Host: This is the podcast from *Clinical Chemistry.* I am Bob Barrett. It is known that thyroid dysfunction is a major public health concern, although thyroid stimulating hormone, or TSH, is generally considered the primary test for evaluating thyroid status, clinicians also need to have access to accurate measurements of total and free thyroid hormone concentrations.

The June issue of *Clinical Chemistry* included three reports from an IFCC Working Group for Standardization of Thyroid Function Tests. The first report relates to TSH. The second addresses free thyroxine and free triiodothyronine. The third discusses total thyroxine and total triiodothyronine.

We have as our guest today Dr. Linda Thienpont, Professor and Analytical Chemistry Statistics and Quality Control at the Faculty of Pharmaceutical Sciences of Ghent University in Belgium.

As an active participant in several international standardization projects, she currently chairs the IFCC Working Group for Standardization of Thyroid Function Tests.

Dr. Thienpont, could you maybe add to the background of justifying the need for standardization of thyroid function testing?

Dr. Linda Thienpont: Well, Bob, it is a fact that thyroid dysfunction represents an increasing public health burden, similar to that of cardiovascular disorders and diabetics.

For example, in U.S. alone, it is estimated that roughly 27 million people suffer from hyper or hypothyroidism, although even 50% remain undiagnosed. This is because the clinical manifestations of thyroid dysfunction vary considerably among patients in the character and severity. For example, symptoms and signs are often nonspecific and progress very slowly.

Consequently, clinical diagnosis alone has limited accuracy. The final diagnosis has to rely on biochemical laboratory analysis.

The importance of laboratory testing is even more obvious when subclinical thyroid diseases are considered. The scarcity or even complete absence of symptoms means that thyroid dysfunction in the subclinical phase is by its very nature dependent of laboratory diagnosis. And these considerations taken together explain that worldwide the

yearly volume of request by clinicians for first-line TSH testing is estimated in the order of 180 million.

In addition, we have to realize that for each third SH test, one free default test is done. So currently, the typical thyroid panel that is primarily ordered by clinicians to help in evaluation of thyroid glands function includes testing for TSH free and total thyroid hormones before we treat.

This testing may be extended to thyroid antibodies and thyroglobulin. The former to diagnose or monitor autoimmune thyroid diseases and to distinguish these from other forms of thyroid disease. The latter to monitor treatment of thyroid cancer and to detect recurrence.

Well, despite widespread use and clinical importance of this typical thyroid function testing, it is a fact that the standardization status of thyroid hormone assays in general is a matter of big debate. The reason is that the lack of uniformity of results hampers the use of common reference intervals and clinical practice decision limits. Also, the controversy about analytical and clinical validity of some of the assays, in particular those for free thyroid hormone measurement, is continuing.

Therefore, the IFCC considered it of priority to document the quality and standardization stages of thyroid testing, and when necessary, to do something about.

Host: Now, could you explain the IFCC Working Group and what are its terms of reference?

Dr. Linda Thienpont: An IFCC Working Group comes under the scientific division and together they aim at instigating and promoting theoretical and practical developments in the field of standards and standardization in clinical chemistry, and this in its broadest sense.

I really emphasize on in its broadest sense, and this is exactly what we try to do in our Working Group for Standardization of Thyroid Function Tests. We really want to expand our work from the methodological foundation of laboratory test to that patient-centered application.

Let me, in this context, explain the work that has been done prior to the results presented in the recent reports published by the Working Group.

(00:04:52)

But first of all, with the implementation of the work in mind, we sought a broad representation of stakeholders in our group. This was achieved by appointing members from the

academic world, metrological institutes, producers of reference materials, the in vitro diagnostic industry, and last but not least, from the International Thyroid Associations.

Second, work was undertaken to lay the metrological foundation to standardization, which is based on a complete reference measurement system, which is preferably an SI-traceable system, and comprising the measurant, reference materials, reference measurement procedures, and routine procedures.

And in the conviction that the reference measurement system needs international recognition, close cooperation with the Joint Committee for Traceability in Laboratory Medicine is required. The abbreviation for that committee is JCTLM.

The JCTLM is indeed a worldwide platform to promote and give guidance on internationally recognized and accepted equivalence and traceability of measurements in laboratory medicine.

Making the link to the JCTLM was rather easy for our Working Group because I head the co-chair in one of the JCTLM Working Groups.

So what metrological tools were available at the start? Well, only for total thyroid hormones was a complete SI-traceable reference measurement system available. In addition, the reference materials and isotope dilution, liquid chromatography–tandem mass spectrometry reference measurement procedures were already listed in the database of the JCTLM.

For free thyroid hormones, the working group proposed to define operation and the measurant as T4 or T3 and a dialysate from equilibrium dialysis of serum or plasma prepared under defined conditions.

For the technical realization of the measurant at a hierarchically higher order, our Working Group chose the concept of an international conventional reference measurement procedure. The convention was about the condition for equilibrium dialysis, but the quantification of the hormone in the dialysate was to be done with an isotope dilution mass spectrometry reference measurement procedure.

Finally, for TSH measurement, the only metrological tool that was available was immunoanalysis, and because of the heterogeneity of this glycoprotein hormone, this is really mixture analysis, expressed in international units, defined

by reference preparation from the World Health Organization.

Therefore, the Working Group realized that TSH should be considered as being in a kind of continuum from its discovery into translation to an SI analyte, with different needs for the reference measurement system at each stage in the continuum.

In this stage, the Working Group proposed a new standard for immunoassays, namely a panel of native materials assigned by the old procedure, tromethamine, a surrogate reference measurement procedure.

So what you see in our recently published reports is just the start of the work and maybe the most easy part. Nevertheless, it is a foundation of all steps that will follow.

The Working Group wanted to create solid evidence of what the status of thyroid testing is in terms of quality and standardization. And to put even more emphasis on these two points, I want to repeat that standardization requires the knowledge of quality first.

My personal opinion is that this requirement is underestimated in many other standardization projects, and that's a pity. Therefore, I will try to explain the relevance of the results by addressing both points.

Host: Well, in this regard a technical question, could you highlight the conclusions and clinical relevance of the published achievements of the Working Group?

Dr. Linda Thienpont: Well, Mr. Barrett, thank you very much for that question and for so kindly referring already to our achievements. Indeed, our reports contain a lot of numbers, tables, figures, and not to forget the supplements, which contain extremely useful information.

As explained before, I will summarize and discuss the results, on the one hand in terms of the quality of the assays, on the other hand of their standardization.

When one looks into the three papers, one will see a great span in overall quality. To give an example, the total imprecision of the total T4 assays ranges from roughly 3-9%; for free T4, from 1.5-14%; for TSH, from 2.5-8%. Well therefore, I think it is a great mistake to argue about the quality of, for example, free T4 immunoassays in general, because as you have seen from these figures, there are good ones as well as assays that need improvement.

The data also shows that precision should always be viewed together with general specificity. Well, as you know, we did not do studies dedicated to seeing the performance on challenging samples, one can observe already that even we sampled from apparently healthy people, considerable sample related effects are present in particular for free thyroid hormone assays.

In other words, the most precise assays are not always the ones with the best accuracy. This brings me to another important point, I mean, the big strength of the study. It not only compares assays with each other, but also with a solid reference.

Our working group was extremely fortunate to have reference method values for the total and free thyroid hormones. This method for example in case of poor quality of performance reflected by low correlation coefficient, we knew exactly that this was caused by the tested assays. However, even for TSH assays, we had an excellent reference.

This statement may be surprising to many, but it is a fact that the trimmed mean from all assays gets an excellent reference point. Another strength of the data interpretation is that we also compare the results with objective, quality criteria derived from the biological evaluation of the components.

I admit there may be better criteria such as outcome related. But the criteria we used proved quite effective in giving us a first insight in the analytical quality of the assays. So if we leave the standardization issue behind for a moment, we can conclude that TSH measurement has reached quite a mature level, and we see that several free and total hormone assays deliver highly precise and accurate results.

It is up to the other in vitro diagnostic manufacturers to catch up with the development or improvement of their assay, and this touches another strength of the study. We had a chance to investigate between 11 and 16 different assays. This really allows diagnostic manufacturers to identify benchmarks of what is technically achievable today.

It's not about goals anymore, but it is really about comparing two competitors.

Host: You mentioned the in vitro diagnostic manufacturers. Can you expand on their role in the working group?

Dr. Linda Thienpont: Well, thank you again for that question. While it is a fact that the responsiveness of the manufacturers was

overwhelming, somewhat more disappointing maybe was the participation of only two most asymmetric assays for free thyroid hormones.

But here I will also take the opportunity to thank once again all the participants. In fact, they sponsored the study in terms of sample procurement and funding for the assignment with reference method values. Also, manufacturers were very responsive in offering their help to solve technical problems, such as in the case of the problematic samples for T3. For me as Chair, it was really a pleasure to work with the in vitro diagnostic manufacturers.

Host: Okay. Well, coming to the standardization issue, could you comment on the standardization status of the different thyroid assays and the impact standardization would have? And could you briefly also describe the process of standardization?

Dr. Linda Thienpont: For total T4, our studies showed that the mean of the results from all immunoassays compared well with the IDMS reference method values. In other words, the two high values seemed to balance the two lower ones. In consequence, it is now the responsibility of the individual manufacturers to eliminate the assay-specific biases. To solve this problem, manufacturers should follow the normal rule towards establishing traceability, which is cooperation with the JCTLM listed reference laboratory.

For total T3, the situation was quite different. Standardization would cause a shift of the whole market to lower values. Current values would decrease in average by roughly 15%.

(00:14:56)

For free hormones, the situation was even more pronounced, quite a the difference among the assays was seen and all were negatively biased in comparison to equilibrium dialysis mass spectrometry.

In consequence, standardization of free thyroid hormones will require a huge effort and will have a major clinical impact. For free T3, the values would increase in average by roughly 15% but as for free T4, even by 35%. The TSH was special in a sense. It was amazing to see the agreement of so many assays, but very unfortunate to see that three assays caused up to 40% difference in results on average, even more pronounced in the low range.

One may question, whether this observation is clinically relevant? Personally, I think it is. The problem with TSH is that so called universal cut-off are used for defining sub-

clinical hypothyroidism, mainly around about five milli international units per liter. However, as the market is not standardized, assay specific rather than universal values should be used. Also the discussion about lowering this value to 3 or even 2.5 milli international units per liter does not address this standardization problem. In my opinion, it is one of the tragedies that these discussions neglect the standardization problems.

And to come now to your question about the process of standardization, well I can tell you that our studies demonstrated that the way to standardization is, in fact, straightforward. It consists of performing a method comparison and using the relationship between the results by participating assays with reference method values or all procedure for recalibration.

In the first study, we did the recalibration mathematically. However, in the meantime, we were able to show for free T4 and TSH, the proof of the concept. Recalibration by manufacturers themselves on the basis of master calibrators included in a second method comparison resulted in a similar outcome as by mathematical recalibration.

Host: Well, let me interrupt you for a second. Do you think the measurement of TSH can be standardized at all? I have always heard about the difficulties to standardize mixture analysis as you called TSH measurement by immunoassays.

Dr. Linda Thienpont: Well, Bob, you have touched on a very important and sensitive point. I mean the very ability in the protein structure and in the glycosylation form that indeed contributes to a major extend to the TSH heterogeneity. Not withstanding this heterogeneity, the results of our first study, we sampled from apparently healthy donors were very promising with regard to the feasibility of standardization of TSH measurement.

However, this brings me to another very important point. Our work is just in the very first beginning. So that actually, we have to be cautious to not over-optimistically interpret the outcome. Our project needs in the next phase to repeat the method comparison study with clinical samples to see in how far they reproduce the observations, maybe with samples from apparently healthy individuals.

Therefore, in my opinion, sustainable standardization must proceed with so-called “normal” and “clinical” samples. But this brings me unfortunately to a disappointing phase in the project because just now we are stuck with the problem that it is extremely difficult to obtain clinical samples in the volumes we need. This is 10 to 20 ml from individual patients. And let me take the opportunity to address myself

to clinicians that are able to provide such samples. Our working group really needs that help.

Host: You emphasize several times that the working group recently started its efforts. Could you outline what its future work will be about?

Dr. Linda Thienpont: Well, indeed all observations done up to now are based on the performance of the assays on samples from apparently healthy individuals but from the technical point of view, it is now of utmost importance to make the switch to clinical samples. It might, for example, be that the correlations observed in this study give an over-optimistic impression about test performance. Therefore, interpretation of method quality parameters needs assessment over the full clinical range from hypo to hyperthyroidism.

(00:20:00)

Also, a much greater variety of challenging samples shall be included to explore the full impact of standardization, for example, specimens obtained from different patient categories or individuals with different pathologies, thyroid pathologies, and others and/or patients and the therapy.

This is of particular importance for standardization of free thyroid hormones and TSH measurements. For the latter, there is indeed evidence of different glycoforms associated with hypothyroidism.

Thus, we really have to investigate whether certain assays perform differently with clinical samples. Another point of concern for our working group is that it cannot be overlooked that standardization, for example, of free thyroid hormone immunoassays will cause a shift of the whole market to higher values, for some even by a factor of two.

In consequence, it will be necessary to involve all stakeholders in a sufficiently early stage. This will be the next difficult task to fulfill. I mean we have to solicit clinicians for active participation and development of an implementation strategy.

My dream is to have an organization such as the National Cholesterol Education Program or the National Institute of Diabetes and Digestive and Kidney Diseases, but then on a global scale.

Host: Well, touching on your dreams and your involvement in so many other standardization projects, what message would you like to give beyond that project?

Dr. Linda Thienpont: Well, Bob, I would like to end by quoting Cali for one of his landmark publications in 1973. He wrote, “My personal opinion is that it is almost certain that the Product Standards soon to be promulgated by the Food and Drug Administration with regard to the labeling requirements of clinical kits, chemicals, and devices will, where applicable, somehow relates to and/or require the use of referee methods or state-of-the-art interim methods where referee methods have not yet been developed to provide the baseline against which other methods or products will be assessed.”

Cali wrote this in a publication in Clinical Chemistry in '73, and this article was entitled *An Idea Whose Time Has Come*. I want to emphasize that he wrote this already in 1973. We have now 2010 and still we have to do something about standardization.

Host: Dr. Linda Thienpont is a professor and Analytical Chemistry Statistics and Quality Control at the faculty of Pharmaceutical Sciences of Ghent University in Belgium and she has been our guest in this podcast from *Clinical Chemistry*.

I am Bob Barrett. Thanks for listening.

Total Duration: 23 Minutes