



**Article:**

H. Stepman, U. Tiikkainen, D. Stöckl, H. Vesper, S. Edwards, H. Laitinen, J. Pelanti, and L. Thienpont.

*Measurements for 8 Common Analytes in Native Sera Identify Inadequate Standardization among 6 Routine Laboratory Assays.*

Clin Chem 2014;60:855-863.

<http://www.clinchem.org/content/60/6/855.abstract>

**Guest:**

Dr. Linda Thienpont is Professor in Analytical Chemistry, Statistics, and Quality Control at Ghent University in Belgium.

Bob Barrett:

This is a podcast from *Clinical Chemistry* sponsored by the Department of Laboratory Medicine at Boston Children's Hospital. I'm Bob Barrett.

Performing measurements that are comparable over time and location and across assays is essential for insuring appropriate clinical and public health practice. One step toward achieving this goal is using assays that are traceable to a higher-order reference measurement system or harmonized by using internationally recognized procedures.

Proficiency testing, also called external quality assessment, plays a key role in this regard and has earned a well-deserved position in the quality management system of laboratories. Unfortunately, proficiency testing materials are often non-commutable, meaning that they do not behave exactly like real patient samples and conclusions drawn from using such materials are limited. The June 2014 issue of *Clinical Chemistry* published a study of a proficiency testing exercise using freshly collected patient serum, that sheds a light on problems that we may have assumed we did not have, and possibly help make things better.

We're joined by the corresponding author of that paper, Dr. Linda Thienpont. She is Professor in Analytical Chemistry, Statistics, and Quality Control at Ghent University in Belgium.

Professor, there have been a number of recent reports on proficiency testing surveys with voluntary participation by laboratories. Of course, your study is one of them. Can you comment on this trend?

Dr. Thienpont:

Well, laboratory medicine has an increasing share in the development, implementation, and control of global healthcare policies. I only refer to a few recent developments such as evidence-based clinical practice guidelines for application of consistent standards of medical care, translation of research into patient care and disease prevention activities, and introduction of electronic patient

health records. Of course, this brings along new needs for medical laboratory practice. Some of the major ones are accessibility of the data, standardization, and interoperability between laboratories.

As you said, many initiatives respond to these new demands, only to mention the IFCC standardization and AACC harmonization projects. But also, proficiency testing recognizes these needs and therefore increasingly focuses on high quality of samples and target values. Examples are the CAT accuracy surveys in the US, or also for example, the Dutch external-quality assessment scheme, all with the final goal of reaching interchangeability of data across laboratories and assays. And I think our study perfectly fits into the picture.

Bob Barrett: What is common to these recent proficiency testing services and what were the specific characteristics of your study?

Dr. Thienpont: Common to the recent trend in the proficiency testing services is the use of high quality samples and targets. Nevertheless, our approach has some distinctively different features. For example, we consequently use a panel of 20 fresh-frozen single-donation samples prepared by the CLSI C37-A protocol, but without filtering and pooling. This design offers the additional potential to assess sample-related effects and total error. Indeed, pooling potentially dilutes matrix interferences present in individual samples.

Different from other services, we of course cannot serve each individual laboratory in a region or country. But instead, we have to go for the "pars pro toto," as we call it. This means that we work, we have a restricted of laboratories. The C37 protocol yields a volume of a maximum 200ml of sample which only allows assessment of the most prominent test systems on the market, and participation of maximally 160 laboratories, given we provide each of them with 1ml of sample per survey. The remaining volume is used for target setting.

We also restrict participation to laboratories using homogenous systems, which means platform, reagent, and calibrator from the same manufacturer. In spite of this limitation, we believe that the inferred performance is sufficiently representative of the test systems in the hands of other laboratories.

And please allow me a special remark related to the targets we use, particularly because we got some critiques on that. As reference laboratories, we surely know that referenced measurement values are the preferred target.

However, you also know that providing reference method values for eight analytes on 20 samples is extraordinarily expensive. Therefore, we decided to go for the most cost-effective solution, which means that we use those targets only when needed.

Let me explain that a little bit better. Suppose we see a very good comparability between different peer groups for a certain analyte, well then, we can take this as an indication that the market is sufficiently standardized and we can work with the all-method trimmed mean. In such a situation, that mean indeed proves a valuable target for the assessment of assay comparability, at least for what concerns the common analytes we investigated. Another very special aspect of our study was that the IVD manufacturers were a genuine part of the survey.

Bob Barrett: Well, doctor, you mentioned the importance of in vitro diagnostic manufacturers being a part of the study. What were the major contributions of IVD manufacturers to this work?

Dr. Thienpont: Well, Bob, I'm glad you gave me the opportunity to better explain this special aspect of our study, because indeed, there was no room to include it in the paper.

First, I would like to emphasize that our group considers it essential to directly work with the IVD industry towards standardization of methods or improvement of quality. This was the reason why, for this study, we again invited the IVD manufacturers of the assessed peer group. And we were fortunate because all of them accepted. This was very positive, first, because it showed their principle interest in experimentally assessing how well the intrinsic quality of their assays is reproduced in a clinical setting; but secondly, because it was promising to us that we both have the opportunity to establish a bottom-up collaboration of ourselves with manufacturers and laboratories, if the study would show the need for a global solution of certain problems. This is exactly what we currently do after closing the study. Therefore, we cannot sufficiently underline how grateful we are for the participation of the manufacturers.

This being said, I also see it as an ideal place to highlight how much appreciation we have for the motivation of the participating laboratories. Indeed, the challenge and willingness to measure eight analytes in 20 samples on a voluntary basis should not be underestimated. We currently are preparing the next survey, and I'm very proud that I can already tell you that the eagerness of the laboratories to participate is, again, overwhelming.

Bob Barrett: Okay. Well, let's go back to your recent paper in *Clinical Chemistry*. Could you summarize the overall experimental design and highlight the main findings of the study?

Dr. Thienpont: Well, as I mentioned already, we use a panel of 20 fresh-frozen single-donation serum samples to assess assays for the measurement of creatinine, glucose, phosphate, uric acid, triglycerides, total cholesterol, HDL and LDL cholesterol. The commercial random-access platforms we included were the Abbott Architect, the Beckman Coulter AU, the Ortho Vitros, Roche Cobas, Siemens Advia, and Thermo Scientific Konelab.

The assessment was done first at the peer group level, and second, by comparison against the all-method trimmed mean. Respectively, the reference method values for cholesterol, creatinine, and uric acid. Quality indicators were intra-assay imprecision, combined imprecision including sample matrix interference, bias, and total error. Fail/pass decisions were based on limits reflecting state-of-the-art performance, but also limits related to biological variation.

The outcome was that most assays showed excellent peer performance attributes. However, assays for HDL and LDL cholesterol were an exception. In contrast, comparison of the results between peer groups revealed individual assays that had biases exceeding the used state-of-the-art limits. This was particularly the case for creatinine, phosphate, triglycerides, uric acid, and LDL cholesterol.

In addition, it was striking to find differences between assays up to 8%, for example, for creatinine and HDL cholesterol. Also, concentration-related biases were frequently observed, again, particularly for HDL cholesterol. And although not shown, the latter, let's presume that more serious biases on samples with pathological concentrations will be seen.

The data also showed that for several analytes, most of the assays do not yet meet the optimal bias limit necessary for clinical use, except for phosphate, LDL cholesterol, triglycerides, and uric acid. It was particularly striking to see these applied also for cholesterol and creatinine in spite of the efforts done in dedicated standardization programs such as the National Cholesterol Education Program and the National Kidney Disease Education Program.

These unexpected observations about standardization problems or a lack of interchangeability of results, even for the here-assessed simple clinical chemistry analytes, perfectly highlights, in our opinion, the utility of organizing these specially designed surveys.

In virtue of the design with 20 single-donation samples, our study also exceptionally gave insight into so-called, "sample-related effects." These were inferred from estimating the combined random error components from comparison of the results against the all-method trimmed mean or reference method targets, and accounting for the method imprecision at the peer group level.

The sample-related effects are determining the total error at the level of the individual sample and reflect a method-analytical specificity. For cholesterol, creatinine, glucose, phosphate, and uric acid, the latter showed sufficient. In contrast, again, for the HDL and LDL cholesterol assays, the estimates for the combined imprecision was much higher, pointing to considerable sample-related effects. These observations, complemented with the afore-mentioned peer group attributes, showed that the HDL and LDL cholesterol assays have not yet reached the quality present for the other analytes.

Finally, the study showed large differences in quality of performance between laboratories. Particularly compelling was the observation of huge inter-laboratory differences, admittedly partly caused by the observed calibration differences between manufacturers. However, the magnitude of differences above 30% pointed to the additional existence of large laboratory effects.

In conclusion, although our study demonstrated in general robustness of the assays intrinsic quality for satisfactory performance in a daily laboratory context, it also pointed to the need for improvement of certain quality aspects even for the simple clinical chemistry analytes assessed in our study. In particular, the interchangeability of results remains jeopardized both by assays standardization issues and individual laboratory effects.

Bob Barrett: Looking ahead, doctor, do you plan to conduct these surveys regularly? And how do you position them in the overall proficiency testing landscape?

Dr. Thienpont: Well, Bob, before looking ahead, please allow me a look back. It's now 20 years since we proposed in literature the concept of two complementary proficiency testing approach in a publication entitled "The combined-target approach: a way out of the proficiency testing dilemma." We tried to convince the community to embrace this way of thinking. However, I must admit we had limited success. And it took us several years to understand why. But we now think the problem was in the missing link between the original scientific idea and its realization and practice.

We looked and found similar problems and solutions in other disciplines. For example, in medicine, the term “translational medicine” was coined to speed up the process from scientific discovery into routine physician practice. And we think that the magic bullet for speeding up a process is the development of a product that can be sold on the market. Yes, we now see it as easy as that.

We realized, however, that we had to take the risk of developing and marketing the product our concept needed. So we decided to do it and called it the “Master Comparisons.” But these are only a part of a greater empower project that additionally monitors outpatient percentiles across laboratories and manufacturers. That part, we called, “The Percentiler”.

You should understand that the Master Comparisons are only a point estimate of quality and comparability, while outpatient percentiles are a potential for online monitoring of the comparability of IVD tests on samples that are as commutable as can be; namely, real patient samples.

Bob Barrett: This sounds very exciting. But as I understand, all of these products are currently available free to participants. So you’ll face challenges to make that sustainable. How will you address these issues? In plain words, will such products sell?

Dr. Thienpont: Well, you are right. This is a real challenge ahead of us. And I can tell you the costs are sky-high, in particular, for what concern the Master Comparisons. While several business models can be envisaged, in our opinion, the most straightforward one would be the integration of these products into the portfolio of one of the bigger proficiency testing providers. This is because we see these new products complementary to the currently offered proficiency testing service; or to repeat the title of our 20-year-old publication in a slightly adapted way, “The combined target, combined material approach: a way out of the proficiency testing dilemma.”

Moreover, online outpatient percentile monitoring holds the promise to move away from point estimation to continuous monitoring of assay comparability and stability.

Anyway, we are tremendously enthusiastic about the Empower Project. We see it as a color not of a translation of laboratory medicine pantheistic painting, making big laboratory data acceptable and assuring better interoperability of laboratories. And I refer here to the pantheism style by the Belgian painter, Théo van Rysselberghe, born in Ghent.

Bob Barrett:

Dr. Linda Thienpont is Professor in Analytical Chemistry, Statistics, and Quality Control at Ghent University in Belgium. She has been our guest in this podcast from *Clinical Chemistry*. I'm Bob Barrett, thanks for listening.