Standards Services Seminar Series

Comparing Key Comparisons: Issues and Examples

Thursday, July 10, 2014

ABSTRACTS

David Duewer — *Comparing Key Comparisons*

"Key Comparisons (KCs)" are inter-laboratory comparison studies conducted by national and international metrology institutions under the supervision of the International Committee for Weights and Measures (CIPM). All KCs share two attributes that differentiate them from other inter-labs in which NIST occasionally participates: (1) all KC participants are our metrological peers, and (2) fully attributed KC results become publicly available on the acceptance of the study's Final Report. Beyond this commonality, KC study designs and preferred evaluation methods differ widely both within and among the nine major study areas recognized by the CIPM. Regardless of study area, NIST's performance in any KC impacts our reputation for measurement excellence and our leadership role within the international community. It is thus important that we all have confidence in the reliability of the data analysis and statistical methods used to evaluate KC results, and that these methods are among the most scientifically appropriate for the given study. The following presentations illuminate some of these issues as raised in two study areas.

Lisa Karam — *Example(±s) from Radionuclide Metrology: The Système International de Référence*

The major international comparison of gamma-emitting radionuclides is the Système International de Référence (SIR), piloted by the BIPM. In this comparison, data is accumulated over many years, adding to the Key Comparison Reference Value (KCRV) and its uncertainty over time. The uncertainty on a radioactivity measurement is most heavily influenced by technical aspects (i.e., non-statistical means of evaluation), which will vary among the different laboratories participating in the SIR key comparison. An optimal method for determining the KCRV and its uncertainty has been recently adopted by the Consultative Committee on Ionizing Radiation Section on Radionuclide Metrology [CCRI(II)], which is an expanded application of the Mandel-Paule evaluation method, and which balances efficiency with reliability. Using this approach, the CCRI(II) has linked results from an international (K2) comparison of a solution of Zn-65 with the KCRV from the SIR (K1) comparison for that nuclide.

Greg Strouse — *Key Comparisons in Thermometry*

The increase in the number of Regional Metrology Organization Key Comparisons (RMO-KCs), particularly within SIM, is a natural extension of completed and ongoing Consultative Committee Key Comparisons. These RMO-KCs are used to meet the MRA requirements to substantiate calibration measurement capability statements (CMCs). The Consultative Committee for Thermometry Working Group for KCs (CCT WG-KC) implements a stringent review and approval criteria as defined by the MRA for a comparison from the start (protocol generation) to completion (acceptance of the Final Report). A recent NIST-piloted humidity RMO-KC (RMO K6a) was found to be unacceptable by the CCT KC-WG on the grounds that a Bayesian uncertainty analysis is not GUM compliant and that a Frequentist uncertainty analysis was required. The results of two Bayesian and one Frequentist results are presented. As the results of the three methods show, the differences in the uncertainty values do not appear significant nor do the results impact the acceptance of the proposed CMCs.

Antonio Possolo — *Statistical Data Reductions for Key Comparisons*

Inter-lab comparisons may serve to evaluate measurement performance, to estimate a consensus value to be assigned to a measurand, or both. They provide ingredients for top-down characterizations of measurement uncertainty: for example, to detect and quantify lurking components of *dark uncertainty*. Selecting which data to use, and which to set aside, is principally a substantive, not a statistical task. The choice of statistical model to describe the dispersion of replicated values within labs, and of measured values between labs, and the purpose of the study, together determine how the data should be reduced, and how the results should be presented. Each comparison demands and deserves customized data reduction, and in fact there is no single statistical model, or data analysis method, that is universally best. In some cases, a consensus value is of no interest, and the focus is on how the labs perform collectively, or on how the values they measure differ from one another. The probability distributions that are used to qualify these differences, and measurement uncertainty generally, may be interpreted as conveying states of individual or collective knowledge about a measurand, or about measurement performance.