and LDL turnover by stimulating the synthesis of LDL receptors and the degradation of LDL (4). Therefore, overt hypothyroidism can increase the concentrations of TC and (or) TG (5). Subclinical hypothyroidism would thus constitute a risk factor for coronary failure by the lipidemic modifications it induces (1). Some authors have reported that subclinical hypothyroidism could be the isolated increase of TSH, increase TC (1), LDL-c (6), and TC/ HDL-c and decrease HDL-c (7); however, this was not fully confirmed by other authors (6,8). The effects of T3 on lipid-related variables in patients with subclinical hypothyroidism are also controversial: Some authors have shown that hormone replacement therapy, besides decreasing serum TSH, could also decrease concentrations of blood apo B and TG (9), LDL-c (10), and TC/HDL-c (7) and increase HDL-c (7, 9) or apo A-I (7,11); these notions, however, were not confirmed by other studies (11–14). Our study has confirmed previous results involving lipids, i.e., that the increase of TSH 9 months after surgery was related to increased atherogenic lipids (1,6) and that the benefit of T3 therapy would be restricted to increased HDL-c (7,9). Our study also showed that partial thyroidectomy had little influence on thyroid functions in the mid-term (9 months): At that stage, only 2 patients of 42 exhibited persistent subclinical hypothyroidism, whereas it had spontaneously disappeared between 3 and 9 months after surgery in the other affected patients.

Hormone replacement therapy after partial thyroidectomy should therefore be indicated only in the presence of obvious persistent hypofunction 9 months after surgery.

References
5. Abrams JJ, Grundy SM, Ginsberg H. Metabolism of plasma triglycerides in hypo- 


Loic Campion 1,2
Gildas Gallou 2
Annie Ruellan 2
Lise Cloarec 2
Hubert Alainnic 1

1 Dept. of Diabetol., Endocrinol., and Metabol. Diseases
2 Med. Biochem. A Lab. Rennes Teaching Hospital Rennes, France

2 Address correspondence to this author at: Clinique Medicale B-F, Hopital-Sud, 16 Bd de Bulgarie, BP 56129, 35056 Rennes Cedex, France

Functional Sensitivity of Thyrotropin Assays

To the Editor:

The interlaboratory performance of thyrotropin (TSH) assays has been the focus of several recent studies. In this issue Spencer et al. (1) examine the “functional sensitivity” of six TSH methods under typical laboratory conditions and assess the reliability of subnormal TSH measured concentrations by a broader range of 16 TSH methods. This study raises many questions about results from a variety of clinical sites when compared with results obtained by the manufacturer.

Improved performance by manufacturers compared with that at clinical sites may be attributed to many variables, such as calibrations and sample conditions rather than the “sensitivity potential” ascribed to them by Spencer et al. Published results have shown that there are many ways to improve in between-run precision (2).

The study design and data interpretation are questionable. Although the development of precision profiles under laboratory conditions is reasonably presented, some conclusions are unsupported. Objective interpretation would be facilitated by including the number of reagent lots and calibrations used by individual laboratories, as well as confidence intervals for the precision profiles.

Although not surprising in light of previous publications where K-30 (Kodak TSH-30) and NIC (Nichols) have failed, in practice, to yield “third-generation” functional sensitivity (3,4), Spencer et al. correctly point out that all three third-generation TSH assays cited fail reliably to meet this criterion. Indeed, two of the assays (K-30, NIC only) marginally meet the third-generation criteria in the manufacturers’ hands. Contrary to the authors’ conclusions, the precision profiles generated in clinical laboratories by these two methods are essentially equivalent to that of Ciba Corning Diagnostics Corp. (CCD; Medfield, MA) ACS™TSH (referred to as COR by Spencer et al.). The CCD insert claims a minimal detectable sensitivity of 0.03 mIU/L, a claim that has been independently substantiated (see Fig. 1), reliably yielding better than second-generation functional sensitivity. The results reported by Spencer et al. for COR agree with CCD low-end precision studies on new instruments and those of the College of American Pathologists (CAP), the French Pro-Qual-Bio Programme and two UK EQAS studies.
The number of uncontrolled variables makes it difficult to draw conclusions on the relative performance of 16 TSH assays from the study on human sera (or pools) with incremental addition of TSH. Two methods [ACC (Sanofi Access), US pool 2; BIO (Bio-Rad CoTube), US pool 1] are evaluated on only one of the two different US pools. Two other methods [DPC (Diagnostic Products Immunilite) and NIC] are evaluated on only the two US pools. Ten methods [BD (Becton-Dickinson Simultrac), BM (Boehringer Mannheim TSH), DAK (Dako Novocline), DEL (Wallac Delfia), IDS (IDS TSH), K-AM (Kodak Amerlite), K-CT (Kodak Coated Tube), K-30, NET (Nertra IRMA), and SER (Serono Maiacline)] are assessed on only the UK pools. One method (ABB; Abbott IMX) is assessed on one of the US pools and both of the UK pools, but only one method (COR) is assessed on all four pools. Preparation of the pools was performed independently. Only the UK pools were treated with sodium azide. Some of the pools (it is unclear which) were prepared from sera screened by one of the TSH methods (NIC), and some were prepared from triiodothyronine-depressed sera. This apparently random experimental design is extremely confusing.

In addition to the differences in the mean concentrations of TSH noted, the spread of results in the two US pools appears to be significantly different for the three methods that were used to study both pools. An F-test on the variance would confirm this observation; however, there is a noticeable lack of statistical analysis.

Fully one-third of the results for the first US pool are >0.1 mIU/L as measured by the most sensitive of the six methods evaluated in the first part of the study (DPC). The high incidence of inconsistent results in the studies with the US pools, the surprising differences in between-laboratory precision between pools, and the incomplete study design (only three assays were evaluated on both US pools) means that no valid conclusions can be drawn from the results on the US pools. The subjective comparison by Spencer et al. of the prospects for good sensitivity of assays evaluated on US pools with those evaluated on UK pools is unjustified.

In contrast, the pools prepared and distributed by the UK EQAS yield few inconsistent results. However, even the authors' interpretation of the results on the UK pools is lacking support from appropriate analysis of the data. Simply stating that the performance of the DAK and DEL assays is superior to that of ABB on these UK samples is a subjective judgment. Again, statistical analysis, typical of that demonstrated in UK EQAS reports, would eliminate subjectivity. Analysis of the quantitative recovery of TSH in the UK pools would also be of interest, given that some methods apparently achieve lower “functional sensitivity” by significantly underestimating TSH content. This study would be improved by disregarding the results for the US pools while analyzing in depth the UK pool results.

Suspicious results for the US pools led Spencer et al. to examine the effects of 22°C storage on a single sample by three assays and on two samples by three assays. It is very misleading to conclude from these results that the CCD ACS TSH assay yields aberrant results in sera stored under conditions recommended by Ciba Corning or commonly encountered in laboratories. The CCD lactate product insert specifies “refrigerate specimens at 2–8°C if testing is not done immediately, or tightly cap and freeze the specimens below −20°C if testing is not done within two days. Freeze specimens only once and mix thoroughly after thawing.” A more detailed study (5) of a greater number of samples (n = 19) stored under more strenuous conditions concluded that there was no clinically relevant effect on CCD ACS TSH results. Additional studies by CCD (prior to commercial release of the CCD ACS TSH assay in 1991) and by independent laboratory scientists (personal communications) have confirmed the reliability of CCD ACS TSH results on samples stored under recommended conditions. We think it inappropriate for Spencer et al. to suggest that any of the assays tested had inferior specificity, on the basis of their results for two samples and in spite of published results to the contrary.

In addition, after questioning laboratory personnel in the US who actually received and evaluated samples for this study, we are even more concerned about the conclusions drawn. In contrast to the statement that the serum pools were "prealiquoted into small volumes for frozen shipment (−10°C) to participating laboratories for analysis," we have learned that the serum pools (sample volumes of 500 μL) were sent to clinical laboratories through the regular surface mail without ice or cold packs and arrived in a warm liquid state, appearing cloudy. Spencer mentioned this in a letter to a participating laboratory, stating “since we determined that TSH in normal human sera is relatively stable for short periods of time at ambient temperatures, we felt that it was not a problem sending these pools by mail.”

Critical protocols for manufacturers of proficiency and control materials and CAP consider sample integrity a priority. Liquid samples and controls are handled as a sterile product, frozen at −80°C until shipped, shipped on solid CO2, and shipped by priority delivery so that the shipment is received the next morning (D. McCluskey, Manager, OEM Control Products, CCD, Irvine, CA; personal communication). The study data are confounded by the fact that pools to clinical laboratories were shipped through surface mail, whereas the sample pools received by Ciba Corning arrived frozen on solid CO2 (T. Tyler, Customer Evaluations Laboratory, CCD, Oberlin, OH; personal communication). It is therefore no surprise that the results obtained by the Ciba Corning Laboratory reflected better results than those from the clinical laboratories.

Critical evaluation of immunoassays is an important force that drives significant improvement of performance. However, it is equally important that the studies performed to evaluate these issues be objective, well designed, and analyzed statistically in a rigorous manner. Ciba Corning Diagnostics is always re-
responsive to legitimate concerns regarding the performance of CCD products and completely committed to introducing new products and improving existing products to satisfy CCD customers' need for reliable immunodiagnostic reagents. Without absolute certainty that the samples used in these studies were handled appropriately, it is impossible to substantiate a report such as this.

References


Julie S. Williams

Thomas M. Jackson

Ciba Corning Diagnostics Corp.
Dept. of Tech. Marketing and
Worldwide Marketing, Med.
Affairs
115 Norwood Park South
Norwood, MA 02062-4694

1 Author for correspondence.

Two of the authors of the study referred to reply:

To the Editor:

The capability of TSH immunometric assays to distinguish normal from subnormal TSH concentrations is essential for a cost-effective, TSH-centered thyroid testing strategy. Because technical publications typically highlight the optimal performance of a method, which may not be met in clinical practice, our study was designed to provide a realistic assessment of the reliability of subnormal TSH measurement across a wide range of methods used in clinical practice. The study was funded solely from research sources, and we made considerable efforts to objectively adhere to the study goal and analyze the data with no bias toward or against any one method.

Clearly, the variability in performance of some methods such as the Ciba Corning Diagnostics ACS:180 (COR) has clinical impact. Our study could not address the reasons for the variable performance. Williams and Jackson suggest that some laboratories might differ in compliance with the manufacturer's instructions. This is probably the case, but merely underscores the manufacturers' responsibility to produce a robust product.

The pools were measured across the laboratories with several different reagent lots (13 for COR). Because the laboratories were asked to treat the pools as routine clinical specimens, the runs used to generate the pool values presumably met the laboratory's criteria for acceptance, irrespective of reagent lot.

We agree with Williams and Jackson that "improved performance by manufacturers compared with that at the clinical sites may be attributed to many variables." It was with this rationale that we proposed that the manufacturer's precision profiles would provide the most effective assessment of inherent functional sensitivity differences between methods. On this basis there appears to be a significant difference between the so-called third-generation assays (DPC, K-30, and NIC) and the COR. Presumably this difference is the rationale for developing the future third-generation COR TSH assay they mention. We agree with Williams and Jackson that the mean functional sensitivities across the clinical laboratories using COR, K-30 and NIC (0.07, 0.06, and 0.05 mIU/L, respectively) are more comparable than a comparison of the manufacturers' functional sensitivity implies; however, only three laboratories were included for COR, whereas seven and six laboratories used K-30 and NIC, respectively. The precision data shown in Williams and Jackson's Fig. 1 is compatible with that generated from our study and suggests that COR is a second-generation assay.

The need to evaluate the methods with different pools did indeed complicate the study design. The pools were prepared and evaluated over 1-2 years, which gives the data a broader basis of clinical relevance. There was insufficient material from any one pool to conduct the entire interlaboratory, intermethod study. Indeed, the use of different pools made the data more applicable to the clinical practice of evaluating multiple different patients' sera. We tried to match the US and UK pools on the basis of the mean pool A value established from multiple interassay measurements made by the most sensitive assays available (K-30 in the UK and both NIC and DPC in the US). Although there were statistically significant differences (mean pool A values ranged between 0.01 and 0.04 mIU/L), this difference was less than the interlaboratory differences observed; and in any case, the Pools A were considered to be clinically equivalent; i.e., they should all have produced subnormal TSH values. Methods that were evaluated only in the US or in the UK had insufficient representation in the other region to include evaluation. The ABB and COR methods are widely used on both sides of the Atlantic and thus were assessed with both US and UK pools. The study of COR was expanded to include all four pools because both the authors and the manufacturer were concerned that the results obtained with the first US pool A (open circles, our Fig. 3) were not comparable with those already available from UK NEQAS. The second US pool A (crosses, our Fig. 3) was made up and distributed in response to these concerns. It was this second US pool that displayed the preanalytical temperature effects shown in our Fig. 4 (serum 2).

We were also surprised to see outlying values in 2 of 14 pool A DPC values and have no explanation for this finding. Limited serum volumes, together with the intent to discourage confirmatory measurements, led us to select the 500-µL pool volume. The spread of values across laboratories that assayed the US pools as compared with those for the UK pools was not significantly different for the methods with similar functional sensitivity (NIC and K-30). We thus believe that the US pool data are valid and the outliers seen with the second set of US pools with the COR method reflect the preanalytical temperature effect on this pool (serum 2, Fig. 4).

We decided not to use extensive statistical analyses of the pool data because our principal goal was to assess the reliability of subnormal range measurement relative to the clinical question as to whether an appropriately subnormal TSH value would have been reported for that specimen. We felt that the interlabor-