Achieving and Assessing Acceptable Analytical Performance: The Challenge of Matrix Effects

Fred D. Lasky, Ph.D.*
Stephen G. Daggett, Ph.D.
Johnson & Johnson Clinical Diagnostics, Inc.
Rochester, New York

***Presenting Author**

Abstract: Accurate and reliable routine *in vitro* diagnostic testing is needed by physicians to help provide appropriate care for their patients. These goals, which have been common to both health care professionals and manufacturers, are now mandated by law. Achieving and assessing acceptable performance have a number of common challenges. Once acceptable levels of accuracy are defined, performance can be evaluated by component: precision; bias; specificity (random interferences); and stability. Each component affects our ability to achieve and maintain performance within an acceptable window. A key element in achieving and assessing the accuracy of test results is the use of well characterized reference or designated comparison methods that are performed under stringent process control. A feature often overlooked in the analytical process is the specification of sample matrices to be analyzed. Differences in composition between patient samples and processed fluids present a challenge for calibrating and assessing performance of routine methods. To manage these differences in sample matrices, we have demonstrated success in establishing calibration of routine methods with patient samples, so that routine methods correlate with designated methods. From these correlations, values are assigned to calibrators. Effective process control contributes to vial-to-vial uniformity. Calibration fluid stability is enhanced by saccharide stabilizers that displace outer-sphere protein-bound water, which aids effective lyophilization. Calibrator set points can be established efficiently when analyte concentrations or activities are adequately recovered after reconstitution of lyophilate. Compatibility of components (analytes, stabilizers and additives) is also desired to prepare economic, multi-purpose fluids. This is a significant challenge that also faces proficiency testing providers, whose fluids are similar in preparation and composition. Alternative strategies to traditional proficiency testing schemes (using lyophilized fluids) are achieving good success. Individual or pooled patient samples may better demonstrate method performance at clinically significant concentrations, although they do increases risk of biohazard exposure. This approach is also consistent with manufacturers' efforts to develop methods that perform well with patient samples. Although processed fluids may provide an adequate tool to assess consistency of results across laboratories for similar methods and instruments, assessing accuracy will continue to require the use of patient samples.

Introduction

Accurate and reliable *in vitro* diagnostic testing is needed by physicians to help provide appropriate care for their patients. These goals are now mandated by law! Achieving and assessing acceptable levels of accuracy have several common challenges, especially when artificial, processed fluids

			Source of Variability	
Analyte	Target	Total C.V.	Lyophilizer	Vial & Rep
CHOL	139.3 mg/dL	1.58%	1.93%	98.07%
GLU	90.9 mg/dL	0.69%	6.12%	93.88%
Na^+	122.9 mmol/L	0.44%	2.79%	97.20%
ALKP	98.8 U/L	0.94%	13.81%	86.18%
CK	168.7 U/L	2.33%	7.43%	92.56%

Table 1:Lyophilizer Qualification: Variance Component Analysis

are used. Although there is no single, agreed-upon standard, acceptable levels of accuracy must be established for meaningful evaluation. To better understand and control the total allowable error (or acceptable accuracy), we use a model in which error components are identified as precision, bias, specificity (random interferences) and stability.[1]

Desirable characteristics of processed fluids include uniformity, stability, analyte recovery and compatibility with the reagent and instrument. When used as calibrators, processed fluids affect bias, laboratory-to-laboratory precision and stability. The same characteristics affect the perception of system performance when they are used in proficiency testing (PT) programs.

Achieving Accuracy

A key element in achieving the accuracy of test results is the use of well characterized reference or designated comparison methods that are performed under stringent process control, such as described in the ISO 25 Guidance for *General Requirements for the Competence of Calibration and Testing Laboratories*. A feature often overlooked in the analytical process is specification of sample matrices to be analyzed. Additionally, performance limits for reference methods must be established and maintained at significantly more stringent levels than what is expected of routine methods.[2,3] International Reference Preparations (IRP) may be used where reference methods are not available or are unlikely to be developed. Differences in matrices (matrix effects) between IRP and patient samples as well as between IRP batches present an additional challenge in calibration.

We have demonstrated success in maintaining acceptable levels of accuracy in routine methods by establishing calibration through correlation with our designated methods using patient samples.[4] Using these correlations, values are assigned to calibrators to transfer comparable performance from the factory to the field. Calibrator properties, therefore, affect performance.

Fluid Manufacturing; Process Challenges

Manufacturing fluids that meet the requirements of calibrators is a challenge that requires careful product design and process control. Some characteristics, such as uniformity, clarity and reconstitution time, are related to controlling the lyophilization process. We demonstrated that our process is capable of acceptable vial-to-vial uniformity, with volumetric transfers and analyte measurements contributing more

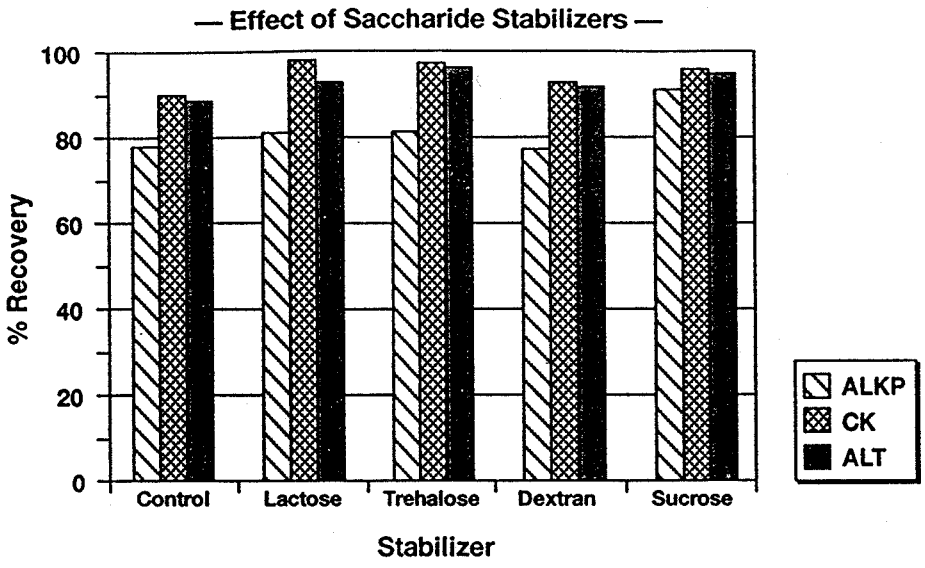

Figure 1. Enzyme Recovery in Reconstituted Fluid

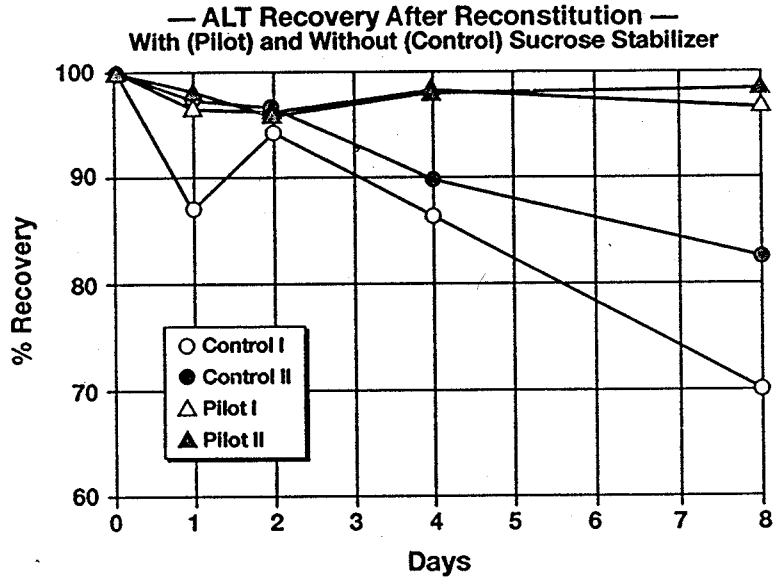

Figure 2. Lyophilate Stability at 50°C

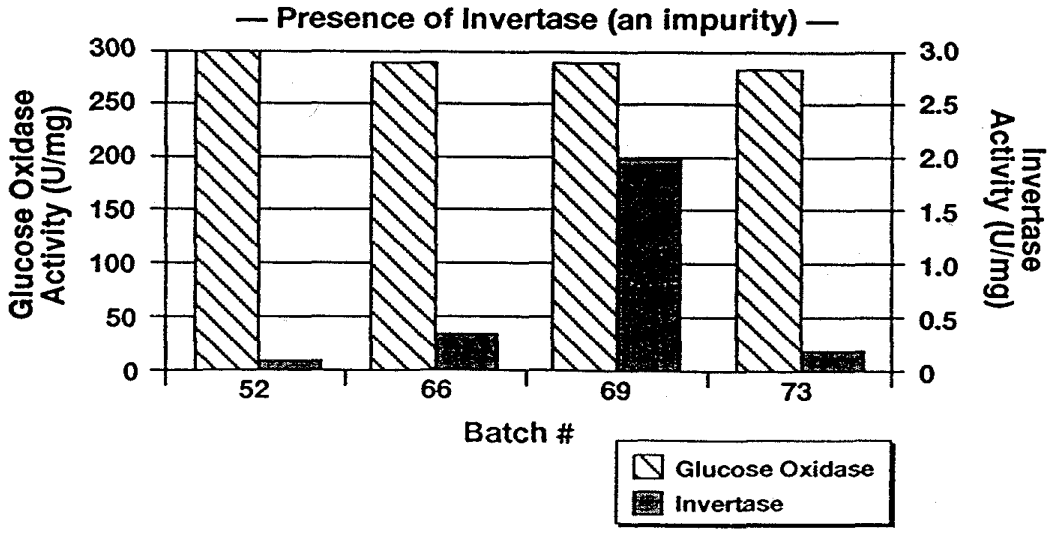

Figure 3. Glucose Oxidase Batch Analysis

variability than lyophilization (Table 1).[5] As expected, protein analytes are more sensitive to manipulation and processing.

Caution must be used when supplements are added, e.g., analytes, stabilizers and additives, to prepare economical, multi-purpose fluids. During development of an enzyme calibrator, unexpected inhibition of CK was observed when amylase was added to the pilot mix. CK activity dropped from 620 U/L (control) to 35 U/L when amylase was added. An alternative supplier's material was satisfactory; CK = 622 U/L. Porcine was the source for all enzymes, heart for CK and pancreas for amylase.

Calibrator stability is enhanced by adding saccharides, which displace outer-sphere protein-bound water and reduce collapse of the cake during lyophilization. Sucrose was selected as a candidate because of better enzyme recovery after reconstitution (Figure 1). Furthermore, lyophilate-enhanced stability was observed in an accelerated storage test at 50°C (Figure 2).

Significant variability, however, was noted between reagent lots of glucose slides using a pilot calibrator preparation (differences up to 50 mg/dL at 130 mg glucose/dL). Raw material batch analysis of glucose oxidase (GO), the active reagent, determined the presence of invertase, an impurity in GO, varying by batch (Figure 3). Invertase converts sucrose to glucose and fructose, thus causing an artifactual increase in the amount of substrate measured.

These same challenges face PT providers, whose fluids are similar in manufacture, preparation and composition to our calibrators. These experiences demonstrate some potential pitfalls in validating the

			Lyophilized			Fresh Serum			
Analyte	Units	AMM*	Interval	AMM	Interval	AMM	Interval	AMM	Interval
Cl^-	mmol/L	99	15	125	16	101	9	102	10
Na^+	mmol/L	134	15	155	16	141	10	141	10
Creatinine	µmol/L	89	75	280	75	89	30	188	50

*AMM=All Method Mean

Table 2. Interval Covering 95% of Participants' Results

	Cholesterol				HDL-Cholesterol	
Sample:	1	2	3	4	1	2
NY State Wadsworth Center	5.67	6.47	4.82	6.14	1.44	1.17
AMM	5.61	6.47	4.75	6.27	1.41	1.15

Table 3: Participants' Results for Cholesterol and HDL-cholesterol (mmol/L)

compatibility of fluid components and formulations with reagent lot composition.

Assessing Accuracy

Several strategies have been attempted to better achieve PT's objective, which is to provide a measure of test system reliability and accuracy. CAP evaluated lyophilized bovine and human serum matrices. Protein-based analytes showed no decrease in variability in the human matrix, while human serum occasionally had greater variability than bovine.[6] We had similar experiences where the matrix (human, bovine or goat serum, or BSA) has less effect on obtaining desirable fluid characteristics than effective control of manufacturing processes.

Fresh human samples (with and without supplementation) have been tried by several PT providers with good success. CAP has a study ongoing; the Veterans Administration (VA) has a program in which both lyophilate and fresh samples are used. Ontario's Laboratory Proficiency Testing Program has published results showing that consistency was significantly better for fresh serum than for lyophilate. [7](see Table 2)

The all-method means for fresh serum cholesterol and HDL cholesterol were remarkably close to results from a CDC-network reference laboratory. (see Table 3)

Additionally, all methods demonstrated cholesterol performance with biases less than 1.5% which was well within the NCEP goal of 3%.

Patient samples better demonstrate method performance at clinically significant concentrations, although they do increase the risk of biohazard exposure.

Improving Program Utility

Using patient samples in PT programs is also consistent with manufacturers' efforts to develop methods that perform well with clinical samples. System changes designed to improve performance with patient samples might, by serendipity, also result in better PT performance. For example, method-specific

means moved closer to the all-method mean when a manufacturer enhanced the on-analyzer stability of its phosphorus method; another changed the read wavelength to improve analyzer model-to-model consistency for glucose. In both cases, however, the objective was to improve method operational characteristics or performance.

From time to time, changes are made in PT fluid manufacture to better assess performance of a method or method group. We worked with CAP to include bicarbonate diluents. A 10% negative bias was eliminated from our urea results when the broad physiologically expected range of CO_2 was present. Hitachi's diluted Cl^- ISE method also improved.

Improving, validating and controlling changes in test systems for and with processed fluids is a daunting task. Reagent changes (suppliers, raw materials, process improvements), combined with changes in fluid batches, matrices (from program to program) and the variety of component additions, make the number of independent variables that must be controlled unmanageable. Differences in results between processed and "native" serum are not surprising when test systems are optimized for use with patient samples; a frequent reminder that defining the appropriate sample -- patient samples -- is a fundamental principle of metrology.

Quality Opportunities

To determine where opportunities lie, we surveyed 17 clinical chemists with the following question: What are the five most important quality issues that are associated with (1) overall clinical laboratory operations (from test request to result utilization), and (2) are broadly related to the analytical services provided by your laboratory which would most benefit the patient if resolved now?

Analytical element concerns were only 13% of the responses, while pre- and post-analytical opportunities were noted 40% and 46%, respectively. Within the laboratory, responses were more diffuse. Of 78 responses, the top issues (with frequency) were: Equivalent results across methods (12); Enhance reagent stability & reduce variability (10); and Improve personnel training & competency. Improved PT was far down the list, mentioned only twice.

We asked representatives of five major manufacturers of chemistry systems about the objectives of their improvement and development programs. Their responses were consistent: Such programs are aimed to improve performance with patient specimens. Not one manufacturer could recall an improvement program being initiated solely because of PT results.

Observations

Because processed fluid manufacturing is so dependent on external factors, such as matrix and variable attempts to mimic the physiologic composition of human serum, we believe its use in assessing method accuracy is fraught with insurmountable limitations. Despite this, PT continues to provide a good assessment of laboratory-to-laboratory *consistency* for total test systems.

Fresh specimens provide better assessments of repeatability and accuracy because methods are designed for these samples. Additionally, assessing accuracy at the clinically important concentrations is manageable with the use of fresh patient samples.

Finally, the benefits of any improvements in reagent specificity or fluid processing to

enable traditional PT programs, which use lyophilized materials, to better assess method accuracy and reliability must be weighed against the costs of diverting resources needed to develop new tests. We surveyed five leading manufacturers of clinical systems, and *none* has ever directed improvement efforts to anything but patient performance. Not surprising, especially now that health care priorities are being critically scrutinized!

Recommendations

Programs are needed to assess the reliability of laboratory tests. With limited laboratory resources, however, priorities must be established *analyte by analyte* to determine which we deal with first. Only then should limits be established for acceptable repeatability and accuracy. Then, costs of establishing a Reference Laboratory Network for the critical analytes need to be determined. This network should be international in scope; include manufacturers, which have resources and often special expertise; require stringent process and procedural control; and require participating laboratory performance to be *significantly better* than routine methods. Determine the effectiveness of alternative strategies, such as having the manufacturer verify *accuracy* of its systems through the network asfrequently as necessary, while clinical laboratories continue to verify *consistency* across laboratories for similar methods and instruments through traditional PT programs. Occasionally, fresh samples (or pools) should be included in PT programs to verify accuracy, especially at critical concentrations. Fresh samples can be relied upon by the laboratory, manufacturer, and government agencies to assess results on the same samples for which the systems are designed and on which the physician depends for patient evaluation.

References

1. Lawton WH, Sylvestre EA, Young-Ferraro BJ. Statistical comparison of multiple analytic procedures: Application to clinical chemistry. *Technometrics* 1979:21;397-409.

2. Bennett ST, Eckfeldt JH, Belcher JD, Connelly DP. Certification of cholesterol measurements by the National Reference Method Network with routine clinical specimens: Effect of network laboratory bias and imprecision. *Clin Chem*. 1992:38;651-657.

3. Fricker RF, Ritter M, Lasky FD, Greenberg N. Comparison of alternative statistical approaches to external quality assessment of cholesterol performance. *Clin Chem*. 1992;38:1028 [Abstract]

4. Lasky FD. Achieving accuracy for routine clinical chemistry methods by using patient specimen correlations to assign calibrator values: A means of managing matrix effects. *Arch Pathol Lab Med*. 1993;117:412-419.

5. Henderson RC, Peters C. Performance of calibrator and control fluids manufactured using high performance lyophilizers. *Clin Chem*. 1995:41;S215 [Abstract]

6. Arkin C. Man v. beast: Comparing human and bovine QC materials. *CAP Today*. 1994;8:49-51.

7. Ontario Laboratory Proficiency Testing Program (LPTP). Review of activities.

1994.

8. Laboratory Proficiency Testing Program. Review of activities. Toronto; LPTP, 1994:36.