.

- If both *x* and *y* are untransformed

$$y = \left(\beta_0 \pm u(\beta_0)\right) + \left(\beta_1 \pm u(\beta_1)\right) \times x\,. \tag{20}$$

- If log(*x*) is the independent variable (plotted along a scattergram X-axis) and *y* is untransformed

$$y = \left(\beta_0 \pm u(\beta_0)\right) + \left(\beta_1 \pm u(\beta_1)\right) \times \log(x)\,. \tag{21}$$

- If log(*y*) is the dependent variable (plotted along a scattergram Y-axis) and *x* is untransformed

$$\log(y) = \left(\beta_0 \pm u(\beta_0)\right) + \left(\beta_1 \pm u(\beta_1)\right) \times x\,.$$

$$y = 10^{\left(\beta_0 \pm u(\beta_0)\right) + \left(\beta_1 \pm u(\beta_1)\right) \times x}\,. \tag{22}$$

- If log(*x*) is the independent variable and log(*y*) is the dependent variable

$$\log(y) = \left(\beta_0 \pm u(\beta_0)\right) + \left(\beta_1 \pm u(\beta_1)\right) \times \log(x)$$
$$y = 10^{(\beta_0 \pm u(\beta_0))} \times x^{(\beta_1 \pm u(\beta_1))}$$
$$y = \left(\beta_0' \pm u(\beta_0')\right) \times x^{(\beta_1 \pm u(\beta_1))} \quad (23)$$

where $\beta_0' = 10^{\beta_0}$ and $u(\beta_0') \cong \frac{10^{\beta_0 + u(\beta_0)} - 10^{\beta_0 - u(\beta_0)}}{2}$.

Equation 23 defines a "power function," a straight line when plotted on a scattergram with log-log axes.

3.2.4. Interpreting Linear Trends

An intercept parameter (β_0) specifies the Y-axis value for an X-axis value of zero and is not directly relevant for charts where the relevant X-axis interval is far from zero.

While the slope parameters (β_1) in the above four equations are not directly comparable, the sign of the slope, +β_1 or -β_1, is a good indicator of whether on average *y* increases or decreases with increasing *x*. And an absolute value of the *t*-statistic

$$|t| = |\beta_1| / u(\beta_1) \quad (24)$$

greater than at least two suggests that the slope may be "significantly" different from zero with the significance level increasing as $|t|$ increases.

In this document, the relevant aspect of the trend and its significance are therefore reported as

$$|t| = |\pm \beta_1| / u(\beta_1) \approx n \quad (25)$$

where ± is the sign of β_1, β_1 and $u(\beta_1)$ are rounded to two significant digits of the $u(\beta_1)$ value, and *n* is rounded to the nearest integer.

4. Overall NIST Engagement

NIST reported at least one measurement result in 264 studies sponsored by the CCQM or one of the RMOs between 1992 and 2023. Since many of these studies involved more than one measurand, NIST results are in 789 datasets. Table 5 lists the numbers of studies and datasets that NIST contributed to by WG, study type, and sponsoring body.

Table 5. Number of Finalized Studies and Datasets by Focus Area and Study Type.

	Number Studies							Number Datasets						
WG	All	KC	SC	PPS	PS	CCQM	RMO	All	KC	SC	PPS	PS	CCQM	RMO
IAWG	84	41	0	6	37	84	0	242	115	0	15	112	242	0
OAWG	66	33	0	7	26	66	0	193	81	0	15	97	193	0
GAWG	56	48	1	6	1	48	8	172	132	29	9	2	129	43
EAWG	20	12	0	1	7	20	0	51	40	0	1	10	51	0
SAWG	7	3	0	3	1	7	0	48	12	0	20	16	48	0
NAWG	11	2	0	2	7	11	0	26	4	0	2	20	26	0
PAWG	8	1	0	4	3	8	0	25	1	0	6	18	25	0
IRWG	6	1	0	3	2	5	0	24	8	0	11	5	24	0
CAWG	6	0	0	2	4	4	0	8	0	0	4	4	8	0
	264	141	1	34	88	256	8	789	393	29	83	284	746	43

The pie charts of Fig. 5 display the proportion of NIST's participations, visualizing the relative effort that NIST has invested in the focus areas.

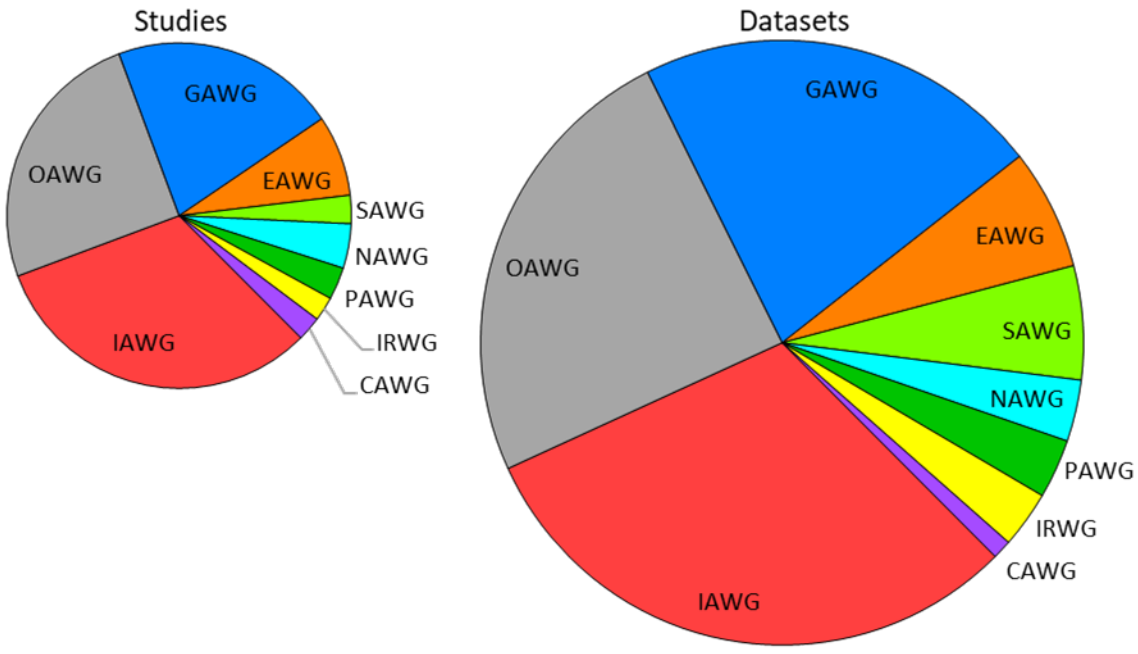


Fig. 5. Proportion of NIST's CCQM Studies and Datasets by Focus Area.
The ratio of areas of the two circles equals the average number of datasets per study: 761/257 ≈ 3.

More than 86 % of NIST's CCQM participations have been in the inorganic, organic, gas mixture, electroanalytical, and isotope ratio arenas that are the responsibility of NIST's Chemical Sciences Division (CSD).

4.1. Participations

The cumulative distributions in Fig. 6 display the time course of all NIST's participations by study type and sponsoring body, where "participation" is defined as providing at least one measurement result. Each participation increases the count by one unit regardless of the number of measurands involved.

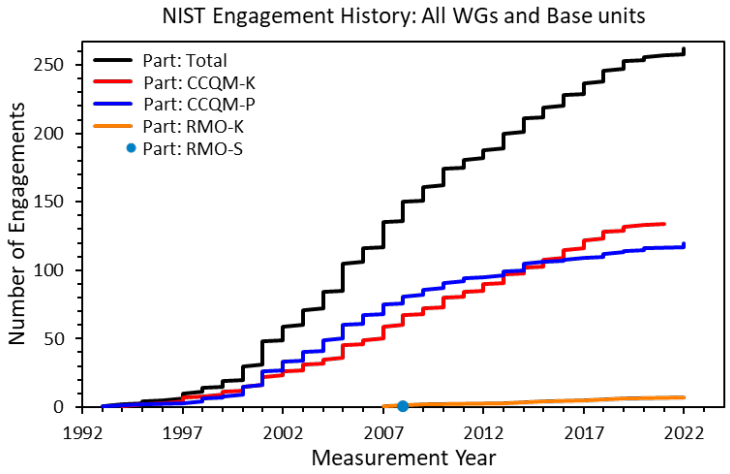


Fig. 6. Cumulative Distributions of NIST Participations in CCQM Studies.

NIST participated in KCs at a roughly constant rate (number of studies per year) from about 2000 through 2019. The PS participation rate was roughly constant to about 2008 and then sharply declined. The decline likely is a result of procedural changes by the IAWG and OAWG, replacing stand-alone studies with KC-associated "parallel pilots" for training and method validation.

The histogram traces of Fig. 7 display the time course of participation relative to the total number of CCQM studies, along with the approximate percentage of CCQM studies that NIST has participated in.

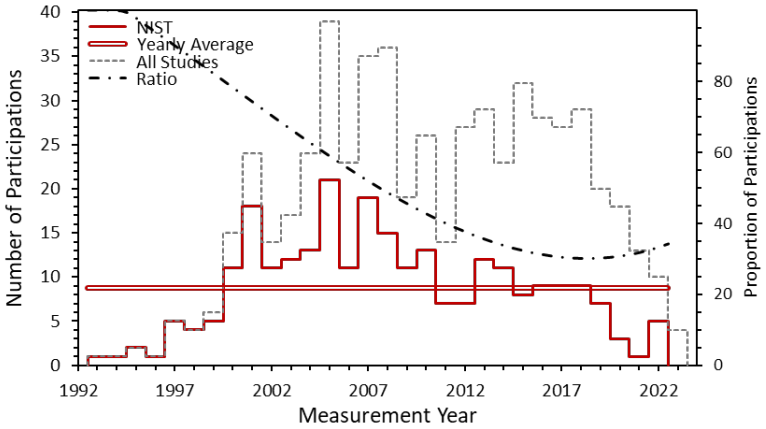


Fig. 7. Number of NIST and of Possible CCQM Study Participations Per Year.

The proportion of participation trendline is a cubic polynomial fit to the ratio between the number of NIST participations and the number of CCQM studies during each yearly interval.

NIST participated in most CCQM studies from the early 1990s to the mid-2000s with participation slowly declining to ≈20 % by 2020. The decline is likely driven by 1) the increased number of CCQM studies, 2) their increased diversity, 3) CSD's strategy on supporting Calibration and Measurement Capabilities (CMCs), and 4) fewer available CSD resources (time and people). Most early CCQM studies addressed issues of general interest; many recent studies address fairly specialized issues in which NIST has limited interest.

CMCs are self-declared but peer-reviewed and approved statements of the "calibration and measurement capability available to customers under normal conditions" [26]. While the original policy was to participate in all KCs that support analyte/matrix combinations delivered by NIST Standard Reference Materials® (SRMs®), CSD now only participates in KCs that support broad-based claims or those that directly address metrological traceability.

4.2. Coordinations

In addition to engagement through participation, NIST has coordinated or co-coordinated many studies. Coordination requires considerably greater and more sustained resource expenditure than does participation. The tasks include material characterization, packaging and shipping samples, data assemblage and analysis, and several stages of report generation and presentation.

The cumulative distributions in Fig. 8 display the time course of NIST's coordination efforts by study type and sponsoring body. Unlike participations where the count increases by 1 unit for each study in which at least one result is contributed, the count of coordinations increases by only 1/*n* units where *n* is the number of co-coordinators sharing the resource burden.

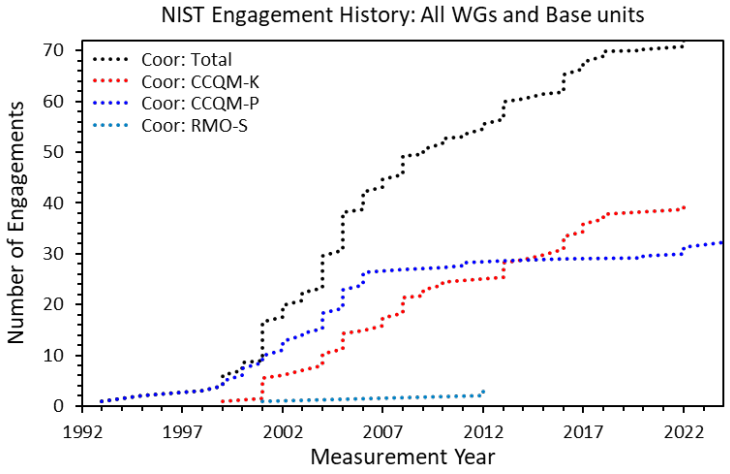


Fig. 8. Cumulative Distributions of NIST Coordinations of CCQM Studies.

The rate of KC coordinations has been relatively constant but the rate of PS coordinations sharply declined in mid 2000s. The decline in PS coordinations results from the increased use of "parallel pilot" studies and the way the *CCQM_Retrospectoscope* system counts participations and coordinations. Parallel studies that use the same samples and schedule as a KC only marginally increase the burden placed upon the study coordinator and so are

not counted with respect to coordinations. However, participants in KCs are free to also submit results to an associated parallel, typically to help evaluate alternate measurement techniques. Therefore parallel studies are considered separate studies with respect to participations.

The histogram traces of Fig. 9 display the time course of the coordinations relative to the total number of studies along with the approximate percentage of CCQM studies that NIST has coordinated.

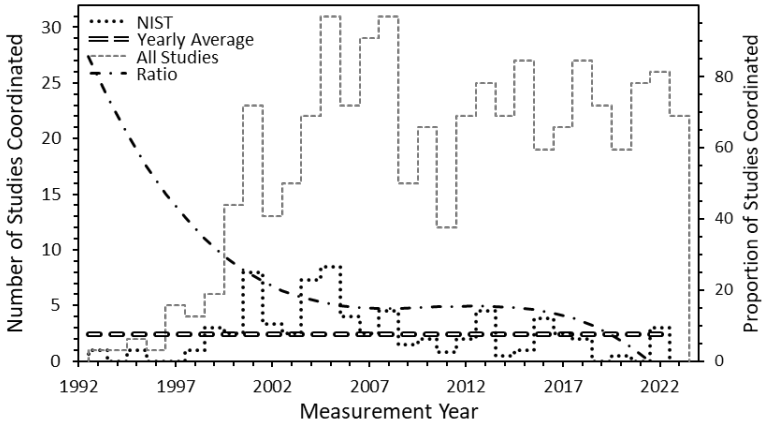


Fig. 9. Number of NIST and of Possible CCQM-Study Coordinations Per Year.

The proportion of studies trendline is a cubic polynomial fit to the ratio between the number of NIST coordinations and the number of CCQM studies during each yearly interval.

The decline in NIST's coordination rate is steeper than that of participation. Since coordination requires considerably more resources than participation, this may reflect a decrease in NIST's capacity as well as an increase in the number of other NMIs and DIs willing to coordinate or co-coordinate studies as they gain experience and demonstrate measurement competence. Some of these NMIs and DIs use the studies they coordinate as mechanisms for value assigning candidate certified reference materials (CRMs). NIST does not use this approach [27].

Note: Since KC-associated parallel pilot studies are considered as separate studies with regard to participation but are ignored with regard to coordinations, the "All Studies" histogram for participations (Fig. 7) describes more studies than it does for coordinations (Fig. 9).

Note: Where practical, the scattergrams used in this document use consistent X- and Y-axis scales to facilitate comparisons among the various WGs.

4.3. Leadership

In addition to participating and coordinating CCQM studies, NIST staff and management have engaged with the CCQM in a variety of ways – not least of which have been presentations, discussions, and informal conversations during (and after) meetings. While widely recognized as important, such transient engagements are difficult to properly acknowledge.

The following engagements are somewhat more readily documented.

4.3.1. International Comparisons Database

Beginning in early 1998, NIST developed the International Comparisons Database to provide electronic access to the results of CCQM comparison studies [28]. The project was developed by a team of NIST management and staff: Robert Watters, Jr., Dinis Camara, Jeanne Springmann, Geraldine Dalton, Marian McCurley, and Charles Sturrock. The BIPM joined the effort in 1999; it is the basis of the current KCDB.

4.3.2. Uncertainty of Measurement

Efficient characterization of sources of bias and variability in chemical measurements has been and continues to be challenging – and often a touch mysterious to chemists unfamiliar with statistical science. A major focus of the IDMS protocol NIST developed in 1993 for CIPM Study III [7] was development and use of an uncertainty budget [9§4.3]. Many of NIST's presentations at CCQM meetings have been designed to illustrate our approaches to uncertainty estimation.

Developed for quantitative measurements of all kinds, NIST's Barry N. Taylor largely wrote the Guide to Uncertainty in Measurement (GUM) [29] which was endorsed by the CIPM in 1992 [30§8]. Access to the GUM was originally fee-based; NIST's Technical Note 1297 [31], developed by Taylor and Chris E. Kuyatt, largely parallels the GUM and was much more accessible until the GUM was made freely available on a BIPM website in 2008. NIST has continued to develop accessible and freely available documents [32,33,34] and web-based tools [35,36,37] relevant to estimating the uncertainty in chemical measurement results. Challenges posed by CCQM participations and evaluations of CCQM comparison datasets have motivated much of the development of these documents and tools.

4.3.3. Workshops, Guest Researchers, Conferences, and Symposia

In addition to contact at meetings of the CCQM and CCQM WGs, NIST has strongly encouraged the development of chemical metrology in other NMI and DIs through presentations at conferences and symposia that emphasize its impact on national economies and international trade.

NIST has more directly influenced other NMIs and DIs through international workshops (e.g., [38,39]), site visits, and Guest Researcher training. Many of these typically year-long secondments are initiated through CCQM contacts.

4.3.4. Individuals

The following individuals have had direct formal influence on the CCQM's organization and which, how, and when studies are conducted.

- Robert L. Watters, Jr. In addition to his work with the CIPM during the CCQM's formulative stages (see Sections 1.2 and 1.3), he was NIST's representative to the first meeting of the CCQM in 1995, Chair of the IAWG from its creation in 1997 through 1998, and managed the development of the International Comparisons Database.
- Hratch G. Semerjian. As the new Director of NIST's Chemical Science and Technology Laboratory (CSTL) and later as NIST Deputy Director, he represented NIST to the CCQM from 1996 to 2003. He helped create and Chaired the Key Comparison Working Group (KCWG) from 1998 through 2003 (the KCWG is responsible for evaluating and keeping track of proposed CCQM studies.) As NIST Acting Director and later Chief Scientist, he represented NIST to the CIPM from 2003 to 2007.
- Willie E. May. Chief of NIST's Analytical Chemistry Division in CSTL, later CSTL Director, then Associate Director for Laboratory Programs prior to becoming NIST Director in 2014. He was Chair of the OAWG from 1998 through 2013, President of the CCQM from 2013 to 2019, and Vice-President of the CIPM from 2011 through 2018.
- Reenie M. Parris. An expert in chemical analysis and measurement quality assurance, she coordinated the CCQM's early studies of methods for assessing the purity of high-purity organic compounds. She led the formulation of protocols used by the KCWG for evaluating the rigor of the CMCs used by NMIs to document their metrological traceability claims.
- Gary L. Gilliland. Chief of NIST's Biotechnology Division in CSTL, he was Co-Chair of the newly created BAWG from 2001 through 2003.
- Laurie Locascio. Chief of NIST's Biotechnology Division in CSTL, in 2011 she was first Chair of the MBSG.
- Jayne B. Morrow. An expert in microbial detection and characterization, she Chaired the MBSG from 2013 to its disbanding in 2017 and then Chaired the CAWG from 2017 through 2019.
- Katrice A. Lippa. Group Leader of Chemical Sciences Division's (CSD's) Organic Chemical Metrology Group, Deputy Chair of the OAWG from 2019 to 2022.
- Michael R. Winchester, Group Leader of CSD's Inorganic Chemical Metrology Group, Chair of the IAWG from 2019 to the present.

5. Inorganic Analysis Working Group (IAWG)

The Inorganic Analysis Working Group (IAWG) has responsibility for “measurements of the elements; cations and anions; inorganic compounds; and organo-metallic compounds” [40]. The IAWG is one of four WGs established by the CCQM during 1997 to 1998 [1].

The interlaboratory comparison (“Study I”) was “organized during the period 1990 [to] 1993 under the aegis of the CIPM” [1] on the measurement of six metals in acidic water using a reference method believed to be capable of providing accurate measurement results. Many of the results from Study I were not within the desired target intervals around the gravimetric preparative values. A highly prescriptive follow-on comparison (“Study III”) demonstrated that the method was capable of providing results that are metrologically traceable to a higher-order reference system. These studies were discussed at the first meeting of the CCQM in 1995 [41]. These studies, identified as CCQM-P0 and CCQM-P1 in the master copy of the *CCQM_Retrospectoscope* and attributed to the IAWG, were coordinated by NIST.

5.1. Engagements in IAWG Studies

NIST participated in 82 IAWG-conducted Key Comparisons (KCs), published pilot studies (PPSs), or confidential pilot studies (PSs) that were finalized by this report’s publication date. NIST has results in 234 datasets generated by these studies. NIST has not participated in any IAWG-related Supplementary Comparisons (SCs).

Table 6 lists in chronological order the finalized IAWG studies in which NIST participated.

Table 6. NIST Engagement in IAWG-Attributed Studies.

Measurements in all studies reported in (or converted to) units of mass fraction, g/g.

Study	Description	Sets [a]	Year	Coordinator(s)
CCQM-P001	Trace elements in water	1	1994	NIST
CCQM-K002	Cadmium and lead in natural water	2	1998	JRC
CCQM-P007	KCl, NaCl, and $K_2Cr_2O_7$ purity	3	1999	NIST
CCQM-K008	Monoelemental calibration solutions of Al, Cu, Fe and Mg	4	2000	EMPA,LNE
CCQM-K013	Amount content of cadmium and lead in sediment	2	2000	JRC
CCQM-P012	Lead in wine	1	2000	JRC
CCQM-P019	Hydrochloric acid	2	2000	NIST
CCQM-P030	Monoelemental calibration solutions of Al, Cu, Fe and Mg	4	2000	EMPA,LNE
CCQM-K024	Cadmium in rice	1	2001	JRC,NMIJ
CCQM-P011	Arsenic in shellfish	1	2001	NIST
CCQM-P014	Trace elements in serum	1	2001	JRC,RISE
CCQM-P026	Sulfur in fuels	2	2001	JRC,NIST
CCQM-Q029	Cadmium and zinc in rice	2	2001	JRC,NMIJ
CCQM-Q032	Anion calibration solutions	2	2001	EMPA
CCQM-P019.1	HCl purity	2	2002	NIST
CCQM-P025	Minor elements in steel	4	2002	NMIJ,NIST,BAM
CCQM-P036	Potassium hydrogen phthalate (KHP) purity	2	2002	SMU,NIST
CCQM-Q013	Metals in synthetic food digest	2	2002	LGC
CCQM-K014	Calcium in human serum	1	2003	JRC
CCQM-K028	Tributyltin (TBT) in sediment	2	2003	LGC,NRC
CCQM-K031	Arsenic in shellfish	1	2003	NIST
CCQM-P033	Boron in silicon	1	2003	PTB

Study	Description	Sets [a]	Year	Coordinator(s)
CCQM-P034	Trace and minor elements in aluminum alloy	8	2003	BAM
CCQM-P046	Inorganic calibration solutions: Preparation	1	2003	NIST
CCQM-Q039	Toxic metals and methymercury in tuna	3	2003	JRC
CCQM-K033	Minor elements in steel	4	2004	NMIJ,NIST,BAM
CCQM-K034	Potassium hydrogen phthalate (KHP) purity	1	2004	SMU
CCQM-K035	Sulfur in diesel fuel	1	2004	NIST
CCQM-P026.1	Sulfur in diesel fuel	2	2004	NIST
CCQM-K042	Trace and minor elements in aluminum alloy	5	2005	BAM
CCQM-K043	Organo-mercury in salmon	4	2005	JRC
CCQM-P062	Nickel purity	6	2005	BAM
CCQM-P064	Trace elements in soyabean powder	4	2005	NIM
CCQM-P066	Metals in fertilizer	5	2005	NIST
CCQM-P074	Nitrogen and trace elements in silicon nitride powder	5	2006	NMIJ,BAM
CCQM-K043.1	Arsenic, mercury, selenium and methylmercury in marine fish	1	2007	NMIJ
CCQM-K049	Toxic and essential elements in bovine liver	7	2007	NIST
CCQM-K056	Trace elements in whole-fat soybean powder	4	2007	NIM
CCQM-K057	Chemical composition of clay	5	2007	CENAM
CCQM-K058	Nitrogen and trace elements in silicon nitride power	4	2007	NMIJ,BAM
CCQM-K059	Nitrite and nitrate in calibration solutions	1	2007	SMU,NRC
CCQM-P074.1	Nitrogen and trace elements in silicon nitride power	1	2007	NMIJ,BAM
CCQM-P096	As and arsenobetaine content in marine fish	0 [b]	2007	NMIJ,NIM
CCQM-P100.1	Mercury in pure water	1	2007	PTB,BAM,JRC,LNE
CCQM-P100.2	Mercury in natural water	1	2007	PTB,BAM,JRC,LNE
CCQM-Q086	Total selenium and selenomethionine in pharmaceuticals	3	2007	LGC,NRC
CCQM-K048	Potassium chloride (KCl) purity	1	2008	NIM
CCQM-P097	Cadmium and lead in herb *Desmodium styracifolium*	2	2008	GLHK
CCQM-P107	Zinc purity	7	2008	BAM
CCQM-K073	HCl purity	1	2009	NIST,CENAM
CCQM-K075	Toxic metals in algae	2	2009	IAEA
CCQM-P106	Toxic metals in polypropylene	4	2009	NIM,NMIJ,KRISS
CCQM-P118	Toxic metals in algae	5	2009	IAEA
CCQM-P119	Lead in solder	1	2009	NMIJ,NIM,KRISS
CCQM-K070	Mercury in natural water: low levels	1	2010	PTB,BAM,LNE
CCQM-K088	Lead in lead-free solder containing silver and copper	1	2010	NMIJ,NIM,KRISS
CCQM-P096.1	As and arsenobetaine in calibration solution	0 [b]	2010	NMIJ,NIM
CCQM-K087	Monoelemental calibration solutions of Cr, Co and Pb	9	2011	PTB
CCQM-K089	Trace and essential elements in *Herba Ecliptae*	5	2011	GLHK
CCQM-P126	Trace and essential elements in *Herba Ecliptae*	4	2011	GLHK
CCQM-K096	Potassium dichromate ($K_2Cr_2O_7$) purity	1	2012	SMU,KRISS
CCQM-K097	Arsenobetaine in standard solution and in tuna tissue	2	2012	NMIJ,NIM
CCQM-K072	Zinc purity	7	2013	BAM
CCQM-K107	Elements and selenium speciation in human serum	5	2013	LGC
CCQM-K108	Total arsenic, arsenic species and cadmium in brown rice flour	2	2013	NMIJ
CCQM-K048.2014	Chloride (Cl^-) in potassium chloride (KCl)	1	2014	NIM
CCQM-K123	Trace elements in biodiesel fuel	1	2014	NMIJ,NIST
CCQM-P156	Element-based quantification and purity analysis	1	2014	KRISS
CCQM-P157	Trace elements in biodiesel fuel	1	2014	NMIJ,NIST
CCQM-P149	Zinc purity	11	2014	BAM
CCQM-Q149	Zinc purity (published subset of CCQM-P149 results)	3	2014	BAM
CCQM-K108.2014	Total arsenic and arsenic species in brown rice flour	3	2015	NMIJ
CCQM-K124	Trace elements and chromium speciation in drinking water	5	2015	NMIJ,GLHK
CCQM-K127	Toxic and trace elements in soils	7	2015	CENAM,IJS
CCQM-P158	Trace elements and chromium speciation in drinking water	2	2015	NMIJ,GLHK
CCQM-K125	Iodine and other elements in infant formula	3	2016	GLHK
CCQM-K034.2016	Amount of acid in a solid weak acid	1	2017	NIM
CCQM-K139	Elements in human serum	3	2017	HSA
CCQM-K143	Copper calibration solutions	1	2018	NIST

Study	Description	Sets [a]	Year	Coordinator(s)
CCQM-K145	Toxic and essential elements in bovine liver powder	2	2018	NIM
CCQM-P183	Toxic and essential elements in bovine liver powder	7	2018	NIM
CCQM-P191	Sulfur in synthetic insulin standard solution	1	2018	KRISS
CCQM-K034.2016.1	Amount of acid in a solid weak acid	1	2019	NIM

a The number of datasets derived from the study that contain a NIST result.
b Quantitative results have not been published.

The cumulative distributions in Fig. 10 display the time course of NIST's engagements with finalized IAWG studies, where "engagements" applies to participations and coordinations.

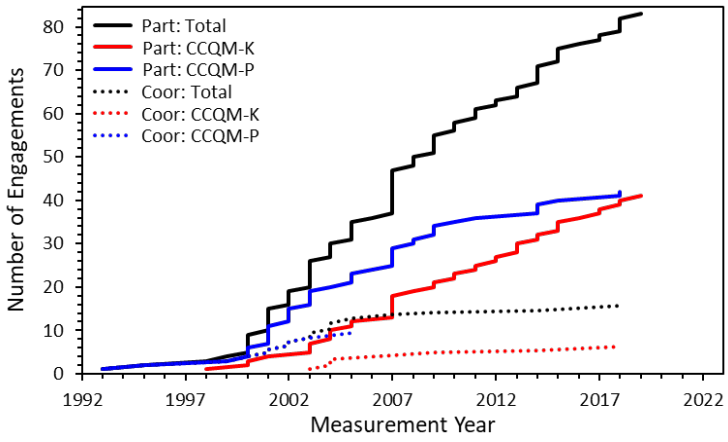


Fig. 10. IAWG Studies: Cumulative Distributions of NIST Engagements.

The histogram traces of Fig. 11 display the engagement time course in higher resolution, along with the average number of participations and coordinations from the first to the most recent engagement.

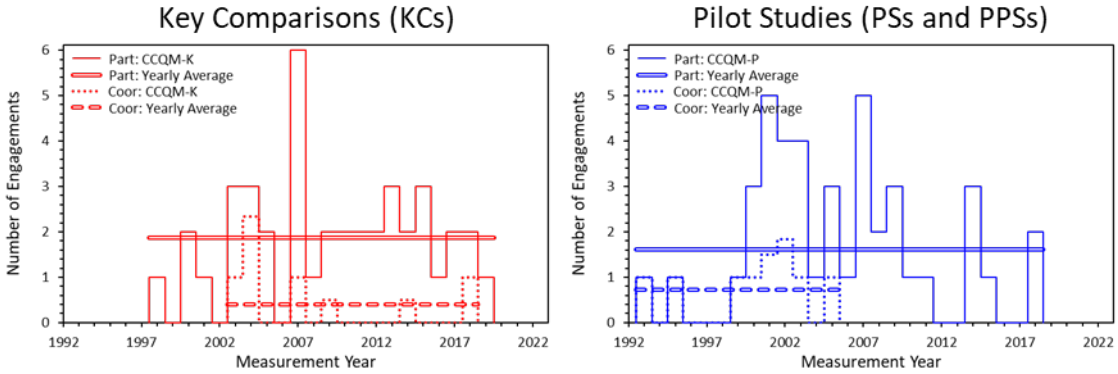


Fig. 11. IAWG Studies: Histogram Traces of NIST Engagements.

5.2. Performance in IAWG Key Comparisons

NIST participated in 41 IAWG-sponsored KCs that were finalized by this report's publication date. NIST has results in 115 datasets generated by these studies. The relationship between the absolute zeta-score bias of these results, $|\zeta_{NIST}|$, and the standard uncertainty of the reported values relative to the median of the participant uncertainties, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, is visualized in Fig. 12.

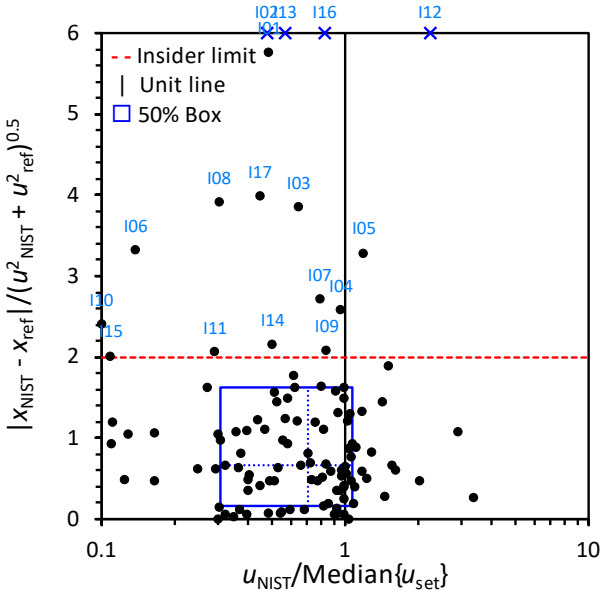


Fig. 12. IAWG Key Comparisons: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one Key Comparison (KC) dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$. The "×" denote values above the chart's Y-axis. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties. The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$).

The median absolute bias is 0.66. More than 85 % of the absolute biases are within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})/\text{Median}\{u(x_i)\}$ is 0.71. The $u(x_{NIST})$ is less than the participant median for 15 of the 17 "outsiders" ($|\zeta_{NIST}| > 2$). This implies that in most cases, lack of equivalence with the RV is a result of underestimating the uncertainty on the reported measurement. The "outsider" datasets are documented in Section 5.2.4.

5.2.1. Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for IAWG KC datasets are displayed in Fig. 13 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.0218|/0.0026 \approx 8$, suggest that the absolute bias increased over time.

This increase may be due to the $U_{95}(RV)$ decreasing over time, perhaps reflecting improvements in participant capabilities or changes in how the $\{RV, U_{95}(RV)\}$ are evaluated.

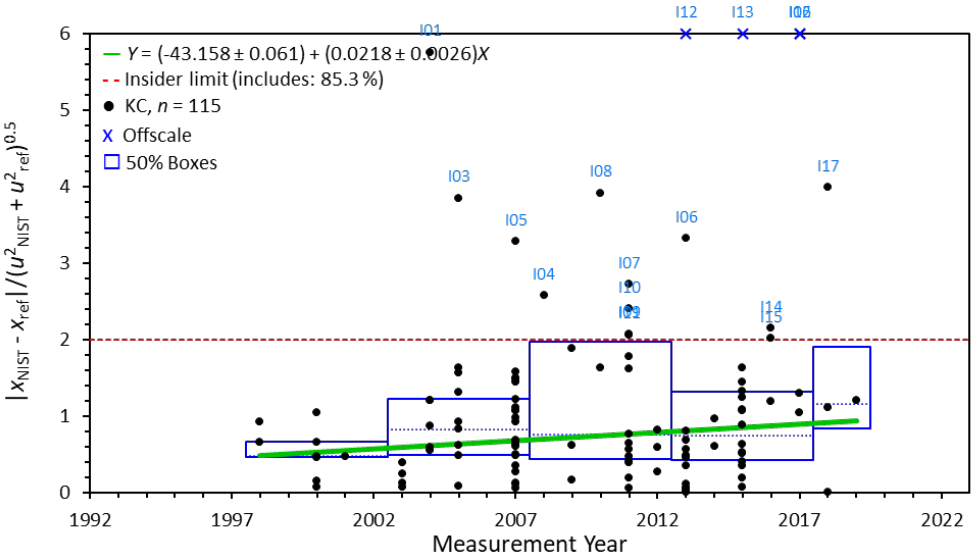


Fig. 13. IAWG Key Comparisons: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one Key Comparison (KC) dataset as a function of measurement year. The "×" denote values larger than the chart's Y-axis maximum. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value. The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit", and the number of results.

The $u(x_{NIST})$/Median$\{u(x_i)\}$ values are displayed in Fig. 14 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.00617|/0.00090 \approx 7$, suggest that the uncertainty ratios decreased over time.

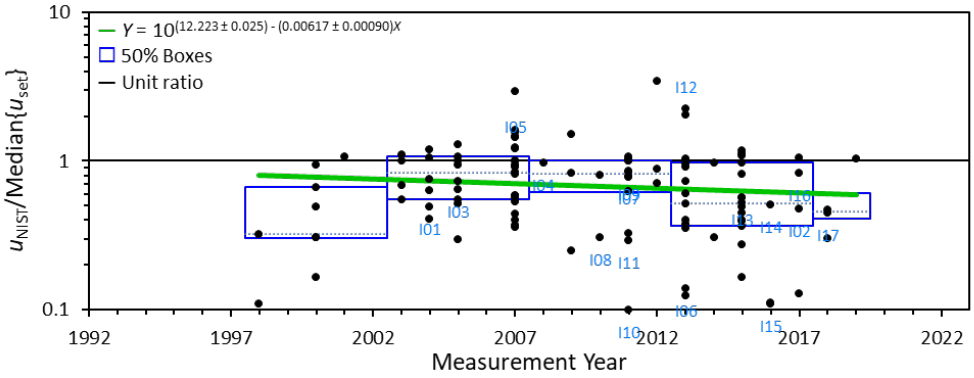


Fig. 14. IAWG Key Comparisons: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})$/Median$\{u(x_i)\}$, in one Key Comparison (KC) dataset as a function of measurement year. The labels identify datasets which contain "outsider" results ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of the values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation.

5.2.2. Bias and Relative Uncertainty Across Measurand Level

The $|\zeta_{NIST}|$ values for IAWG KC datasets are displayed in Fig. 15 as a function of measurand level. While the linear trend, $|t| = |+0.0318|/0.0069 \approx 5$, suggests that the absolute bias increases with measurand level, the pattern of the "50% boxes" suggests that there is a "sweet spot" from (10^{-7} to 10^{-3}) g/g with slight increases on either side. The distribution of "outsider" results appears to be fairly uniform across level.

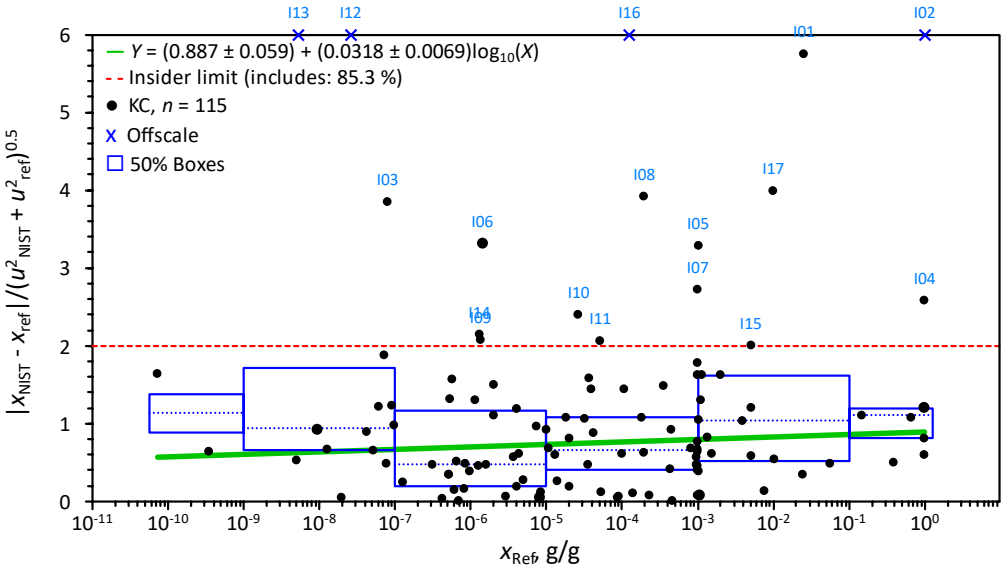


Fig. 15. IAWG Key Comparisons: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one Key Comparison (KC) dataset as a function of mass fraction. See Fig. 13 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 16 as a function of measurand level. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.0049|/0.0024 \approx 2$, suggest that the uncertainty ratios weakly increase with increasing mass fraction.

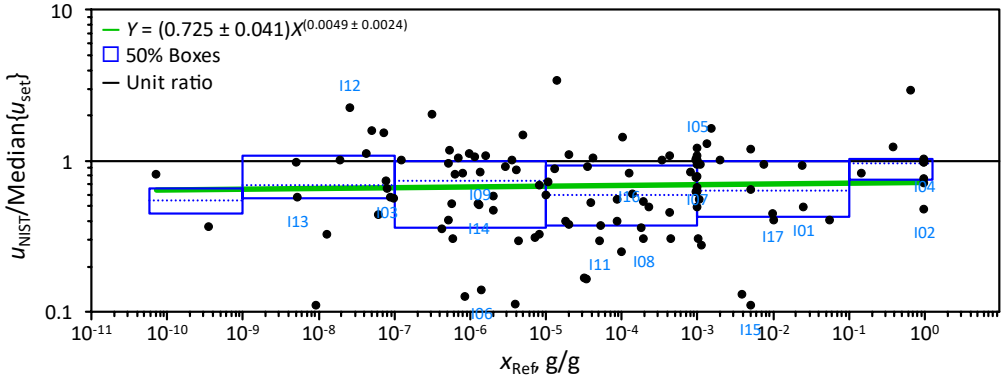


Fig. 16. IAWG Key Comparisons: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, in one Key Comparison (KC) dataset as a function of mass fraction. See Fig. 14 for description of the graphical elements.

5.2.3. Measurand Level Over Time

The KC reference values, x_{ref}, are displayed in Fig. 17 as a function of measurement year. The linear trend, $|t| = |+0.0465|/0.0077 \approx 6$, suggests that the mass fraction level of the measurands has decreased over time. However, the pattern of the "50% boxes" suggests that the level changes are episodic, likely reflecting differences in the nature of the measurands studied.

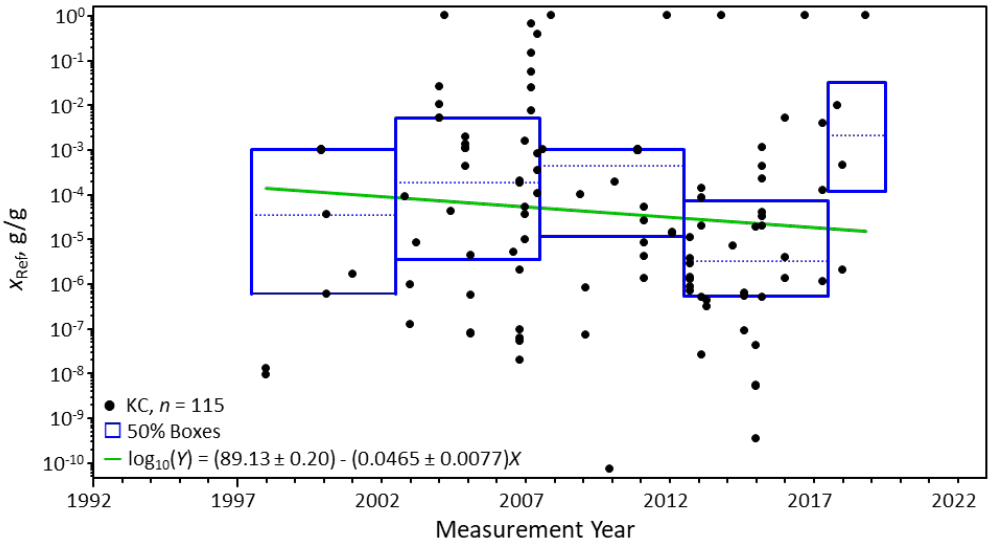
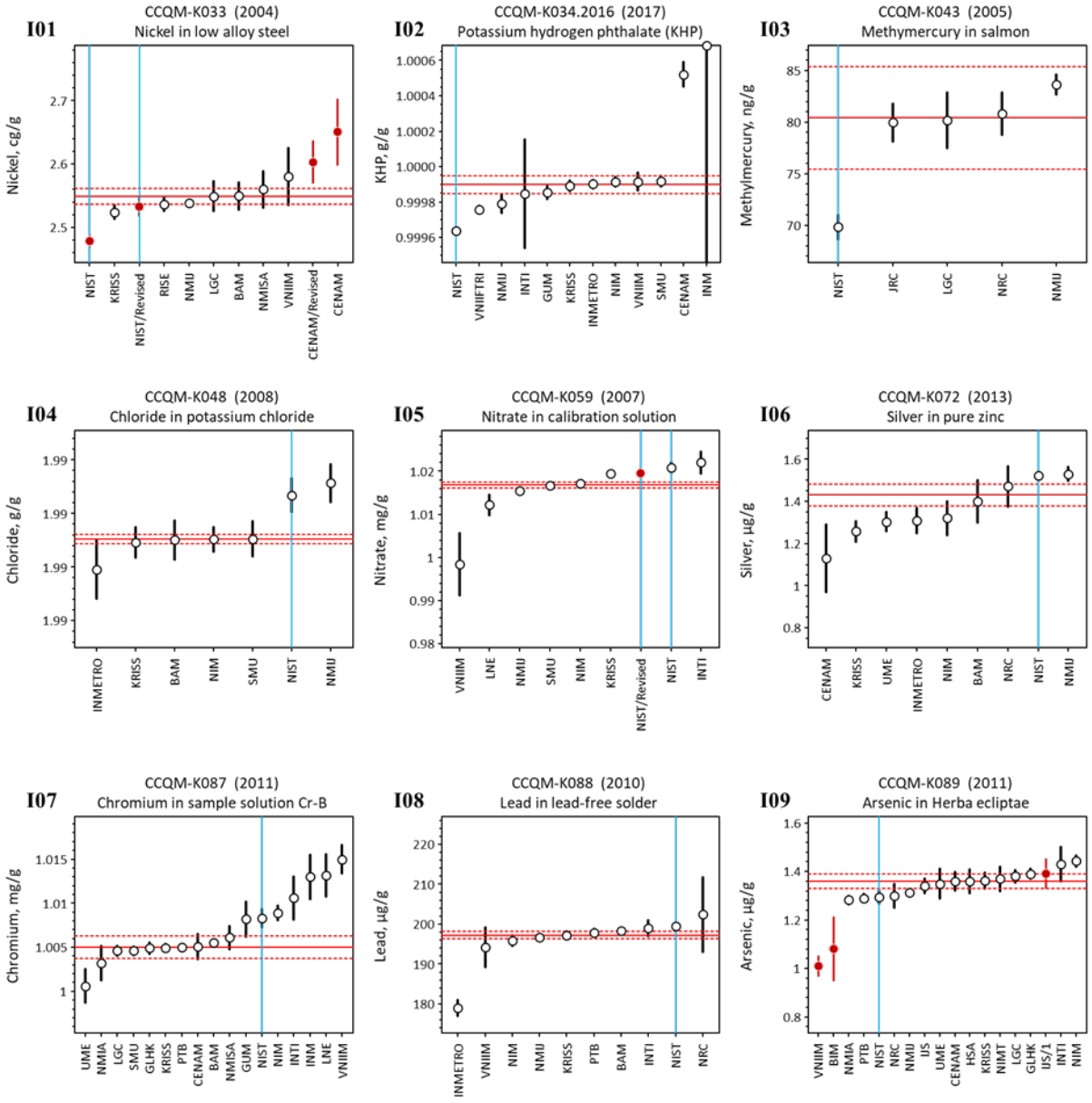


Fig. 17. IAWG Key Comparisons: Measurand Level Over Time.

Each symbol represents the reference value, x_{ref}, in one Key Comparison (KC) dataset in which NIST contributed a value. Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation and the number of reference values summarized.

5.2.4. ”Outsider” Datasets

The 17 KC datasets that contain “outsider” NIST results are depicted in Fig. 18 as dot-and-bar charts, each panel showing all results for one dataset, the reference value, and the 95 % level of confidence reference interval. The denotation of “outsider” here is not mutually exclusive with the result being considered equivalent (see Section 3.2.1.2). The label in the upper left corner of each panel, running in sequence from “I01” to “I17”, is the code used to identify the “outsiders” in Fig. 12 to Fig. 16.



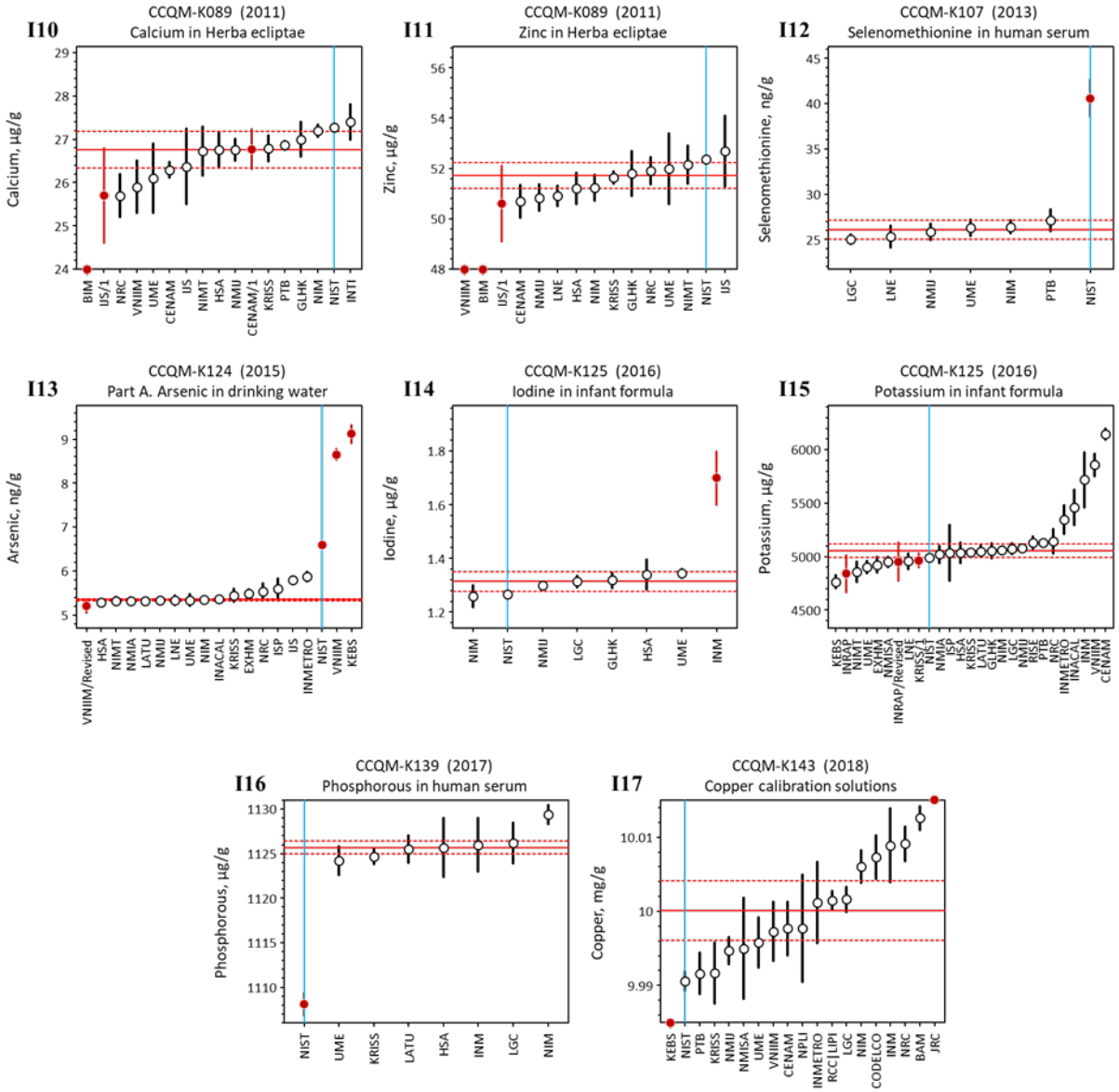


Fig. 18. IAWG Key Comparisons: "Outsider" Datasets.

Each panel shows all reported results for one dataset with a NIST "outsider" ($|\zeta_{NIST}| > 2$) result. Results are sorted in order of increasing magnitude. Participants are identified by code name; the full names are listed in Appendix B. Open circles denote results considered valid; solid red circles denote results excluded from estimating the reference value. Bars represent one standard uncertainty. The horizontal solid red line denotes the reference value; the dashed red lines bound a 95 % level of confidence interval about the reference value. The blue vertical lines mark NIST results.

5.3. Performance in IAWG Pilot Studies

NIST participated in 41 IAWG-sponsored pilot studies (6 PPSs and 35 PSs) that were finalized by this report's publication date. NIST has results in 118 datasets (26 PPSs and 95 PSs) generated by these studies. Most of the participations were in stand-alone studies used for training and method validation during the IAWG's first decade. The relationship between the $|\zeta_{NIST}|$ and $u(x_{NIST})/\text{Median}\{u(x_i)\}$ is visualized in Fig. 19.

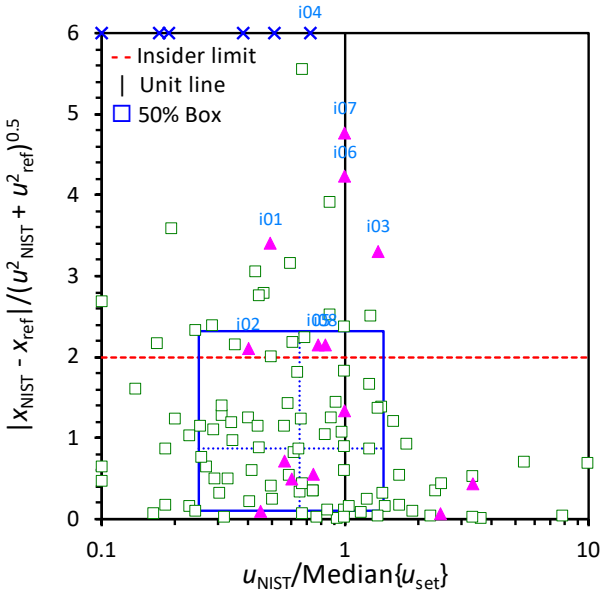


Fig. 19. IAWG Pilot Studies: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$: solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot studies (PS) results. The "×" denote values above the chart's Y-axis. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties. The labels identify "outsider" PPS datasets ($|\zeta_{NIST}| > 2$).

The median absolute bias for IAWG pilot study datasets is 0.87, slightly larger than for the KC datasets. More than 76 % of the absolute bias results are within $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})/\text{Median}\{u(x_i)\}$ is 0.64, slightly smaller than for the KC datasets. The $u(x_{NIST})$ are less than the participant medians in 24 of the 29 "outsiders" implying that an underestimation of measurement uncertainty is a contributing factor to lack of equivalence. The eight "outsider" PPS datasets are documented in Section 5.3.4.

5.3.1. Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for IAWG PPS and PS datasets are displayed in Fig. 20 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.0001|/0.0027 \approx 0$, suggest that on average the absolute bias did not change over time.

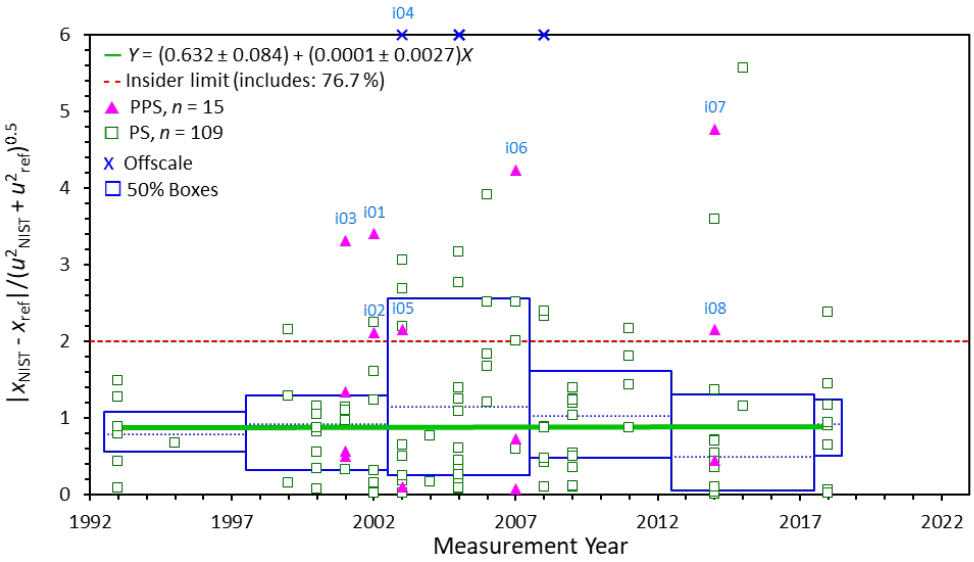


Fig. 20. IAWG Pilot Studies: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of measurement year: solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot studies (PS) results. The "×" denote values above the chart's Y-axis. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The labels identify "outsider" PPS datasets ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the trend. The legend provides the trend equation, the percentage of values within the "insider limit", and the number of PPS and PS results.

The $u(x_{NIST})$/Median$\{u(x_i)\}$ values are displayed in Fig. 21 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.0291|/0.0010 \approx 29$, suggest that the uncertainty ratios increased over time. This may be related to the decrease in the mass fraction levels of the measurands used in pilot studies: see Section 5.3.3. The "low hanging fruit" addressed in the early studies are gone.

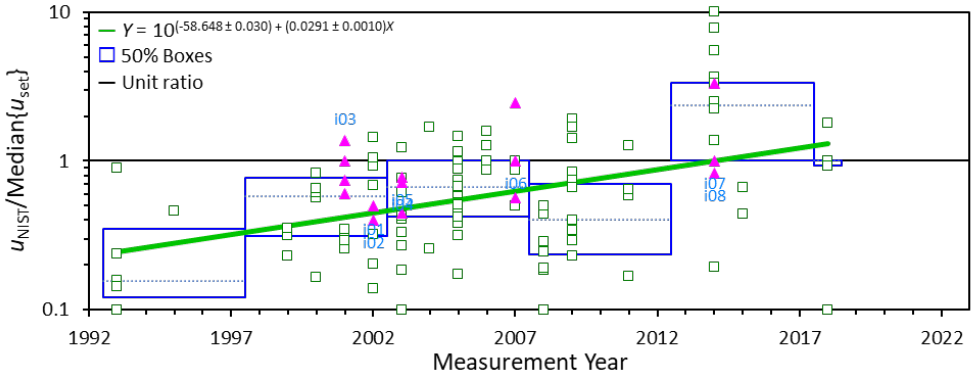


Fig. 21. IAWG Pilot Studies: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})$/Median$\{u(x_i)\}$, in one dataset as a function of measurement year: solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The labels identify "outsider" results ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of the values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation.

5.3.2. Bias and Relative Uncertainty Across Measurand Level

The $|\zeta_{NIST}|$ values for IAWG PPS and PS datasets are displayed in Fig. 22 as a function of measurand level. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.0452/0.0099| \approx 5$, suggest that the absolute bias decreases with increasing level. This differs from the results for KC datasets.

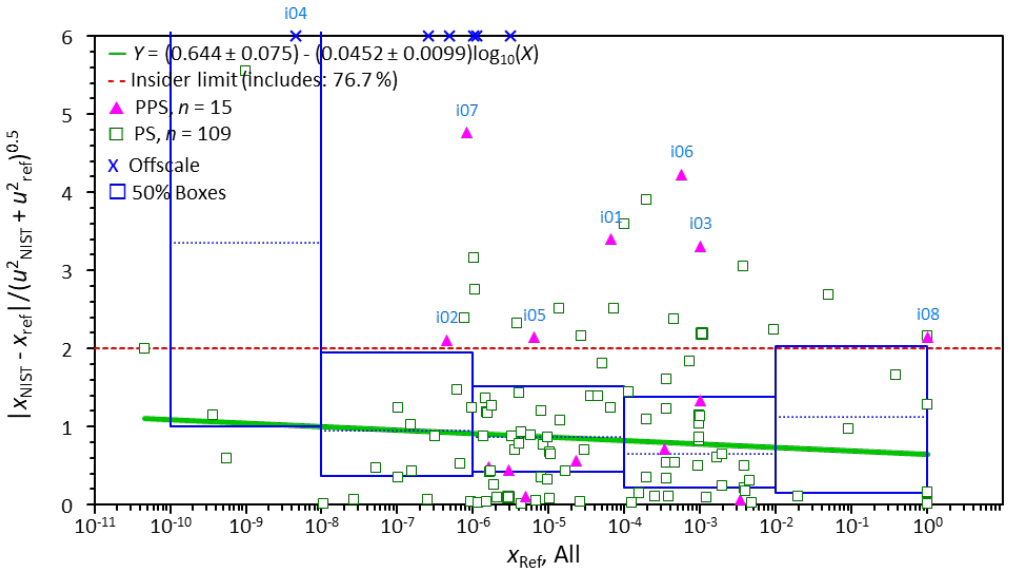


Fig. 22. IAWG Pilot Studies: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of mass fraction. See Fig. 20 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 23 as a function of measurand level. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.0154/0.036| \approx 3$, suggests that the uncertainty ratios for pilot studies weakly increase with increasing mass fraction. This is similar to the results for KC datasets.

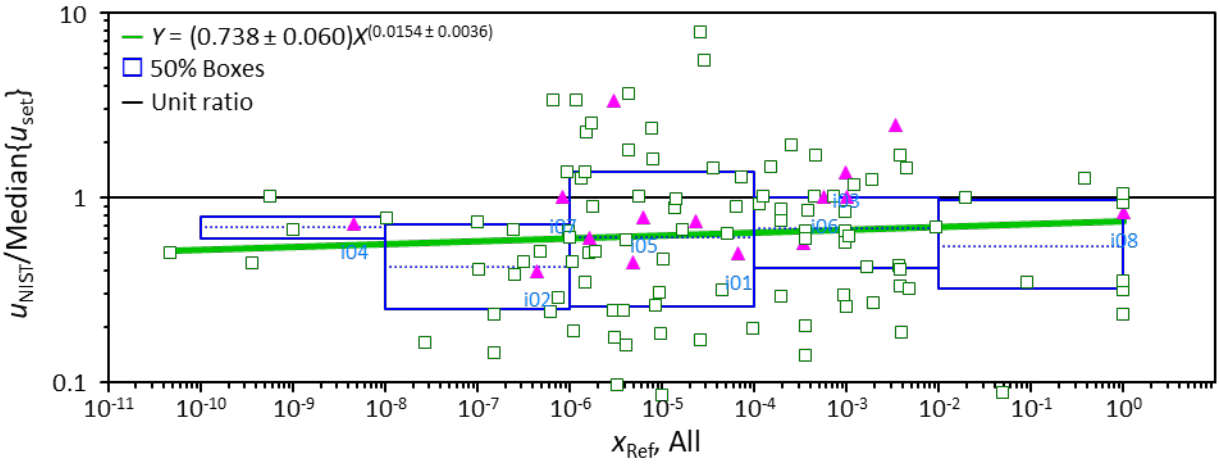


Fig. 23. IAWG Pilot Studies: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, in one dataset as a function of mass fraction. See Fig. 21 for description of the graphical elements.

5.3.3. Measurand Level Over Time

The pilot study reference values, x_{ref}, are displayed in Fig. 24 as a function of measurement year. The linear trend, $|t| = |+0.1065|/0.0053 \approx 20$, suggests that the mass fraction level of the measurands in these pilot studies has decreased over time. The pattern of the "50% boxes" suggests that the decrease has plateaued at about the µg/g level.

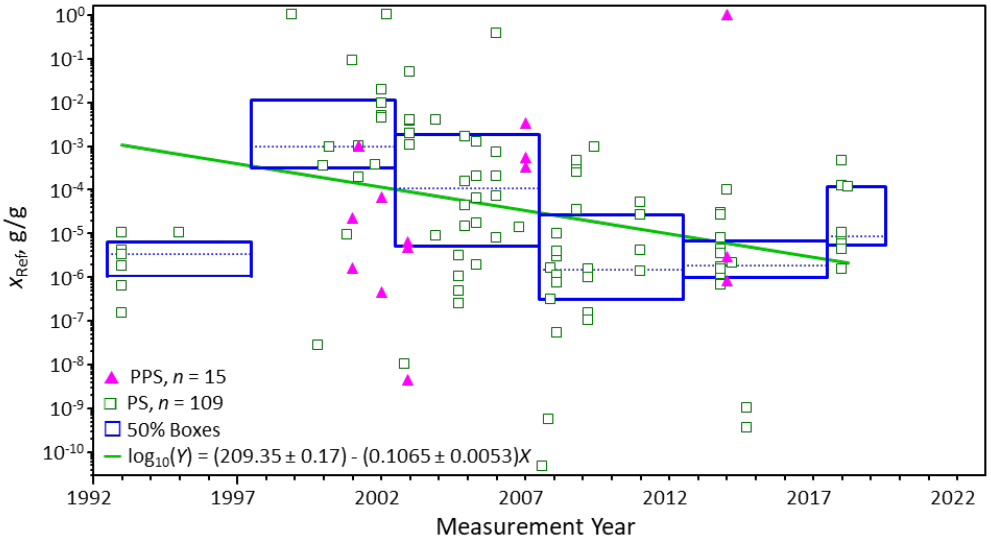


Fig. 24. IAWG Pilot Studies: Measurand Level Over Time.

Each symbol represents the reference value, x_{ref}, in one pilot study (published and confidential) dataset in which NIST contributed a value. See Fig. 18 for description of the graphical elements.

5.3.4. "Outsider" Datasets in Published Pilot Studies,

Results for PS datasets are confidential and cannot be published. The eight PPS datasets that contain "outsider" NIST results are depicted in Fig. 25 as dot-and-bar charts; each panel showing all results for one dataset, the reference value, and the 95 % level of confidence reference interval. The label in the upper left corner of each panel, running in sequence from "i01" to "i08," is the code used to identify the "outsiders" in Fig. 19 to Fig. 23.

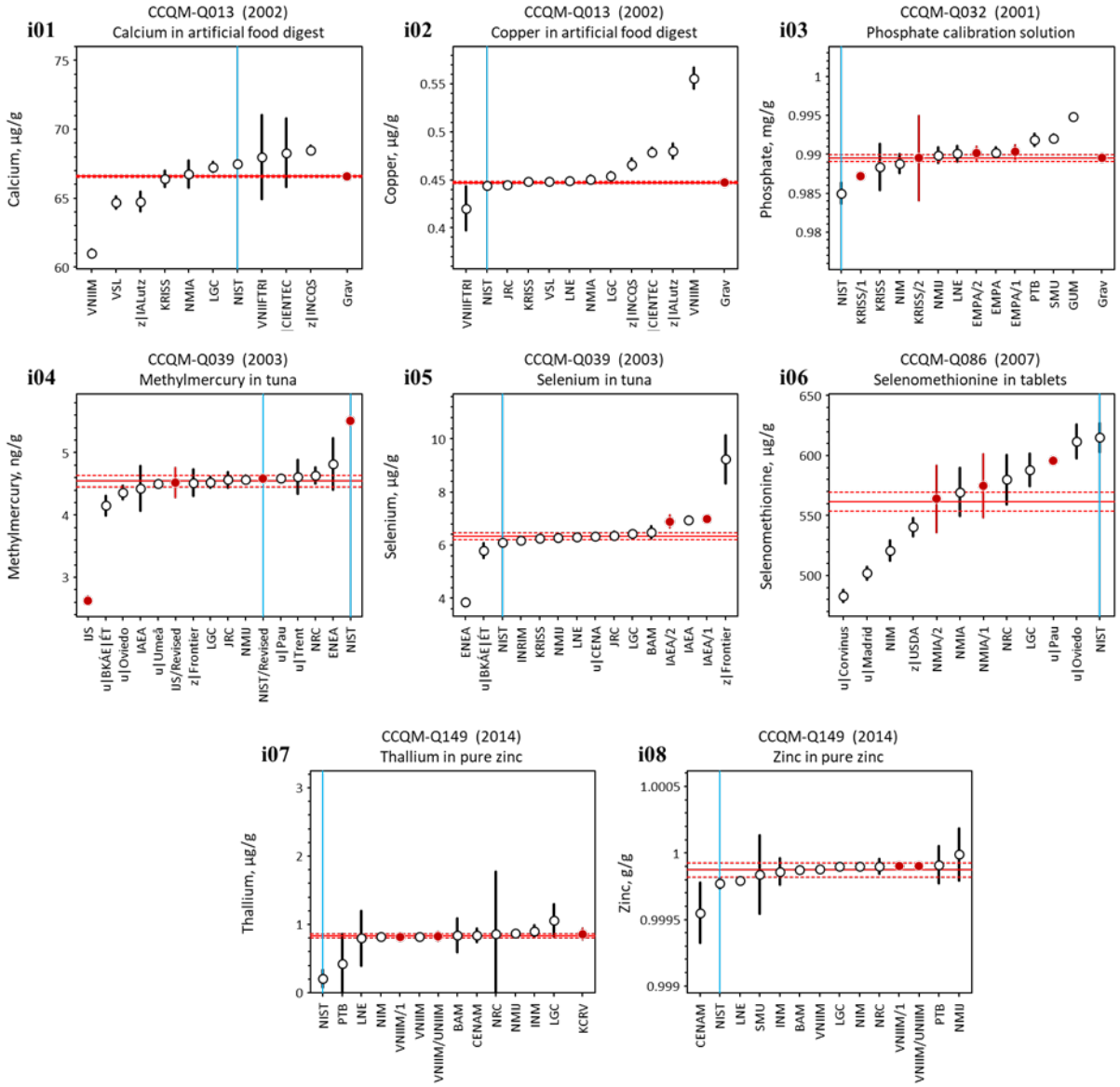


Fig. 25. IAWG Published Pilot Studies: "Outsider" Datasets.

Each panel shows all reported results for one dataset with a NIST "outsider" ($|\zeta_{NIST}| > 2$) result. Results are sorted in order of increasing magnitude. Participants are identified by code name; the full names are listed in Appendix B. Open circles denote results considered valid; solid red circles denote results excluded from estimating the reference value. Bars represent one standard uncertainty. The horizontal solid red line denotes the reference value; the dashed red lines bound a 95 % level of confidence interval about the reference value. The blue vertical lines mark NIST results.

5.4. IAWG Lessons Learned

The history of performance can be useful when looking at the metrological infrastructure and culture that may exist within an NMI, persisting beyond individual staffing changes or updates in measurement instrumentation. It is easy to distill a few general trends from NIST IAWG performance over the past few decades, and the ability to look at past performance can certainly help make CMC claims where a "Record of Performance" [42] may be required such as when broad-based CMC claims are made.

One overall trend that is apparent from looking at the assembled datasets is NIST's tendency to underestimate measurement uncertainty, particularly in KCs. This is shown in Fig. 12 where most outsider results occur when the uncertainty on a key comparison result is smaller than the median comparison. This is also true for pilot studies as shown in Fig. 19.

A microcosm of the general issue of underestimation of measurement uncertainty can be observed in CCQM K123 *Trace Elements in Biodiesel Fuel* [43] results for sulfur. NIST reported results for sulfur using IDMS, which is considered to be a "definitive" technique. Two other institutes also reported IDMS results as well as one result using X-ray fluorescence (XRF) and two results using inductively coupled plasma optical emission spectroscopy (ICP-OES). Curiously, the results from the three "definitive" IDMS method processes did not have overlapping uncertainties, causing much discussion and eventual revision of the results and uncertainties [43, Tables 9 and 10]. In the end, the underestimation of the uncertainty, as well as other issues, resulted in no official RV being generated for sulfur.

Over the years, the IAWG has dealt with this issue in a variety of ways, but one critical component has been in choosing how to calculate the RV for the comparison. In general, the stance of the IAWG has been that the statistical treatment to arrive at an RV should not distract from or obfuscate the measurement challenges present in a comparison. The intent was not to be bogged down in a purely statistical discussion, but it resulted in many datasets with inconsistencies where there is not good overlap between most of the participants, readily apparent in many of the "outsider" datasets in Fig. 18. This inconsistency is a result of participants like NIST underestimating measurement uncertainty in submitted results as well as simplistic statistical treatments of these varied datasets to arrive at an RV and assign its uncertainty. In some cases, the RV is set using either the mean or the median, chosen solely based on the number of participants.

More recently the IAWG has adopted the use of the NIST Decision Tree [36], which applies the appropriate statistical treatment based on the "shape" of the data. The Decision Tree also incorporates "dark" uncertainty into the calculations, which helps account for underestimation of measurement uncertainty, even if that is not attributed as the source. It has been pointed out at IAWG meetings during discussions of this software package that the result will likely be RVs that have larger uncertainties and that as a result the WG will tend to err on the side of showing participants are equivalent when they might not be. Consequently, the results coming out of IAWG comparisons going forward may be more difficult to interpret with regards to NIST's historical performance.

Participation in CCQM exercises ideally improves performance in measurement science and in the delivery of measurement services, such as reference materials. When looking at the historical data not just for the IAWG, but for all WGs, it should be remembered that performance is a bit of a moving target. While pilot studies might move towards more difficult measurands, KCs would more often cover repeated measurements of “core” capabilities.

It is hoped then that the measurement capabilities of the entire WG would improve over time, leading to a tighter uncertainty on any RV from a comparison. As a result, bias versus the RV would be more noticeable (Fig. 13) even while the apparent uncertainty decreased over time (Fig. 14). One question that could be asked is whether the trends observed can be attributed to other participants improving their measurement capabilities at a faster rate than NIST in the inorganic analysis arena. This would not be surprising as improving measurement capabilities tends to give diminishing returns as one eliminates, minimizes, or accounts for the dominant source of imprecision or bias and moves to the next most significant source.

Other points to keep in mind when looking at the historical performance of NIST in IAWG comparisons are the number of analysts and the variety of measurement techniques used. NIST has historically had more analysts and uses a greater variety of methods than do many other NMIs and DIs. As a result, the datasets looking at performance across different mass fraction levels (Fig. 15, Fig. 16, Fig. 22 and Fig. 23) can be difficult to interpret.

Many NMIs and DIs may have few measurements methods available or with which they wish to link CMC claims. As a result, those methods are made to fit the measurement problem presented by the IAWG. NIST has tended to use ICPMS primarily, but has also used ICP-OES, XRF and nuclear methods. Each of these methods have their own “sweet spot” with regards to mass fraction level, suggesting that NIST’s performance of the mass fraction range is a patchwork of different measurement procedures and analyst skill sets while other NMIs may have performance governed by a smaller group of analysts and techniques that hinge on the ability to properly dilute a test sample gravimetrically.

Of particular interest would be CCQM-K57: *Chemical Composition of Clay* [44] which shows a variety of measurement techniques being used including both XRF and nuclear methods by NIST. In this comparison the official values submitted by NIST and other participants were a combination of measurement methods, more accurately capturing how many NMIs and DIs value-assign their reference materials.

6. Organic Analysis Working Group (OAWG)

The Organic Analysis Working Group (OAWG) has responsibility for “well-defined organic molecular entities” excluding “gaseous compounds, organometallic compounds, and large bio-molecules” [45]. The OAWG is one of four WGs established by the CCQM during 1997 to 1998 [1]. Several pilot studies initiated by the CCQM and conducted before the formal WG assignments were formalized are assigned to the OAWG.

6.1. Engagements in OAWG Studies

NIST participated in 66 OAWG-conducted Key Comparisons (KCs), published pilot studies (PPSs), or confidential pilot studies (PSs) that were finalized by this report’s publication date. NIST has results in 193 datasets generated by these studies. NIST has not participated in any OAWG-related supplementary Comparisons (SCs).

Table 7 lists in chronological order the finalized OAWG studies in which NIST participated.

Table 7. NIST Engagement in OAWG-Attributed Studies.

Study	Description	Units	Sets [a]	Year	Coordinator(s)
CCQM-Q002	Dichlorodiphenyldichloroethylene (p,p'-DDE) in isooctane	g/g	2	1997	LGC
CCQM-P003	qNMR for analysis of mixtures	mol/mol	5	1998	BAM
CCQM-P004	Dichlorodiphenyldichloroethylene (p,p'-DDE) in corn oil	g/g	2	1998	LGC
CCQM-Q006	Cholesterol in human serum	g/g	2	1998	NIST
CCQM-P005	Acetanilide, benzoic acid and naphthalene purity	g/g	6	1999	NIST
CCQM-K005	Dichlorodiphenyldichloroethylene (p,p'-DDE) in fish oil	g/g	2	2000	LGC
CCQM-P010	Hexachlorocyclohexane (γ-HCH) in fish oil	g/g	2	2000	LGC
CCQM-P021	Dichlorodiphenyltrichloroethane (p,p'-DDT) in Fish Oil	g/g	2	2000	LGC
CCQM-Q009	Creatinine in human serum	g/g	2	2000	NIST
CCQM-Q017	Polychlorinated biphenyls (PCBs) in sediments	g/g	4	2000	NRC,NIST
CCQM-K006	Cholesterol in human serum	g/g	2	2001	NIST
CCQM-K011	Glucose in human serum	g/g	1	2001	NIST
CCQM-K012	Creatinine in human serum	g/g	2	2001	NIST
CCQM-K021	Dichlorodiphenyltrichloroethane (p,p'-DDT) in fish oil	g/g	2	2001	LGC
CCQM-P010.2	Hexachlorocyclohexane (γ-HCH) in fish oil	g/g	2	2001	LGC
CCQM-P027	Lysergic acid diethylamide (LSD) in urine	g/g	1	2001	LGC
CCQM-P035	Ethanol in aqueous matrix	g/g	2	2001	BAM,LGC
CCQM-Q008	Glucose in human serum	g/g	2	2001	NIST
CCQM-K025	Polychlorinated biphenyls (PCBs) in sediments	g/g	5	2002	NIST,NRC
CCQM-K027.a	Ethanol in aqueous matrix	g/g	2	2002	LGC,BAM
CCQM-K027.b	Ethanol in aqueous matrix	g/g	1	2002	LGC,BAM
CCQM-P020.b	o-Xylene purity	g/g	1	2002	NIST
CCQM-P020.c	Atrazine purity	g/g	2	2004	NMIA
CCQM-P020.d	Chlorpyrifos purity	g/g	1	2004	NMIA
CCQM-P031.a	Polycyclic aromatic hydrocarbon (PAH) calibration solutions	g/g	5	2004	NIST
CCQM-P031.b	Polychlorinated biphenyl (PCB) calibration solutions	g/g	5	2004	NIST
CCQM-P031.c	Organochlorine pesticide calibration solutions	g/g	4	2004	NIST
CCQM-P040	Organic contaminants in mussel tissue	g/g	13	2004	NIST
CCQM-K038	Polycyclic aromatic hydrocarbons (PAHs) in solution	g/g	5	2005	NIST
CCQM-K039	Organochlorine pesticides in solution	g/g	4	2005	NIST
CCQM-K040	Polychlorinated biphenyl (PCB) congeners in solution	g/g	4	2005	NIST
CCQM-P057	Polychlorinated biphenyl (PCB) congeners in tissue extract	g/g	10	2005	NIST
CCQM-P061	Volatile organic compounds (VOCs) in solution	g/g	6	2005	CENAM,NIST
CCQM-P067	Polychlorinated biphenyl (PCB) congeners in tissue	g/g	5	2005	NIST
CCQM-P068	Anabolic steroids in urine	g/g	1	2005	NMIA

Study	Description	Units	Sets [a]	Year	Coordinator(s)
CCQM-K047	Volatile organic compounds (VOCs) in solution	g/g	4	2006	CENAM,NIST
CCQM-P077.a	Progesterone in human serum	g/g	2	2006	NIST
CCQM-P077.b	Cortisol in human serum	g/g	2	2006	NIST
CCQM-P078	Nutrients in infant/adult formula	g/g	3	2006	NIST
CCQM-K050	Polycyclic aromatic hydrocarbons (PAHs) in soil and particulates	g/g	10	2007	CENAM,BAM
CCQM-P069	Polycyclic aromatic hydrocarbons (PAHs) in soil and particulates	g/g	5	2007	CENAM,BAM
CCQM-Q020.e	Theophylline purity	g/g	2	2007	BIPM,LGC
CCQM-K062	Nutrients in infant/adult formula	g/g	3	2008	NIST
CCQM-K063.a	Cortisol in human serum	g/g	1	2008	NIST
CCQM-K063.b	Progesterone in human serum	g/g	1	2008	NIST
CCQM-P114	Brominated flame retardants in plastics	g/g	7	2008	JRC
CCQM-Q020.f	Digoxin purity	g/g	1	2008	BIPM,LGC
CCQM-K055.a	17-β-Estradiol purity	g/g	1	2009	BIPM
CCQM-K080	Creatinine in human serum	g/g	1	2009	NIST
CCQM-K055.b	Aldrin purity	g/g	1	2010	BIPM
CCQM-K079	Ethanol in aqueous matrix	g/g	1	2010	BAM,NIST
CCQM-P129	Water and ethanol in bioethanol fuel	g/g	2	2010	INMETRO,VSL
CCQM-K055.c	L-(+)-Valine purity	g/g	1	2012	BIPM
CCQM-K095	Organochlorine pesticides in tea	g/g	2	2012	GLHK,NIM
CCQM-P150	qNMR purity: Data acquisition and processing	g/g	1	2014	NMIJ
CCQM-K102	Polybrominated diphenyl ethers (PBDEs) in sediment	g/g	3	2015	JRC
CCQM-K132	Vitamin D metabolites in human serum	g/g	3	2015	NIST
CCQM-K055.d	Folic acid purity	g/g	1	2016	BIPM
CCQM-K095.1	Polycyclic aromatic hydrocarbons (PAHs) in tea	g/g	2	2016	NIST
CCQM-K109	Urea and uric acid in human serum	g/g	4	2016	HSA
CCQM-K131	Polycyclic aromatic hydrocarbons (PAHs) in acetonitrile	g/g	3	2016	NIST
CCQM-K142	Urea and uric acid in human serum or plasma	g/g	2	2016	HSA,NIST
CCQM-K078.a	Multi-component amino acids in dilute HCl solution	g/g	4	2017	BIPM
CCQM-K147	Niacin (vitamin B_3) in milk powder	g/g	1	2017	NIST,CENAM
CCQM-K146	Benzo[a]pyrene in olive oil	g/g	1	2018	NIM
CCQM-K148.a	Bisphenol-A purity	g/g	1	2019	BIPM

a The number of datasets derived from the study that contain a NIST result.

The cumulative distributions in Fig. 26 display the time course of NIST's engagements with finalized OAWG studies, where "engagements" applies to participations and coordinations.

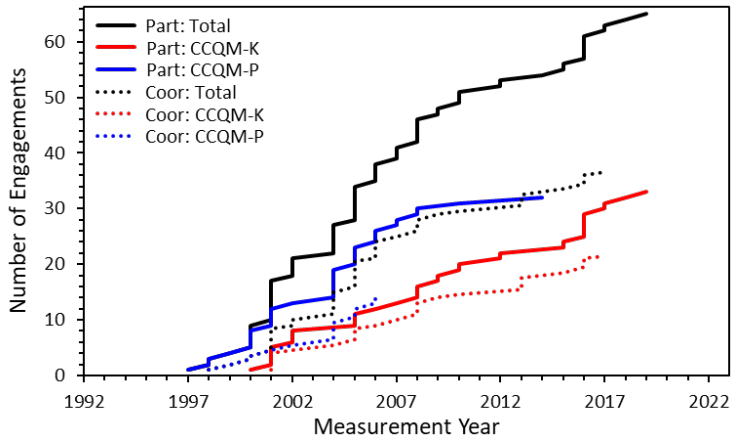


Fig. 26. OAWG Studies: Cumulative Distributions of NIST Engagements.

The histogram traces of Fig. 27 display the engagement time course in higher resolution, along with the average number of participations and coordinations from the first to the most recent engagement.

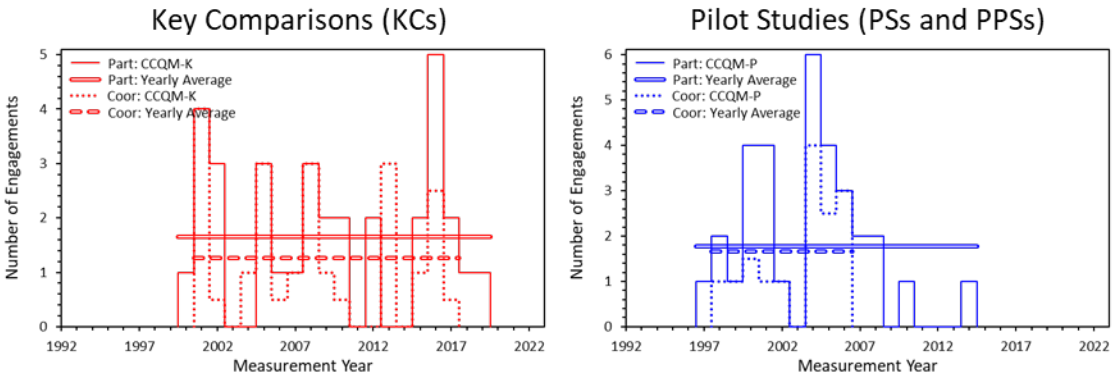


Fig. 27. OAWG Studies: Histogram Traces of NIST Engagements.

6.2. Performance OAWG Key Comparisons

NIST participated in 33 OAWG-conducted KCs that were finalized by this report's publication date. NIST has results in 81 datasets generated by these studies. The relationship between the absolute zeta-score bias of these results, $|\zeta_{NIST}|$, and the standard uncertainty of the reported values relative to the median of the participant uncertainties, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, is visualized in Fig. 28.

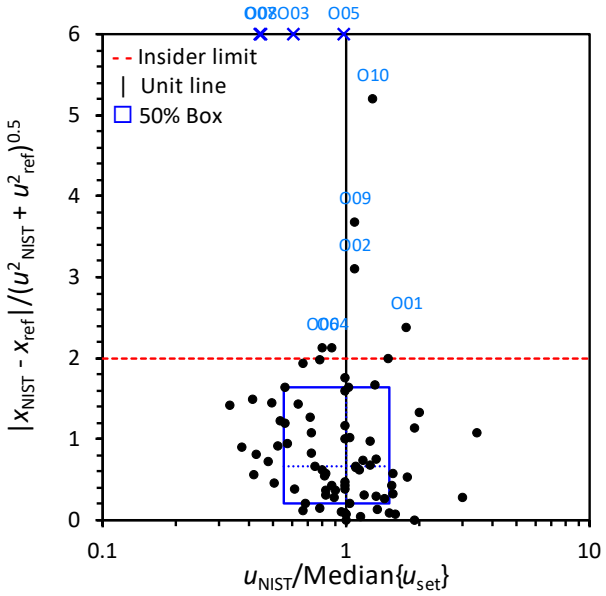


Fig. 28. OAWG Key Comparisons: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one Key Comparison (KC) dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$. The "×" denote values above the chart's Y-axis. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties. The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$).

The median absolute bias is 0.66. More than 87 % of the results are within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})/\text{Median}\{u(x_i)\}$ is 1.00. The $u(x_{NIST})$ is less than the participant median for six of the ten "outsiders" ($|\zeta_{NIST}| > 2$). The "outsider" datasets are documented in Section 6.2.4.

6.2.1. Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for OAWG KC datasets are displayed in Fig. 29 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.0101|/0.0033 \approx 3$, suggest that the absolute bias slightly increased over time.

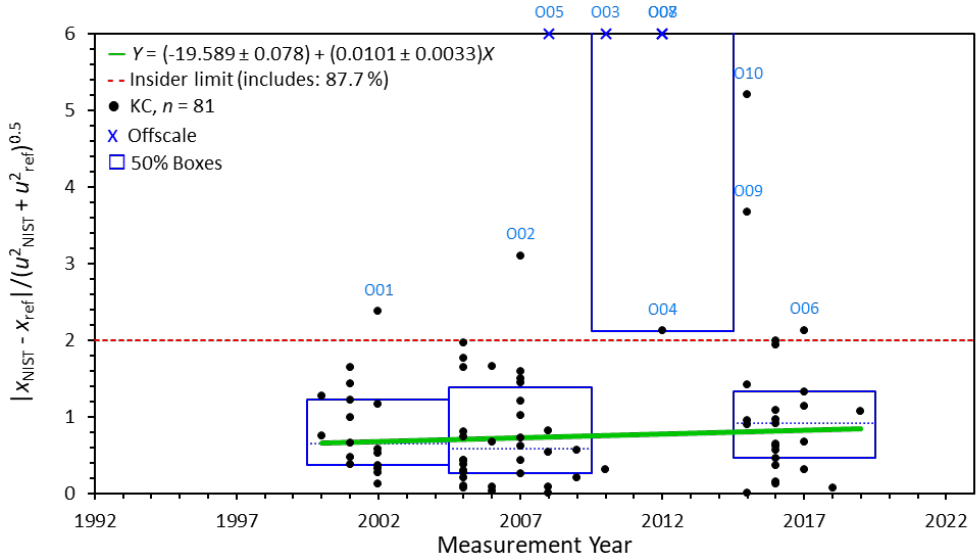


Fig. 29. OAWG Key Comparisons: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one Key Comparison (KC) dataset as a function of measurement year. The "×" denote values above the chart's Y-axis. The dashed horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of values within a sequential five-year interval. The sloping green line denotes the estimated linear trend. The legend provides the change per year, the percentage of values within the "insider limit", and the number of $|\zeta_{NIST}|$.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 30 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.00074|/0.00084 \approx 1$, suggest that the uncertainty ratios did not change over time.

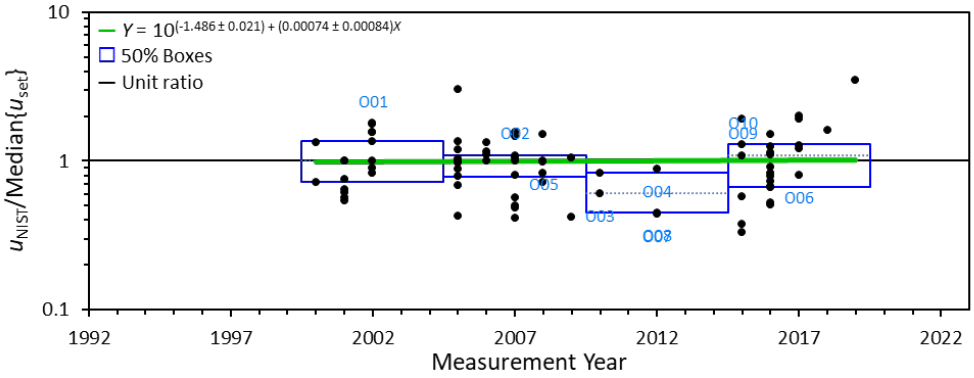


Fig. 30. OAWG Key Comparisons: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, in one Key Comparison (KC) dataset as a function of measurement year. The labels identify datasets which contain "outsider" results ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of the values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation.

6.2.2. Bias and Relative Uncertainty Across Measurand Level

The $|\zeta_{NIST}|$ values for OAWG KC datasets are displayed in Fig. 31 as a function of measurand level. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.012|/0.011 \approx 1$, suggest that the absolute bias is independent of measurand level.

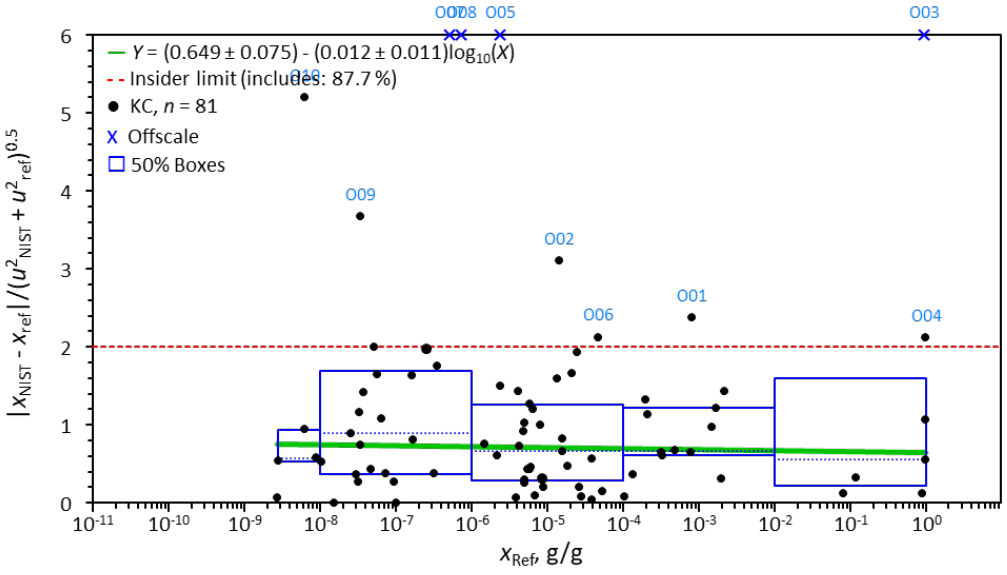


Fig. 31. OAWG Key Comparisons: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of mass fraction. See Fig. 29 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 32 as a function of measurand level. While the linear trend, $|t| = |+0.0068|/0.0025 \approx 3$, suggests that the uncertainty ratios increase slightly with measurand level, the pattern of the "50% boxes" suggests that they are better described as being independent of measurand level.

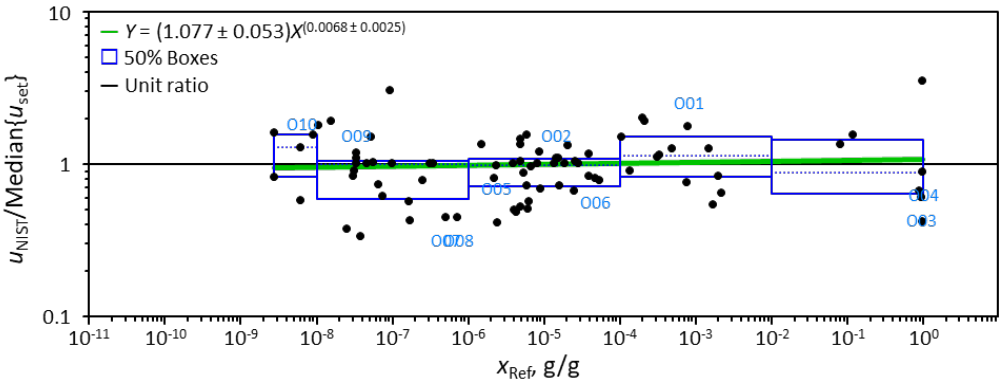


Fig. 32. OAWG Key Comparisons: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, in one dataset as a function of mass fraction. See Fig. 30 for description of the graphical elements.

6.2.3. Measurand Level Over Time

The KC reference values, x_{ref}, are displayed in Fig. 33 as a function of measurement year. The linear trend, $|t| = |+0.0642|/0.0098 \approx 6$, suggests that the mass fraction level of the measurands has increased over time. However, the pattern of the "50% boxes" suggests that the level changes are episodic, likely reflecting differences in the nature of the measurands studied.

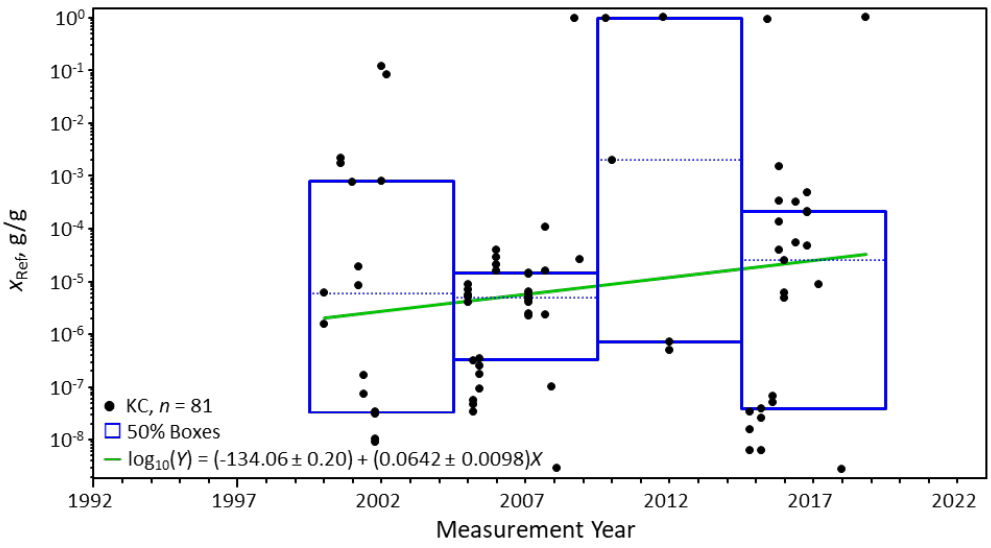
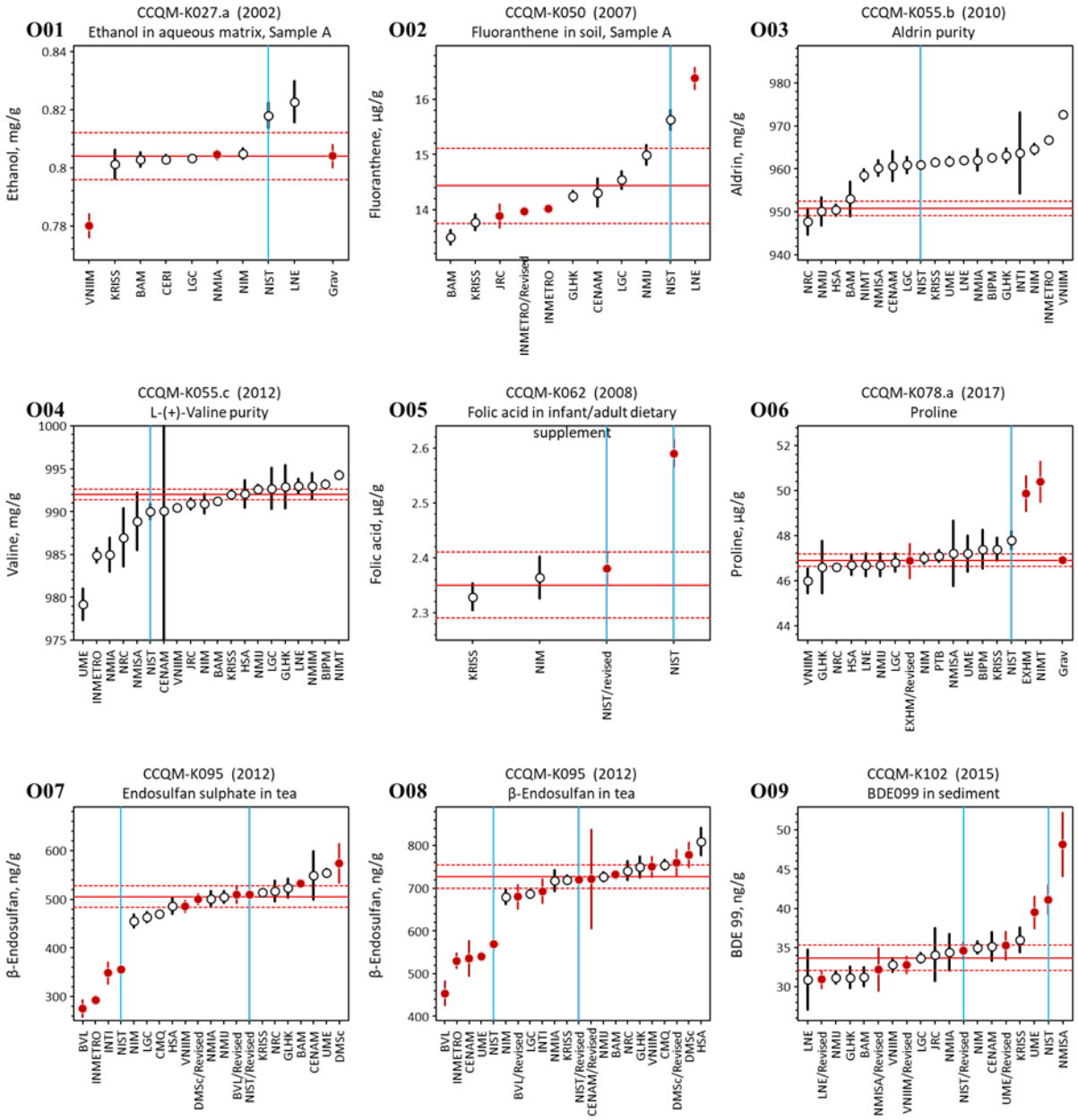


Fig. 33. OAWG Key Comparisons: Measurand Level Over Time.

Each symbol represents the reference value, x_{ref}, in one Key Comparison (KC) dataset in which NIST contributed a value. Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation and the number of reference values summarized.

6.2.4. “Outsider” Datasets

The ten OAWG KC datasets that contain “outsider” NIST results are depicted in Fig. 34 as dot-and-bar charts, each panel showing all results for one dataset, the reference value, and the 95 % level of confidence reference interval. The label in the upper left corner of each panel, running in sequence from “O01” to “O10”, is the code used to identify the “outsiders” in Fig. 28 to Fig. 32.



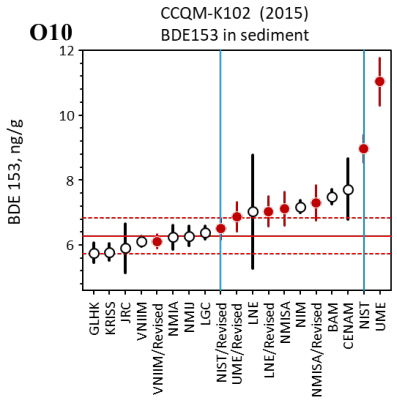


Fig. 34. OAWG Key Comparisons: "Outsider" Datasets.

Each panel shows all reported results for one dataset with a NIST "outsider" ($|\zeta_{NIST}| > 2$) result. Results are sorted in order of increasing magnitude. Participants are identified by code name; the full names are listed in Appendix B. Open circles denote results considered valid; solid red circles denote results excluded from estimating the reference value. Bars represent one standard uncertainty. The horizontal solid red line denotes the reference value; the dashed red lines bound a 95 % level of confidence interval about the reference value. The blue vertical lines mark NIST results.

6.3. Performance in OAWG Pilot Studies

NIST participated in 33 OAWG-conducted pilot studies (7 PPSs and 26 PSs) that were finalized by this report's publication date. NIST has results in 112 datasets (15 PPSs and 97 PSs) generated by these studies. Most of the participations were in the stand-alone studies used for training and method validation during the OAWG's first decade.

The relationship between the absolute zeta-score bias of these results, $|\zeta_{NIST}|$, and the standard uncertainty of the reported values relative to the median of the participant uncertainties, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, is visualized in Fig. 35.

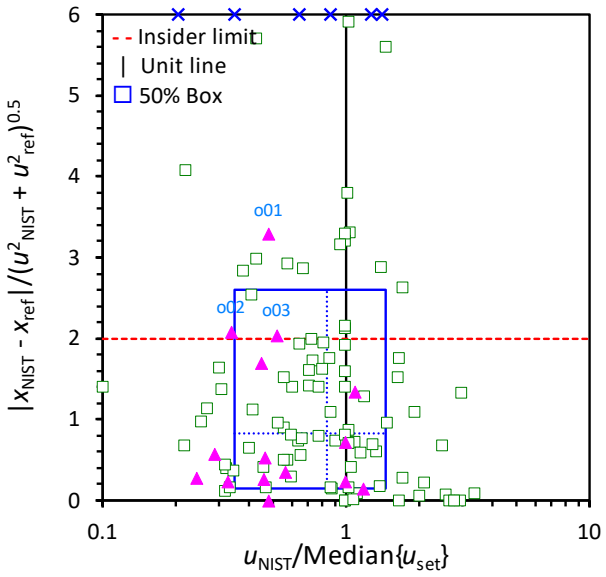


Fig. 35. OAWG Pilot Studies: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$: solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot studies (PS) results. The "×" denote values above the chart's Y-axis. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties. The labels identify "outsider" PPS datasets ($|\zeta_{NIST}| > 2$).

The median absolute bias for OAWG pilot study datasets is 0.83, slightly more than for the KC datasets. More than 75 % of the results are within $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})/\text{Median}\{u(x_i)\}$ for pilot study datasets is 0.83, somewhat less than for the KC datasets. The $u(x_{NIST})$ are less than the participant medians in 15 of the 27 "outsiders." The three PPS "outsiders" are documented in Section 6.3.4.

6.3.1. Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for OAWG PPS and PS datasets are displayed in Fig. 36 as a function of measurement year. Although the linear trend, $|t| = |+0.0150|/0.0065 \approx 2$, suggests that the absolute bias declined slightly over time, the pattern of the "50% Boxes" suggests that the absolute bias in the pilot studies has remained approximately constant.

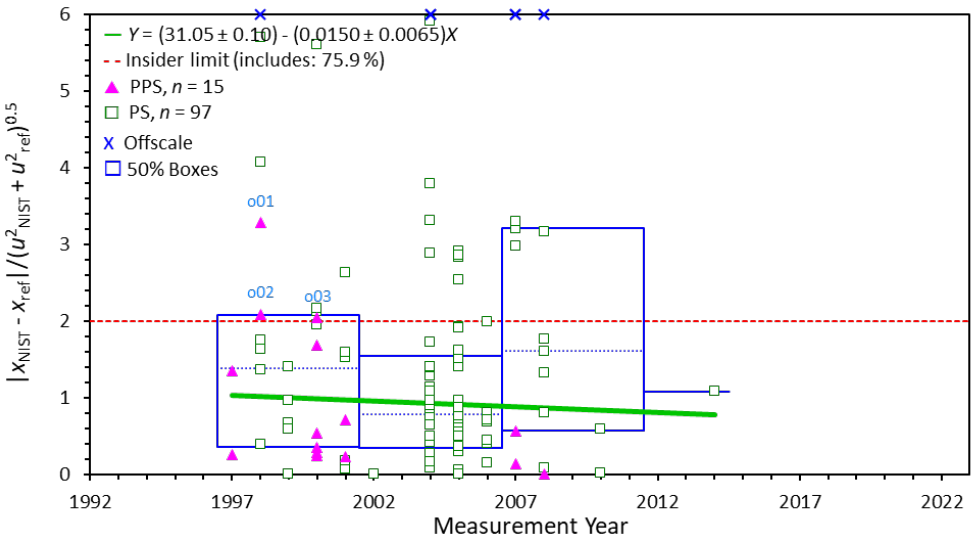


Fig. 36. OAWG Pilot Studies: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of measurement year: solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot studies (PS) results. The "×" denote values above the chart's Y-axis. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The labels identify "outsider" PPS datasets ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit," and the number of PPS and PS results.

The $u(x_{NIST})$/Median$\{u(x_i)\}$ values are displayed in Fig. 37 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.0124|/0.0015 \approx 8$, suggest that the uncertainty ratios increased over time.

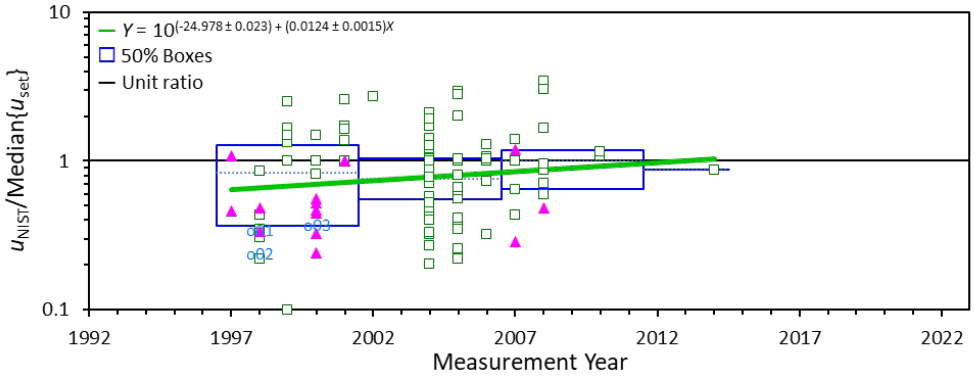


Fig. 37. OAWG Pilot Studies: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})$/Median$\{u(x_i)\}$, as a function of measurement year: solid magenta triangles denote results from published pilot studies (PPSs), open green squares denote results from confidential pilot studies (PSs). The labels identify datasets which contain "outsider" results ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of the values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation.

6.3.2. Bias and Relative Uncertainty Across Measurand Level

The $|\zeta_{NIST}|$ values for OAWG PPS and PS datasets are displayed in Fig. 38 as a function of measurand level. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.0274|/0.0098 \approx 3$, suggest that the absolute bias decreases slightly with increasing level.

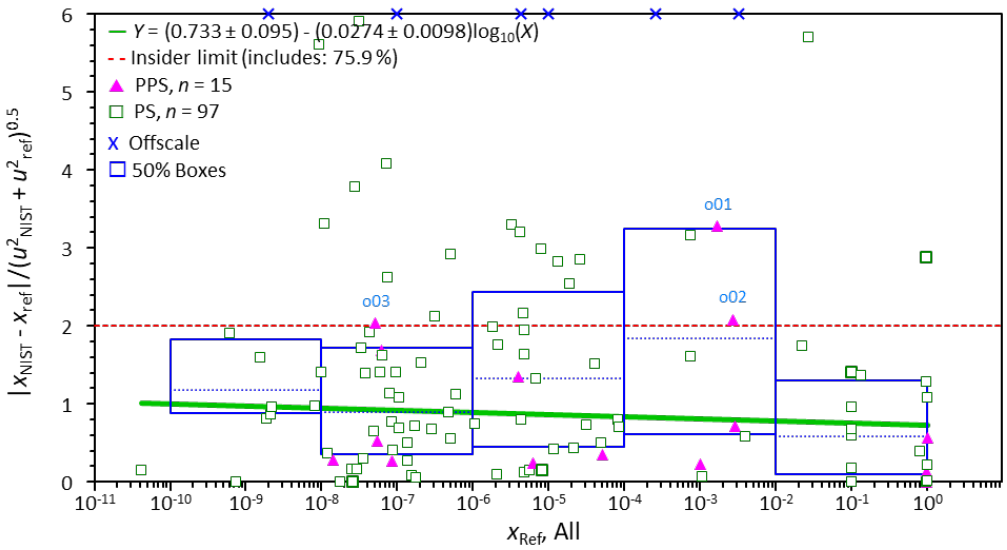


Fig. 38. OAWG Pilot Studies: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of mass or mole fraction. See Fig. 36 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 39 as a function of measurand level. The pattern of the "50% boxes" and linear trend, $|t| = |+0.0003|/0.0021 \approx 0$, suggest the uncertainty ratios are independent of measurand level.

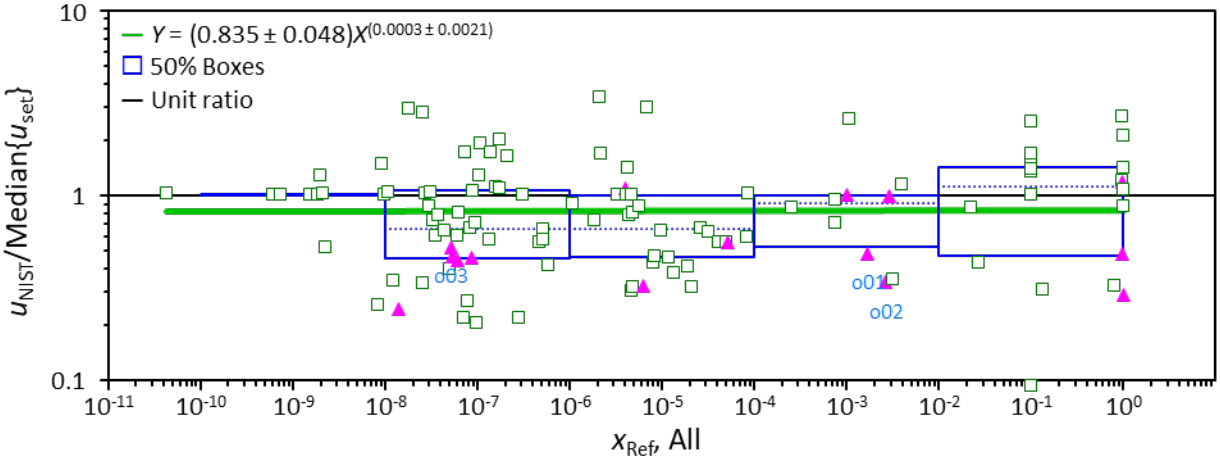


Fig. 39. OAWG Pilot Studies: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of mass or mole fraction. See Fig. 37 for description of the graphical elements.

6.3.3. Measurand Level Over Time

The pilot study reference values, x_{ref}, are displayed in Fig. 40 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.009|/0.018 \approx 0$, suggests that the mass fraction level of the measurands did not change over time.

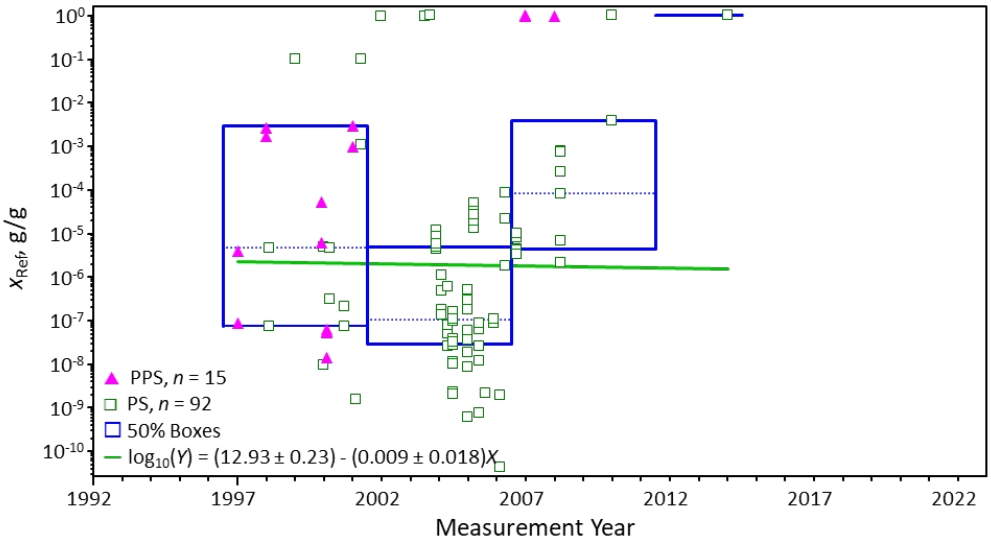


Fig. 40. OAWG Pilot Studies: Measurand Level Over Time.

Each symbol represents the reference value, x_{ref}, in one pilot study dataset in which NIST contributed a value: solid magenta triangles denote results from published pilot studies (PPSs), open green squares denote results from confidential pilot studies (PSs). Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation and the number of reference values summarized.

6.3.4. "Outsider" Datasets in Published Pilot Studies

The three PPS datasets that contain "outsider" NIST results are depicted in Fig. 41 as dot-and-bar charts, each panel showing all results for one dataset, the reference value, and the 95 % level of confidence reference interval. The label in the upper left corner of each panel, running in sequence from "O01" to "O03," is the code used to identify the "outsiders" in Fig. 35 to Fig. 39.

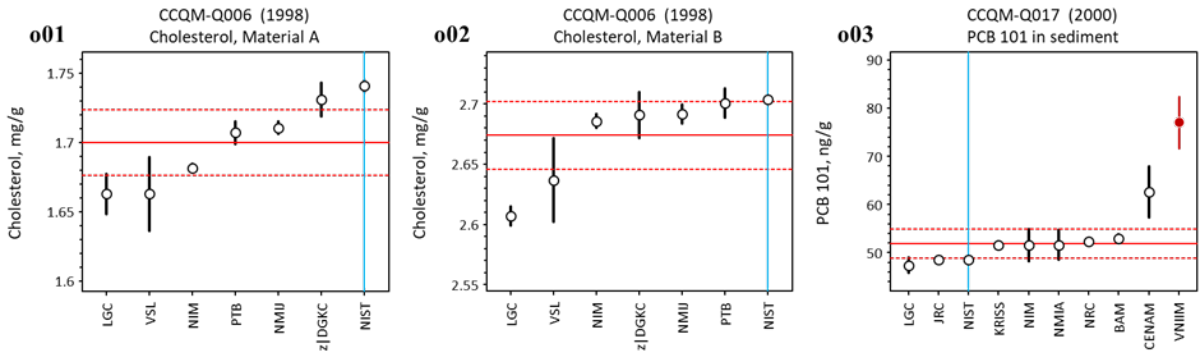


Fig. 41. OAWG Published Pilot Studies: "Outsider" Datasets.

Each panel shows all reported results for one dataset with a NIST "outsider" ($|\zeta_{NIST}| > 2$) result. Results are sorted in order of increasing magnitude. Participants are identified by code name; the full names are listed in Appendix B. Open circles denote results considered valid; solid red circles denote results excluded from estimating the reference value. Bars represent one standard uncertainty. The horizontal solid red line denotes the reference value; the dashed red lines bound a 95 % level of confidence interval about the reference value. The blue vertical lines mark NIST results.

6.4. Lessons Learned

Over the past several years, the OAWG has adopted significant policy changes to procedures used for the reduction and analysis of comparison study results. Notably, the use of relatively advanced statistical methods in KCRV determination is now encouraged to allow coordinators to make reasonable scientific inferences from comparison data sets that feature results that are over-dispersed, distributed in non-Gaussian arrangements, or heteroscedastic. The inclusion of uncertainty components accounting for excessive variance or "dark uncertainty" [22] in the KCRV estimation and in calculating the DoE of participant results is a notable shift in OAWG practice.

This practice was prompted in 2016 with the concurrent NIST-coordinated comparisons CCQM-K95.1 Low-Polarity Analytes in a Botanical Matrix: Polycyclic Aromatic Hydrocarbons (PAHs) in Tea [46] and CCQM-K131 Polycyclic Aromatic Hydrocarbons (PAHs) in acetonitrile [47]. Results for the two measurands (benz[*a*]anthracene and benzo[*a*]pyrene) of K95.1 and K131 were determined using independent, disparate analytical methods. The lack of overlap among the 95 % level of confidence expanded uncertainties, U_{95}, in both studies implied that the reported uncertainties did not fully account for the between-participant variance, often termed "excess variance" or "dark uncertainty" [22]. After extensive discussions within the OAWG, it was agreed that there were no apparent "outlying" results that could be attributed to non-random, systematic effects. The OAWG deemed that the uncertainties reported for all technically valid results were credible.

Following from this decision, statistical procedures for estimating measurand consensus values that account for additional layers of information provided by participants, namely standard uncertainties and degrees of freedom in their values, were implemented. A KCRV estimator (i.e., DerSimonian-Laird variance-weighted mean) that takes excess variance into account was executed via a bespoke spreadsheet implementation that was validated against an early version of the NIST Consensus Builder (NICOB) [35].

The accessibility of the statistically rigorous consensus estimation software such as NICOB and more recently the NIST Decision Tree [36] has facilitated other comparison coordinators and members within OAWG (and other WGs) to fold in more empirically derived information about the measurement into their analysis and meta-analysis of study results. These methods have enabled more comprehensive assessments and possibly reduced the likelihood of improperly flagging results as significantly discrepant.

Recent OAWG comparisons (e.g., CCQM-K148.a [48] and SIM.QM-K27 [49]) report a value for the excess variance, symbolized τ (tau), that expresses the degree of variation in the study data beyond that is in excess to the uncertainties reported with participants' results. While statistical procedures based on the typical random effects model account for τ in calculating RV, knowing its value aids study coordinators in detecting non-random effects on the measurements and elucidating causes for significant dispersion of results.

To assist OAWG study coordinators who have little expertise in statistics, a heuristic guidance document was prepared by NIST that directs them through a series of critical considerations for choosing a suitable KCRV estimator offered by new statistical software. This "KCRV Quick Guide" [50] encourages scrutiny of anomalous results and prompts the reader to consider several factors that influence the suitability of estimator options in the software. These factors include the quality ("reasonability") of reported uncertainty values, as well as the distribution shape and degree of dispersion of the study results. The NIST Decision Tree provides a similar decision-oriented guide that is coupled with statistical procedures; however, this tool should be used with datasets that have qualified as scientifically sound after having been subjected to the chemically-relevant technical scrutiny encouraged by the Decision Guide.

The OAWG is represented in an *ad-hoc* working group (KCRVWG) established to revise the *CCQM Guidance Note: Estimation of a consensus KCRV and associated degrees of equivalence* [16], a document comprising official CCQM-sanctioned guidance for KCRV and degree of equivalence calculations. In light of recent advances in statistical software and practical theories for seeking consensus values, revision of this document is being undertaken by CCQM WG representatives and statistical experts from NMIs around the world. The stated aim of this revision is to incorporate various contemporary viewpoints on KCRV estimation into official CCQM guidance and to accommodate and/or acknowledge technologies and schools of thought (e.g., Bayesian inference) that are already being implemented within the CCQM, albeit with little uniformity across the WGs.

The relationship of excessive variance and impact on CMCs has been recently raised by the CCQM, which has questioned whether a CCQM-wide approach to handle excessive variance should be applied across all WG comparisons. A proposal to have each WG record their understanding of where the dark uncertainty might have arisen is under consideration. This would also include any practical descriptions to any study-specific issues.

A few significant "lessons learned" from the results of OAWG studies have elucidated particular weakness or gaps within NIST's analytical laboratory measurement capabilities.

- Participation in the CCQM-P20 series of purity assessment pilot studies identified the need for improved organic purity evaluation methods and associated uncertainty assessments. This led to the development of a mass-balance approach rather than reliance on separation-based and thermogravimetric methods [51]. However, mass-balance methods suffer from the inherent (and common) assumption that the measurand is the entire proportion of sample material that has not been characterized.
- Poor results CCQM-K55.b Aldrin chemical purity study [52] exemplified the limitations to the mass-balance approach and the potential utility of quantitative NMR (qNMR) methods [51]. The presence of an unsuspected polymeric impurity present at 0.1% in the sample was missed using the suite of mass-balance methods employed while a result (unfortunately discounted) from a then-experimental qNMR method proved accurate. Results from this study underpinned the development and incorporation of qNMR capability into NIST's organic chemical metrology capacity. This capability has since become the core foundation of NIST's organic chemical purity methodology and greatly advanced the measurement services that NIST delivers. The technique has also revolutionized the way in which NIST establishes metrological traceability to the SI for many chemical measurements. Thus, the advantages of qNMR discovered through participation in the OAWG has ultimately enabled NIST to better serve its stakeholders.
- Poor results for CCQM-K055.c L-(+)-valine [53] demonstrated the merit of employing both qNMR and mass balance approaches to gain knowledge about the measurand. The result reported in Fig. 34 was based on only the qNMR result and proved to be slightly low relative to the RV (which included a significant portion of mass balance results).

  Participation in K55.c demonstrated that executing both qNMR and mass balance methods, when necessary and feasible, is a more metrologically sound approach to purity analysis than using only one of the methods. In addition to providing values that can be combined, bias in the result from one method could be inferred or investigated simply through comparison to the result from the other. While this is a basic principle of metrology, the practical inferences gained by such a comparison in chemical purity analyses can be profound. qNMR and mass balance procedures assay different analytes and produce results through completely different sets of assumptions. Checking these assumptions against one another is critical. Furthermore, to realize measurement units for chemical mass fraction, purity analyses aim to achieve complete knowledge about the composition of a sample. Yet, standards or references delivering "true" and complete knowledge about the total chemical make-up of the sample do not exist. A purity analysis is the first link of metrological traceability for virtually any organic chemical measurement and a high level of confidence in the result requires a commensurate degree of evidence that the composition is sufficiently understood.

As a result of this lesson learned through participation in OAWG purity comparisons, employing both qNMR and mass balance procedures for purity analysis has become a routine method for the production of pure organic chemical NIST SRMs [51]. The use of appropriate statistical methods to combine results with varying probability distribution functions is key to providing meaningful fit-for-purpose value assignments.

- The utility of this approach was demonstrated by the excellent results for CCQM-K146 Benzo[a]pyrene in Olive Oil [54] in which results from the two analytical methods were combined using the Linear Pool method provided in NICOB [35].
- Poor results for CCQM-K95 pesticides in tea [55] (Fig. 34) demonstrated the need to use appropriate control samples, i.e. similar matrices to the unknown samples, when evaluating unfamiliar matrices.
- Poor results for CCQM-K102 BDEs in sediment [56] (Fig. 34) demonstrated a missed opportunity to employ appropriate chromatographic measurement techniques together with adequate quality control practices. NIST withdrew its results for BDE 99 and BDE 153 after a follow up study determined that the chromatographic column used did not separate out the interfering species. Comparing results from different chromatographic columns of differing chemistries and/or lengths is important to ensuring the lack of interfering compounds along with using appropriate control materials.

7. Gas Analysis Working Group (GAWG)

The Gas Analysis Working Group (GAWG) is one of four WGs established by the CCQM during 1997 to 1998 [1]. The GAWG has responsibility for “gas composition (including binary and multicomponent mixtures); gas/liquid mixture composition; nanoparticle and aerosol concentration; isotope ratio measurement; [and] concentration of dissolved gases in liquid or solid matrices” [57].

The first segments of a feasibility study(“Study II”) “organized during the period 1990 [to] 1993 under the aegis of the CIPM” [1] on the measurement relatively simple primary standard mixtures of gases were completed in 1994 and discussed at the first meeting of the CCQM in 1995 [41]. The final segment of the study was completed in 1998. The segments of Study II are now known as CCQM-K1.a to CCQM-K1.g and are attributed to the GAWG.

7.1. Engagements in Non-Ozone GAWG Studies

NIST participated in 48 GAWG-affiliated non-ozone Key Comparisons (KCs), Supplementary Comparisons (SCs), published pilot studies (PPSs), or confidential pilot studies (PSs) that were finalized by this report’s publication date. NIST has results in 156 datasets generated by these studies. NIST’s participation in ozone photometer comparisons is discussed in Section 7.3. NIST also participated in the 2022 CCQM-P204 CO_2 isotope ratios in pure carbon dioxide, a pilot study organized by the GAWG in close collaboration with the IRWG. Isotopic measurement results are addressed in Section 12.

Table 8 lists in chronological order the finalized non-ozone GAWG studies with results reported in units of mole fraction (mol/mol) that NIST has participated in.

Table 8. NIST Engagement in GAWG-Attributed Non-Ozone Studies.

Measurements in all studies reported in units of mole fraction, mol/mol.

Study	Description	Sets [a]	Year	Coordinator(s)
CCQM-K001.b	Carbon dioxide in nitrogen	3	1994	VSL
CCQM-K001.a	Carbon monoxide in nitrogen	3	1995	VSL
CCQM-K001.c	Nitric oxide (NO) in nitrogen	2	1996	VSL
CCQM-K001.d	Sulfur dioxide in nitrogen	2	1997	VSL
CCQM-K001.e	Natural gas type I	6	1997	VSL
CCQM-K001.f	Natural gas type II	6	1997	VSL
CCQM-K001.g	Natural gas type III	6	1997	VSL
CCQM-K003	Automotive emission gases in nitrogen	3	1999	VSL
CCQM-K004	Ethanol in air	1	1999	NPL
CCQM-K007	Benzene, toluene and xylene (BTX) in nitrogen	5	1999	NIST
CCQM-K010	Benzene, toluene and xylene (BTX) in nitrogen	3	2001	NIST,NPL
CCQM-P023	Carbon monoxide in nitrogen	2	2001	VSL
CCQM-K016.a	Natural gas type IV	12	2002	BAM,VSL
CCQM-K015	SF_6 and CF_4 in nitrogen	2	2003	KRISS
CCQM-K022	Volatile organic compounds (VOCs) in air	8	2003	NMIJ
CCQM-Q041.1	Methane and carbon dioxide in air: Measurement capability	2	2003	VSL
CCQM-Q041.2	Methane and carbon dioxide in air: PSM comparison	2	2003	VSL
CCQM-K026.a	Nitric oxide (NO) in nitrogen	1	2004	NPL
CCQM-K041	Hydrogen sulfide in nitrogen	1	2005	NIST
CCQM-K052	Carbon dioxide in air	1	2006	VSL
CCQM-K054	n-Hexane in methane	1	2006	VSL
CCQM-Q073	Nitric oxide (NO) in nitrogen	1	2006	BIPM
CCQM-K046	Ammonia in nitrogen	1	2007	VSL
CCQM-K053	Oxygen (O_2) in nitrogen	1	2007	KRISS
CCQM-K051	Carbon monoxide in nitrogen	1	2008	NMISA
CCQM-K068	Nitrous oxide (N_2O) in synthetic air	1	2008	KRISS
CCQM-K071	Multi-component gas stack emissions	5	2008	VSL
EUQM-S003	Volatile organic compounds (VOCs) in air	29	2008	NPL
CCQM-K066	Methane purity	4	2009	NMIJ
CCQM-K074	Nitrogen dioxide (NO_2) in nitrogen	1	2010	BIPM
CCQM-K076	Sulfur dioxide in nitrogen	1	2010	NIST
CCQM-Q110.B1	Nitrogen dioxide (NO_2) in nitrogen by FT-IR spectroscopy	1	2010	BIPM
CCQM-Q110.B2	Nitrogen dioxide (NO_2) in nitrogen using synthetic spectra	1	2010	BIPM
CCQM-K084	Carbon monoxide in synthetic air at ambient level	1	2012	KRISS
CCQM-K093	Ethanol in nitrogen or air	1	2012	NPL
CCQM-K082	Methane in air: PSM comparison	2	2013	BIPM
CCQM-K083	Halocarbons in dry whole air	6	2013	NIST
CCQM-K101	Oxygen (O_2) in nitrogen	1	2013	NIM
CCQM-K111	Propane in nitrogen	1	2014	VSL
CCQM-K113	Noble gas mixture	3	2015	KRISS
CCQM-K121	Monoterpenes in nitrogen	4	2016	NIST
CCQM-K120.a	Carbon dioxide at background level	1	2017	BIPM,NIST
CCQM-K120.b	Carbon dioxide at urban level	2	2017	BIPM,NIST
CCQM-K010.2018	Benzene, toluene, ethylbenzene, and xylene (BTEX) in nitrogen	7	2018	NIST
CCQM-K137	Nitric oxide (NO) in nitrogen	2	2018	BIPM
CCQM-K003.2019	Automotive emission gases in nitrogen	4	2019	VSL
CCQM-K117	Ammonia in nitrogen	1	2019	VSL
CCQM-K068.2019	Nitrous oxide (N_2O) in air at ambient levels	1	2020	BIPM,KRISS

a The number of datasets derived from the study that contain a NIST result.

The cumulative distributions in Fig. 42 display the time course of NIST's engagements with finalized non-ozone GAWG studies, where "engagements" applies to participations and coordinations.

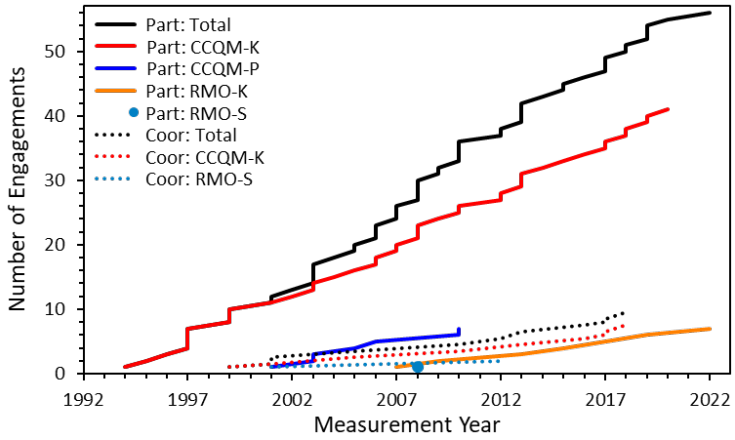


Fig. 42. Non-Ozone GAWG Studies: Cumulative Distributions of NIST Engagements.

The histogram traces of Fig. 43 display the engagement time course in higher resolution, along with the average number of participations and coordinations from the first to the most recent engagement.

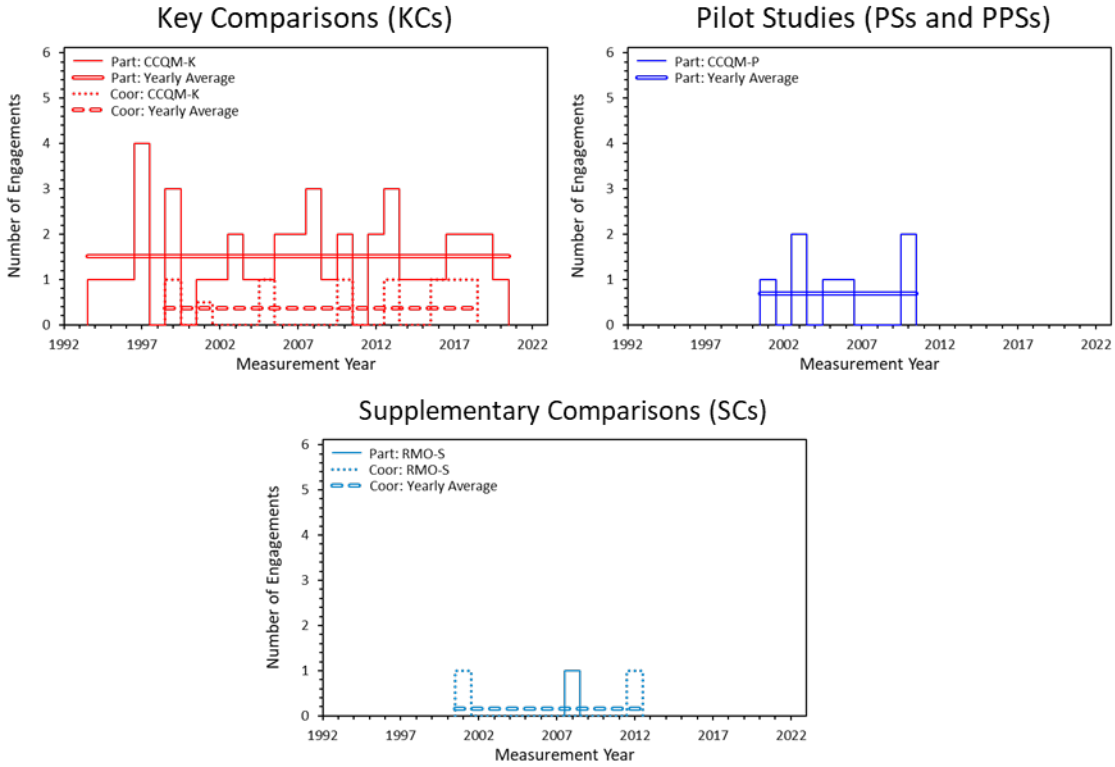


Fig. 43. Non-Ozone GAWG Studies: Numbers of NIST Engagements.

7.2. Performance in Non-Ozone GAWG Studies

The following documents NIST's measurement performance in non-ozone key and Supplementary Comparisons (KCs and SCs) separately from performance in pilot studies (PSs and PPSs).

7.2.1. GAWG Key and Supplementary Comparisons

NIST participated in 47 GAWG-related non-ozone KCs and SCs that were finalized by this report's publication date. NIST has results in 154 datasets generated by these studies. The relationship between the absolute zeta-score bias of these results, $|\zeta_{NIST}|$, and the standard uncertainty of the reported values relative to the median of the participant uncertainties, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, is visualized in Fig. 44.

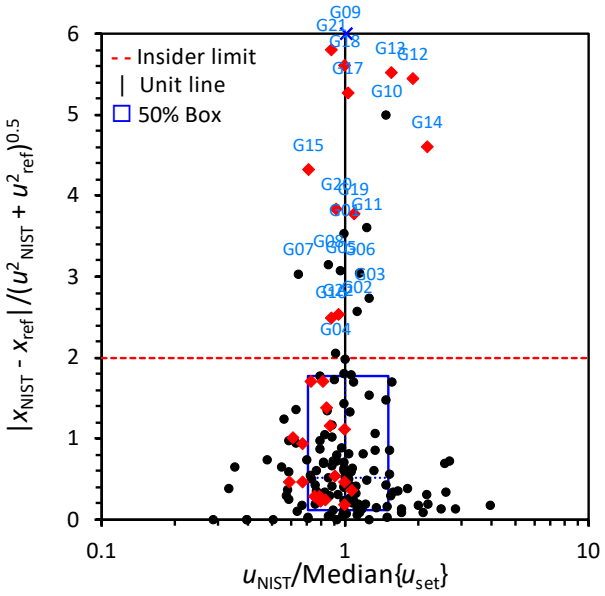


Fig. 44. GAWG Key Comparisons: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the result's standard uncertainty relative to the participant median $u(x_{NIST})/\text{Median}\{u(x_i)\}$: solid black circles denote Key Comparisons (KC) results, solid red diamonds denote Supplementary Comparisons (SC) results. The "×" denotes a value above the chart's Y-axis. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties. The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$).

The median absolute bias is 0.59. More than 85 % of the dataset results are within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})/\text{Median}\{u(x_i)\}$ is 1.00. The $u(x_{NIST})$ is less than the participant median for 10 of the 22 "outsider" datasets ($|\zeta_{NIST}| > 2$). The "outsider" datasets are documented in Section 7.2.1.4.

7.2.1.1. GAWG Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for GAWG KC and SC datasets are displayed in Fig. 45 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.0201|/0.0013 \approx 15$, suggest that the absolute bias increased over time.

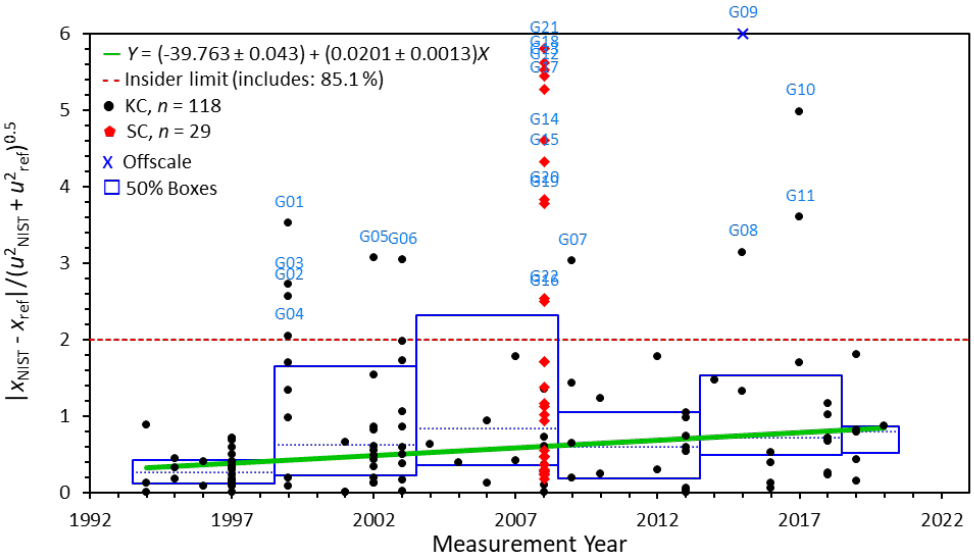


Fig. 45. GAWG Key Comparisons: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of measurement year: solid black circles denote Key Comparisons (KC) results, solid red diamonds denote Supplementary Comparisons (SC) results. The "×" denotes a value larger than the chart's Y-axis maximum. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit," and the number of KC and SC results displayed.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 46 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.00665|/0.00030 \approx 22$, suggest that the uncertainty ratios decreased over time.

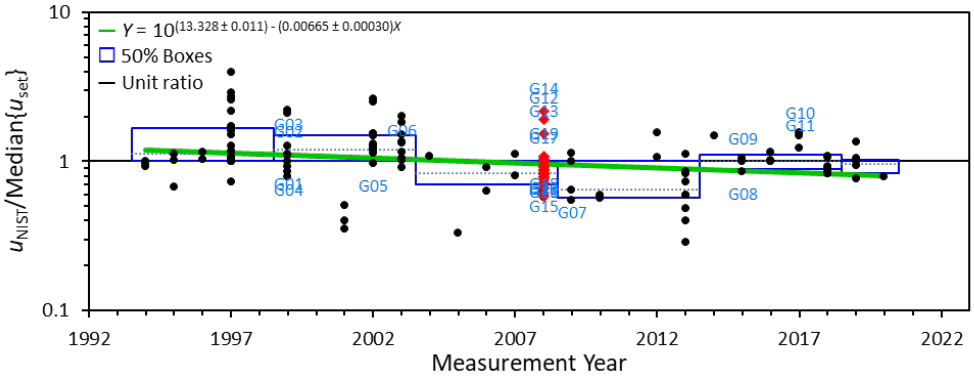


Fig. 46. GAWG Key Comparisons: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of measurement year. The labels identify datasets which contain "outsider" results ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of the values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation.

Because of its temporal location about midway between the initial and most recently completed studies, the multitude of results from the 2008 SC do not dominate the analysis. Excluding the 29 results from EUQM-S003 (EUROMET.QM-S3) VOCs in air results very modestly lowers the slope and decreases the significance of the trends: $|t| = |+0.0174|/0.0013 \approx 13$ for absolute bias and $|t| = |-0.00580|/0.00037 \approx 16$ for relative uncertainty. However, these SC results warn that there can be qualitative differences between KC and SC participations.

7.2.1.2. GAWG Bias and Relative Uncertainty Across Measurand Level

The $|\zeta_{NIST}|$ values for GAWG KC and SC datasets are displayed in Fig. 47 as a function of measurand level. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.0434|/0.0041 \approx 11$, suggest that the absolute bias decreases with increasing measurand level.

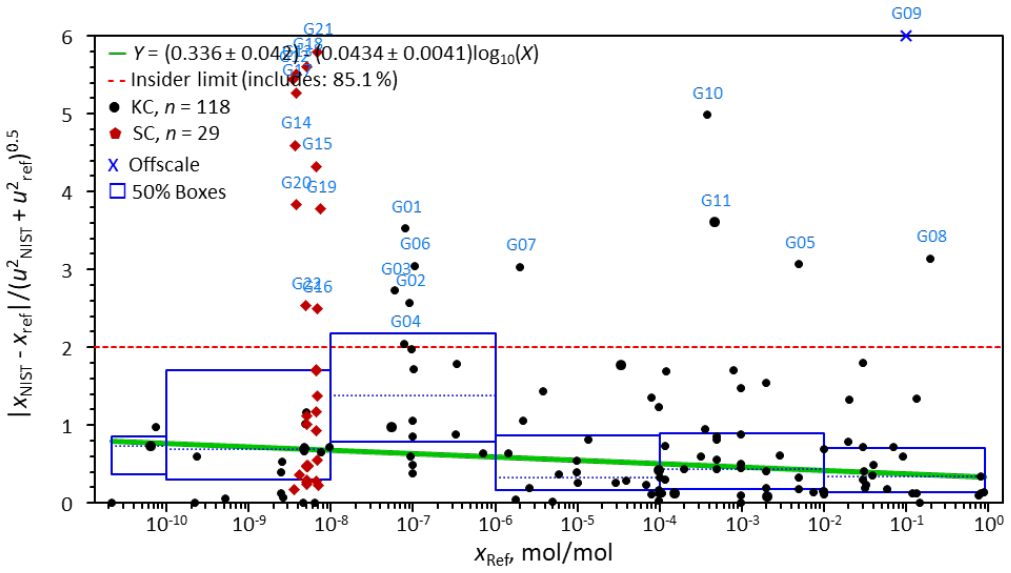


Fig. 47. GAWG Key Comparisons: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of mole fraction. See Fig. 45 for description of the graphical elements.

The $u(x_{NIST})$/Median$\{u(x_i)\}$ values are displayed in Fig. 48 as a function of measurand level. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.02537|/0.00076 \approx 34$, suggest that the uncertainty ratios increase with increasing measurand level.

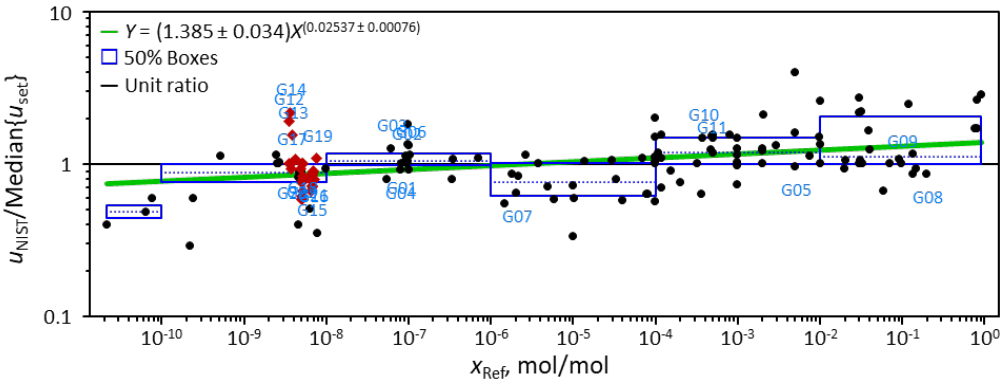


Fig. 48. GAWG Key Comparisons: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of mole fraction. See Fig. 46 for description of the graphical elements.

Because the multitude of results from the 2008 SC are all at relatively low measurand level (30 nmol/mol to 100 nmol/mol), they impact the trend across measurand level analysis more than they effect the trend over time. Excluding the 29 EUQM-S003 lowers the slope and decreases the significance of the absolute bias trend: $|t| = |+0.0103|/0.0033 \approx 3$. However, it increases the slope and significance of the relative uncertainty trend: $|t| = |+0.03061|/0.00087 \approx 35$.

7.2.1.3. Measurand Level Over Time

The KC and SC reference values, x_{ref}, are displayed in Fig. 49 as a function of measurement year. The linear trend, $|t| = |+0.2076|/0.0041 \approx 50$, suggests that the mole fraction level of the measurands has dramatically decreased over time. However, the pattern of the "50% boxes" suggests that the decrease plateaued in the mid-2010s and has since increased. This likely reflects changes in the nature of the measurands studied.

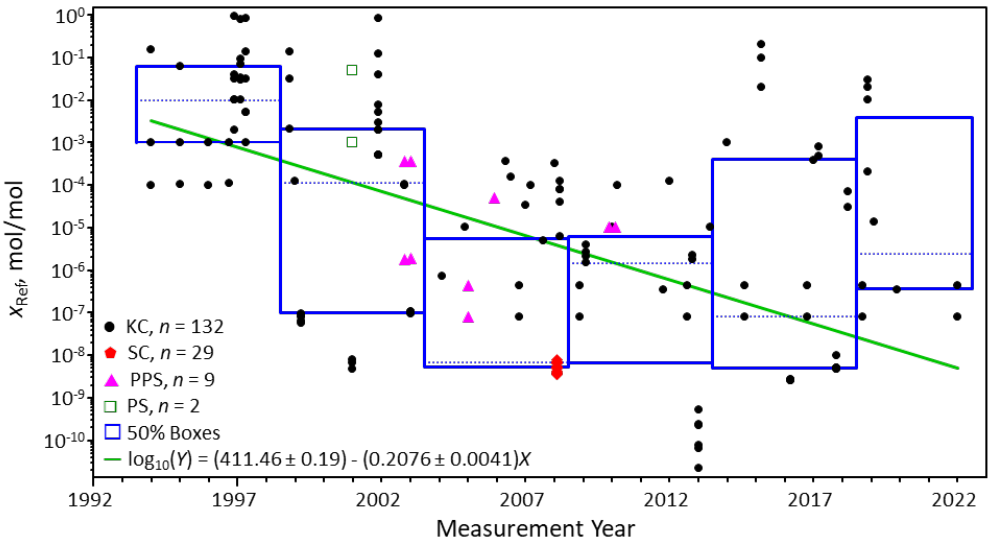
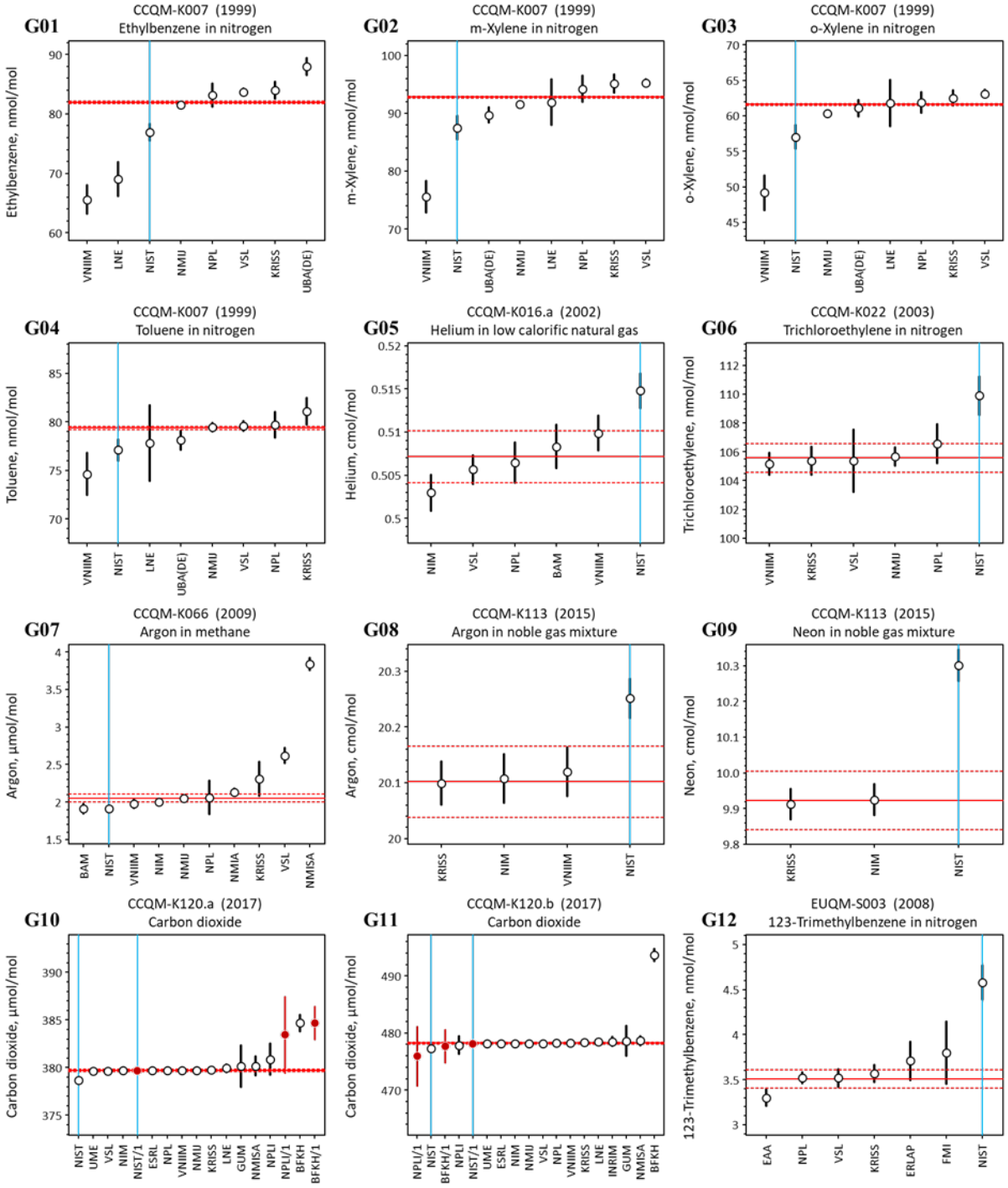


Fig. 49. GAWG Key Comparisons: Measurand Level Over Time.

Each symbol represents the reference value, x_{ref}, in one dataset in which NIST contributed a value: solid black circles denote Key Comparisons (KC) results, solid red diamonds denote Supplementary Comparison (SC) results, solid magenta triangles denote published pilot study (PPS) results. Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation and the number of reference values summarized.

7.2.1.4. “Outsider” Datasets

The 22 GAWG KC and SC datasets that contain “outsider” NIST results are depicted in Fig. 50 as dot-and-bar charts, each panel showing all results for one dataset, the reference value, and the 95 % level of confidence reference interval. The label in the upper left corner of each panel, running in sequence from “G01” to “G22,” is the code used to identify the “outsiders” in Fig. 44 to Fig. 48.



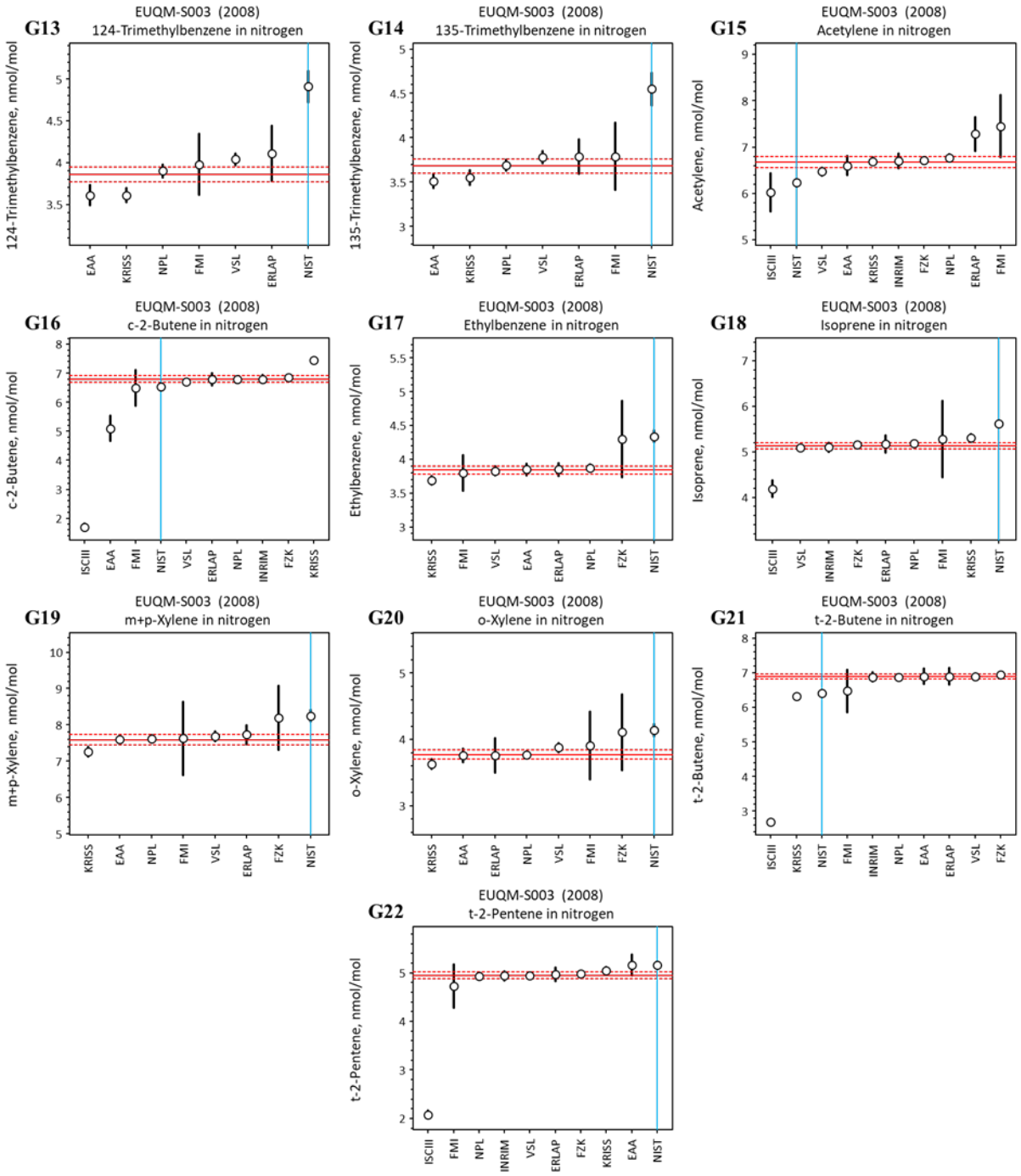


Fig. 50. GAWG Key Comparisons: "Outsider" Datasets.

Each panel shows all reported results for one dataset with a NIST "outsider" ($|\zeta_{NIST}| > 2$) result. Results are sorted in order of increasing magnitude. Participants are identified by code name; the full names are listed in Appendix B. Open circles denote results considered valid; solid red circles denote results excluded from estimating the reference value. Bars represent one standard uncertainty. The horizontal solid red line denotes the reference value; the dashed red lines bound a 95 % level of confidence interval about the reference value. The blue vertical lines mark NIST results.

7.2.2. GAWG Pilot Studies

NIST participated in only five non-ozone GAWG-conducted pilot studies (four PPSs and one PS) that were finalized by this report's publication date. NIST has results in nine datasets (seven PPSs and two PSs) generated by these studies. The relationship between the $|\zeta_{NIST}|$ and $u(x_{NIST})/\text{Median}\{u(x_i)\}$ is visualized in Fig. 51

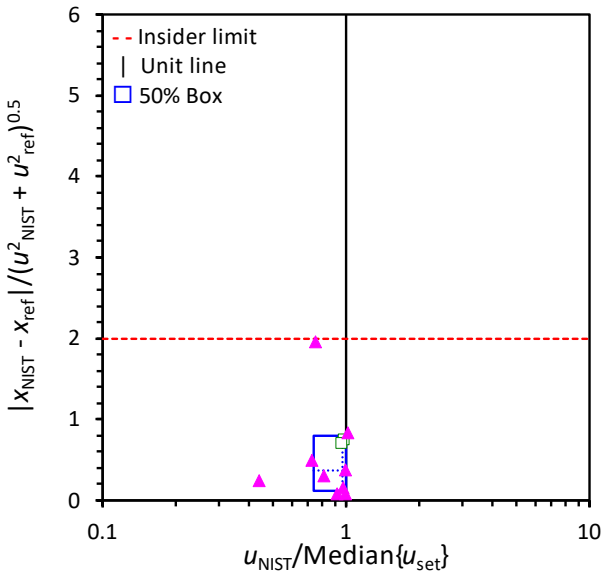


Fig. 51. GAWG Pilot Studies: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$: solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties.

The median absolute bias for GAWG pilot study datasets is 0.37. All of the results are within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})/\text{Median}\{u(x_i)\}$ is 0.98.

7.2.2.1. GAWG Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for the few GAWG PPS and PS non-ozone pilot study datasets are displayed in Fig. 52 as a function of measurement year. The results are too few for confident analysis, but both the pattern of the "50% boxes" and the linear trend, $|t| = |-0.039|/0.041 \approx 1$, suggest that the absolute bias did not change over time.

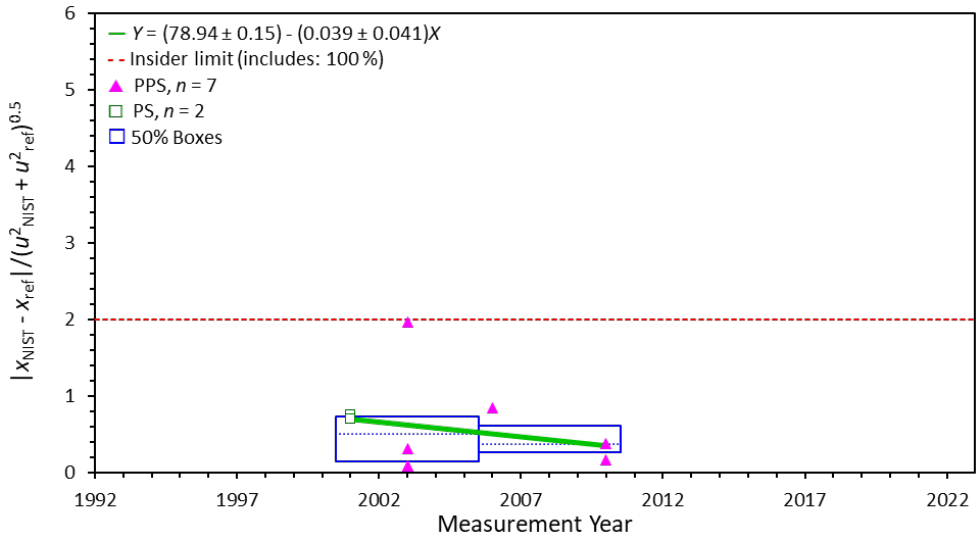


Fig. 52. GAWG Pilot Studies: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of measurement year: solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit", and the number of PPS and PS results displayed.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 53 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.0020|/0.0017 \approx 1$, suggest that uncertainty ratios did not change over time.

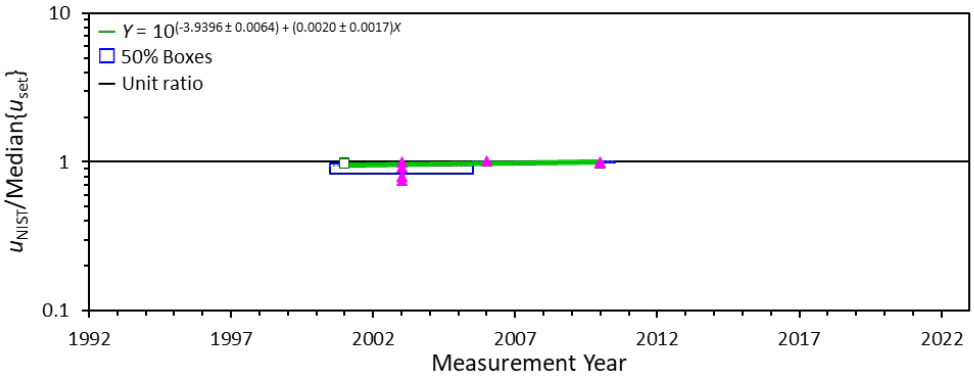


Fig. 53. GAWG Pilot Studies: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of measurement year: solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. Each blue box encloses the central 50 % of the values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation.

7.2.2.2. GAWG Bias and Relative Uncertainty Across Measurand Level

The $|\zeta_{NIST}|$ values for GAWG PPS and PS non-ozone datasets are displayed in Fig. 22 as a function of measurand level. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.155|/0.063 \approx 2$, suggest that the absolute bias increases with increasing level. However, since the higher measurand levels are for the earlier studies, the trend across level is confounded with the trend over time.

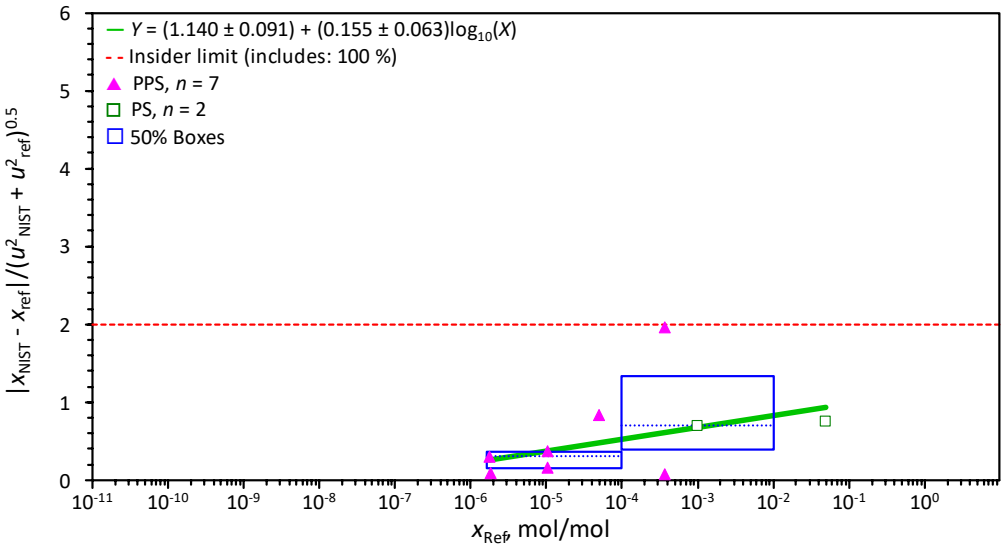


Fig. 54. GAWG Pilot Studies: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of mole fraction. See Fig. 52 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values for the GAWG PPS and PS datasets are displayed in Fig. 55 as a function of measurand level. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.0008|/0.0052 \approx 0$, suggest that the uncertainty ratios are independent of measurand level.

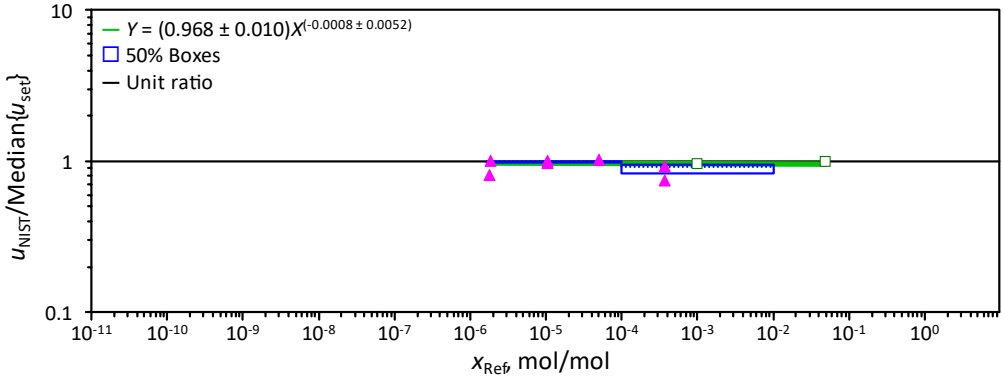


Fig. 55. GAWG Pilot Studies: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of mole fraction. See Fig. 53 for description of the graphical elements.

7.2.3. Measurand Level Over Time

The pilot study reference values, x_{ref}, are displayed in Fig. 56 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.218|/0.056 \approx 4$, suggests that the mole fraction level of the measurands in the few GAWG pilot studies decreased over time.

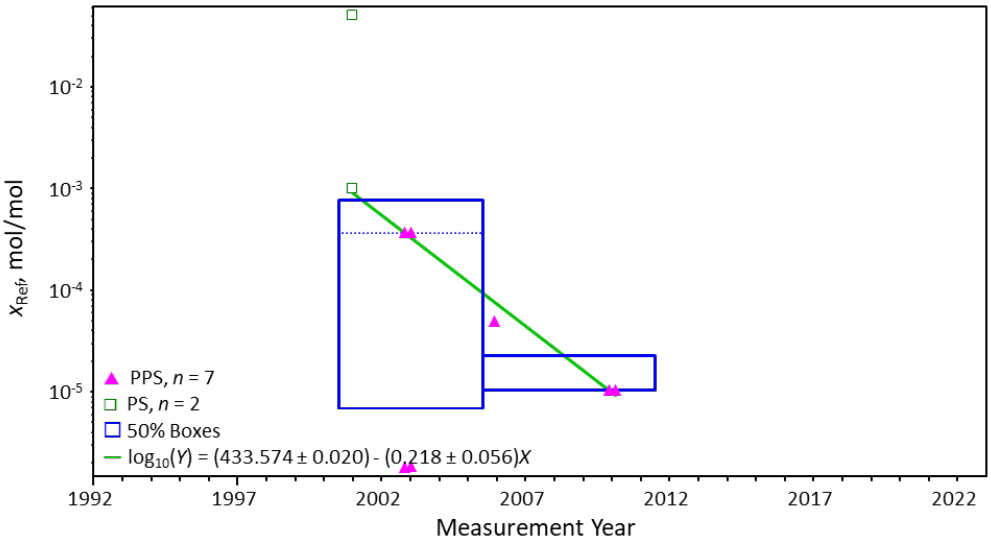


Fig. 56. GAWG Pilot Studies: Measurand Level Over Time.

Each symbol represents the reference value, x_{ref}, in one pilot study dataset in which NIST contributed a value: solid magenta triangles denote results from published pilot studies (PPSs), open green squares denote results from confidential pilot studies (PSs). Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation and the number of reference values summarized.

7.3. GAWG (Non-Ozone) Lessons Learned

The Gas Sensing Metrology Group (GSMG) at NIST has extensive uncertainty calculation programs for the preparation and verification of gravimetric standards, and the analytical procedures used for mole fraction assignments to gas mixtures. Additionally, uncertainties that reflect experience-based expert opinion are also included in the total uncertainty budget.

Many factors play into the total uncertainty, such as gas standards preparation systems and the balances used for those mass measurements. The GSMG utilizes some of the more advanced analytical instruments available, which allow for very low measurement uncertainty, including gas chromatographs, laser-based systems, and preconcentration units.

Not all NMIs that participate in KCs may have the most advanced instruments available to them, in particular those NMIs that are just developing. Thus, their uncertainties can be much larger, or their preparation methods may not be as accurate as those of the more established NMIs. As such, there can exist large differences in uncertainty and standards

accuracy among participating NMIs, which can skew results in KCs, resulting in an NMI that accurately reports a small uncertainty being penalized.

The “Outsider Datasets” G01 (ethylbenzene), G02 (*m*-xylene), G03 (*o*-xylene) and G04 (toluene) are from the CCQM-K7 BTEX in nitrogen (N_2) comparison [58]. Since NIST had probably the longest history in developing gas standards of organic compounds, it was assigned as coordinator for this comparison. Therefore, two GSMG analysts were assigned: one to gravimetrically prepare and verify the BTEX comparison mixtures (the coordinator) and the other to assign the NIST comparison values to its CCQM-K7 sample. After the NIST sample was analyzed and values assigned, the sample was turned back over to the coordinator. Based on the low results, it appeared that the mole fractions of the target organic compounds had decreased. However, there was not enough remaining gas pressure in the sample cylinder to adequately reanalyze to assess stability. It is known that the VOC concentration delivered by a cylinder can increase or decrease at low pressures, although the sorption mechanism(s) haven’t been determined [59]. In the later CCQM-K10 BTEX in N_2 comparison [60], which NIST also coordinated, the NIST values were within the KC determined values. An additional comparison between several NMIs showed good agreement for these same compounds at the pmol/mol level [61].

The G05 “Outsider” for helium (He) is from CCQM-K16, composition of natural gas [62]. The GSMG used binary He in N_2 standards to assign the value, as He was not included in the suite of NIST’s primary natural gas standards. One explanation for NIST’s high value is that the NIST primary He in N_2 standards were not mixed for long enough time periods after preparation. Improper mixing, such as not applying mild heat and rolling the cylinders for days after preparation, results in stratification of the gases, and He, being significantly lighter than N_2, can remain more concentrated in the top portion of the gas mixture if the cylinder is left in a standing position for long periods of time. Laying gas cylinders in a horizontal position in storage aids in creating “gas currents” that assist in keeping the gases consistently mixed.

The GSMG has made a practice of only participating in KCs for components that they support in their programs. In some cases, that has meant that if there is a multi-component KC, for which NIST only has developed standards for some of the components, NIST will only value assign those components that they support. This has resulted in “outsider” datasets caused by matrix effects. For those KCs where NIST has only value assigned some of the components in a multicomponent sample, binary primary standards were used instead of multicomponent standards.

The primary standards used to analyze an unknown sample must have the same components present in the gas matrix, since certain analytical instruments will show biases between different gas matrices. This was the case for the G08 and G09 “Outsiders” in CCQM-K113 noble gas mixture [63]. NIST only supported argon, neon, and krypton, and the GSMG only had binary standards for Ar in N_2, Ne in N_2 and Kr in N_2. After the results were revealed, the GSMG suspected that there was a strong matrix effect. On request, the CCQM-K113 coordinator returned the cylinder that NIST had analyzed. Post-study research was performed using several multicomponent primary standards that contained Ne, Ar, Kr,

and Xe in a He matrix [64]. The results from the original and re-analysis of NIST's CCQM-K113 sample are shown in Table 9. The relative differences in the mole fractions were all < ±0.05 % compared to the original differences of (+0.74 to +3.8) %. This post KC research confirmed that the gas matrix was the primary factor in the original NIST results.

Table 9. NIST Amount-of-Substance Composition of CCQM-K113 Mixtures. [a]

		Binary [b]		Multicomponent [c]	
Gas	Assigned, cmol/mol [d]	NIST, cmol/mol	Δ, % [e]	NIST, cmol/mol	Δ, % [e]
Neon	9.920 ± 0.082	10.298 ± 0.022	+3.8	9.925 ± 0.012	+0.05
Argon	20.079 ± 0.064	20.228 ± 0.028	+0.74	20.088 ± 0.026	+0.04
Krypton	1.989 ± 0.080	2.0639 ± 0.0022	+3.8 [f]	1.9900 ± 0.0014	+0.05
Xenon	1.017 ± 0.102	Not measured		1.0169 ± 0.0012	-0.01

a ± values are expanded uncertainties at approximately 95 % level of confidence.
b Original NIST results using binary (individual noble gas in N_2) primary standards.
c NIST re-analysis results using multicomponent (Ne, Ar, Kr, Xe in He) standards.
d Value assigned by CCQM-K113 Coordinator.
e Percent difference between the NIST and the Assigned values: 100×(NIST/Assigned - 1).
f Result not considered an "Outsider" by the *CCQM_Retrospectoscope*'s analysis criterion.

Adsorption/desorption of gas components to/from the walls in aluminum gas cylinders has long been a concern, and it reared its ugly head in G10 and G11 CCQM-K120 carbon dioxide (CO_2) in air [65]. Each participant in this KC was asked to submit to the coordinator two primary gas standards containing CO_2 in air. NIST chose to use two of its 10+ gravimetric standards from its new 2012 suite of CO_2 in air. All standards in the suite had gas pressures of (5500 to 6900) kPa ((800 to 1000) psi) remaining, and the suite was completely reanalyzed to make sure the results were consistent with expected values.

NIST was informed that its two submitted standards were giving low values when compared to all other participants, which could suggest that there was desorption of CO_2 from the cylinder walls into the gas mixture, resulting in a larger analytical instrument response compared to NIST's reported gravimetric mole fraction. The GSMG undertook an extensive research program to investigate adsorption/desorption of CO_2 in aluminum cylinders with different cylinder treatments. In general, it was found that very small amounts of CO_2 adsorb onto the cylinder wall during initial preparation [66]. As the gas pressure in the cylinder goes down with use, CO_2 begins to desorb from the walls into the gas mixture, effectively changing the CO_2 mole fraction. The research results show that this event can start occurring at pressures as low as 9000 kPa (1300 psi) and is not consistent from cylinder to cylinder. As a result, NIST requested to pull its results from the KC but was denied. An additional uncertainty due to adsorption/desorption effects will be included in the total uncertainty budget in the future. Also, any future participation in KCs for CO_2 in air will require the NIST samples to be freshly prepared gravimetric standards.

There are several possible causes for the G12 to G22 "Outsiders" from the EUROMET.QM-S3 volatile organic compounds in air [67] comparison. NIST is consistently high in G12 (1,2,3-trimethlylbenzene), G13 (1,2,4-trimethlylbenzene) and G14 (1,3,5-trimthylbenzene); however, in some later non-CCQM comparisons, NIST agreed well with some of the same labs that had participated in this EUQM-S003 comparison. One possible

explanation is that the labs reassessed their preparation techniques, which resulted in increased accuracy in preparing organic gas mixtures.

Double and triple bonded compounds degrade over time, at least in the "Aculife" treated aluminum cylinders that NIST has historically used for organic gas mixtures [59]. This may have influenced the G15 (acetylene), G16 (*c*-2-butene), G18 (isoprene), G21 (*t*-2-butene) and G22 (*t*-2-pentene) results. *Cis* and *trans* compounds can also isomerize in a gas mixture, which can increase or decrease the mole fraction of one or the other. Acetylene is also a safety concern as it is highly explosive and requires great care in transferring pure acetylene into a cylinder. For these reasons, NIST has stopped including acetylene in its hydrocarbon and organic gas standards, and no longer supports the value assignment of acetylene in gas mixtures.

The G17 (ethylbenzene) and G20 (*o*-xylene) results are difficult to explain since good agreement has been achieved in CCQM-K10 and other comparisons [60,68], including one comparison at pmol/mol levels [61]. The result for G19 (*m*+*p*-xylene) is also difficult to explain. Depending on what analytical gas chromatographic (GC) column is used, one may or may not be able to baseline resolve these two compounds in a GC analysis. NIST uses a column that fully separates these compounds and as such value assigns each component. Most all the other participants in this comparison used columns that produced only one peak containing both xylenes. Whether this creates a bias has not been ascertained.

7.4. Performance in Ozone Photometer Comparisons

NIST also participates in BIPM-conducted comparisons of ozone (O_3) measurements at ambient levels in purified air. These are bilateral comparisons of reference photometers (mostly NIST-produced standard reference photometers (SRPs)) with the NIST-produced SRP27 maintained by and at the BIPM as a common reference. While comparisons are made at numerous O_3 mole fraction levels, the only results included in the CCQM-P28 PPS report and the series of continuous BIPM.QM-K1 reports are at (nominal) levels of (80 and 420) nmol/mol.

The measurements for the 23 participants in CCQM-P28 were carried out from 2003 to 2005 [69]. NIST's SRP2 has been evaluated seven times (2007, 2009, 2013, 2015, 2017, 2019, and 2022) in BIPM.QM-K1 [70]. The median absolute bias of these results is 0.37. All of the results are within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})$/Median$\{u(x_i)\}$ is 0.85.

The $|\zeta_{NIST}|$ values for these datasets are displayed in Fig. 57 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.0031|/0.0035 \approx 0$, suggest that the absolute bias has not changed over time.

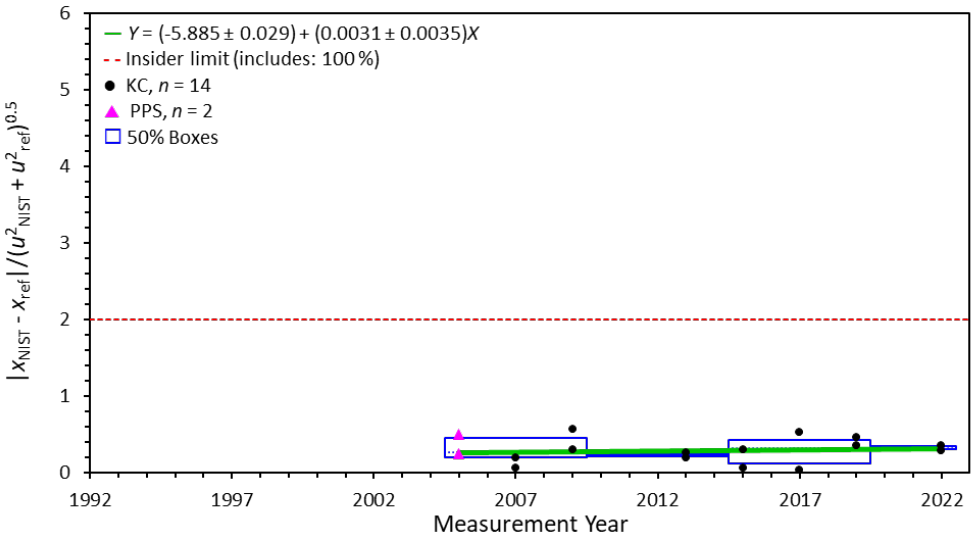


Fig. 57. GAWG Ozone Studies: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset: solid black circle denote BIPM.QM-K1 Key Comparison (KC) results, solid magenta triangle denote CCQM-P28 published pilot study (PPS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit," and the number of KC and PPS results displayed.

The $u(x_{NIST})$/Median$\{u(x_i)\}$ values are displayed in Fig. 58 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.0005|/0.0016 \approx 0$, suggest that the uncertainty ratios did not change over time.

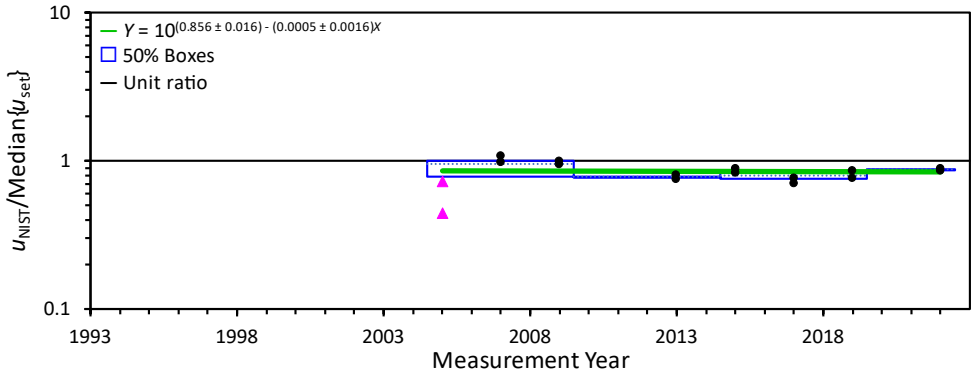


Fig. 58. GAWG Ozone Studies: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})$/Median$\{u(x_i)\}$, in one dataset as a function of measurement year: solid black circles denote BIQM-K001 Key Comparison (KC) results, solid magenta triangles denote CCQM-Q028 published pilot study (PPS) results. Each blue box encloses the central 50 % of the values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation.

7.4.1. GAWG (Ozone) Lessons Learned

NIST (then the National Bureau of Standards, NBS) began producing SRPs for US laboratories in the early 1980s. The SRPs provided and continue to provide metrological traceability for the US Environmental Protection Agency's ozone measurement programs [71]. NIST started producing SRPs for laboratories in other countries in the late 1980s and thus became a global provider of ozone measurement traceability. The World Meteorological Organization's Global Atmospheric Watch Program (WMO/GAW) [72] program still lists NIST as the Central Calibration Laboratory (CCL) for ozone traceability.

NIST was not staffed for handling global demand. By 2001 26 NIST SRPs had been built and disseminated within the U.S., Canada, Europe, and Australia, and all SRP users regularly asked NIST to provide measurement traceability for their SRPs via side-by-side comparisons to NIST's master SRP at NIST or to NIST's traveling SRP in their laboratories. Discussions with one of the European SRP users as to whether they would serve as a European reference point for SRP ozone traceability were abandoned due to intra-European national concerns. However, shortly thereafter when BIPM staff visited NIST to discuss what their new Chemistry Section should work on, ozone measurement traceability was firmly suggested, and it became the BIPM Chemistry Section's first project. This proved to be a perfect fit and the demands on NIST's resources were greatly reduced.

After purchasing two NIST SRPs, the BIPM did a complete study and determined that the design of the NIST SRPs had two clear sources of bias [73]. NIST then developed an upgrade kit and quickly began modifying all existing SRPs to minimize these biases [74]. This work helped improve the overall agreement of SRPs from around 0.7 % to 0.3 %, which has been maintained with all new SRP built since then.

The ongoing BIPM.QM-K1 ozone key comparison has demonstrated that the NIST SRPs and other instruments maintained by NMIs are very stable and in close agreement to the SRPs maintained at the BIPM and at NIST. The original participation interval requirement was every two years. The repeated close agreements enabled loosening the requirement first to every four years and now to every (four to eight) years. However, NIST and the BIPM intend to continue to compare their SRPs every two years to better maintain continued confidence in the global ozone reference system.

8. Electrochemical Analysis Working Group (EAWG)

The Electrochemical Analysis Working Group (EAWG) is one of the four WGs created by the CCQM during the 1997 to 1998 period [1]. The EAWG is responsible for "pH, electrolytic conductivity, and coulometry" [75]. The EAWG shares responsibility with the IAWG for measurements of inorganic purity; however, for purposes of retrospective analysis all results expressed as mass fractions from these jointly conducted studies are attributed to the IAWG.

8.1. Engagements in pH-Related Studies

Table 10 lists in chronological order the pH-related studies in which NIST participated that were finalized by this report's publication date. Comparisons CCQM-Q037.1 and CCQM-P037.2 addressed qualitative issues and did not produce quantitative results.

Table 10. NIST Engagement in EAWG Studies of pH.

Measurements in all quantitative studies reported in units of pH.

Study	Name (EAWG)	Sets [a]	Year	Coordinators
CCQM-K009	pH of phosphate buffer	13	2000	PTB
CCQM-K017	pH of phthalate buffer	3	2001	PTB
CCQM-P037	pH: Fundamental studies of standards	1	2002	SMU
CCQM-K019	pH of borate buffer	3	2005	PTB
CCQM-K018	pH of carbonate buffer	1	2006	SMU
CCQM-K020	pH of tetroxalate buffer	3	2008	NIST
CCQM-K019.1	pH of borate buffer	3	2010	PTB
CCQM-Q037.1	pH: Problems in primary measurement	0 [b]	2010	PTB,NPL
CCQM-K091	pH of an unknown phthalate buffer	5	2011	PTB
CCQM-P037.2	pH: Preparation of Ag|AgCl electrodes	0 [b]	2012	NPL
CCQM-K099	pH of phosphate buffer	5	2014	PTB
CCQM-K018.2016	pH of carbonate buffer	1	2017	NIST
CCQM-P093	pH: Preparation study at pH 7	1	2018	SMU

a The number of datasets derived from the study that contain a NIST result.
b Study did not generate quantitative results suitable for retrospective analysis.

The cumulative distributions in Fig. 59 display the time course of NIST's engagements with the EAWG's finalized pH-related studies, where "engagements" applies to participations and coordinations.

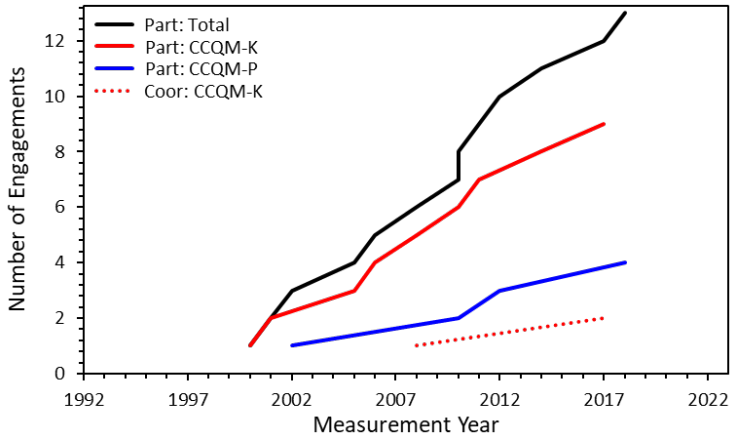


Fig. 59. EAWG pH-Related Studies: Cumulative Distributions of NIST Engagements.

The histogram traces of Fig. 60 display the engagement time course in higher resolution, along with the average number of participations and coordinations from the first to the most recent engagement.

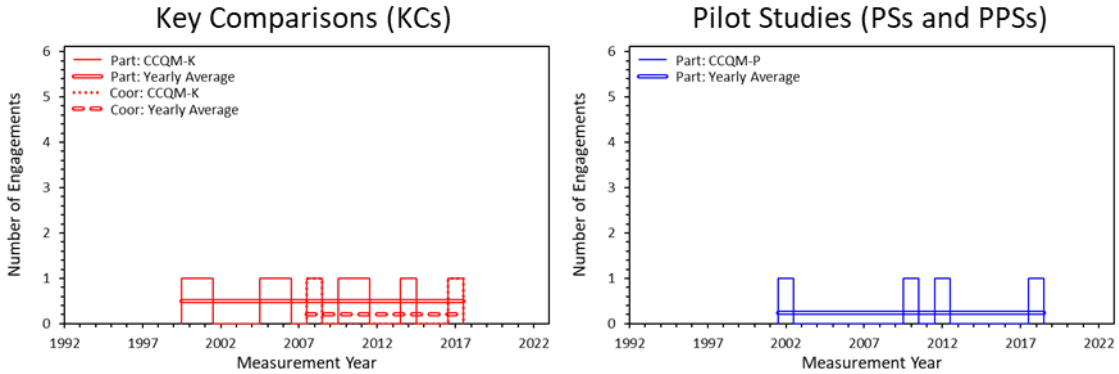


Fig. 60. EAWG pH-Related Studies: Histogram Traces of NIST Engagements.

8.1.1. Performance in pH-Related Studies

NIST participated in nine pH-related Key Comparisons (KCs), one published pilot study (PPS), and three confidential pilot studies (PSs) that were finalized by this report's publication date. NIST has results in 39 datasets generated by these studies (37 from KCs and two from PSs). The relationship between the absolute zeta-score bias of these results, $|\zeta_{NIST}|$, and the standard uncertainty of the reported values relative to the median of the participant uncertainties, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, is visualized in Fig. 61.

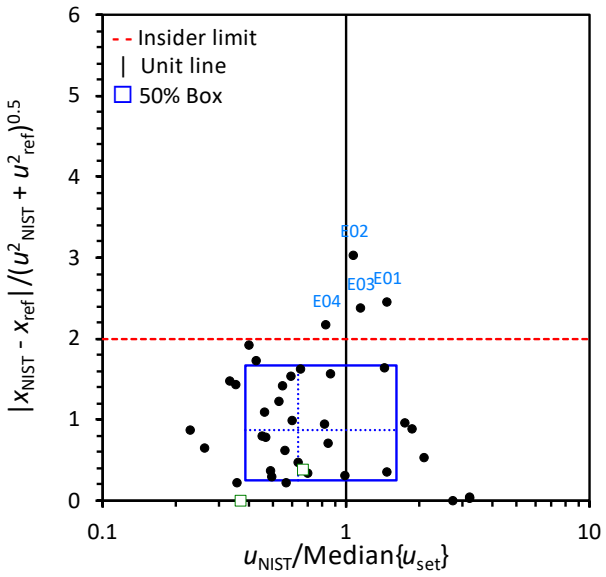


Fig. 61. EAWG pH-Related Studies: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})$/Median{$u(x_i)$}: solid black circles denote Key Comparisons (KC) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the results; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties. The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$).

The median absolute bias is 0.87. More than 89 % of the absolute biases are within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})$/Median{$u(x_i)$} is 0.64. The $u(x_{NIST})$ is less than the participant median for only one of the four "outsiders" ($|\zeta_{NIST}| > 2$). The "outsider" datasets are documented in Section 8.1.1.3.

8.1.1.1. Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for the EAWG pH-related datasets are displayed in Fig. 62 as a function of measurement year. The linear trend, $|t| = |+0.0140|/0.0058 \approx 3$, suggests that the absolute bias slightly increased over time. However, the pattern of the "50% boxes" suggests episodic rather than linear changes with time.

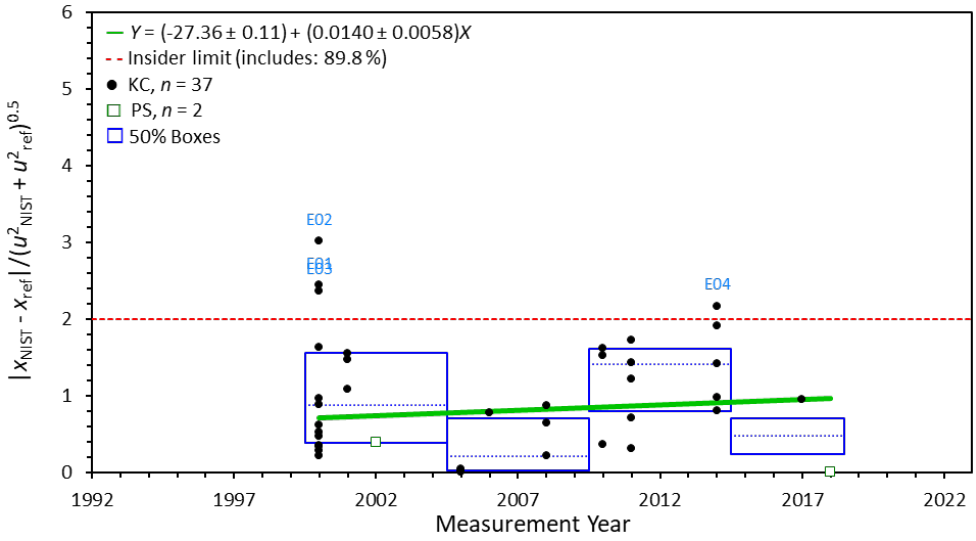


Fig. 62. EAWG pH-Related Studies: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of measurement year: solid black circles denote Key Comparisons (KC) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit," and the number of KC and PS results displayed.

The $u(x_{NIST})$/Median$\{u(x_i)\}$ values are displayed in Fig. 63 as a function of measurement year. The linear trend, $|t| = |+0.0181|/0.0022 \approx 8$, suggests that the uncertainty ratios decreased over time. The pattern of the "50% boxes" suggests that they reached an asymptotic constant proportion of about 0.6 beginning in 2010.

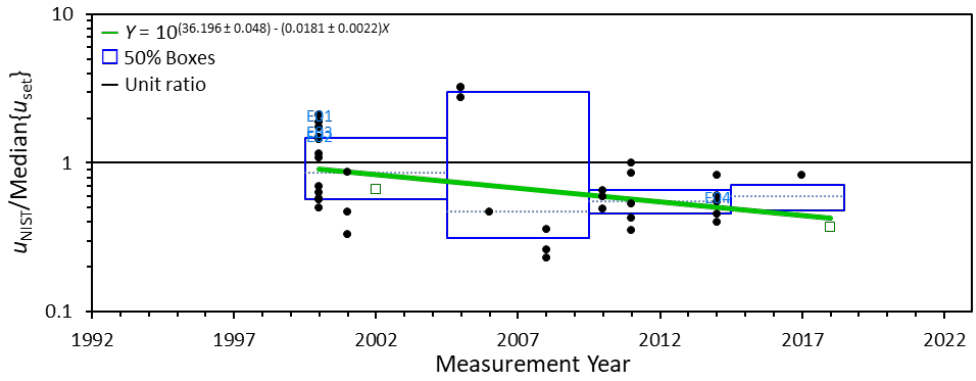


Fig. 63. EAWG pH-Related Studies: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})$/Median$\{u(x_i)\}$, as a function of measurement year: solid black circles denote Key Comparisons (KC) results, open green squares denote confidential pilot study (PS) results. The labels identify datasets which contain "outsider" results ($|\zeta_{NIST}| > 2$). Each blue box encloses the central 50 % of the values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation.

8.1.1.2. Bias and Relative Uncertainty Across Measurand Level

The $|\zeta_{NIST}|$ values for pH datasets are displayed in Fig. 64 as a function of pH. Each of the five buffer systems evaluated (tetroxalate, phthalate, phosphate, borate, and carbonate) is useful as a standard over only a relatively narrow pH interval. While the linear trend, $|t| = |-0.036|/0.020 \approx 2$, suggests that the absolute bias decreases slightly as pH increases, this is likely related more to the experimental conditions used for each of the buffer systems rather than the pH value *per se*.

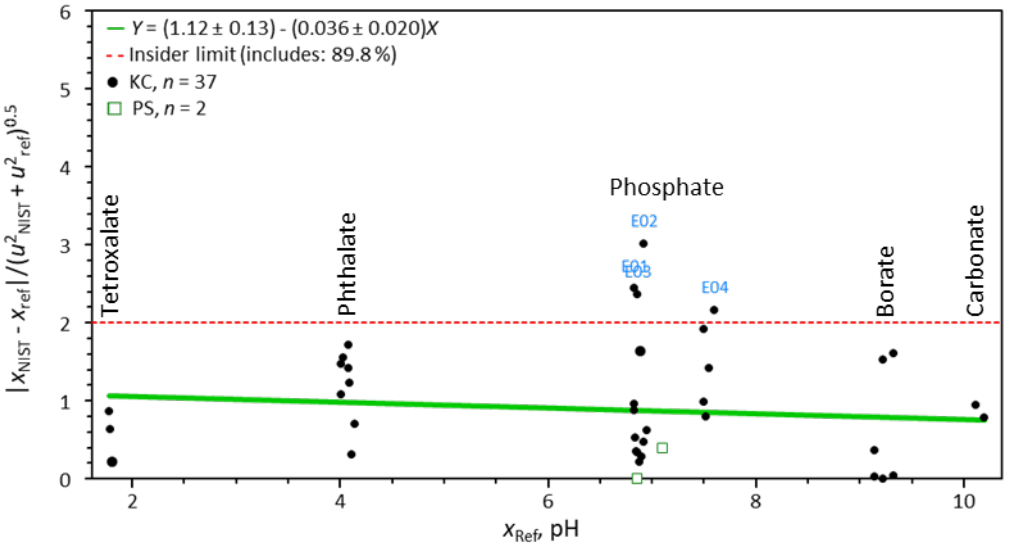


Fig. 64. EAWG pH-Related Studies: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of pH. The text above each cluster of values identifies the five buffer systems evaluated by the EAWG. See Fig. 62 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 65 as a function of pH. The linear trend, $|t| = |+0.0416|/0.0068 \approx 6$ suggests that the uncertainty ratios may increase with increasing pH.

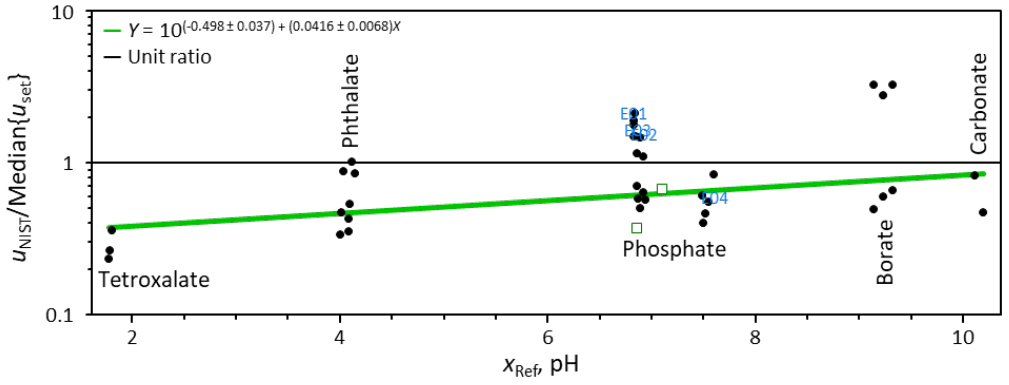


Fig. 65. EAWG pH-Related Studies: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of pH. See Fig. 63 for description of the graphical elements.

8.1.1.3. "Outsider" Datasets

The four pH-related datasets that contain "outsider" NIST results are depicted in Fig. 66 as dot-and-bar charts, each panel showing all results for one dataset, the reference value, and the 95 % level of confidence reference interval. The label in the upper left corner of each panel, running in sequence from "pH1" to "pH4," is the code used to identify the "outsiders" in Fig. 61 to Fig. 65.

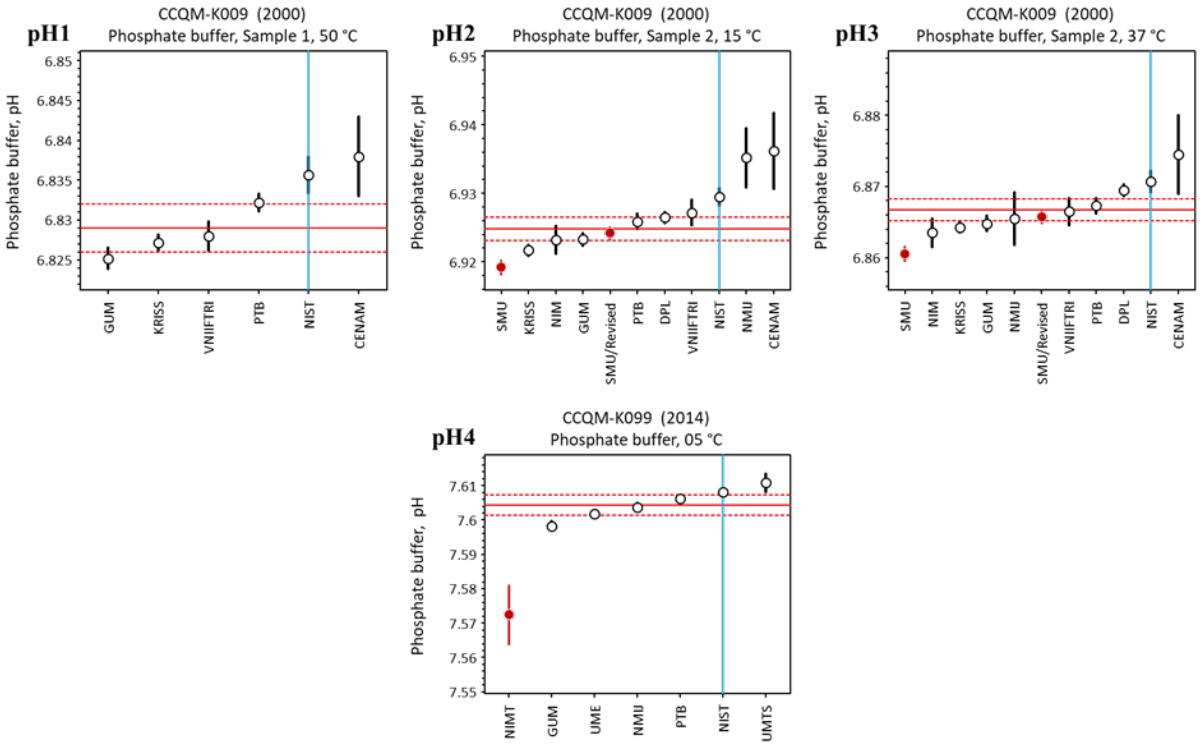


Fig. 66. EAWG pH-Related Studies: "Outsider" Datasets.

Each panel shows all reported results for one dataset with a NIST "outsider" ($|\zeta_{NIST}| > 2$) result. Results are sorted in order of increasing magnitude. Participants are identified by code name; the full names are listed in Appendix B. Open circles denote results considered valid; solid red circles denote results excluded from estimating the reference value. Bars represent one standard uncertainty. The horizontal solid red line denotes the reference value; the dashed red lines bound a 95 % level of confidence interval about the reference value. The blue vertical lines mark NIST results.

8.2. Engagement with Electrolytic Conductivity Studies

Table 11 lists the electrolytic conductivity studies in which NIST participated.

Table 11. NIST Engagement in EAWG Studies of Conductivity.

Measurements in all studies reported in units of Siemens per meter, S/m.

Study	Name (EAWG)	Sets [a]	Year	Coordinators
CCQM-P022	Electrolytic conductivity (0.1 S/m)	2	2001	CCQM-P022
CCQM-P047	Electrolytic conductivity (5 mS/m and 50 mS/m)	2	2004	CCQM-P047
CCQM-K036.a	Electrolytic conductivity (0.5 S/m)	1	2005	CCQM-K036.a
CCQM-K036.b	Electrolytic conductivity (5 mS/m)	1	2005	CCQM-K036.b
CCQM-P083	Electrolytic conductivity (0.5 mS/m)	2	2009	CCQM-P083
CCQM-K092	Electrolytic conductivity (0.05 S/m and 20 S/m)	1	2011	CCQM-K092
CCQM-P143	Electrolytic conductivity: Preparative comparison	1	2014	CCQM-P143

a The number of datasets derived from the study that contain a NIST result.

The cumulative distributions in Fig. 67 display the time course of NIST's engagements with the EAWG's conductivity studies, where "engagements" applies to participations and coordinations.

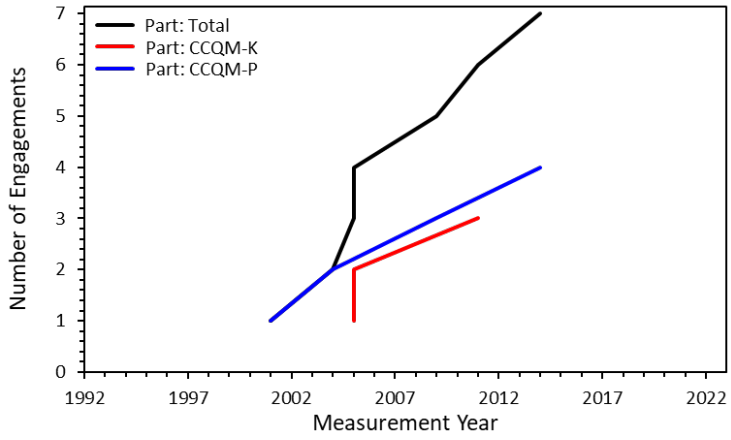


Fig. 67. EAWG Conductivity Studies: Cumulative Distributions of NIST Engagements.

The histogram traces of Fig. 68 display the engagement time course in higher resolution, along with the average number of participations and coordinations from the first to the most recent engagement.

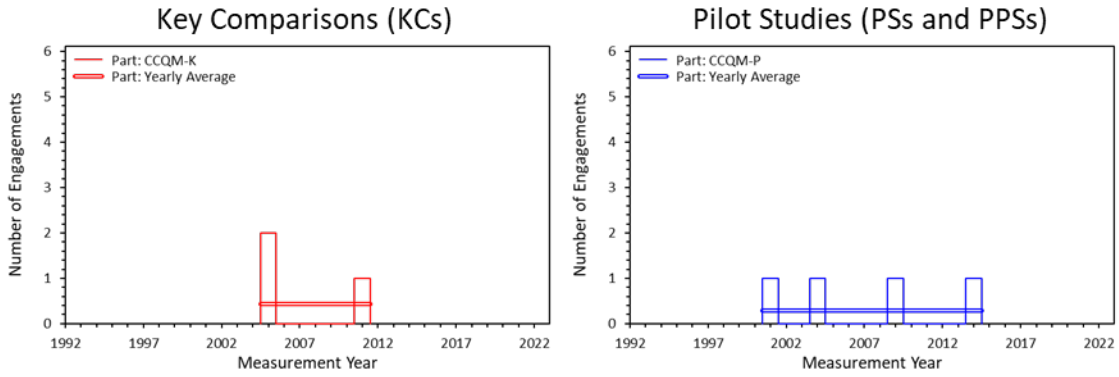


Fig. 68. EAWG Conductivity Studies: Histogram Traces of NIST Engagements.

8.2.1. Performance in Electrolytic Conductivity Studies

NIST participated in seven conductivity studies (three KCs and four PS) between 2001 and 2014. NIST has results in 10 datasets generated by these studies (three from KCs and seven from PSs). The NIST results in all the pre-2014 CCQM conductivity studies were traceable to calibration of a working cell using a KCl solution of specified molality prepared at the time of the given study. The conductivity of this formulation had been assigned previously [76] in a primary cell at NIST using separate batches of nominally identical composition prepared at the time of these primary measurements. The uncertainty of the comparison measurements thus relied on the accuracy of the subsequent preparation of the standard formulation for the given CCQM study. Study CCQM-P143 in 2014 was performed to validate this "recipe-based" approach. Solutions were prepared by participants and shipped to the coordinating NMI, which measured them there using a primary cell.

The median absolute bias the conductivity datasets is 0.90. 80 % of the results were within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. Two of the PS datasets contain "outsider" ($|\zeta_{NIST}| > 2$) NIST results. The median $u(x_{NIST})/\text{Median}\{u(x_i)\}$ was 0.74. The relationship between the absolute zeta-score bias of these results, $|\zeta_{NIST}|$, and the standard uncertainty of the reported values relative to the median of the participant uncertainties, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, is visualized in Fig. 69.

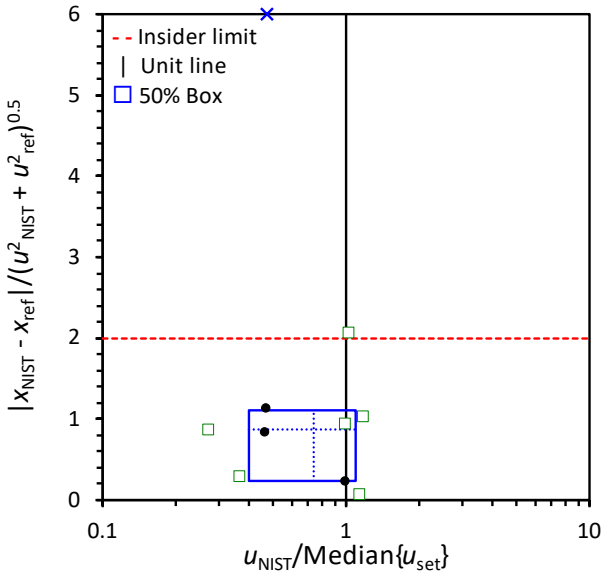


Fig. 69. EAWG Conductivity Studies: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$: solid black circles denote Key Comparisons (KC) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties.

8.2.1.1. Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for the EAWG conductivity datasets are displayed in Fig. 70 as a function of measurement year. While the linear trend, $|t| = |-0.046|/0.041 \approx 1$, is not statistically significant, the pattern of the "50% boxes" suggests that it decreased from an early average of about 1.0 to an asymptotic average of about 0.5 after 2005.

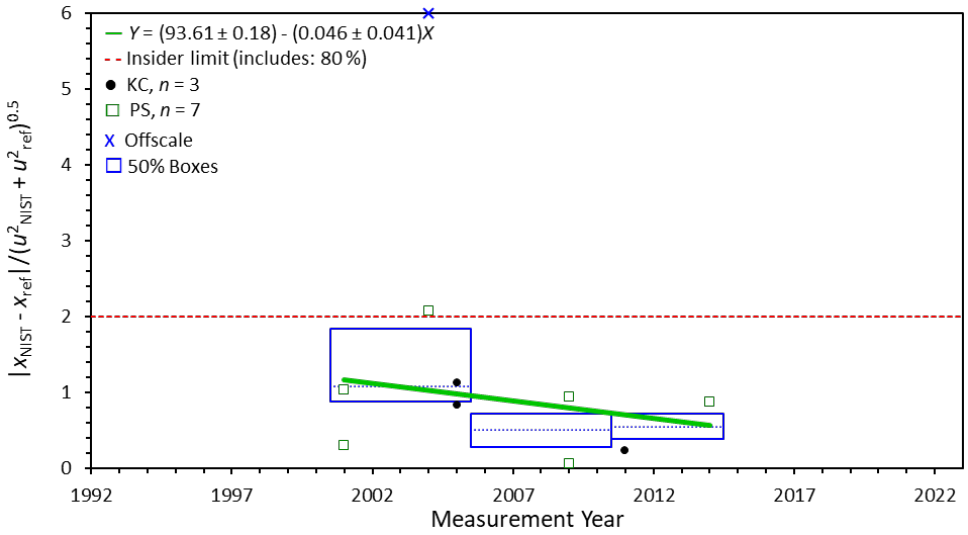


Fig. 70. EAWG Conductivity Studies: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of measurement year: solid black circles denote Key Comparisons (KC) results, open green squares denote confidential pilot study (PS) results. The "×" denotes a value larger than the chart's Y-axis maximum. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). Each blue box encloses the central 50 % of values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit", and the number of KC and PS results displayed.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 71 as a function of measurement year. The pattern of the "50% boxes" and the linear trend, $|t| = |-0.003|/0.011 \approx 0$, suggests that uncertainty ratios did not change over time.

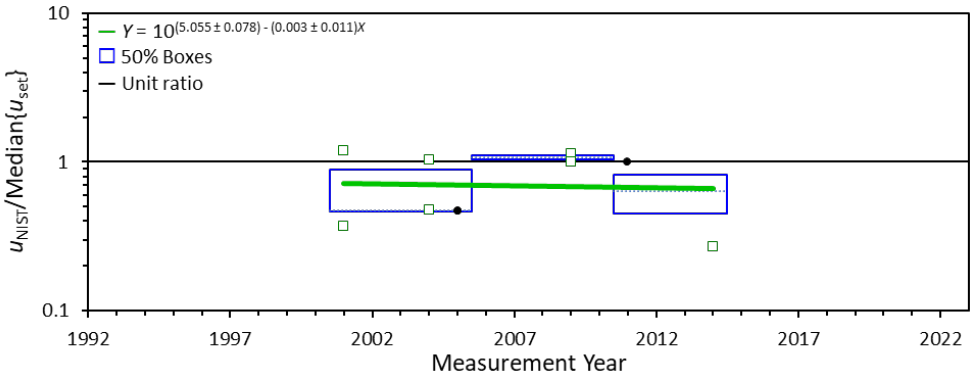


Fig. 71. EAWG Conductivity Studies: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of measurement year: solid black circles denote Key Comparisons (KC) results, open green squares denote confidential pilot study (PS) results. Each blue box encloses the central 50 % of the values within a sequential five-year interval. The thick green line represents the estimated linear trend. The legend provides the trend equation.

8.2.1.2. Bias and Relative Uncertainty Across Measurand Level

The $|\zeta_{NIST}|$ values for the EAWG electrolytic conductivity datasets are displayed in Fig. 72 as a function of the electrolytic conductivity in S/m. The pattern of the "50% boxes" and the linear trend, $|t| = |+0.10|/0.15 \approx 1$ suggest that that the absolute bias is independent of conductivity level.

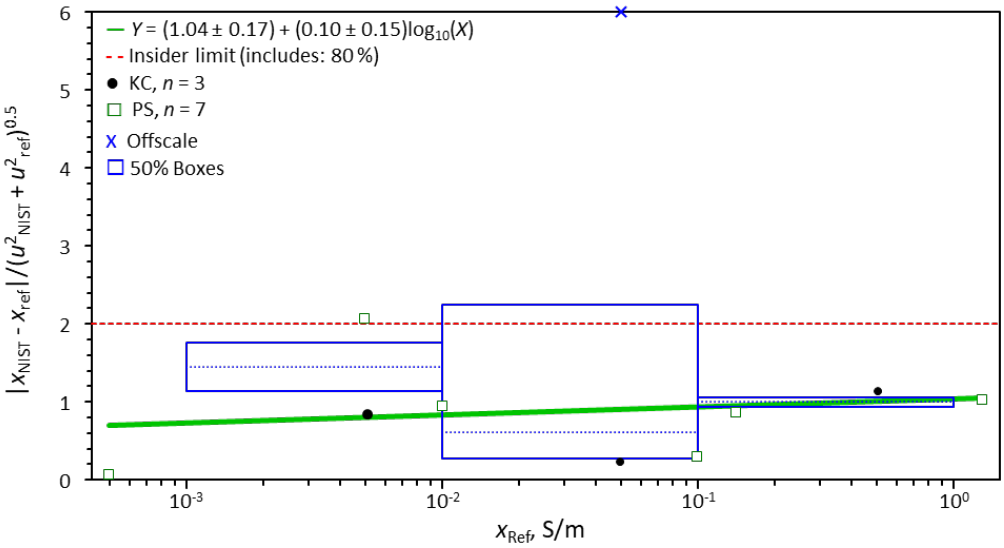


Fig. 72. EAWG Conductivity Studies: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the electrolytic conductivity. See Fig. 70 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 73 as a function of S/m. While the linear trend, $|t| = |0.043|/0.047 \approx 1$, suggests that the uncertainty ratios are independent of conductivity level, the pattern of the "50% boxes" suggests they may be somewhat smaller at higher conductivities.

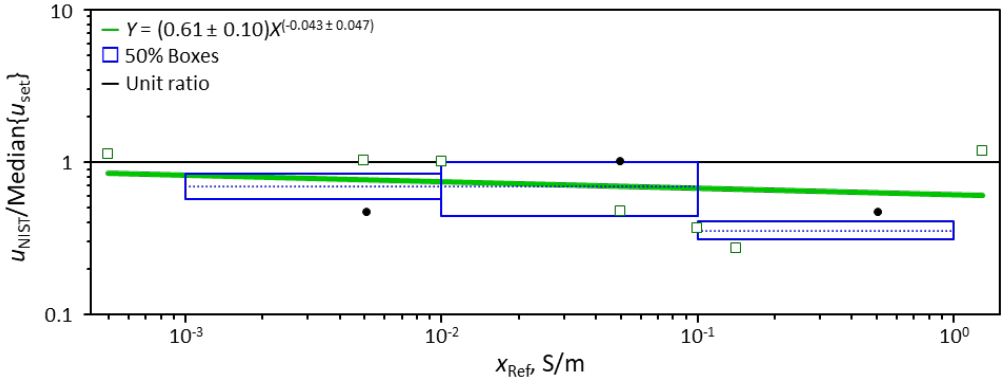


Fig. 73. EAWG Conductivity Studies: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of electrolytic conductivity. See Fig. 71 for description of the graphical elements.

8.3. EAWG Lessons Learned

The early pH KCs and pilot studies directly led to several improvements in the primary measurement technique as realized at NIST [77] and used for pH KC comparisons and certifications of pH SRMs. These improvements were implemented in 2006 and led to significantly reduced uncertainties for NIST in subsequent KCs (see Fig. 63) and pH SRM certifications.

Problems with the integrity of bottled solutions in shipment, particularly in international air shipments subject to extended periods of reduced pressure, first emerged in 2000 in the first pH KC, CCQM-K9 [78]. From this experience, NIST developed a protocol to verify the integrity of shipped solutions by participants in subsequent KCs. Versions of this NIST protocol have been used by all coordinating laboratories for all pH and conductivity KCs and pilot studies starting from about 2005 (e.g., see the Sample Delivery and Verification of Mass Stability of Shipped Bottles" section for CCQM-K18.2016 [79]).

Even though the "recipe-based" calibration used in the earlier conductivity studies was validated in CCQM-P143, a more direct metrological approach is preferred today. In this approach, the conductivity of a calibrant solution of nominal composition is value-assigned in a primary cell. This same solution is then directly used as the calibrant for the working cell. The direct approach eliminates the uncertainty associated with replicate preparation of a specified formulation of calibrant solution and provides more robust traceability to the SI.

NIST ceased participating in the EAWG's conductivity studies after terminating the SRM 3190 to 3199 "Aqueous Electrolytic Conductivity" series in 2013. The main factor in this decision to terminate the service was large bottle-to-bottle variations in the long-term (months) stability of the conductivity SRMs on storage, relative to the small inherent measurement uncertainty assessed in the conductivity KCs. The KCs did not address this stability-related uncertainty.

9. Surface Analysis Working Group (SAWG)

The Surface Analysis Working Group (SAWG) originated as an *ad hoc* WG in 2000 and became fully fledged in 2002 [1,80]. The SAWG has responsibility for "spatially resolved chemical surface analysis at the micro and nanoscale" [81].

9.1. Engagements in SAWG Studies

NIST participated in seven SAWG-conducted Key Comparisons (KCs), published pilot studies (PPSs), or confidential pilot studies (PSs) that were finalized by this report's publication date. NIST has results in 46 datasets generated by these studies. NIST has not participated in any SAWG-related Supplementary Comparisons (SCs).

Table 12 lists in chronological order the finalized SAWG studies in which NIST participated. The CCQM-Q038 method comparison study results are complex measurands (slopes, intercepts, and RMSE) not suited for meaningful inclusion as *CCQM_Retrospectoscope* datasets. Results from CCQM-Q140 were published without attribution.

Table 12. NIST Engagement in SAWG Studies.

Study	Name (SAWG)	Units	Sets [a]	Year	Coordinators
CCQM-Q038	Thickness of silicon dioxide on silicon film	nm	0 [b]	2002	NPL
CCQM-K032	Thickness of silicon dioxide on silicon film	nm	4	2005	NPL
CCQM-P130	WD and ED electron probe microanalysis on Au-Cu	k-values	16	2013	NIST,BAM
CCQM-Q140	CIGS layer composition	nm	0 [b]	2013	KRISS
CCQM-K129	Cu-In-Ga (di)Selenide (CIGS) thin layer composition	mol/mol	4	2014	KRISS
CCQM-Q190	Thickness of hafnium oxide film	nm	18	2017	KRISS
CCQM-K157	Thickness of hafnium oxide film	nm	4	2021	KRISS

a The number of datasets derived from the study that contain a NIST result.
b No attributed quantitative results.

The cumulative distributions in Fig. 74 display the time course of NIST's engagements with finalized SAWG studies, where "engagements" applies to participations and coordinations.

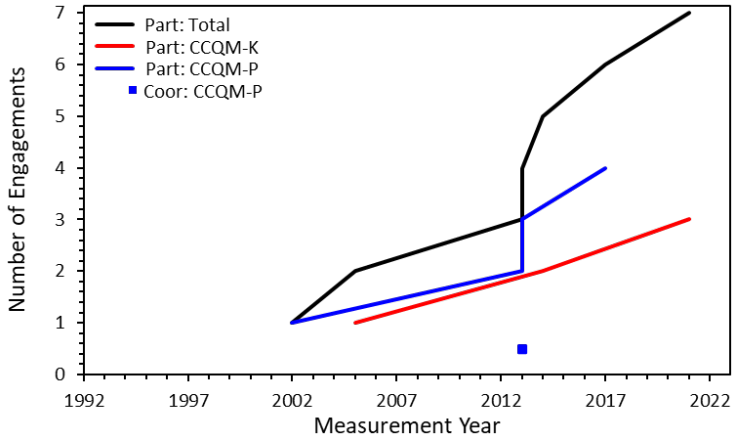


Fig. 74. SAWG Studies: Cumulative Distributions of NIST Engagements.

The histogram traces of Fig. 75 display the engagement time course in higher resolution, along with the average number of participations and coordinations from the first to the most recent engagement.

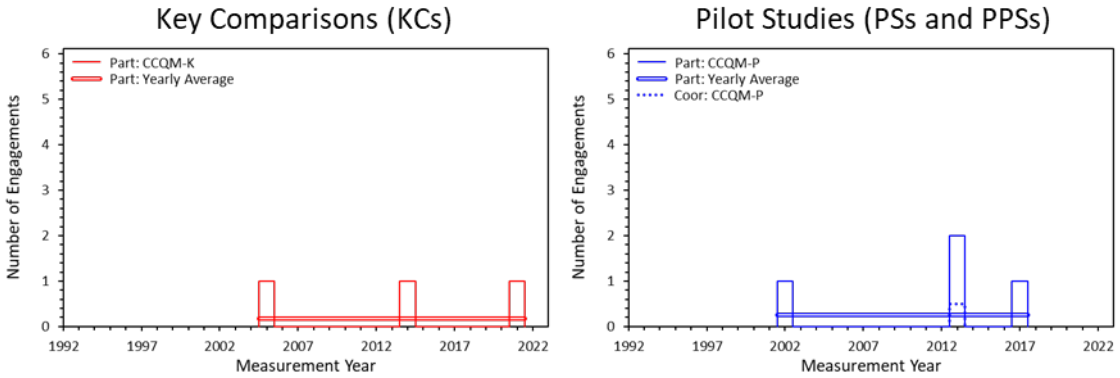


Fig. 75. SAWG Studies: Histogram Traces of NIST Engagements.

9.2. Performance in SAWG Studies

The relationship between the absolute zeta-score bias of these results, $|\zeta_{NIST}|$, and the standard uncertainty of the reported values relative to the median of the participant uncertainties, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, is visualized in Fig. 76.

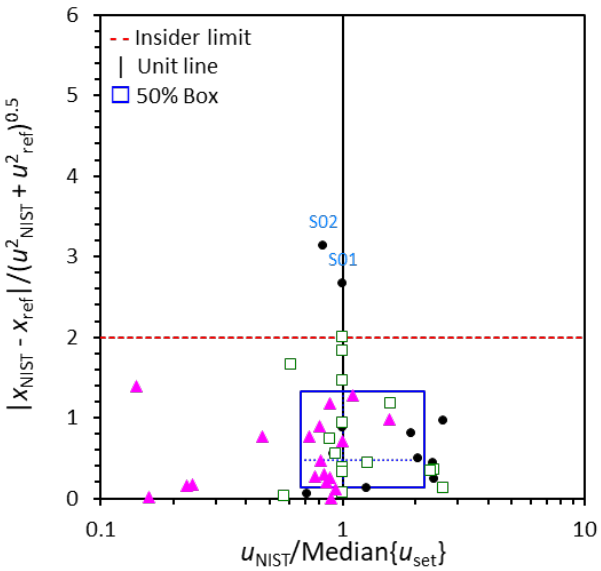


Fig. 76. SAWG Studies: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$: solid black circles denote Key Comparisons (KC) results, solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties. The labels identify "outsider" KC datasets ($|\zeta_{NIST}| > 2$).

The median absolute bias in the SAWG datasets is 0.48. More than 93 % of the absolute biases are within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})$/Median$\{u(x_i)\}$ is 1.00. The two "outsider" ($|\zeta_{NIST}| > 2$) KC datasets are documented in Section 9.2.4.

9.2.1. Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for the SAWG datasets are displayed in Fig. 77 as a function of measurement year. While the data are too few and measurands too diverse for confident analysis, the linear trend, $|t| = |-0.0234|/0.0052 \approx 4$, suggests that on average the absolute bias has declined over time.

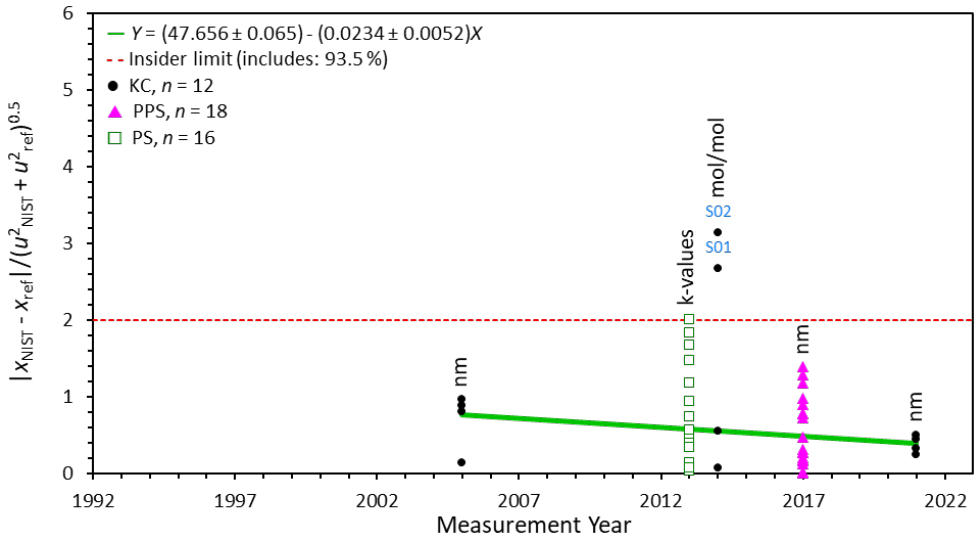


Fig. 77. SAWG Studies: Absolute Bias Over Time.

Each symbol represents the absolute bias of NIST's result in one dataset as a function of measurement year: solid black circles denote Key Comparisons (KC) results, solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$). The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit", and the number of KC and PS results displayed. The text above each cluster of values identifies the reporting units.

The $u(x_{NIST})$/Median$\{u(x_i)\}$ values are displayed in Fig. 78 as a function of measurement year. While the data are too few and measurands too diverse for confident analysis, the linear trend, $|t| = |-0.0129|/0.0027 \approx 5$, suggests that on average the uncertainty ratios have decreased over time.

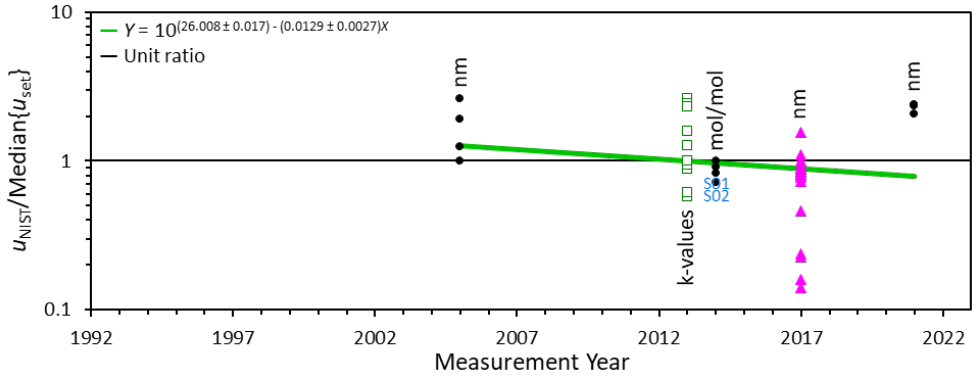


Fig. 78. SAWG Studies: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of measurement year: solid black circles denote Key Comparisons (KC) results, solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$). The thick green line represents the estimated linear trend. The legend provides the trend equation.

9.2.2. Bias and Relative Uncertainty Across Film Thickness

The $|\zeta_{NIST}|$ values for the SAWG film thickness datasets are displayed in Fig. 79 as a function of measurand level. The linear trend, $|t| = |-0.34|/0.027 \approx 1$, suggests that the absolute bias is independent of film thickness.

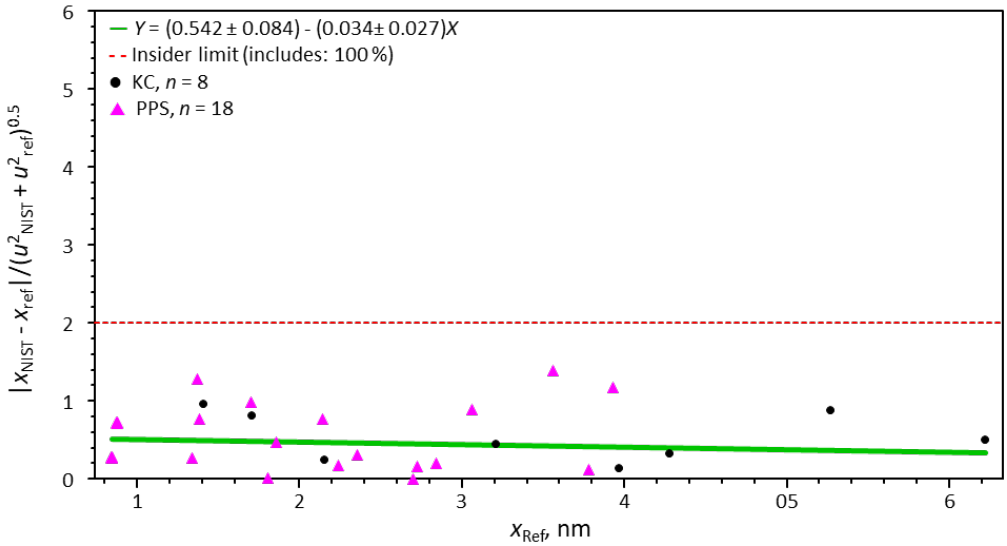


Fig. 79. SAWG Film Thickness: Absolute Bias Across level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of film thickness. See Fig. 77 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 80 as a function of measurand level. The linear trend, $|t| = |0.023|/0.021 \approx 1$, suggests that the uncertainty ratios are independent of film thickness.

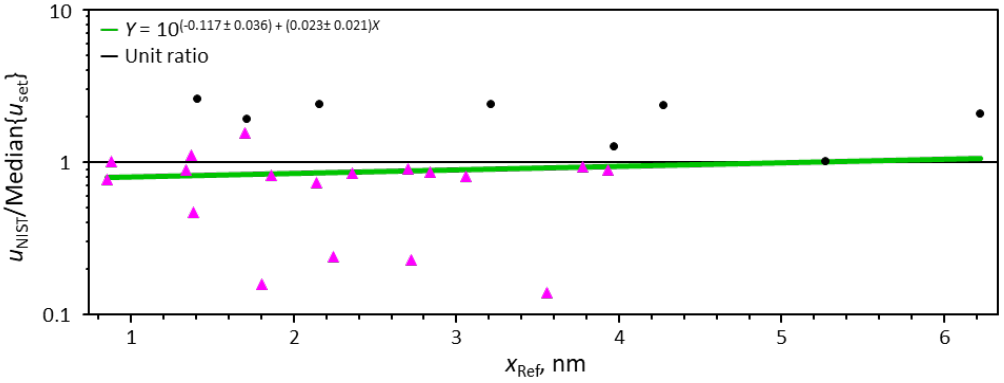


Fig. 80. SAWG Film Thickness: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of film thickness. See Fig. 78 for description of the graphical elements.

9.2.3. Bias and Relative Uncertainty Over k-Value

In quantitative electron probe microanalysis (EMPA), k-values are the ratio of intensities between the unknown sample composition and a pure standard [82]. In CCQM-P130, k-values were the determined measurand for gold and copper in four binary gold-copper alloys [83] at two probe energies [84].

The $|\zeta_{NIST}|$ values for the k-value datasets are displayed in Fig. 81 as a function of measurand level. The linear trend, $|t| = |-0.13|/0.35 \approx 0$, suggests that the absolute bias is independent of the k-value.

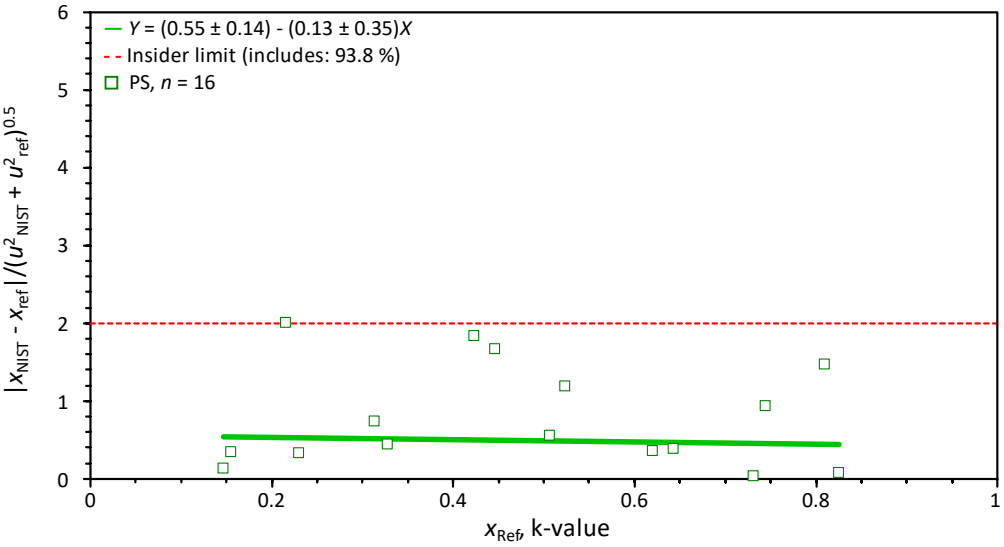


Fig. 81. SAWG k-Value: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of k-value. See Fig. 77 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 82 as a function of measurand level. The linear trend, $|t| = |-0.097|/0.090 \approx 1$, suggests that the uncertainty ratio is independent of the k-value.

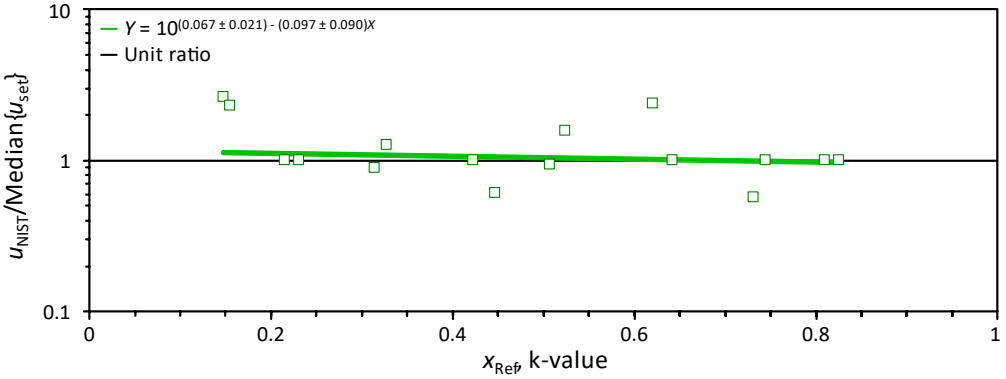


Fig. 82. SAWG k-Value: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of k-value. See Fig. 78 for description of the graphical elements.

9.2.4. Bias and Relative Uncertainty Over Thin Film Composition

CCQM K-129 measured mole fractions of copper, indium, gallium, and selenium in a $Cu(In,Ga)Se_2$ film. The $|\zeta_{NIST}|$ values for these four datasets are displayed in Fig. 83 as a function of measurand level. While the linear trend, $|t| = |-2.7|/3.2 \approx 1$, suggests that on average the absolute bias is independent of mole fraction but the pattern of the available data suggests it may be element-specific.

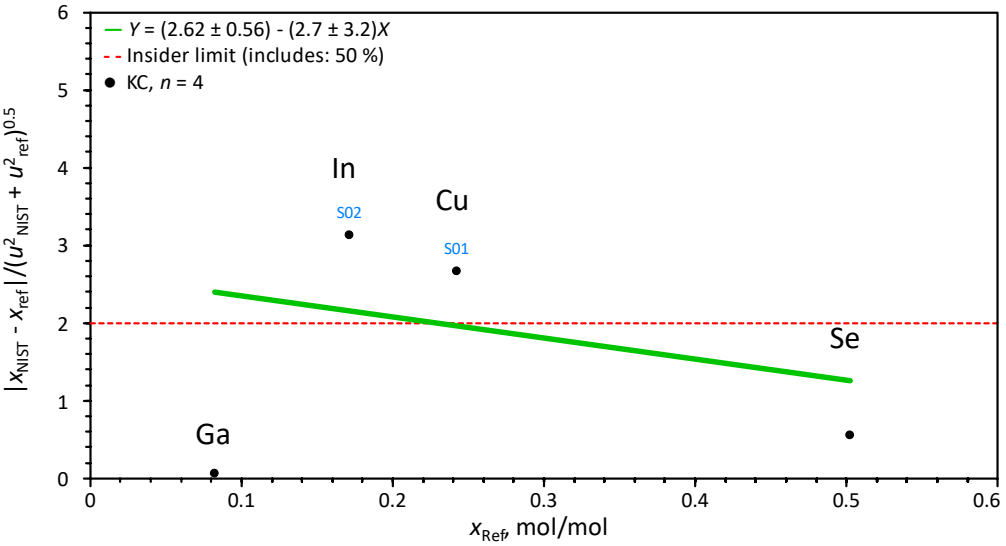


Fig. 83. SAWG Film Composition: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of mole fraction. See Fig. 77 for description of the graphical elements. The text above each cluster of values identifies the analyte.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 84 as a function of measurand level. The linear trend, $|t| = |+0.49|/0.24 \approx 2$, suggests that the uncertainty ratio slightly increases with increasing mole fraction.

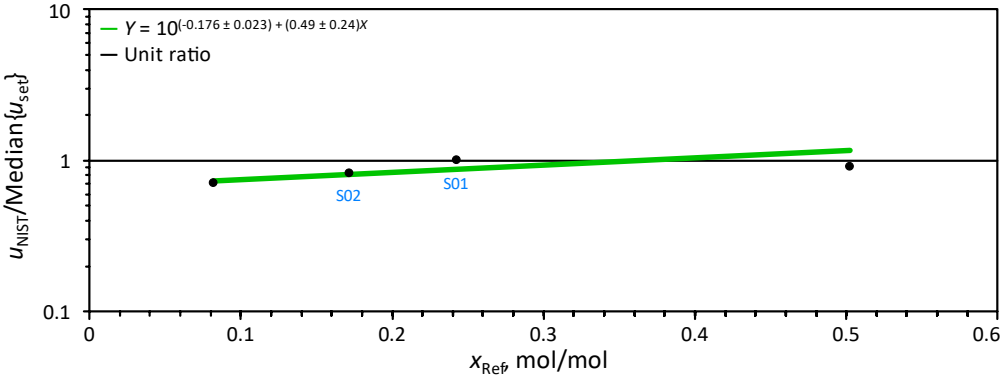


Fig. 84. SAWG Film Composition: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of mole fraction. See Fig. 78 for description of the graphical elements.

9.2.5. "Outsider" Datasets

The two "outsider" SAWG datasets are displayed in Fig. 85. The NIST value for copper is biased high and that for indium is biased low. This may suggest the presence of an analytical interaction between copper and indium in NIST's analytical method. However, this high/low pattern is expected since, in the absence of contaminants, the sum of the four elemental mole fractions is constrained to be 1 mol/mol.

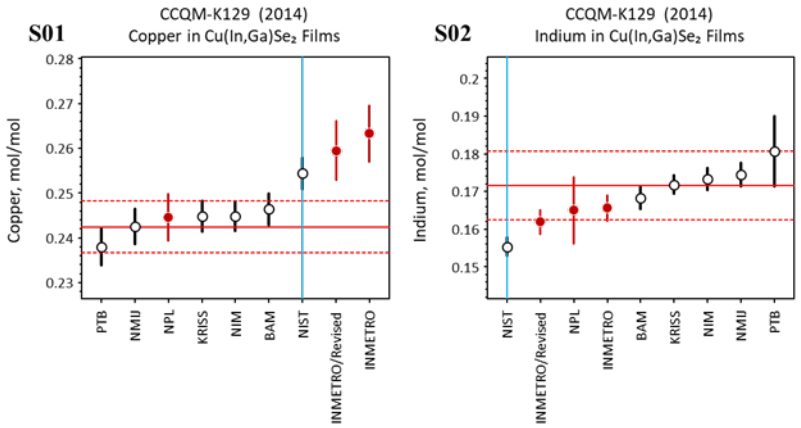


Fig. 85. SAWG Studies: "Outsider" Datasets.

Each panel shows all reported results for one dataset with a NIST "outsider" ($|\zeta_{NIST}| > 2$) result. Results are sorted in order of increasing magnitude. Participants are identified by code name; the full names are listed in Appendix B. Open circles denote results considered valid; solid red circles denote results excluded from estimating the reference value. Bars represent one standard uncertainty. The horizontal solid red line denotes the reference value; the dashed red lines bound a 95 % level of confidence interval about the reference value. The blue vertical lines mark NIST results.

9.3. SAWG Lessons Learned

Understanding the composition and chemistry of surfaces is essential for a broad sector of the US and global economies, with applications in health, energy, and electronics. NIST is often approached with difficult surface chemistry questions. When studying these interfaces between material surfaces and the environment, there often exists competing physical models that can describe the same measurement results. Multiple measurement methods must be compared to arrive at a clearer understanding of surface composition and morphology.

Working with partner NMIs through method comparison pilot studies (PPS and PS) helps NIST determine measurement bias and model interdependencies between different measurement methods which often employ very different measurement probes, such as optical light, X-rays, electrons, and ions, and different lateral length scales. The conclusions drawn from these studies enable us to provide our customers with multiple metrology solutions for determining composition at a surface and the amount of substance for thin (nm-scale) interface layers.

The SAWG KC for amount of substance at a surface has allowed us to find uncertainty limits for specific measurement methods, including: electron microprobe analysis (EMPA), secondary-ion mass spectrometry (SIMS), X-ray reflectivity (XRR) and X-ray photoelectron spectroscopy (XPS). Accurate measurements using these methods are essential for the US semiconductor industry as well as other market sectors.

One limitation illustrated in the SAWGs work-to-date is the relative difficulty inherent in constructing low heterogeneity and environmentally stable test and reference materials for our surface analysis studies. We have had only a limited number of studies over our 20-year effort, with individual studies taking, on average, (3 to 4) years to complete.

A future challenge for the SAWG is developing studies that can cover multiple classes of materials, enabling broad CMC claims across multiple measurement methods and elemental ranges. The currently underway pilot study CCQM-P229 for determining the elemental composition ratio of PtNi thin films is allowing us and our partner NMIs to test methodologies for performing surface analysis on multiple element systems.

SAWG is currently collaborating with the IAWG on nanoparticle concentration studies. This nanomaterial focus ties in well with existing SAWG effort on determining surface area to mass ratios for nano- and micro particles essential to catalysis and gas capture sectors. A potential growth area for SAWG will be investigating measurements of surfaces of nanoparticles for nanotechnology and biological applications, with a focus on surface composition and morphology. Collaborations with other WGs where surface analysis may be essential to understanding results may be another area for growth,

More outputs such as peer reviewed papers from participating NMIs, CMCs, and incorporation of our studies into documentary standards may enhance the impact of NISTs work with SAWG.

10. Protein Analysis Working Group (PAWG)

The Protein Analysis Working Group (PAWG) was one of three WGs created in 2015 from what had been the Bioanalysis Working Group (BAWG) [1]. Several studies initiated by the BAWG addressed protein-related measurands; for continuity these studies are attributed to the PAWG.

10.1. Engagements in PAWG Studies

Table 13 lists the finalized PAWG studies in which NIST participated. The design and textual discussion of results CCQM-Q058, CCQM-Q059, and CCQM-Q059.1were published without attributed quantitative data.

Table 13. NIST Engagement in PAWG Studies.

Study	Name (PAWG)	Units	Sets [a]	Year	Coordinators
CCQM-Q058	Fluorescence in ELISA	µg/g	0 [b]	2005	NPL,NIST
CCQM-Q059	Protein structural measurements by CD (repeat)	a.u.	0 [b]	2006	NPL,NIST
CCQM-P055	Peptide / protein quantification	µg/g	13	2007	LGC
CCQM-Q059.1	Protein structural measurements by CD (repeat)	a.u.	0 [b]	2008	NPL,NIST
CCQM-P055.1	Peptide / protein quantification (repeat)	nmol/g	4	2009	LGC
CCQM-Q058.1	Immunoassay quantitation of human cardiac troponin I	ng/g	3	2011	NPL,NIST
CCQM-K151	Quantification of insulin analog protein solution	µg/g	1	2018	KRISS
CCQM-P201	Total haemoglobin concentration in human whole blood	mg/g	1	2019	PTB

a The number of datasets derived from the study that contain a NIST result.
b No attributed quantitative results available

The cumulative distributions in Fig. 86 display the time course of NIST's engagements with the PAWG's finalized studies, where "engagements" applies to participations and coordinations.

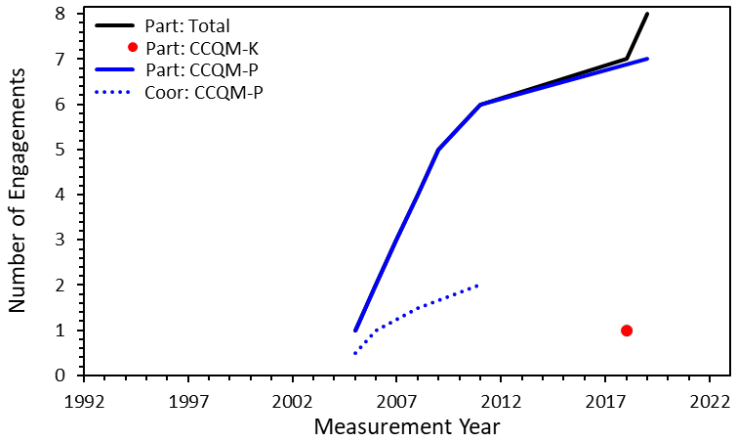


Fig. 86. PAWG Studies: Cumulative Distributions of NIST Engagements.

The histogram traces of Fig. 87 display the engagement time course in higher resolution, along with the average number of participations and coordinations from the first to the most recent engagement.

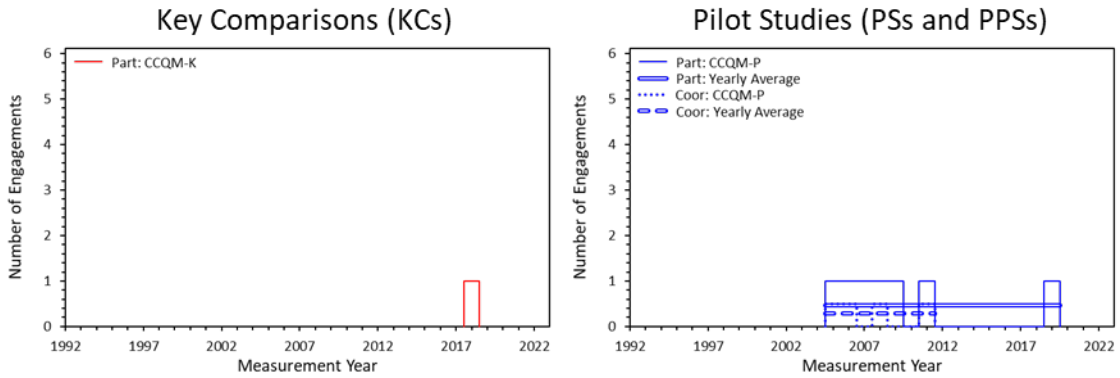


Fig. 87. PAWG Studies: Histogram Traces of NIST Engagements.

10.2. Performance in PAWG Studies

The relationship between the absolute zeta-score bias of the NIST results, $|\zeta_{NIST}|$, and the standard uncertainty of the reported values relative to the median of the participant uncertainties, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, is visualized in Fig. 88.

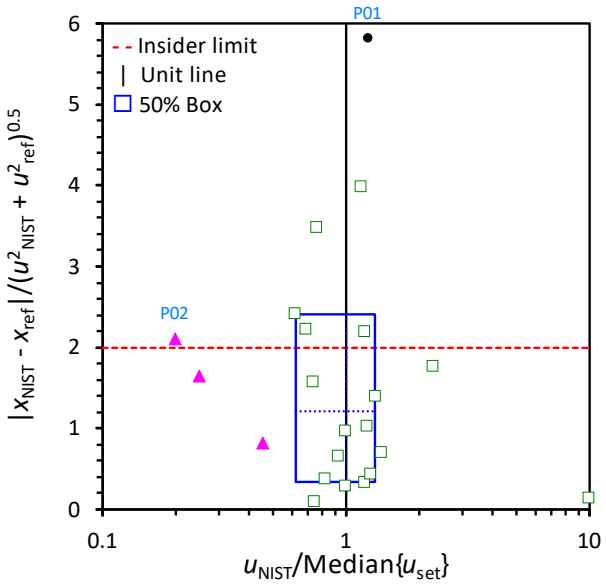


Fig. 88. PAWG Studies: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$: the solid black circle denotes a Key Comparisons (KC) result, solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties. The labels identify "outsider" PPS datasets ($|\zeta_{NIST}| > 2$).

The median absolute bias is 1.2. More than 68 % of the absolute biases are within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})/\text{Median}\{u(x_i)\}$ is 1.0. The $u(x_{NIST})$ is less than the participant median for one of the two "outsiders" ($|\zeta_{NIST}| > 2$). The "outsider" Key Comparison (KC) and published pilot study (PPS) datasets are documented in Section 10.2.3.

10.2.1. Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for the PAWG datasets are displayed in Fig. 89 as a function of measurement year. While based upon too few studies for confident analysis, the linear trend, $|t| = |+0.075|/0.033 \approx 2$, suggests that the absolute bias slightly increased over time.

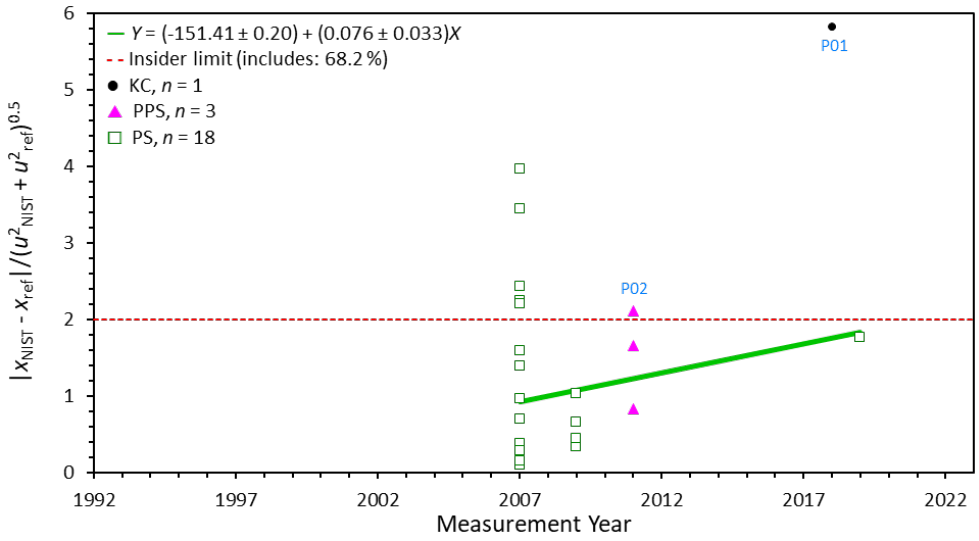


Fig. 89. PAWG Studies: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of measurement year: the solid black circle denotes a Key Comparisons (KC) result, solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The labels identify "outsider" datasets ($|\zeta_{NIST}| > 2$). The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit", and the number of KC, PPS, and PS results displayed.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 90 as a function of measurement year. While the linear trend, $|t| = |+0.001|/0.011 \approx 0$, suggests that the uncertainty ratios have not changed over time, the clustering by study suggests that the ratios are characteristic of the measurand and/or experimental conditions.

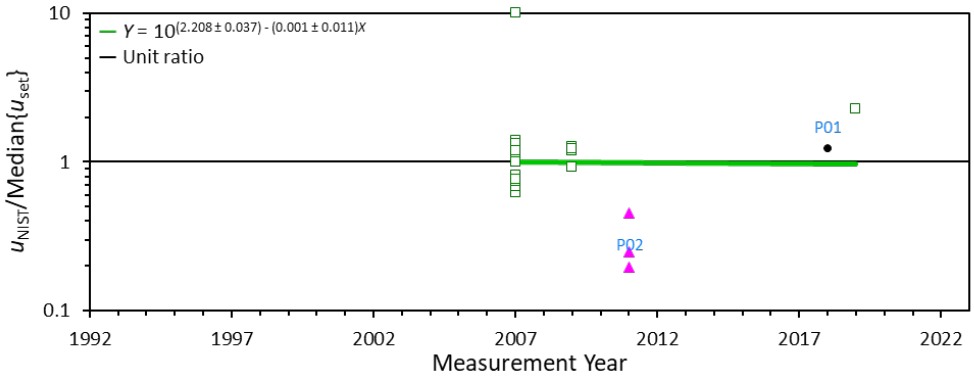


Fig. 90. PAWG Studies: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of measurement year: the solid black circle denotes a Key Comparison (KC) result, solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The thick green line represents the estimated linear trend. The legend provides the trend equation.

10.2.2. Bias and Relative Uncertainty Across Measurand Level

The $|\zeta_{NIST}|$ values for the PAWG datasets are displayed in Fig. 91 as a function of mass fraction. While based upon too few studies for confident analysis, the linear trend, $|t| = |-0.05|/0.13 \approx 0$, suggests that the absolute bias is independent of measurand level.

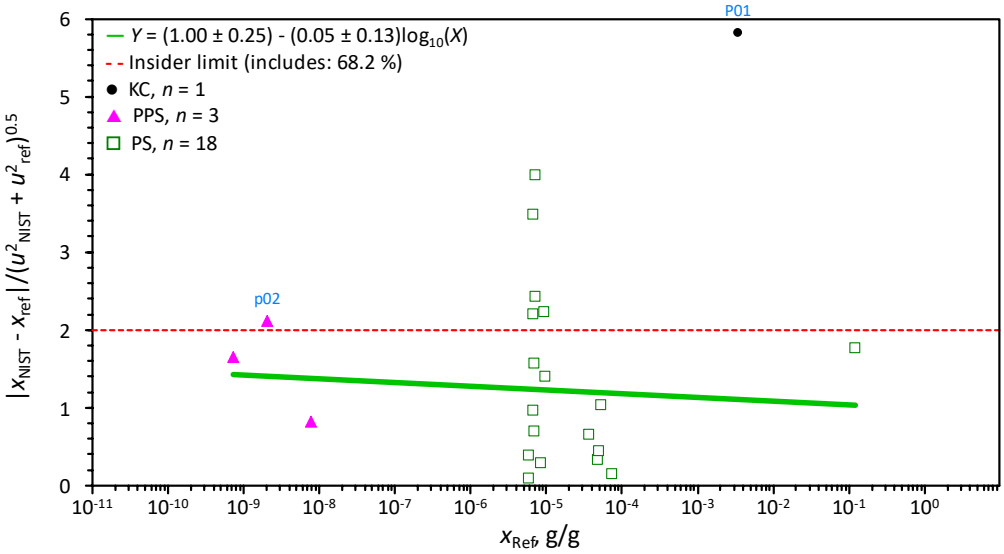


Fig. 91. PAWG Studies: Absolute Bias Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of measurement year. See Fig. 89 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 92 as a function of mass fraction. The linear trend, $|t| = |+0.132|/0.013 \approx 10$, suggests that the uncertainty ratios increase with increasing mass fraction.

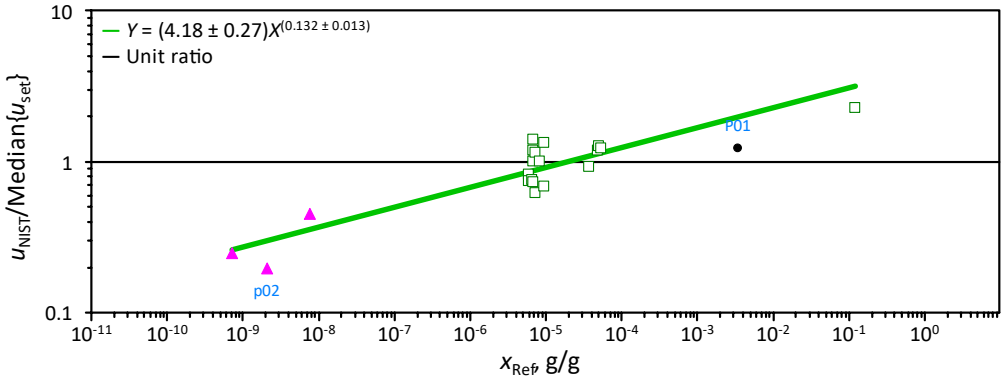


Fig. 92. PAWG Studies: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of mass fraction. See Fig. 90 for description of the graphical elements.

10.2.3. "Outsider" Datasets

The two datasets that contain "outsider" NIST results are depicted in Fig. 93 as dot-and-bar charts, each panel showing all results for one dataset, the reference value, and the 95 % level of confidence reference interval. The label in the upper left corner of each panel, running in sequence from "P01" to "P02", is the code used to identify the "outsiders" in Fig. 88 to Fig. 92.

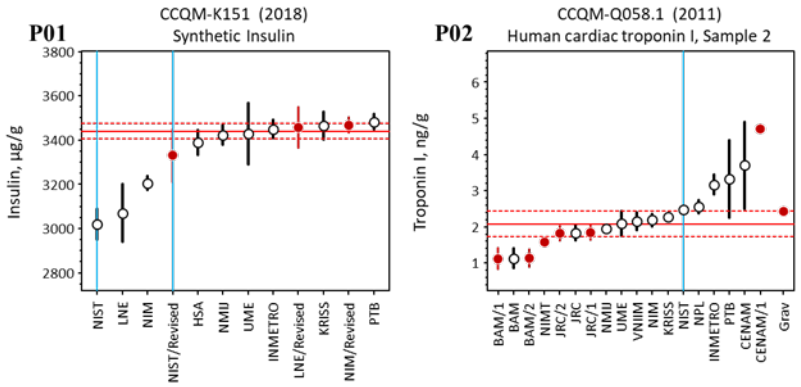


Fig. 93. PAWG Studies: "Outsider" Datasets.

Each panel shows all reported results for one dataset with a NIST "outsider" ($|\zeta_{NIST}| > 2$) result. Results are sorted in order of increasing magnitude. Participants are identified by code name; the full names are listed in Appendix B. Open circles denote results considered valid; solid red circles denote results excluded from estimating the reference value. Bars represent one standard uncertainty. The horizontal solid red line denotes the reference value; the dashed red lines bound a 95 % level of confidence interval about the reference value. The blue vertical lines mark NIST results.

10.3. PAWG Lessons Learned

Before the PAWG emerged from the BAWG, NIST was the driving force behind protein analysis. Guest researchers from PTB and LGC came to NIST for the sole purpose of learning the basics of protein analysis. These three NMIs then transferred that knowledge to the other NMIs. The later insulin study reinforced that hydrolysis conditions in amino acid analysis need to be optimized for each protein. NIST's first set of data was collected using parameters that had been used successfully for the analysis of other proteins.

One of the biggest lesson learned from these studies is that shipping samples internationally on dry ice is always going to be a nightmare!

11. Nucleic Acid Analysis Working Group (NAWG)

The Nucleic Acid Analysis Working Group (NAWG) was one of three WGs created in 2015 from what had been the Bioanalysis Working Group (BAWG) [1]. Several studies initiated by the BAWG addressed nucleic acid-related measurands; for continuity these studies are attributed to the NAWG.

11.1. Engagements in NAWG Studies

Table 14 lists the finalized NAWG studies in which NIST participated. Results from CCQM-P044, CCQM-P044.1, CCQM-Q060, and CCQM-P113 were reported without attributed quantitative data.

Table 14. NIST Engagement in NAWG Studies.

Study	Name (PAWG)	Units	Sets [a]	Year	Coordinators
CCQM-P044	Real-time qPCR of plasmid DNA	ng/L	0 [b]	2003	NIST,LGC
CCQM-P053	DNA profiling by AFLP	bp	8	2004	NMIA
CCQM-P044.1	Real-time qPCR of plasmid DNA	ng/L	0 [b]	2005	NIST,LGC
CCQM-Q060	DNA extraction: Reference method	cn/n	0 [b]	2005	JRC
CCQM-K061	Real-time qPCR of a linearized plasmid DNA	ng/L	2	2007	NIST,LGC
CCQM-P113	DNA fragments in ground maize seed	cn/n	0 [b]	2008	JRC
CCQM-K086	DNA fragments in ground maize seed	cn/n	2	2010	JRC
CCQM-Q103.1	Multiplexed biomarker panel of RNA	n/n	1	2013	LGC
CCQM-P154	Absolute quantification of linearized plasmid DNA	mg/L	1	2014	KRISS,NMIA,JRC,LGC
CCQM-P199	Copy [number] of HIV-1 RNA genomic sequences	Mn/L	3	2019	LGC,NIBSC
CCQM-P199.b	Copy [number] of SARS-CoV-2 RNA genomic fragments	Mn/L	5	2020	LGC,NIM,NIBSC,NIST

a The number of datasets derived from the study that contain a NIST result.
b No attributed quantitative results available

The cumulative distributions in Fig. 94 display the time course of NIST's engagements with the NAWG's finalized studies, where "engagements" applies to participations and coordinations.

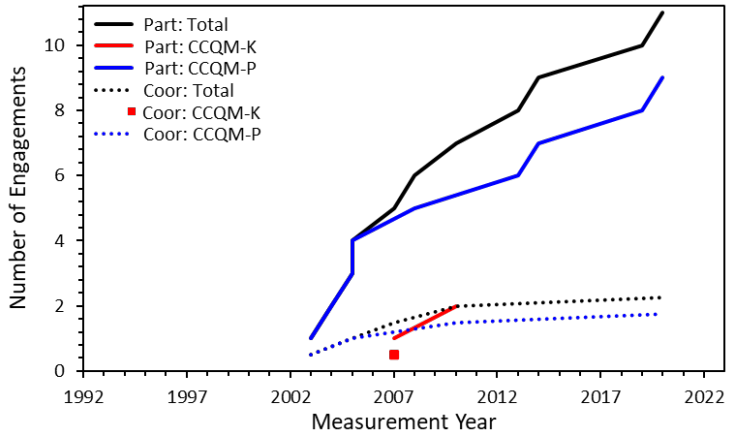


Fig. 94. NAWG Studies: Cumulative Distributions of NIST Engagements.

The histogram traces of Fig. 95 display the engagement time course in higher resolution, along with the average number of participations and coordinations from the first to the most recent engagement.

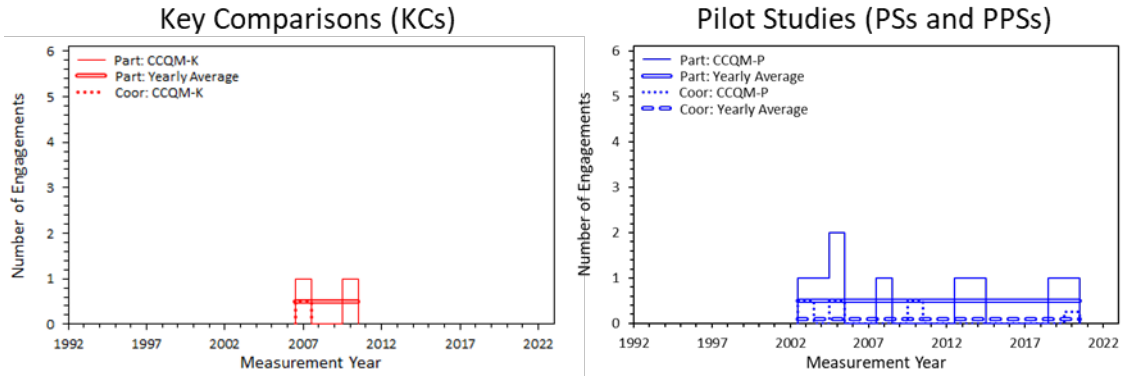


Fig. 95. NAWG Studies: Histogram Traces of NIST Engagements.

11.2. Performance in NAWG Studies

The relationship between the absolute zeta-score bias of these results, $|\zeta_{NIST}|$, and the standard uncertainty of the reported values relative to the median of the participant uncertainties, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, is visualized in Fig. 96.

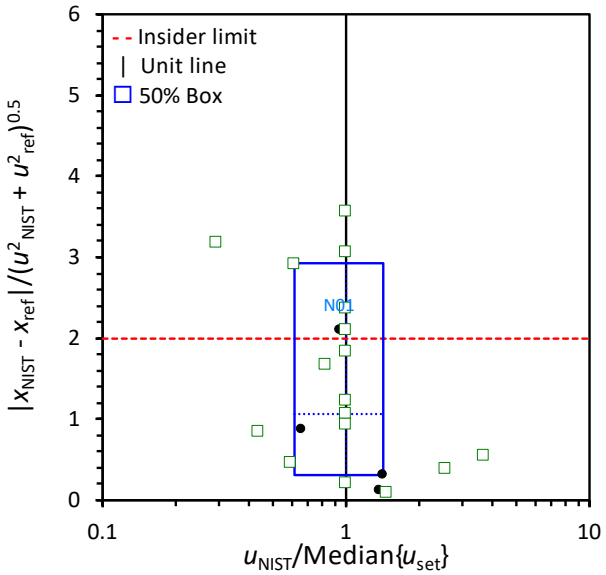


Fig. 96. NAWG Studies: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$: solid black circles denote Key Comparisons (KC) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties.

The median absolute bias is 1.1. More than 66 % of the absolute biases are within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})$/Median$\{u(x_i)\}$ is 1.0. The $u(x_{NIST})$ is about equal to the participant median for the one "outsider" ($|\zeta_{NIST}| > 2$). The "outsider" Key Comparison (KC) dataset is documented in Section 11.2.3.

11.2.1. Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for NAWG datasets are displayed in Fig. 97 as a function of measurement year. While the linear trend, $|t| = |-0.057|/0.012 \approx 5$, suggests that the absolute bias decreased over time, the, diversity of the reporting units suggests that the bias is more likely to be related to the nature of the measurand and/or the experimental conditions rather than a linear trend.

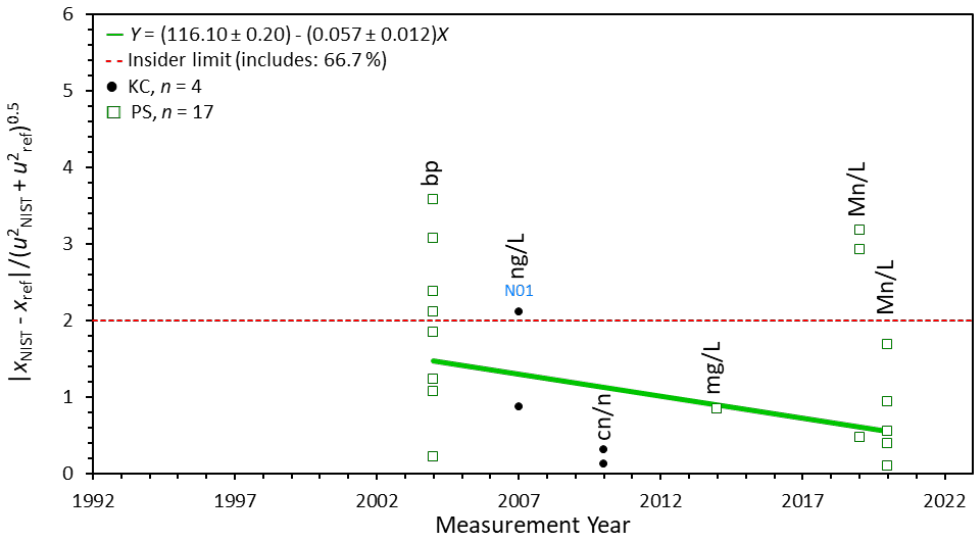


Fig. 97. NAWG Studies: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of measurement year: solid black circles denote Key Comparisons (KC) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The label identifies an "outsider" ($|\zeta_{NIST}| > 2$) KC dataset. The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit", and the number of KC and PS results displayed. The text above each cluster of values identifies the reporting units.

The $u(x_{NIST})$/Median$\{u(x_i)\}$ values are displayed in Fig. 98 as a function of measurement year. The linear trend, $|t| = |-0.0000|/0.0029 \approx 0$, suggests that the uncertainty ratios did not change over time.

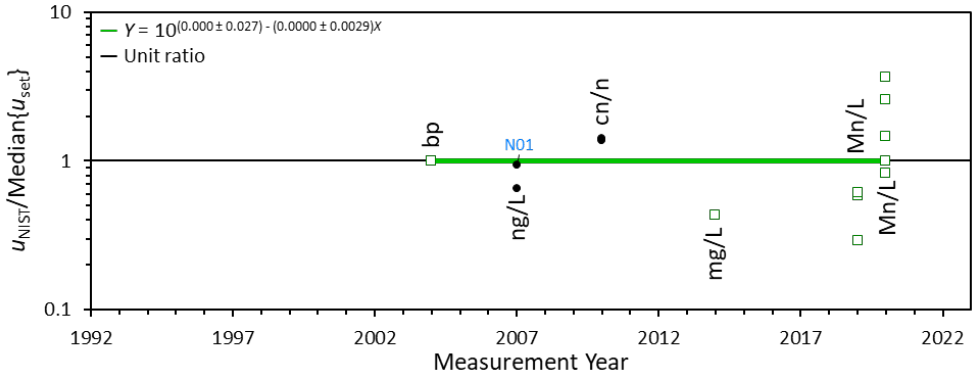


Fig. 98. NAWG Studies: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of measurement year: the solid black circles denote Key Comparison (KC) results, open green squares denote confidential pilot study (PS) results. The label identifies an "outsider" ($|\zeta_{NIST}| > 2$) KC dataset. The thick green line represents the estimated linear trend. The legend provides the trend equation.

11.2.2. Bias and Relative Uncertainty Across Measurand Level

The $|\zeta_{NIST}|$ values for NAWG datasets are displayed in Fig. 99 as a function of measurand level. The linear trend, $|t| = |+0.006|/0.022 \approx 0$, suggests that the absolute bias is independent of the measurand level.

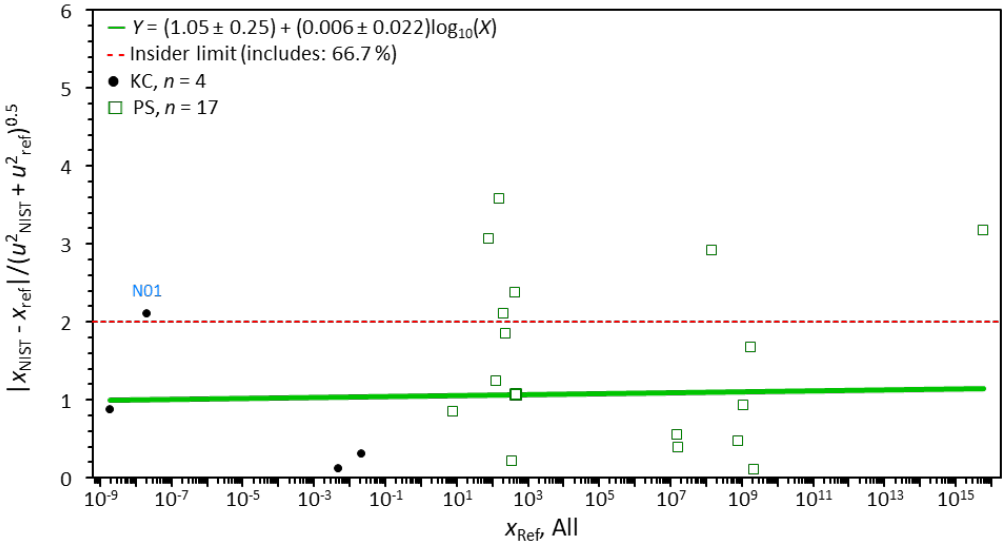


Fig. 99. NAWG Studies: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of measurand level. See Fig. 97 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 100 as a function of measurand level. The linear trend, $|t| = |+0.0000|/0.0041 \approx 0$, suggest that the uncertainty ratios are independent of measurand level.

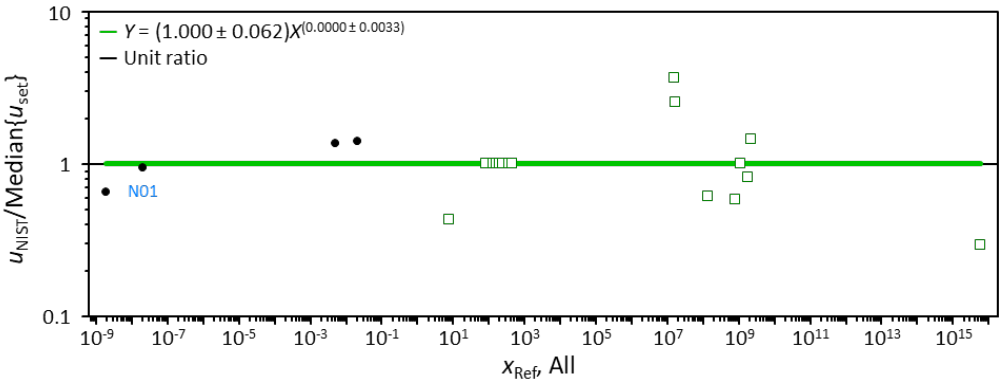


Fig. 100. NAWG Studies: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of measurand level. See Fig. 98 for description of the graphical elements.

11.2.3. "Outsider" Dataset

The NAWG dataset that contains an "outsider" NIST result is depicted in Fig. 101 as a dot-and-bar chart. The chart shows all results for the dataset, the reference value, and the 95 % level of confidence reference interval. The label in the upper left corner, "N01," is the code used to identify the "outsider" in Fig. 96 to Fig. 100.

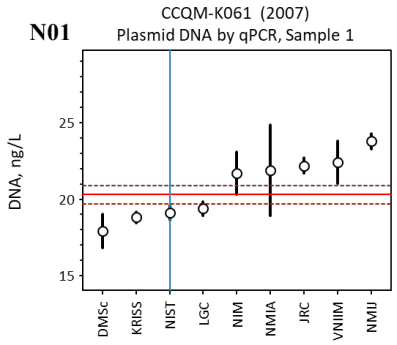


Fig. 101. NAWG Studies: "Outsider" Dataset.

The chart shows all reported results for the dataset with a NIST "outsider" ($|\zeta_{NIST}| > 2$) result. Results are sorted in order of increasing magnitude. Participants are identified by code name; the full names are listed in Appendix B. Open circles denote results considered valid. Bars represent one standard uncertainty. The horizontal solid red line denotes the reference value; the dashed red lines bound a 95 % level of confidence interval about the reference value. The blue vertical line marks the NIST result.

11.3. NAWG Lessons Learned

Participation in the P199 (HIV-1 RNA copy number quantification) in 2019 prepared NIST for the production of Research Grade Test Material (RGTM) 10169 SARS-CoV-2 Synthetic RNA Fragments during the COVID-19 pandemic in 2020. Prior to the P199 study, NIST had little experience measuring RNA via reverse transcription (RT) digital polymerase chain reaction (dPCR). The protocol for the production of RNA (synthesis, purification, and storage in tubes) for RGTM 10169 was with only minor modifications based on the P199 coordinating lab's protocol, which they used to make the RNA for the P199 study. During the COVID-19 pandemic, NAWG ran the P199b study for measurement of SARS-CoV-2 RNA. Many NMIs participating in that study requested units of RGTM 10169 to use as a positive control and returned data to NIST regarding their dPCR values for the material.

Based on results from the P199 and P199b studies, NIST is looking into biases in RNA measurement using reverse transcription digital polymerase chain reaction (RT-dPCR).

12. Isotope Ratio Working Group (IRWG)

The Isotope Ratio Working Group (IRWG) was established by the CCQM in April 2018. The IRWG has responsibility for "stable isotope ratio measurement science" [85]. Several studies initiated by the IAWG address isotope ratio-related measurands, for continuity these studies are attributed to the IRWG. One study, CCQM-P204 Isotope ratios in pure carbon dioxide, was conducted by the GAWG in consultation with the IRWG but is here attributed to the IRWG.

12.1. Engagements in IRWG Studies

Table 15 lists the finalized IRWG studies in which NIST participated. The design and textual discussion of results CCQM-Q160 were published without attributed quantitative data.

Table 15. NIST Engagement in IRWG Studies.

Study	Name (IRWG)	Units	Sets [a]	Year	Coordinators
CCQM-P075	Stable isotope delta values in methionine	‰ [b]	1	2006	JRC,IAEA
CCQM-K098	Lead isotope ratios in solution and in bronze	n/n	8	2013	BAM
CCQM-P134	Lead isotope ratios in solution and in bronze	g/mol,‰ [b]	4	2013	BAM
CCQM-Q160	Isotope ratios / molar mass of Si	g/mol	0 [c]	2016	PTB
CCQM-Q204	CO_2 isotope ratios ($\delta^{13}C$ and $\delta^{18}O$) in pure CO_2	‰ [b]	8	2022	BIPM,IAEA
CCQM-Q213	Copper isotope ratios in high purity copper	‰ [b]	2	2022	NRC,NIST,BAM,PTB

a The number of datasets derived from the study that contain a NIST result.
b The symbol "‰" denotes "part per thousand". It is analogous to "%" denoting "part per hundred".
c No attributed quantitative results available

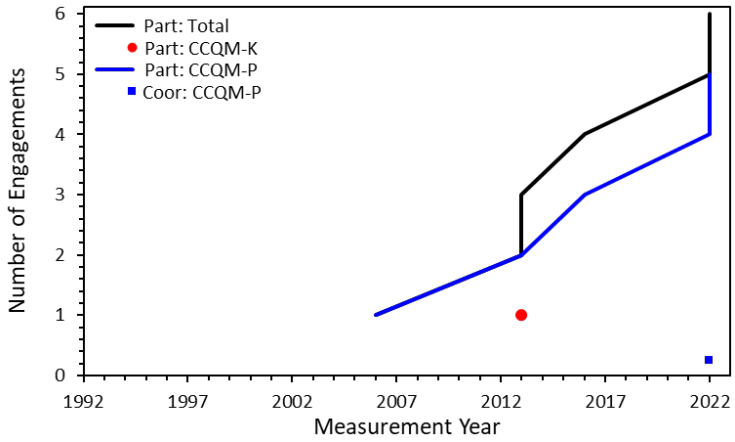


Fig. 102. IRWG Studies: Cumulative Distributions of NIST Engagements.

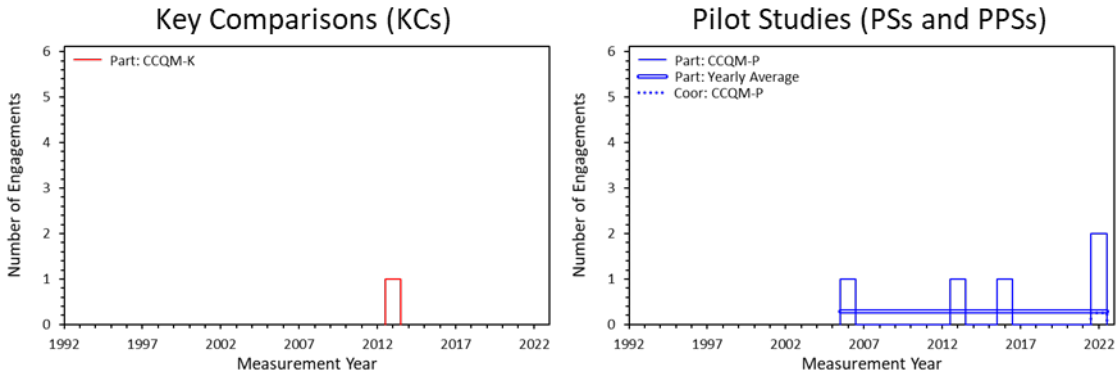


Fig. 103. IRWG Studies: Histogram Traces of NIST Engagements.

12.2. Performance in IRWG Studies

The relationship between the absolute zeta-score bias of these results, $|\zeta_{NIST}|$, and the standard uncertainty of the reported values relative to the median of the participant uncertainties, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, is visualized in Fig. 104.

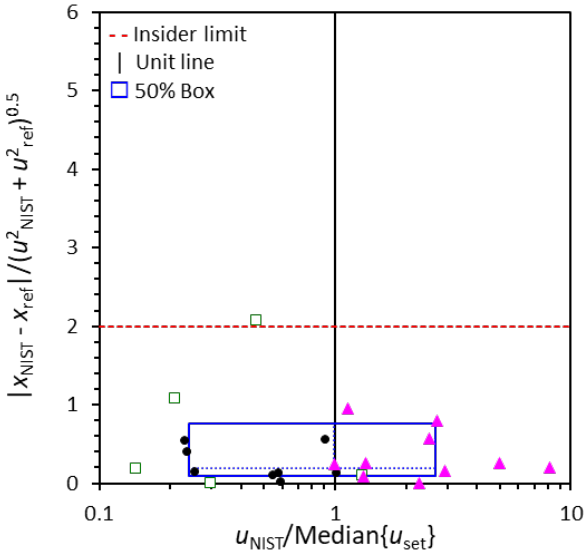


Fig. 104. IRWG Studies: Bias as a Function of Relative Uncertainty.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of the result's standard uncertainty relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$: solid black circles denote Key Comparisons (KC) results, solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The solid black vertical line marks where NIST's standard uncertainty is equal to the participant median. The solid blue "50% box" lines bound the central 50 % of the datasets; the dotted blue horizontal and vertical lines within the box mark the median bias and relative uncertainties.

The median absolute bias is 0.19. More than 95 % of the absolute biases are within the nominal "equivalent to the reference value" limit: $|\zeta_{NIST}| \leq 2$. The median $u(x_{NIST})/\text{Median}\{u(x_i)\}$ is 0.99. The $u(x_{NIST})$ is less than the participant median for the one "outsider" ($|\zeta_{NIST}| > 2$).

12.2.1. Bias and Relative Uncertainty Over Time

The $|\zeta_{NIST}|$ values for the IRWG datasets are displayed in Fig. 105 as a function of measurement year. While the linear trend, $|t| = |+0.0024|/0.0044 \approx 0$, suggests that the absolute bias did not change over time, the single dataset from the 2006 study shows a relatively high bias.

In this early study, participants used differing isotope-delta values for the reference materials and in some cases did not report the value that they used. This resulted not only in differences in the final values reported but also in an inaccurate assessment of bias. Different types of instrumentation having differing achievable uncertainties were used, further confounding the results.

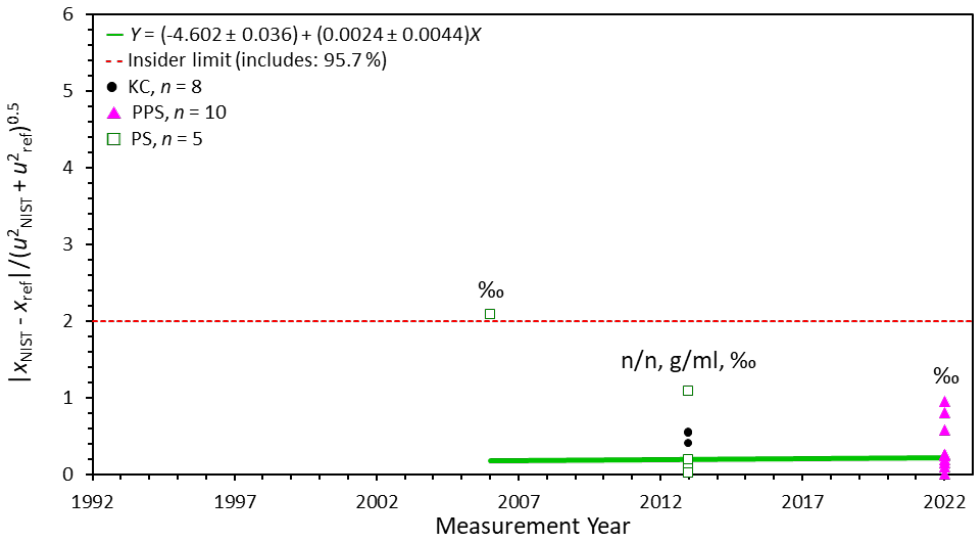


Fig. 105. IRWG Studies: Absolute Bias Over Time.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of measurement year: solid black circles denote Key Comparisons (KC) results, solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The dashed red horizontal "insider limit" line bounds results considered equivalent to the reference value ($|\zeta_{NIST}| \leq 2$). The thick green line represents the estimated linear trend. The legend provides the trend equation, the percentage of values within the "insider limit", and the number of KC, PPS, and PS results displayed. The text above each cluster of values identifies the reporting units.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 106 as a function of measurement year. The linear trend, $|t| = |+0.0732|/0.0043 \approx 17$, suggests that the uncertainty ratios have increased over time. However, many different approaches to isotope ratio measurements have been used as well as addressing different isotope systems (S, Pb, Si, C, and O). Differing analytical capabilities and inconsistent data reporting policies contribute to the observed trend. Unlike elemental measurands where the data reporting methods are essentially the same, uncertainty assessments for isotopic measurands differ due to the differing ways in how the measurand is assessed.

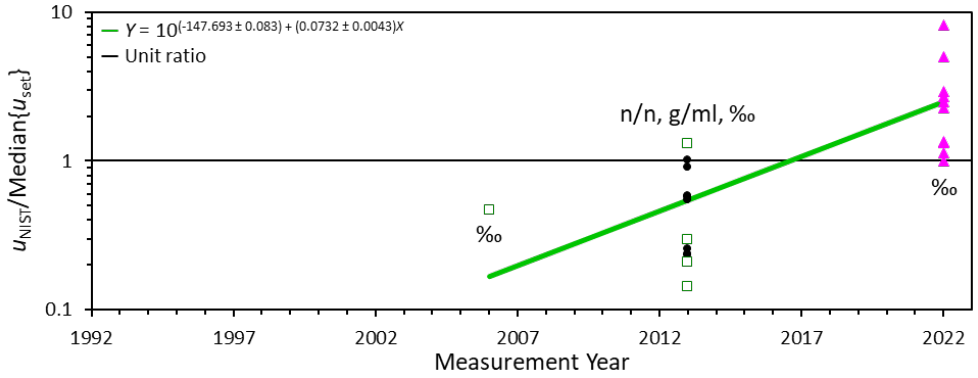


Fig. 106. IRWG Studies: Relative Uncertainty Over Time.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of measurement year: the solid black circles denote a Key Comparison (KC) result, solid magenta triangles denote published pilot study (PPS) results, open green squares denote confidential pilot study (PS) results. The thick green line represents the estimated linear trend. The legend provides the trend equation.

12.2.2. Bias and Relative Uncertainty Over Isotopic Ratio

The $|\zeta_{NIST}|$ values for IRWG measurements are displayed in Fig. 107 as a function of per mille (‰) difference from a reference isotopic composition. The linear trend, $|t| = |+0.0038|/0.0035 \approx 1$, suggests that the absolute bias is effectively independent of the magnitude of the difference from the reference.

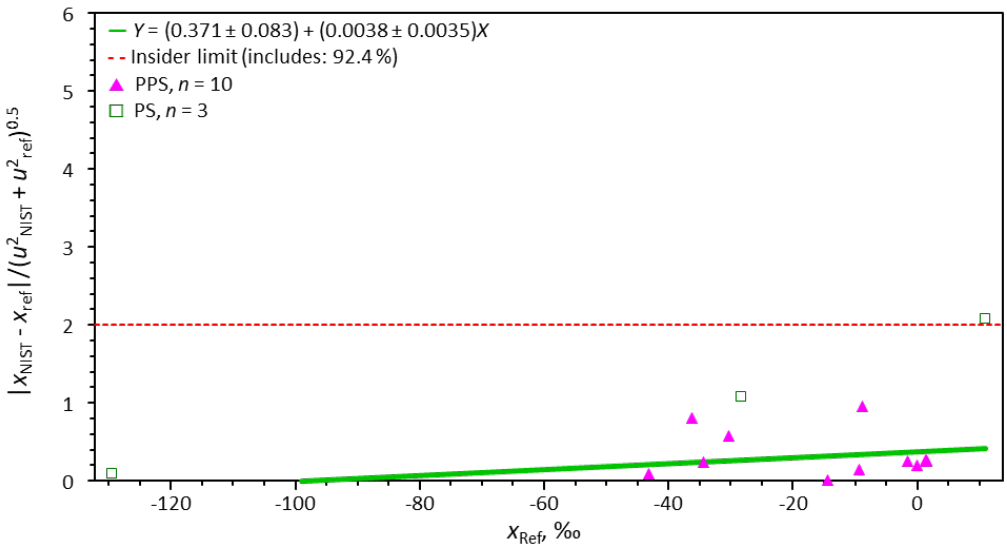


Fig. 107. IRWG Isotopic Ratio Studies: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function of per mille (‰) difference from a reference composition. See Fig. 105 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 108 as a function of ‰. The linear trend, $|t| = |+0.0025|/0.0028 \approx 1$, suggests that the uncertainty ratios are independent of the magnitude of the differences.

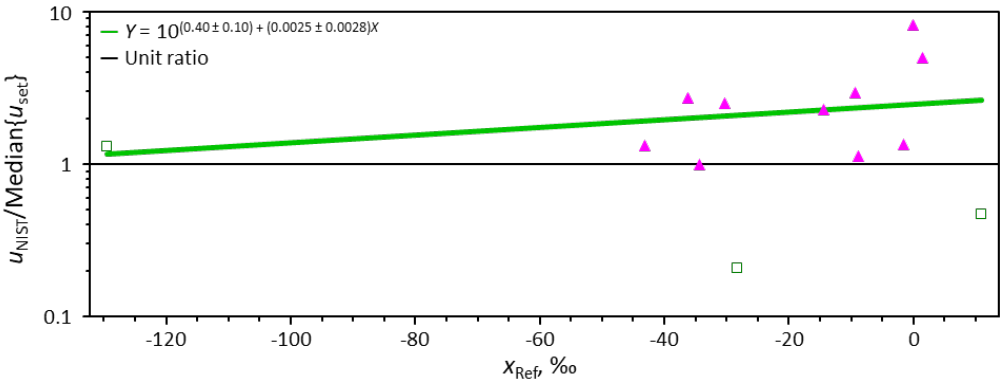


Fig. 108. IRWG Isotopic Ratio Studies: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of per mille (‰) difference from a reference composition. See Fig. 106 for description of the graphical elements.

12.2.3. Bias and Relative Uncertainty Over Isotope Amount Ratio

The $|\zeta_{NIST}|$ values for IRWG measurements are displayed in Fig. 109 as a function of isotope amount ratio, the amount of a given isotope relative to the amount of a more abundant isotope of the same element expressed as n/n. The linear trend, $|t| = |+0.0045|/0.0031 \approx 1$, suggests that the absolute bias is independent of the relative amounts.

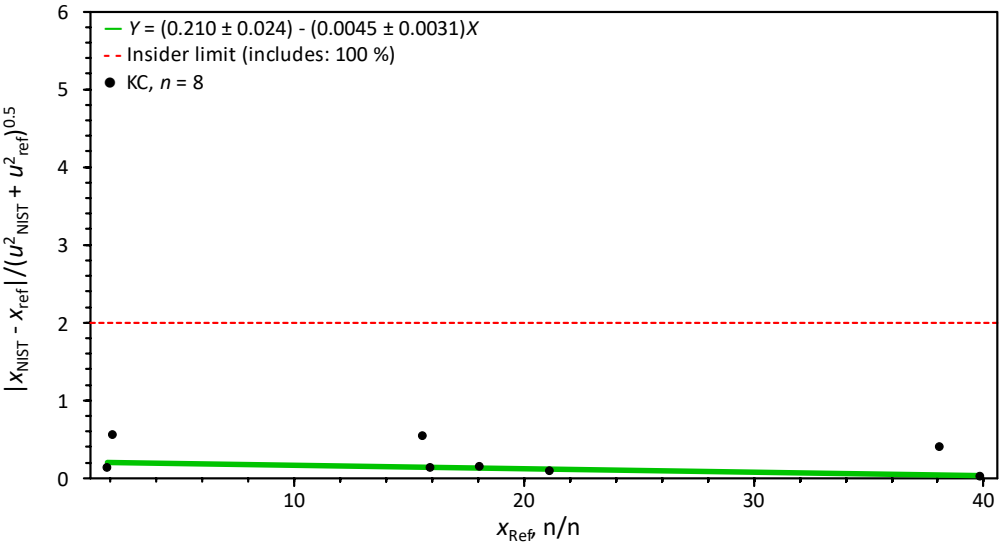


Fig. 109. IRWG Isotope Amount Ratio Study: Absolute Bias Across Level.

Each symbol represents the absolute bias of a NIST result, $|\zeta_{NIST}|$, in one dataset as a function isotope amount ratio. See Fig. 105 for description of the graphical elements.

The $u(x_{NIST})/\text{Median}\{u(x_i)\}$ values are displayed in Fig. 110 as a function of n/n. The linear trend, $|t| = |-0.090|/0.0032 \approx 3$, suggests that the uncertainty ratios slightly decrease with increasing amount ratio.

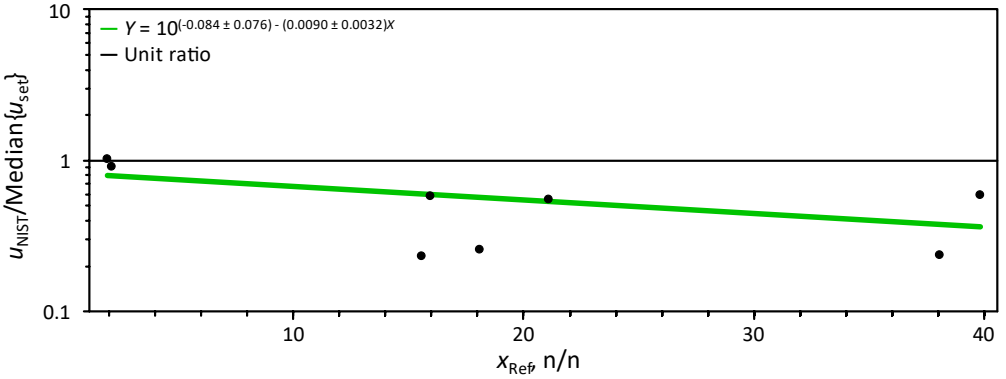


Fig. 110. IRWG Isotope Amount Ratio Study: Relative Uncertainty Across Level.

Each symbol represents the standard uncertainty of a NIST result relative to the participant median, $u(x_{NIST})/\text{Median}\{u(x_i)\}$, as a function of isotope amount ratio. See Fig. 106 for description of the graphical elements.

12.3. IRWG Lessons Learned

To date the number of pilot and key comparison studies involving isotope ratio analysis has been small. The measurement approaches used for isotopic analysis have been quite varied and assessing changes in bias, uncertainties, and/or deviations from a reference over time is not straightforward. The results from these assessments must be understood in this context.

In some cases, the validity of the assessment result is questionable due to specific challenges/issues with the design of the studies themselves. To enhance the ability to assess capabilities in isotope ratio measurements over time, more rigorously designed studies are required. For example, a simple change would be to have all participants use the same reference materials for calibration and reporting, as was the case with CCQM-Q213. This would eliminate potential unknowns and biases caused by using different materials. At a minimum, participants should be required to state the reference materials and the values they are using for these comparisons. This holds true for instrumentation as well, as some instruments perform internal calculations using values for reference materials that may differ from other instruments resulting in biases.

Isotope ratio analysis is complex and details of how measurements are made is critical if there is anything to learn from the data results.

13. Cell Analysis Working Group (CAWG)

The Cell Analysis Working Group (CAWG) is one of three WGs created in 2015 from what had been the Bioanalysis Working Group (BAWG) [1]. The CAWG has responsibility for "identification and quantification of intact cells and cell properties indicative of function as a result of emergent behavior in complex matrices and mixtures"[86]. Several studies initiated by the BAWG addressed cell-related measurands; for continuity these studies are attributed to the CAWG.

The CAWG covers cellular analysis in both prokaryotic and eukaryotic domains, each having particular measurement needs and challenges that impact participation and the prioritization of comparison studies [87].

The CAWG studies are currently designed to demonstrate capabilities in novel areas and build technical expertise for new or emergent techniques. The present focus is on cell counting metrology for a range of entities in both suspension and adherent cell formats, including flow cytometry, manual microscopy, and automated technologies. These services will ultimately enable a measurement framework supported by traceable measurement systems for particle analysis, supporting instrument calibration, and the measurement of specific biological macromolecules of fundamental importance to cellular analysis [87].

The CAWG is working to develop follow-on comparisons that will underpin existing measurement services including the quantification of blood cells and quantification of bacteria in food and water matrices. Areas of future interest include reference value assignments for complete blood count, adherent cells on two-dimensional supports, antimicrobial susceptibility, cell viability and the quantification of synthetic cell analogues [87].

Table 16 lists the CAWG studies, both finalized and in-progress, in which NIST has participated. Results from the finalized studies have been reported without attributed quantitative data. As of this document's publication date, no attributed results are available for any CAWG study.

Table 16. NIST Engagement with CAWG Studies.

Study	Name	Units	Sets[a]	Year	Coordinators
CCQM-Q102	Prelabeled CD4+ cell concentration	n/L,EFF	0[b]	2011	NIBSC,NIST,PTB
CCQM-P123	Cell quantification on solid substrate	a.u.	0[b]	2015	INRIM,LGC,NIST
CCQM-Q165	CD34+ stem cell concentration	a.u.	0[b]	2016	NIBSC,NIST,PTB
CCQM-P197[c]	Count of proliferated stem cells per unit area				NIST,NPL
CCQM-P205[c]	Membrane-intact *E. coli* in drinking water				NIM,NIST
CCQM-P217[d]	Count of fixed blood mononuclear cells			2020	LGC,NIBSC,NIST,PTB

a The number of datasets derived from the study that contain a NIST result.
b No attributed quantitative results available.
c Study in progress as of March 2024.
d Report not finalized as of March 2024.

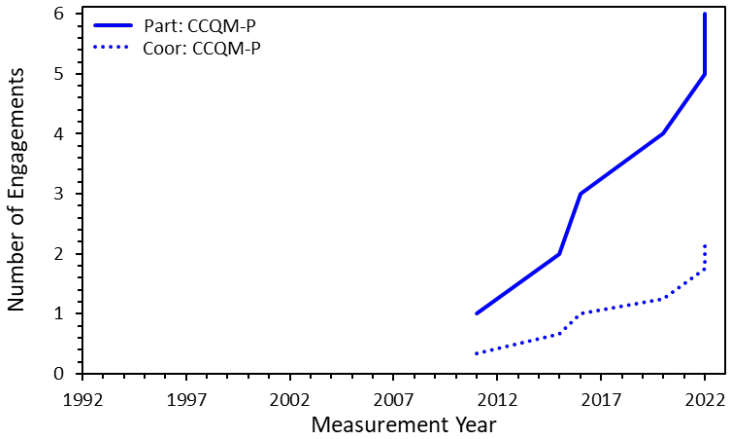


Fig. 111. CAWG Studies: Cumulative Distributions of NIST Engagements.

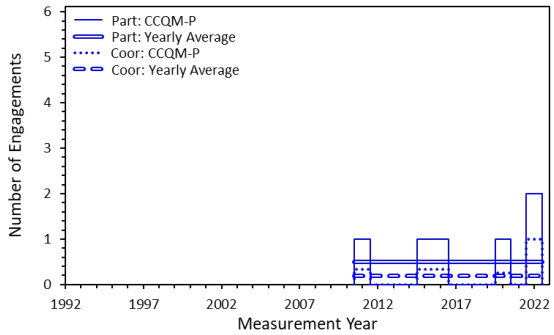


Fig. 112. CAWG Studies: Histogram Trace of NIST Engagements.

13.1. NIST's Role In CAWG Studies

The CAWG's focus on enumeration of biological entities (e.g., prokaryotic and eukaryotic cells) is well aligned with current NIST programmatic areas in regenerative medicine, microbiome, and engineering biology. NIST has been leading international efforts in cell counting within the International Organization for Standardization (ISO) [88,89,90] and has therefore been well positioned to take a leadership role in current CAWG studies. While other NMIs have initiated the CAWG's studies, NIST has shared coordination responsibilities for the count-based pilot studies to help assure the incorporation of best practices.

- NIST expertise in microscopy led to leadership in CCQM-P197 where NIST led the incorporation of critical control materials (bead-based control materials for benchmarking fluorescence intensity) to support data analysis and confirmation of imaging quality.
- NIST supported the CCQM-P217 initiators in incorporating standardized dilution series study designs into the sample preparation to improve the quality of the data and conclusions. NIST took leadership of the data analysis, applying ISO statistical strategies for cell counting as well as developing additional analysis for evaluating sources of variability and developing consensus value estimates.
- NIST supported the design of CCQM-P205 by recommending a standardized dilution series study design for the flow cytometry component of the study. NIST is also in the process of proposing ISO statistical strategies for comparing cell count data.

13.2. CAWG Lessons Learned

Because NIST has developed advanced measurement assurance strategies in cell-based measurements, NIST often takes the lead in providing ideas for incorporation of appropriate control materials and study designs. This has resulted in a burden on NIST for further data analysis and report generation support for studies that other NMIs initiated. Clear plans should be established prior to study execution for the data analysis and reporting. Study initiators should be capable of incorporating the experimental designs and control materials within their reports with minimal NIST support. If this is not possible, then NIST will need to consider its availability to contribute to the study analysis.

It has been a positive experience in participating within the CAWG studies. It is clear that NIST is working on high-priority areas for cell metrology, and the interlaboratory data collected by CAWG has been valuable in gaining a better understanding into important measurands for biotechnology. In the future NIST will need to consider whether it would like to propose and initiate studies; however, most measurement areas for cell characterization are not mature enough for CAWG.

Outputs such as peer-reviewed manuscripts are needed if the CAWG is to support the broader cell analysis community.

14. Microbiology Steering Group (MBSG)

In 2011 the *ad hoc* Steering Group on Microbial Measurements (MBSG) was formed, following the CCQM Workshop on Metrology and the Need for Traceable Microbiological Measurements to Ensure Food Quality and Safety. Led by NIST, the MBSG focused on defining key properties of microbial quantity and identity, considering how to establish traceability and measurement uncertainty for these properties, and determining when the technologies and the field would be ready for robust measurements. MBSG consisted of two working groups, a Quantitative Working Group and an Identity Working Group, that focused on existing and emerging measurement methods and technologies.

The MBSG conducted one investigative study in each working group. These studies were investigative, rather than pilot, because microbial measurement capabilities were not yet established across NMIs, and informative studies regarding comparability and measurement uncertainty across institutions were needed. The MBSG completed its mission and was disbanded in 2016 [91].

14.1. Microbial Quantity

In 2013, the MBSG Quantity Working Group completed an investigative study led by NMIA to establish comparability for cell count via the plate count method. The four participating NMIs procured and characterized commercially available reference materials of *Listeria monocytogenes*, a microorganism relevant to food safety and public health. Participants reported comparable results and measurement uncertainty, with increasing uncertainty at low microbe numbers, as expected. The plate count method does not have a well-defined measurand, and the MBSG recognized the need for higher order measurements for microbial quantity. This discussion was carried over to the CAWG when it was formed, with NIM proposing a study that became CCQM-P205.

14.2. Microbial Identity

In 2012 NIST proposed a microbial identity investigative study to establish comparability across laboratories using 16S rRNA sequencing for identity via different sequencing technologies, including two next-generation sequencing (NGS) platforms. As microorganisms relevant to food safety, *L. monocytogenes* and *Escherichia coli* were sequenced. Participating laboratories contributed raw sequencing data, and all data were analyzed at NIST via a novel high-resolution method to compare data across laboratories and sequencing platforms for a multi-copy gene. The results demonstrated that increased sequencing depth and longer read lengths (via high-throughput sequencing methods) were needed for reproducible variant copy ratios at biologically variable positions. Participants in this forward-looking study included one designated institute and one stakeholder laboratory, in addition to four NMIs, as expertise in NGS technologies was not widespread within the NMIs at the time. This work resulted in a NIST-led peer-reviewed publication in 2015 [92].

14.3. MBSG Lessons Learned

The investigative studies in the MBSG demonstrated a proof of principle for comparable microbial metrology. However, many NMIs did not have the necessary technical capabilities to participate.

When the BAWG transitioned to the CAWG, NAWG, and PAWG, the microbial quantification efforts became part of the CAWG and transitioned to higher-order methods rather than colony counting. The microbial identification efforts based on sequencing were not continued, as both the technologies and most NMIs were not ready for additional studies. NIST took on the major responsibility for the study design and data analysis and chose to continue their work in NGS-based measurements with non-CCQM collaborators who were experts in the field.

To date, the 16S rRNA sequencing effort is the only genomic sequence analysis study conducted by the CCQM, perhaps reflecting the belief that genome sequence key comparisons are not currently necessary to support international commerce. The experience and methods developed by the MBSG have been captured in the NAWG.

References

[1] Kaarls R (2018) The Consultative Committee for Metrology in Chemistry and Biology – CCQM. J Chem Metrol 12(1):1-16. https://doi.org/10.25135/jcm.11.17.12.060

[2] Key Comparison Database (KCDB). https://www.bipm.org/kcdb/. Accessed: 5/17/2024.

[3] Duewer DL (2023) CCQM_Retrospectoscope. https://doi.org/10.18434/mds2-2952. See also: Duewer DL, Nelson MA, Cecelski CE. CCQM_Retrospectoscope Quick Start: A beginner's guide to a data analysis system for visualizing results from CCQM studies (2023) NIST Internal Report (NIST IR) 8487. https://doi.org/10.6028/NIST.IR.8487.

[4] BIPM (1990) Comite International des Poids et Mesures (CIPM) 79th Meeting. Proceedings of the Sessions Held at Sèvres (25-27 September 1990). https://www.bipm.org/documents/20126/17315032/CIPM79.pdf/e1d53b2a-5fb2-e839-e909-96f4692d690b. Accessed: 5/17/2024.

[5] BIPM (1991) Comite International des Poids et Mesures (CIPM) 80th Meeting. Proceedings of the Sessions Held at Sèvres and Paris (26 September – 2 October1991). https://www.bipm.org/documents/20126/17315032/CIPM80.pdf/031aac1d-1d9a-9370-6897-bb28366ea3dc. Accessed: 5/17/2024.

[6] BIPM (1993) Comite International des Poids et Mesures (CIPM) 82nd Meeting. Proceedings of the Sessions Held at Sèvres (21-23 September 1993). https://www.bipm.org/documents/20126/17315032/CIPM82.pdf/74d3ed82-22b1-ae99-9d1d-05e0b6ae4e03. Accessed: 5/17/2024.

[7] Beary ES, Beck CM, Fassett JD, Paulsen PJ, Watters RL Jr. (1994) Preparation and Calibration of Isotopic Spikes for the CIPM Metrology in Chemistry Round Robin. In: CSTL Technical Activities 1993 (National Institute of Standards and Technology, Gaithersburg, MD), NIST Internal Report (NIST IR) 5303. https://doi.org/10.6028/NIST.IR.5303.

[8] Alink A (2000) The first key comparison of primary standard gas mixtures. Metrologia 37(1):35-49. https://doi.org/10.1088/0026-1394/37/1/5.

[9] BIPM (1994) Comite International des Poids et Mesures (CIPM) 83rd Meeting. Proceedings of the Sessions Held at Sèvres (27-29 September 1994). https://www.bipm.org/documents/20126/17315032/CIPM83.pdf/04c6be4e-c7c4-2c29-a911-93d9ca9bb7bf. Accessed: 5/17/2024.

[10] Watters RL Jr, Eberhardt KR, Beary ES, Fassett JD (1997) Protocol for isotope dilution using inductively coupled plasma-mass spectrometry (ICP-MS) for the determination of inorganic elements. Metrologia 34(1):87-96. https://doi.org/10.1088/0026-1394/34/1/13.

[11] Papadakis I, Taylor PDP, De Bievre P (2001) CCQM-K2 key comparison: cadmium and lead content in natural water. Metrologia 38(6):543-547. https://doi.org/10.1088/0026-1394/38/6/9.

[12] BIPM. Final reports – QM: Pilot studies. https://www.bipm.org/en/committees/cc/ccqm/pilot-studies. Accessed: 5/17/2024.

[13] CIPM-MRA-2003. Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, Paris, 14 October 1999. Technical Supplement revised in October 2003 (pages 38-41). https://www.bipm.org/documents/20126/43742162/CIPM-MRA-2003.pdf. Accessed: 5/17/2024.

[14] CIPM MRA-G-11, Measurement comparisons in the CIPM MRA (2021) https://www.bipm.org/documents/20126/43742162/CIPM-MRA-G-11.pdf. Accessed: 5/17/2024.

[15] Wikipedia. EURAMET. https://en.wikipedia.org/wiki/EURAMET. Accessed: 5/17/2024.

[16] CCQM Working Document CCQM/2013-22 (2013). CCQM Guidance note: Estimation of a consensus KCRV and associated Degrees of Equivalence (author: Ellison SLR) https://www.bipm.org/documents/20126/28430045/working-document-ID-5794/49d366bc-295f-18ca-c4d3-d68aa54077b5.

[17] ISO 13528:2015 (2015) Statistical methods for use in proficiency testing by interlaboratory comparison. International Standards Organization International Standard. Geneva, Switzerland.

[18] Andrews DF, Bickel PJ, Hampel FR, Huber PJ, Rogers WH, Tukey JW. Robust Estimates of Location. Princeton University Press, 1972

[19] Rousseeuw PJ, Croux C. Alternatives to Median Absolute Deviation. J Am Stat Assoc. 1993;88:1273-83. https://doi.org/10.1080/01621459.1993.10476408

[20] Wikipedia. Robust measures of scale. https://wikipedia.org/wiki/Robust_measures_of_scale. Accessed: 5/17/2024.

[21] Wilrich P-Th. Robust estimates of the theoretical standard deviation to be used in interlaboratory precision experiments. Accred Qual Assur. 2007;12:231–40. https://doi.org/10.1007/s00769-006-0240-7.

[22] Thompson M, Ellison SLR. (2011) Dark uncertainty. Accred Qual Assur 16:483-487. https://doi.org/10.1007/s00769-011-0803-0.

[23] Meija J, Possolo A (2023) Interlaboratory comparisons of chemical measurements: *Quo Vadis*? Accred Qual Assur 28, 89–93. https://doi.org/10.1007/s00769-022-01505-y.

[24] Sen PK (1968) Estimates of the Regression Coefficient Base on Kendall's Tau, JASA 63(324):1379-1389. https://doi.org/10.1080/01621459.1968.10480934.

[25] Theil-Sen estimator. https://en.wikipedia.org/wiki/Theil-Sen_estimator. Accessed: 5/17/2024.

[26] CIPM-MRA-G-13, Calibration and measurement capabilities in the context of the CIPM MRA (2022) https://www.bipm.org/documents/20126/43742162/CIPM-MRA-G-13.pdf. Accessed: 5/17/2024.

[27] Beauchamp CR, Camara JE, Carney J, Choquette SJ, Cole KD, DeRose PC, Duewer DL, Epstein MS, Kline MC, Lippa KA, Lucon E, Molloy J, Nelson MA, Phinney KW, Polakoski M, Possolo A, Sander LC, Schiel JE, Sharpless KE, Toman B, Winchester MR, Windover D. Metrological Tools for the Reference Materials and Reference Instruments of the NIST Material Measurement Laboratory. NIST Special Publication (NIST SP) 260-136-2021. 2021. https://doi.org/10.6028/NIST.SP.260-136-2021.

[28] NIST (1999) International Measurement Database Goes Online. https://www.nist.gov/news-events/news/1999/12/international-measurement-database-goes-online. Accessed: 5/17/2024.

[29] JCGM 100:2008; Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement (ISO GUM 1995 with Minor Corrections); Joint Committee for Guides in Metrology (JCGM) (2008)

[30] BIPM (1992) Comite International des Poids et Mesures (CIPM) 81th Meeting. Proceedings of the Sessions Held at Sèvres (29 September 1992). https://www.bipm.org/documents/20126/17315032/CIPM81.pdf/db7956db-43c2-4976-aa3d-c900fe276f16. Accessed: 5/17/2024.

]31] Taylor BN, Kuyatt CE (1994) Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, NIST Technical Note (NIST TN) 1297. https://doi.org/10.6028/NIST.TN.1297.

[32] Possolo A (2020) Evaluating, Expressing, and Propagating Measurement Uncertainty for NIST Reference Materials. NIST Special Publication (NIST SP) 260-202. https://doi.org/10.6028/NIST.SP.260-202.

[33] Possolo A (2015) Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results. NIST Technical Note (NIST TN) 1900. http://doi.org/10.6028/NIST.TN.1900.

[34] Possolo A, Hibbert D, Stohner J, Bodnar O, Meija J (2024) A brief guide to measurement uncertainty (IUPAC Technical Report). Pure Appl Chem 96(1): 113-134. https://doi.org/10.1515/pac-2022-1203.

[35] NIST Consensus Builder (NICOB). https://consensus.nist.gov. See also: Koepke A, LaFarge T, Possolo A, Toman B (2020) NIST Consensus Builder User's Manual. https://consensus.nist.gov/app_direct/nicob/NISTConsensusBuilder-UserManual.pdf.

[36] NIST Decision Tree (NDT) https://decisiontree.nist.gov/. See also: Possolo A, Koepke A, Newton D, Winchester M (2021) Decision Tree for Key Comparisons. JResNIST 126:126007. https://doi.org/10.6028/jres.126.007. See also: Newton D, Koepke A, Possolo A, Winchester M (2023) NIST Decision Tree User's Manual, 2nd ed. Available via the NDT's "About" webpage at https://decisiontree.nist.gov/.

[37] NIST Uncertainty Machine. https://uncertainty.nist.gov. See also:
Lafarge T, Possolo A (2015) The NIST Uncertainty Machine, NCSL Int meas 10(3):20-27. https://doi.org/10.1080/19315775.2015.11721732.
Lafarge T, Possolo A (2020) NIST Uncertainty Machine – User's Manual. https://uncertainty.nist.gov/NISTUncertaintyMachine-UserManual.pdf.

[38] SIM (1998) Metrology in Chemistry: A New Challenge for the Americas. Proceedings of the First Interamerican Workshop on Metrology in Chemistry. Rio de Janeiro, RJ, Brazil. November 3 to 4, 1997. Editors: Frota MN, Muller TH, Semerjian HG. Qualitymark Editora Ltda. Rio de Janeiro

[39] Kuselman I (2000) UME – NIST Workshop on Establishment of Chemical and Medical Metrology Laboratories at UME, Gebze, Turkey,19–20 June 2000. Accred Qual Assur 5:453-454. https://doi.org/10.1007/s007690000221.

[40] BIPM. CCQM Working Group on Inorganic Analysis (CCQM-IAWG) Terms of Reference. https://www.bipm.org/en/committees/cc/ccqm/wg/ccqm-iawg.
Accessed: 5/17/2024.

[41] COMITÉ CONSULTATIF POUR LA QUANTITÉ DE MATIÈRE, SESSION DE 1995 (1995). https://www.bipm.org/en/committees/cc/ccqm/1-1995.
Accessed: 5/17/2024.

[42] IAWG-GD-03, Record of Participation in CCQM Key Comparisons/Pilot Studies (2020) https://www.bipm.org/documents/20126/56018303/IAWG-GD-03-Record_of_Participation.docx/bfca6fea-7b8f-1c9f-dbb1-9c80c6cebc7f.
Accessed: 5/17/2024.

[43] Kuroiwa T, Zhu Y, Inagaki K, Long SE, Christopher SJ, Puelles M, Borinsky M, Hatamleh N, Murby J, Merrick J, White I, Saxby D, Sena RC, Almeida MD, Vogl J, Phukphatthanachai P, Fung WH, Yau HP, Okumu TO, Kang'iri JN, Télle JAS, Campos EZ, Galv´n EC, Kaewkhomdee N, Taebunpakul S, Thiengmanee U, Yafa C, Tokman N, Tunç M, Can S Z (2017) Report of the CCQM-K123: trace elements in biodiesel fuel. 2017 Metrologia 54(1A):08008.https://doi.org/10.1088/0026-1394/54/1A/08008.

[44] Salas A, Ramírez E (2009) Final report on CCQM-K57: Chemical composition of clay. Metrologia 46(1A):08020. https://doi.org/10.1088/0026-1394/46/1A/08020.

[45] BIPM. CCQM Working Group on Organic Analysis (CCQM-OAWG) Terms of Reference. https://www.bipm.org/en/committees/cc/ccqm/wg/ccqm-oawg.
Accessed: 5/17/2024.

[46] Duewer DL, Murray JA, Wood LJ, Wise SA, Sander LC, Lippa K, Hein S, Koch M, Philipp R, Werneburg M, Hackenberg R, Polzer J, Avila MA, Serrano V, Kakoulides E, Alexopoulos C, Giannikopoulou P, Gui EM, Lu T, Teo TL, Hua T, Dazhou C, Chunxin L, Changjun Y, Hongmei L, Nammoonnoy J, Quinn L, Swiegelaar C, Fernandes-Whaley M, Gören AC, Gökçen T. CCQM-K95.1 Low-polarity analytes in a botanical matrix: polycyclic aromatic hydrocarbons (PAHs) in tea (2019) Metrologia 56(1A):08002. https://doi.org/10.1088/0026-1394/56/1A/08002.

[47] Duewer DL, Sander LC, Wise SA, Lippa K, Philipp R, Hein S, Hackenberg R, Polzer J, Avila MA, Serrano V, Kakoulides E, Alexopoulos C, Giannikopoulou P, Chan P-k, Lee H-m, Tang H-s, Tang P-o, Yip Y-c, Lu T, Cheow PS, Teo TL, do Rego ECP, Violante FGM, de Carvalho LJ, Leal RVP, Sega M, Rolle F, Baek S-Y, Kim B, Lee S, Cabillic J, Fallot C, Hua T, Dazhou C, Changjun Y, Chunxin L, Hongmei L, Itoh N, Quinn L, Prevoo-Franzsen D, Fernandes-Whaley M Gören AC, Gökçen T, Gündüz S, Krylov A, Mikheeva A,Baldan A, van der Hout JW,van der Veen AMH. CCQM-K131 Low-polarity analytes in a multicomponent organic solution: polycyclic aromatic hydrocarbons (PAHs) in acetonitrile (2019) Metrologia 56(1A):08003. https://doi.org/10.1088/0026-1394/56/1A/08003.

[48] Westwood S, Martos G, Ralf Josephs1, Choteau T, Wielgosz R, Davies S, Moawad M, Tarrant G, Chan B, Alamgir M, de Rego E, Wollinger W, Garrido B, Fernandes J, de Sena R, Oliveira R, Melanson J, Bates J, Le PM, Meija J, Quan C, Huang T, Zhang W, Ma R, Zhang S, Hao Y, He Y, Song S, Wang H, Su F, Zhang T, Li H, Lam W-h, Wong W-f, Fung W-h, Philipp R, Dorgerloh U, Meyer K, Piechotta C, Riede J, Westphalen T, Giannikopoulou P, Alexopoulos Ch, Kakoulides E, Kitamaki Y, Yamazaki T, Shimizu Y, Kuroe M, Numata M, Pérez-Castorena A, Balderas-Escamilla M, Garcia-Escalante J, Krylov A, Mikheeva A, Beliakov M, Palagina M, Tkachenko I, Spirin S, Smirnov V, Lin TT, Sze CP, Juan W, Lingkai W, Ting L, Qinde L, Yizhao C, Peng SL, Fernandes-Whaley M, Prevoo-Franzsen D, Quinn L, Nhlapo N, Mkhize D, Marajh D, Chamane S, Ahn S, Choi K, Lee S, Han J, Baek S-Y, Kim B, Marbumrung S, Jongmesuk P, Shearman K, Boonyakong C, Bilsel M, Gündüz S, Ün I, Yilmaz H, Bilsel G, Gökçen T, Clarkson C, John Warren J, Achtar E (2021) CCQM-K148.a Mass fraction assignment of Bisphenol-A high purity material. Metrologia 58(1A):08015. https://doi.org/10.1088/0026-1394/58/1A/08015.

[49] do Rego ECP, Rodrigues JM, de Carvalho LJ, Monteiro TM, da Cunha VS, Roa LJB, Morales LV, Ferreira E, Silva A, Barriga R, Sandoval S, Marajh D, Fernandes-Whaley M, Visser R, Sudsiri N, Boonyakong C, Gören AC, Ün I, Gündüz S, Yilmaz H, Niederhauser B, Stalder M, Cheow PS, Lu T, Teo TL, Liu Q, Melnyk D, Diumin V, Karbovska R, Levbarg O, Ovchar L, Rozhnov M, Tsyparenko S, Kadir HA, Nasir NH, Kakoulides E, Garcia SA, Calderon MA, Serrano V, Arce M (2019) Final Report on Key Comparison SIM.QM-K27: Ethanol in Aqueous Matrix. Metrologia 56(1A):0816. https://doi.org/10.1088/0026-1394/56/1A/08016.

[50] Nelson MA, Meija J, Toman B. CCQM-OAWG/100 Decision guide for selecting CCQM Model 1 key comparison reference value estimation procedures, OAWG General Document, Version 2 (2020). https://www.bipm.org/documents/20126/83755283/CCQM-OAWG-100-KCRV+QuickGuide_V2.pdf/396034b2-75ba-2f2a-9a63-8f1cec03141a

[51] Westwood S, Lippa K, Shimuzu Y, Lalerle B, Saito T, Duewer D, Dai X, Davies S, Ricci M, Baldan A, Lang B, Sarge S, Wang H, Pratt K, Josephs R, Mariassy M, Pfeifer D, Warren J, Bremser W, Ellison S, Toman B, Nelson M, Huang T, Fajgelj A, Gören A, Mackay L, Wielgosz R (2023) Methods for the SI-traceable value assignment of the purity of organic compounds (IUPAC Technical Report). Pure Appl Chem 95(1):1-77. https://doi.org/10.1515/pac-2020-0804.

[52] Westwood S, Josephs R, Choteau T, Daireaux A, Mesquida C, Wielgosz R, Rosso A, Ruiz de Arechavaleta M, Davies S, Wang H, Cristina Pires do Rego E, Marques Rodrigues J, de Freitas Guimarães E, Vinicius Barreto Sousa M, Maria Monteiro T, Alves das Neves Valente L, Gustavo Marques Violante F, Rubim Ribeiro Almeida R, Cristina Baptista Quaresma M, Nogueira R, Windust A, Dai X, Li X, Zhang W, Li M, Shao M, Wei C, Wong S-k, Cabillic J, Gantois F, Philipp R, Pfeifer D, Hein S, Klyk-Seitz U-A, Ishikawa K, Castro E, Gonzalez N, Krylov A, Lin TT, Kooi LT, Fernandes-Whaley M, Prévoo D, Archer M, Visser R, Nlhapo N, de Vos B, Ahn, Preeyaporn Pookrod, Kanjana Wiangnon, Nittaya Sudsiri, Kittiya Muaksang, Chainarong Cherdchu S, Gören AC, Bilsel M, LeGoff T, Bearden D, Bedner M, Duewer D, Hancock D, Lang B, Lippa K, Schantz M, Sieber J. Final report on key comparison CCQM-K55.b (aldrin): An international comparison of mass fraction purity assignment of aldrin. Metrologia 2012;49(1A):08014. https://doi.org/10.1088/0026-1394/49/1A/08014.

[53] Westwood S, Josephs R, Choteau T, Daireaux A, Wielgosz R, Davies S, Moad M, Chan B, Muñoz A, Conneely P, Ricci M, do Rego ECP, Garrido BC, Violante FGM, Windust A, Dai X, Huang T, Zhang W, Su F, Quan C, Wang H, Lo M-f, Wong W-f, Gantois F, Lalerle B, Dorgerloh U, Koch M, Klyk-Seitz U-A, Pfeifer D, Philipp R, Piechotta C, Recknagel S, Rothe, Taichi Yamazaki, Osman Bin Zakaria, E Castro, M Balderas, N González, C Salazar, L Regalado R, Valle E, Rodríguez L, Laguna LA, Ramírez P, Avila M, Ibarra J, Valle L, Pérez M, Arce M, Mitani Y, Konopelko L, Krylov A, Lopushanskaya E, Teo Lin TT, Liu Q, Kooi LT, Fernandes-Whaley M, Prevoo-Franzsen D, Nhlapo N, Visser R, Kim B, Lee H, Kankaew P, Pookrod P, Sudsiri N, Shearman K, Gören AC, Bilsel G, Yilmaz H, Bilsel M, Çergel M, Çoskun FG, Uysal E, Gündüz S, Ün I, Warren J, Bearden DW, Bedner M, Duewer DL, Lang BE, Lippa KA, Schantz MM, Sieber JR. (2014) Final report on key comparison CCQM-K55.c (L-(+)-Valine): characterization of organic substances for chemical purity. Metrologia 51(1A):08010. https://doi.org/10.1088/0026-1394/51/1A/08010.

[54] Li X, Li H, Zhang Q, Lu X, Li S, Koch M, Polzer J, Hackenber R, Moniruzzaman Md, Khan M, Kakoulides E, Richy KP-w, Chi-shing N, Lu T, Gui EM, Cheow PS, Teo TL, Rego E, Garrido B, Carvalho L, Leal R, Violante F, Baek S-Y, Lee S, Choi K, Kim B, Bucar-Miklavcic M, Hopley C, Nammoonnoy J, Murray J, Wilson W, Toman B, Itoh N, Gokcen T, Krylov A, Mikheeva A (2020) CCQM-K146 low-polarity analyte in high fat food: benzo[a]pyrene in olive oil. Metrologia 57(1A):08017. https://doi.org/10.1088/0026-1394/57/1A/08017.

[55] Sin DWM, Li H, Wong DSK, Lo MF, Wong YL, Wong YC, Mok CS (2015) Final report on mid-polarity analytes in food matrix: mid-polarity pesticides in tea (CCQM-K95). Metrologia 52(1A):08007. https://doi.org/10.1088/0026-1394/52/1A/08007.

[56] Ricci M, Shegunova P, Conneely P, Becker R, Torres MM, Osuna MR, On TP, Man LH, Baek S-Y, Kim B, Hopley C, Liscio C, Warren J, Le Diouron V, Lardy-Fontan S, Lalere B, Mingwu S, Kucklick J, Vamathevan V, Matsuyama S, Numata M, Brits M, Quinn L, Fernandes-Whaley M, Gören AC, Binici B, Konopelko L, Krylov A, Mikheeva A (2017) CCQM-K102: polybrominated diphenyl ethers in sediment. Metrologia 54(1A):08026. https://doi.org/10.1088/0026-1394/54/1A/08026.

[57] BIPM. CCQM Working Group on Gas Analysis (CCQM-GAWG) Terms of Reference. https://www.bipm.org/en/committees/cc/ccqm/wg/ccqm-gawg. Accessed: 5/17/2024.

[58] Guenther FR, Rhoderick GC, Miller WR, Marschal A, Medem A, Kato K, So H-Y, de Leer EWB, Baldan A, Hafkenscheid TL, Nieuwenkamp GN, van der Veen AMH, Konopelko L, Milton MJT (2002) Final report on key comparison CCQM-K7: benzene, toluene, m-xylene, o-xylene and ethylbenzene in nitrogen. Metrologia 39(1A):08006. https://doi.org/10.1088/0026-1394/39/1A/24.

[59] Rhoderick GC, Cecelski CE, Miller WR, Worton DR, Moreno S, Brewer PJ, Viallon J, Idrees F, Moussay P, Kim YD, Kim D, Lee S, Baldan A, Li J (2019) Stability of gaseous volatile organic compounds contained in gas cylinders with different internal wall treatments. Elementa (Wash DC) 7:10.1525/elementa.366. https://doi.org/10.1525/elementa.366.

[60] Guenther FR, Rhoderick GC, Marschal A, Medem A, Kato K, Heo GS, de Leer EWB, Baldan A, Hafkenscheid TL, Nieuwenkamp GN, van der Veen AMH, Konopelko L, Brookes C, d'Souza H, Milton MJT (2002) Final report on key comparison CCQM-K10: benzene, toluene, *o*-xylene in nitrogen. Metrologia 39(1A):8007. https://doi.org/10.1088/0026-1394/39/1A/25.

[61] Rhoderick GC, Duewer DL, Apel E; Baldan A, Hall B, Harling A, helmig D, Heo GS, Hueber J, Kim ME, Kim YD, Miller B, Montzka S, Reimer D (2014) International Comparison of a Hydrocarbon Gas Standard at the Picomol per Mol Level. Anal Chem 86(5):2580-2589. https://doi.org/10.1021/ac403761u.

[62] van der Veen AMH, Heine H-J, Brinkmann FNC, Ziel PR, de Leer EWB, Zhen WL, Kato K, Konopelko LA, Popova TA, Alexandrov YI, Kortchagina EN, Kustikov YA, Musil S, Milton MJT, Guenther F, Rhoderick G (2005) International Comparison CCQM-K16: Composition of natural gas types IV and V. Metrologia 42(1A):08003. https://doi.org/10.1088/0026-1394/42/1A/08003.

[63] Lim JS, Lee J, Moon D, Tshilongo J, Lee J, Qiao H, Shuguo H, Tiqiang Z, Kelley ME, Rhoderick GC, Konopelko LA, Kolobova AV, Vassermann II, Zavyalov SV, Gromova EV, amd Efremova OV (2017) International comparison CCQM-K113—noble gas mixture. Metrologia 54(1A):08017. https://doi.org/10.1088/0026-1394/54/1A/08017.

[64] Rhoderick GC, Kelley ME, Gameson L, Harris KJ, Hodges JT (2018) Issues with analyzing noble gases using gas chromatography with thermal conductivity detection. Anal Bioanal Chem 410:6247-6255. https://doi.org/10.1007/s00216-018-1235-z.

[65] Flores E, Viallon J, Choteau T, Moussay P, Idrees F, Wielgosz RI, Lee J, Zalewska E, Nieuwenkamp G, van der Veen A, Konopelko LA, Kustikov YA, Kolobova AV, Chubchenko YK, Efremova OV, Zhe BI, Zhou Z, Miller WR Jr, Rhoderick GC, Hodge JT, Shimosaka T, Aoki N, Hall B, Brewer P, Cieciora D, Sega M, Macé T, Fükő J, Szilágyi ZN, Büki T, Jozela MI, Ntsasa NG, Leshabane N, Tshilongo J, Johri P, Tarhan T (2019) CCQM-K120 (Carbon dioxide at background and urban level) Metrologia 56(1A):08001. https://doi.org/10.1088/0026-1394/56/1A/08001.

[66] Miller WR Jr, Rhoderick GC, Guenther FR (2014) Investigating Adsorption/Desorption of Carbon Dioxide in Aluminum Compressed Gas Cylinders. Anal Chem 87(3):1957-1962. https://doi.org/10.1021/ac504351b.

[67] Grenfell R, Brookes C, Vargha G, Quiney P, Milton M, Woods P, Harris P, Ballesta P, Hakola H, Sassi M, Patier R, Heo GS, Kim Y-D, Bar H-G, Rhoderick GC, Wessel J, Fröhlich M, Steinbrecher S. NPL Report AS 029 EURO.QM-S3 EURAMET 886 Comparison of multi-component ambient VOC measurements. https://eprintspublications.npl.co.uk/4306/1/AS29.pdf. Accessed: 5/17/2024.

[68] Cecelski CE, Rhoderick GC, Possolo AM, Carney J, Vokoun M, Privoznikova J, Lee S, Kang JH, Kim YD, Kim DH, Macé T, Sutour C, Pascale C, Ntsasa N, Tshilongo J, Jozela M, Leshabane N, Lekoto G, Worton DR, Brewer PJ, Farrow-Dunn F, Moreno S, Wirtz K, Stummer V, Konopelko LA, Kolobova AV, Kustikov YA, Klimov AY, Efremova OV, van Wijk JIT, van der Veen AMH (2022) International comparison CCQM-K10.2018: BTEX in nitrogen at 5 nmol mol^{-1}. Metrologia 59(1A):08003. https://doi.org/10.1088/0026-1394/59/1A/08003.

[69] Viallon J, Moussay P, Esler M, Wielgosz R, Bremser W, Novák J, Vokoun M, Botha A, Van Rensburg MJ, Zellweger C, Goldthorp S, Borowiak A, Lagler F, Walden J, Malgeri E, Sassi MP, Gomez PM, Patier RF, Madruga DG, Woo J-C, Kim YD, Macé T, Sutour C, Surget A, Niederhauser B, Schwaller D, Frig B, Váraljai IG, Hashimoto S, Mukai H, Tanimoto H, Ahleson HP, Egeløv A, Ladegard N, Marsteen L, Tørnkvist K, Guenther FR, Norris JE, Hafkenscheid TL, Van Rijn MM, Quincey P, Sweeney B, Langer S, Magnusson B, Bastian J, Stummer V, Fröhlich M, Wolf A, Konopelko LA, KustikovYA, Rumyanstev DV (2006) International Comparison CCQM-P28: Ozone at ambient level. Metrologia 43(1A):08010. https://doi.org/10.1088/0026-1394/43/1A/08010

[70] BIPM. International comparisons of ozone standards (BIPM.QM-K1). https://www.bipm.org/en/gas-metrology/ozone/bipm.qm-k1#. Accessed: 5/17/2024.

[71] US EPA (2021) Clean Air Status and Trends Network (CASTNET) Quality Assurance Project Plan, Revision 9.5. https://www3.epa.gov/castnet/docs/QAPP_v9-5_Main_Body.pdf.

[72] WMO No. 1143 (2014) The Global Atmosphere Watch Programme 25 years of global coordinated atmospheric composition observations and analyses. World Meteorological Organization. Geneva, Switzerland. https://library.wmo.int/doc_num.php?explnum_id=7886.

[73] Viallon J, Moussay P, Norris JE, Guenther FR, Wielgosz RI (2006) A study of systematic biases and measurement uncertainties in ozone mole fraction measurements with the NIST Standard Reference Photometer. Metrologia 43:441-450. https://doi.org/10.1088/0026-1394/43/5/016.

[74] Norris JE, Choquette SJ, Viallon J, Moussay P, Wielgosz RI, Guenther FR (2013) Temperature measurement and optical path-length bias improvement modifications to National Institute of Standards and Technology ozone reference standards. J Air Waste Manage, 63(5):565-574. http://dx.doi.org/10.1080/10962247.2013.773951.

[75] BIPM. CCQM Working Group on Electrochemical Analysis (CCQM-EAWG) Terms of Reference. https://www.bipm.org/en/committees/cc/ccqm/wg/ccqm-eawg. Accessed: 5/17/2024.

[76] Pratt KW, Koch WF, Wu YC, Berezansky (2001) Molality-Based Primary Standards of Electrolytic Conductivity (IUPAC Technical Report). Pure Appl Chem 73(11):1783-1793. https://doi.org/10.1351/pac200173111783.

[77] Pratt KW (2015) Standard Reference Materials®: a complement to pH key comparisons. Anal Bioanal Chem 407:3193-3203. https://doi.org/10.1007/s00216-014-8326-2.

[78] Spitzer P, Vyskočil L, Máriássy M, Pratt KW, Hongyu X, Dazhou C, Fanmin M, Kristensen HP, Hjelmer B, Rol PM, Nakamura S, Kim M, Torres M, Kozlowski W, Wyszynska J, Pawlina M, Karpov OV, Zdorikov N, Seyku E, Maximov I, Schmidt I, Eberhardt (2000) Final report for CCQM-K9 pH determination on two phosphate buffers by Harned cell measurements. https://www.bipm.org/kcdb/comparison/doc/download/85/ccqm-k9_final_report.pdf.

[79] Waters JF, Máriássy M, Bastkowski F, Asakai T, Prokunin SV, Gonzaga FB, Dimitrova L, Amaya RC, Quezada HT, Vospelova A, Rozikova M, Snedden A, Stoica D, Sander B, Jakusovszky B, Maksimov I, Onuma S, Suzuki T, Ruiz JM, Ortiz-Aparicio JL, Tincona G, Vasquez J, Pawlina M, Kozlowski W, Dobrovolskiy VI, Aprelev AV, Hanková Z, Rodruangthum P, Uysal E, Liv L, Nakiboglu N, Gavrilkin V, Petrenko A, Manska O, Easley RA, Fajardo S (2020) Key Comparison CCQM-K18.2016 pH of carbonate buffer. https://www.bipm.org/utils/common/pdf/final_reports/QM/K18/CCQM-K18.2016.pdf.

[80] BIPM (2000) Consultative Committee for Amount of Substance. Report of the 6th Meeting (6-7 April 2000). https://www.bipm.org/en/committees/cc/ccqm/6-2000.

[81] BIPM. CCQM Working Group on Surface Analysis (CCQM-SAWG) Terms of Reference. https://www.bipm.org/en/committees/cc/ccqm/wg/ccqm-sawg. Accessed: 5/17/2024.

[82] Newbury DE (2004) Quantitative Electron Probe Microanalysis (EPMA): Doing It Right and Doing It Wrong! https://www.ctcms.nist.gov/~cecamp/spring04/EPMA Right&Wrong.pdf. Accessed: 5/17/2024.

[83] National Institute of Standards and Technology (August 25, 1988) Certificate of Analysis, Standard Reference Material 482 Gold-Copper Wires for Microprobe Analysis. https://tsapps.nist.gov/srmext/certificates/482.pdf.

[84] Hodoroaba V-D, Marinenko RB, Small JA, Unger WES (2011) EPMA on Copper-Gold Alloys – A CCQM/SAWG Study. Microsc Microanal 17(S2):856-857. https://doi.org/10.1017/S1431927611005150.

[85] CCQM Working Group on Isotope Ratios (IRWG). Strategy for Rolling Programme Development (2021-2030). https://www.bipm.org/documents/20126/2071059/CCQM-IRWG+Strategy+document+2021-2030.pdf. Accessed: 5/17/2024.

[86] BIPM. CCQM Working Group on Cell Analysis (CCQM-CAWG) Terms of Reference. https://www.bipm.org/en/committees/cc/ccqm/wg/ccqm-cawg. Accessed: 5/17/2024.

[87] BIPM. CCQM Working Group on Cellular Analysis: Strategy 2021-2030. https://www.bipm.org/documents/20126/2071059/CCQM-CAWG Strategy document 2021-2030.pdf. Accessed: 5/20/2024.

[88] ISO 20391-2:2019 Biotechnology, Cell counting Part 2: Experimental design and statistical analysis to quantify counting method performance. International Standards Organization International Standard. Geneva, Switzerland.

[89] Sarkar S, Lund SP, Vyzasatya R, Vanguri P, Elliott JT, Plant AL, Lin-Gibson S (2017) Evaluating the quality of a cell counting measurement process via a dilution series experimental design. Cytotherapy, 19(12), 1509-1521. https://doi.org/10.1016/j.jcyt.2017.08.014.

[90] Pierce L, Newton D, Lund SP, Kurbanov F, Sarkar S (2023) A guided demonstration of the Counting Method Evaluation Tool (COMET) for implementing the ISO 20391-2 Cell Counting standard. Cell & Gene Therapy Insights 2023; 9(5), 581–609. https://doi.org/10.18609/cgti.2023.083.

[91 BIPM (2016) Consultative Committee for Amount of Substance: Metrology in chemistry and Biology (CCQM). Report of the 22nd meeting (21-22 April 2016) to the International Committee for Weights and Measures. https://www.bipm.org/documents/20126/30130094/CCQM22.pdf/2f796985-5bc5-e803-c11c-26a95787fd6e. Accessed: 5/17/2024.

[92] Olson ND, Lund SP, Zook JM, Rojas-Cornejo F, Beck B, Foy C, Huggett J, Whale AS, Sui Z, Baoutina A, Dobeson M, Partis L, Morrow JB (2015) International interlaboratory study comparing single organism 16S rRNA gene sequencing data: Beyond consensus sequence comparisons. Biomol Detect Quantif 3:17-24. https://doi.org/10.1016/j.bdq.2015.01.004.

Appendix A. List of Symbols, Abbreviations, and Acronyms

A.1. Abbreviations and Acronyms

AHWG	CIPM ad hoc Working Group on Chemical and Physico-Chemical Measurements
BAWG	Bioanalysis Working Group
BIPM	Bureau International des Poids et Mesures
CAWG	Cell Analysis Working Group
CCQM	Originally “Consultative Committee for the Quantity of Matter”; now “Consultative Committee for the Amount of Substance: Metrology in Chemistry and Biology”
CIPM	Comité International des Poids et Mesures
CMCs	Calibration and Measurement Capabilities
CRM	certified reference material
CSD	Chemical Sciences Division (an administrative organization within NIST)
CSTL	Chemical Science and Technology Laboratory (a prior administrative organization within NIST responsible for chemical and biological measurement programs)
DI	designated institute, an organization having the responsibility for a specified aspect of a nation’s measurement infrastructure. Only NMIs and DIs can participate in CCQM KCs.
DoE	degree of equivalence
EAWG	Electrochemical Analysis Working Group
GAWG	Gas Analysis Working Group
GSMG	Gas Sensing Metrology Group (an administrative organization with CSD)
IAWG	Inorganic Analysis Working Group
ICP-OES	inductively coupled plasma optical emission spectroscopy
IDMS	isotope dilution mass spectrometry
IRWG	Isotope Ratio Working Group
KC	Key Comparison
KCDB	Key Comparison Database
KCWG	Key Comparison Working Group
MAD_E	adjusted median absolute deviation from the median
MBSG	Microbiology Steering Group
NAWG	Nucleic Acid Analysis Working Group
NIST	National Institute of Standards and Technology
NMI	national metrology institute, an organization having broad responsibility for a nation’s measurement infrastructure not delegated to a DI. Only NMIs and DIs can participate in CCQM KCs.
OAWG	Organic Analysis Working Group
PAWG	Protein Analysis Working Group
PS	pilot study
PPS	published pilot study
qNMR	quantitative nuclear magnetic resonance spectroscopy

RGTM	Research Grade Test Material
RMO	Regional Metrology Organization
RMSE	root-mean-square error (mean residual of a regression model)
RV	reference value
SAWG	Surface Analysis Working Group
SC	Supplementary Comparison
SRM	Standard Reference Material, a NIST-certified CRM
TC	Technical Committee (of an RMO)
WG	Working Group
WGMC	CIPM Working Group on Metrology in Chemistry
XRF	X-ray fluorescence spectroscopy

A.2. Symbols

β_0	intercept of a line
β_1	slope of a line
d	difference between two values; degree of equivalence
$\%d$	relative degree of equivalence expressed as a percentage
D	Cartesian plane distance
i	index
K	constant, locally defined
n	number, locally defined
Q_n	a robust estimate of standard deviation
R	ratio of two values, locally defined
RV	reference value
t	t-statistic, the ratio of an uncertainty of a value to the value
u	standard uncertainty
u_{rel}	relative standard uncertainty
U_{95}	expanded uncertainty providing circa a 95 % level of confidence coverage
x	value of a variable
$\bar{x}$	arithmetic average
y	value of a variable
z	transformed value
ζ	zeta-score

A.3. Functions

cov(·,·)	covariance between two values
log(·)	logarithm of a value to the base 10, aka $\log_{10}(.)$
Median{·}	median value of a set of values
|·|	absolute value of a value

Appendix B. Codenames for National Metrology and Designated Institutes

The following information is derived from the contents of the Datacore_Codes worksheet.

Code	Country	Name
A*STAR	Singapore	Agency for Science, Technology and Research
ANSTO	Australia	Australian Nuclear Science and Technology Organisation
BAM	Germany	Bundesanstalt für Materialforschung und -prüfung
BelGIM	Belarus	Belorussian State Institute for Metrology
BFKH	Hungary	Government Office of the Capital City Budapest
BIM	Bulgaria	Bulgarian Institute of Metrology
BIPM	CIPM	Bureau International des Poids et Mesures
BJCDC	China	Beijing Municipal Center for Disease Prevention and Control
BSJ	Jamaica	Bureau of Standards
BVL	Germany	Federal Office of Consumer Protection and Food Safety
CCHEN	Chile	Comisión Chilena de Energía Nuclear
CEM	Spain	Centro Español de Metrología
CENAM	Mexico	Centro Nacional de Metrología
CERI	Japan	Chemicals Evaluation and Research Institute
CFIA	Canada	Canadian Food Inspection Agency
CHMI	Czech Republic	Ceský hydrometeorologický ústav
CMI	Czech Republic	Ceský metrologický institut
CMQ	Chile	Centro de Metrología Quimica de Fundación Chile
CMS	Chinese Taipei	Center for Measurement Standards
CODELCO	Chile	Chemical Laboratory of Codelco Norte Division
CSIRO(AR)	WMO	Commonwealth Scientific & Industrial Research Organisation Atmospheric Research
CSMMoeKR	Kyrgyz Republic	Center for Standardization and Metrology
DFM	Denmark	Danish Institute of Fundamental Metrology
DHMZ	Croatia	Meteorological and Hydrological Service
DMDM	Serbia	Directorate of Measures and Precious Metals
DMSc	Thailand	Department of Medical Science
DPL	Denmark	RadiometerA/S
DRiCM	Bangladesh	Designated Reference Institute for Chemical Measurements
DSM	Netherlands	Royal DSM N.V.
DWD	WMO	Deutscher Wetterdienst
EAA	Austria	Environment Agency Austria
EIM	Greece	Hellenic Institute of Metrology
EMPA	WMO	Swiss Federal Laboratories for Materials Science and Technology
ENEA	Italy	Ente per le Nuove tecnologie, l'Energia e l'Ambiente
ERLAP	EU	European Reference Laboratory of Air Pollution
ESRL	WMO	National Oceanic and Atmospheric Administration
EXHM	Greece	National Laboratory of Chemical Metrology
FMI	Finland	Finnish Meteorological Institute
FORCE	Denmark	FORCE Technology
FTMC	Lithuania	Center for Physical Sciences and Technology
FZK	WMO	Forschungszentrum Karlsruhe
GEOSTM	Georgia	Georgian National Agency for Standards and Metrology
GLHK	Hong Kong	Government Laboratory (Hong Kong)
GUM	Poland	Główny Urząd Miar
HSA	Singapore	Health Sciences Authority
IAEA	UN	International Atomic Energy Agency
IBMETRO	Bolivia	Instituto Boliviano de Metrología
IH-LQPM	Portugal	Instituto Hidrográfico - Laboratório de Química e Poluição do Meio Marinho
IJS	Slovenia	Jozef Stefan Institute
IMBiH	Bosnia-Herzegovina	Institute of Metrology of Bosnia-Herzegovina
INACAL	Peru	Instituto Nacional de Calidad
INECC	Mexico	Instituto nacional de ecología y cambio climático
INEN	Ecuador	Instituto Ecuatoriano de Normalización

Code	Country	Name
INM	Romania	National Institute of Metrology
INM(CO)	Colombia	Instituto Nacional de Metrología de Colombia
INMETRO	Brazil	Instituto Nacional de Metrologia, Normalização e Qualidade Industrial
INPL	Israel	National Physical Laboratory of Israel
INRAP	Tunisia	National Institute of Research and Physical and Chemical Analysis
INRIM	Italy	Istituto Nazionale di Ricerca Metrologica
INTI	Argentina	Instituto Nacional de Tecnología Industrial
IPQ	Portugal	Instituto Português da Qualidade
IRSN	France	Institut de Radioprotection et de Sûreté Nucléaire
ISCIII	Spain	Instituto de Salud Carlos III
ISP	Chile	Instituto de Salud Pública
ITI	Sri Lanka	Industrial Technology Institute
IW	Bosnia-Herzegovina	Institut za vode d.o.o. Bijeljina
JRC	EU	Joint Research Centre - JRC - European Commission
KazInStandart	Kazakhstan	Kazakhstan Institute of Standardization and Metrology
KEBS	Kenya	Kenya Bureau of Standards
KRISS	Republic of Korea	Korea Research Institute of Standards and Science
LACOMET	Costa Rica	Laboratorio Costarricense de Metrología
LATU	Uruguay	Laboratorio Tecnológico del Uruguay
LGC	UK	formerly “Laboratory of the Government Chemist”
LNE	France	Laboratoire National de Métrologie et d'Essais
METAS	Switzerland	Swiss Federal Office of Metrology and Accreditation
MSL	New Zealand	Measurement Standards laboratory of New Zealand
MUSSD	Sri Lanka	Measurement Units, Standards and Services Department
NBSM	Nepal	Nepal Bureau of Standards and Metrology
NIB	Slovenia	National Institute of Biology
NIBSC	UK	National Institute for Biological Standards and Control
NILU	Norway	Norwegian Institute for Air Research
NIM	China	National Institute of Metrology
NIMT	Thailand	National Institute of Metrology
NIS	Egypt	National Institute for Standards
NIST	USA	National Institute of Standards and Technology
NMIA	Australia	National Measurement Institute of Australia
NMIJ	Japan	National Metrology Institute of Japan
NMIM	Malaysia	National Metrology Institute of Malaysia
NMISA	South Africa	National Metrology Institute of South Africa
NMLPHIL	Philippines	National Metrology Laboratory of the Philippines
NPL	UK	National Physical Laboratory
NPLI	India	National Physical Laboratory of India
NRC	Canada	Institute for National Measurement Standards
OAP	Thailand	Office of Atoms for Peace
PTB	Germany	Physikalisch-Technische Bundesanstalt
RCC|LIPI	Indonesia	Research Center for Chemistry – Indonesian Institute of Sciences
RCM|LIPI	Indonesia	Research Center for Metrology-LIPI
RISE	Sweden	Research Institutes of Sweden AB
SASO-NMCC	Saudi Arabia	National Measurement and Calibration Center
SCK	EU	Studiecentrum voor Kernenergie
SL	Ireland	State Laboratory
SMU	Slovakia	Slovenský metrologický ústav
SNSU-BSN	Indonesia	National Measurement Standards, National Standardization Agency of Indonesia
STD|ITDI	Philippines	Standards and Testing Division, Industrial Technology Development Institute
SYKE	Finland	Finnish Environment Institute
TISTR	Thailand	Thailand Institute of Scientific and Technological Research
TTBS	Trinidad	Trinidad and Tobago Bureau of Standards
UBA(DE)	Germany	Umweltbundesamt
UME	Türkiye	TÜBITAK Ulusal Metroloji Enstitüsü

Code	Country	Name
UMTS	Ukraine	All-Ukrainian State Scientific and Production Center of Standardization, Metrology, Certification and Consumer Protection
VMI	Vietnam	Vietnam Metrology Institute
VNIIFTRI	Russian Federation	Institute for Physical-Technical and Radiotechnical Measurements
VNIIM	Russian Federation	D.I. Mendeleyev Institute for Metrology
VSL	Netherlands	Van Swinden Laboratorium
ZRS_Koper	Slovenia	Science and Research Center Koper

Appendix C. Where the Figures and Tables Came From

All information presented in this document's figures and tables is current as of yhis document's publication date.

C.1. Figures

With the exception of Fig. 4, all of the figures in this document were generated using one of the following *CCQM_Retrospectoscope* subsystems:

- Dataset_Review: Dot-and-bar plots of datasets (e.g., Fig. 1 to Fig. 3, Fig. 18). The "thumbnail" plot size and multiplot versions were produced with the aid of one or more of the "Document_" subroutines in the "C02_OddsnEnds" module.
- Lab_Activity: Relative proportions of datasets and studies in each WG (Fig. 5).
- Lab_Engagements: Cumulative number of engagements (participations and coordinations) with the CCQM as function of time (e.g., with some editing Fig. 6 and Fig. 8, without editing Fig. 10 and Fig. 11). This subsystem uses WG_Participations and WG_Coordinations.
- WG_Participations: Relative proportion of possible participations as a function of time (e.g., Fig. 7). This subsystem is used by Lab_Engagements.
- WG_Coordinations: Relative proportion of possible coordinations as a function of time (e.g., Fig. 9). This subsystem is used by Lab_Engagements.
- Lab_History: Bias and associated uncertainty as functions of time (e.g., Fig. 12 to Fig. 14).
- Lab_Bias: Bias and associated uncertainty as functions of measurand level (e.g., Fig. 15 and Fig. 16).
- Dataset_RefVals: Reference values as a function of time (e.g., Fig. 17).

Figure 4 is from the BIPM's Annual Review: Supplement, Activities of the BIPM Departments, 1 January 2022 — 31 December 2022, file: SP-report-2022 -annual review.pdf.

C.2. Tables

With the exception of Table 9, all of the tables in this document were generated using one of the following CCQM_Retrospectoscope subsystems:

- Database_Checkup: Summaries of the number of datasets and studies by sponsoring body, WG, study type, and base units (Table 1 to Table 4).
- Dataset_Locate: Number of finalized studies and their datasets (e.g., Table 5, Table 6). The information in Table 5 is also produced by the Lab_Activity subsystem.

Table 9 is derived from information in the CCQM_KC worksheet, which in turn is derived from results contained in the CCQM-K113 Final Report [63].