Differences in Radioimmunoassay Results for Thyroglobulin That Affect Management of Patients with Thyroid Cancer

Monika F. Bayer and I. Ross McDougall

We compared two commercial assays for measurement of serum thyroglobulin [Nuclear Medical Systems (NMS) and "CIS" (Damon Diagnostics)] with each other and with one developed at Stanford (J Clin Endocrinol Metab 49:557–564, 1979). The NMS assay is a competitive-binding RIA, the CIS and Stanford assays are two-site immunoradiometric assays. The kit standards varied in thyroglobulin concentration. The NMS standards differed in immunoreactivity from thyroglobulin in clinical specimens and from the other standards. Also, nonparallelism between standards and patients’ sera in the NMS assay suggested a less-specific antisera. Results by the CIS and Stanford assays correlated well (n = 120, r = 0.964), those by the NMS assay less strongly (n = 101, r = 0.855 vs CIS, r = 0.888 vs Stanford). Clinical evaluation in 50 patients treated for differentiated thyroid carcinoma (10 with metastases and 40 currently disease-free) indicated good agreement for positive results by the three assays. The CIS and the Stanford assay both gave high results (≥ 25 µg/L) in all 10 cases with metastases; the NMS RIA identified eight of these patients (thyroglobulin ≥ 30 µg/L), but excluded two with anti-thyroglobulin autoantibodies. In subjects without disease, however, the percentage of undetectable thyroglobulin (negative result), as opposed to low measurable thyroglobulin (inconclusive result) varied considerably: 85% by CIS, 30% by NMS, and 75% by the Stanford assay.

Additional Keyphrases: "kit" methods  competitive-binding RIA  immunoradiometric assay  monitoring disease  thyroid disease  cutoff (reference) values

Thyroglobulin (Tg), a large glycoprotein with a molecular mass of ~650 000 Da, functions as a prohormone in the synthesis of thyroxin (T4) and triiodothyronine (T3) (1). It is contained primarily within the follicles of the thyroid gland. Small quantities of Tg are released into the circulation and sensitive assays can detect it in the serum of most healthy individuals (2, 3). Normal concentrations of Tg in serum range from undetectable to 60 µg/L, depending on the assay used. Moderately higher concentrations are found in pregnancy and in newborns. Distinctly increased concentrations of Tg are associated with goiter and thyroid hyperfunction; with inflammation of, or physical injury to, the thyroid; and with thyroid cancer (4–6).

Tg measurements do not facilitate the initial diagnosis of thyroid cancer because they cannot distinguish between benign and malignant thyroid nodules (9). Nevertheless, data on Tg are clinically important in the long-term management of patients with differentiated thyroid cancer who have undergone thyroidectomy and radiiodine ablation therapy, or both. After complete removal of the thyroid—the only tissue known to synthesize Tg—the Tg concentrations should become undetectable, whereas recurrent or metastatic disease results in measurable to high Tg concentrations in serum.

The Tg assays reported to date are either competitive radioimmunoassays (RIA)—i.e., modifications of the original assay for Tg by Van Herle et al. (3, 10–12)—or solid phase, two-site, immunoradiometric assays (IRMA), similar to the one described first by us (13, 14).

In this study we compared the first two commercial assays for Tg—a competitive binding RIA by Nuclear Medical Systems (NMS) and the "CIS" IRMA-type assay—with each other and with the Tg assay developed at Stanford a few years ago (13). We studied the analytical and clinical performance of these assays and evaluated their clinical utility in the management of patients with thyroid cancer.

Materials and Methods

Procedures

For measurement of Tg in serum, we used the following assays.

NMS RIA kit (Nuclear Medical Systems Inc., Newport Beach, CA 92663). In this assay, a typical competitive-binding RIA, Tg in standards or samples competes with 125I-labeled Tg for a limited number of binding sites on an anti-Tg antisera. Free and bound Tg are then separated by precipitation with a second antibody (3).

"CIS" IRMA kit (Damon Diagnostics, Needham Heights, MA 02194) and the Stanford IRMA, described previously (13). Both of these IRMAs are based on the same principle. Tg in standards or specimens is first bound to a solid surface coated with (rabbit) anti-Tg antibody and then quantified in the second step, after aspiration of the assay mixture, by binding of labeled rabbit anti-Tg antibody. The two assays differ only in the solid phase: microtiter plates in the Stanford assay and plastic tubes in the CIS assay.

Measurement of anti-Tg autoantibodies in sera. Anti-Tg was measured by a solid-phase, sandwich-type RIA (15).

Assay Validation

The Tg standard preparations and the specificity of the antisera were examined by comparing the standard curve of each assay with (a) curves generated with serial dilutions of sera from thyroid-cancer patients that contained high concentrations of Tg, (b) with dose–response curves obtained by use of the Tg standard solutions of the other two methods, and (c) with dose–response curves produced by serial dilutions of the most concentrated standard of each method. Serial dilutions were made in serum from athyreotic patients (whose thyroids had been removed surgically) who were judged clinically to be disease free and who had no measurable Tg or anti-Tg. To rule out differences in the dilutions, we analyzed the same serial dilutions in parallel experiments, following the assay protocols exactly.

Department of Radiology, Division of Nuclear Medicine, Stanford University School of Medicine, Stanford, CA 94305.

1 Nonstandard abbreviations: Tg, thyroglobulin; IRMA, immunoradiometric assay; NMS, Nuclear Medical Systems Inc.; T4, thyroxin; and T3, triiodothyronine.

Received August 1, 1983; accepted September 29, 1983.
Analytical recovery studies involved adding Tg (20–100 μg/L, either the CIS or NMS standards) to serum specimens from thyroid-cancer patients with undetectable to moderately increased Tg (0–100 μg/L) concentrations. The same mixtures of various sera plus exogenous Tg were analyzed by all three assays and the recovery was calculated as follows: \( \frac{\text{Tg}_{\text{found}} - \text{Tg}_{\text{endogenous}}}{100} \times \text{Tg}_{\text{added}}. \)

Intra- and inter-assay coefficients of variation were determined with three serum pools with Tg concentrations in the low, middle, and high assay range. Each pool was measured in at least eight replicates in one assay set-up and (in duplicate) in 10 consecutive assays.

The normal range for individuals with an intact thyroid is ≤ 40 μg/L in the Stanford assay (13) and (as suggested by the manufacturers) ≤ 60 μg/L for the CIS and NMS assays, but these values cannot be applied to athyreotic patients. The reference values for thyroid-cancer patients were determined, retrospectively, in a group of more than 100 patients seen at Stanford during the past years by correlating results for Tg (generated with the Stanford assay) with all the relevant clinical data (16). The reference values, selected to provide maximum clinical sensitivity, were as follows: (a) Tg undetectable (≤ 2.5 μg/L) for patients who have undergone total thyroidectomy and \(^{131}I\) ablation and are on \(T_4\) replacement therapy at the time of Tg measurement, and (b) Tg ≤ 10 μg/L for patients with residual thyroid tissue and (or) currently not on \(T_4\) suppression therapy. The corresponding cutoff values for the two commercial assays, derived by linear regression analysis were: undetectable (≤ 3 μg/L) and ≤ 10 μg/L for the CIS kit and undetectable (≤ 5 μg/L) and 15 μg/L for the NMS kit.

Patients

Group 1 consisted of 50 patients seen at the Stanford University Medical Center, whose history and extent of disease was known clinically, for correlation with Tg measurements. All of these patients had been treated for follicular, papillary, or mixed papillary–follicular thyroid cancer. Thirty-one had undergone total (or near-total) thyroidectomy followed by \(^{131}I\) ablation therapy; nine patients had subtotal thyroidectomy and \(^{131}I\) treatment; and four patients had total and six had subtotal thyroidectomy without \(^{131}I\). All were receiving adequate \(T_4\) replacement and had normal concentrations of thyrotropin (less than 2 milli-int. units/L) in their serum.

Ten of these patients had metastatic thyroid cancer, as judged by a positive \(^{131}I\) whole-body scan (nine of 10 cases), and in some cases by symptoms, palpable lesions, positive chest roentgenogram, or computed tomography scan. The scans and chest roentgenograms for the 10th patient were repeatedly negative, but autopsy (death unrelated) revealed microscopic metastatic papillary thyroid carcinoma in the lungs.

Forty patients were clinically disease free, although one had an \(^{131}I\) uptake of 0.5% in the thyroid bed, measured 48 h after oral administration of 2.0 mCi of \(^{131}I\).

Nineteen (38%) of the 50 sera were found to contain anti-Tg and thus could be analyzed only by the two IRMA assays.

Group 2 consisted of sera from 70 patients, mostly referrals for whom the corresponding clinical data were not available. As far as we could determine, at least 80% of these sera were also from thyroid-cancer patients, but they included also patients with other thyroid disorders and sera drawn preoperatively from patients evaluated for possible thyroid cancer. These sera were selected by one criterion: they did not contain anti-Tg and thus could be subjected to all three Tg assays.

For linear regression analysis, we replaced Tg values greater than 1000 μg/L with the Tg concentrations measured in the same specimens diluted 10-fold.

Results

Analytical Performance of Assays

Table 1 summarizes various assay parameters. The typical standard curve we obtained for the CIS assay (Figure 1) was identical to that suggested by the manufacturer. Furthermore, substitution of the Stanford standard for the CIS standards, or use of serial dilutions of the highest CIS standard in human Tg-free serum, or use of serial dilutions of sera from thyroid-cancer patients with high concentrations of Tg gave response curves essentially parallel to, or superimposable on, the original CIS standard curve. Only the various standards supplied with the NMS kit deviated somewhat from parallelism when measured by the CIS assay.

The NMS assay, carried out according to the manufacturer’s protocol, produced a standard curve and also results for the two control samples in very good agreement with those stated in the kit literature, indicating to us that we used it properly. However, sera from thyroid cancer patients with high Tg, in the same serial dilutions that exhibited parallelism in the CIS and Stanford assays, gave curves with widely differing slopes, most of them significantly different from the slope of the NMS standard curve (Figure 2). Similarly, when we used the NMS kit to determine Tg in the CIS kit standards, a displacement curve resulted that did not parallel the NMS standard curve, but was considerably flatter, intersecting the latter in midrange.

A typical standard curve for the Stanford assay is depicted in Figure 3. Serial dilutions of sera from patients with metastatic thyroid cancer produced indistinguishable curves (13) and analysis by the Stanford assay of the various Tg standard solutions supplied in the CIS or NMS kits gave curves that largely paralleled the Stanford standard curve. Only the least-concentrated standards (≤ 10 μg/L) were shifted somewhat to higher Tg values. Serial dilution curves of the most-concentrated NMS standard (500 μg/L) or CIS standard (1000 μg/L) in human serum, devoid of Tg, were very similar to the curves obtained with the kit standards, except that they were linear and paralleled the Stanford curve, even in the low range of 3 or 5 μg/L.

Mean analytical recovery of Tg added to sera was, by the CIS method, (± SD) 92.4 ± 10.0% (range, 80–109%) for the CIS Tg preparation and 91.9 ± 9.8% (range, 87–105%) for the CIS Tg preparation. Recovery in the NMS assay, on the other hand, strongly depended on the Tg preparation and the quantity used; i.e., 92.0% (SD 14.4%, range, 81–125%) of added NMS Tg standard was recovered, but 141.3% (SD 42.3%, range, 100–220%) of the CIS Tg standard.

Tg up to 2000 μg/L, \(T_3\) up to 20 μg/L, or moniodothyronine or diiodothyronine up to 5000 μg/L did not interfere significantly in any of the three Tg assays.

Correlation between Methods

Results of Tg measurements by both IRMA correlated well. Linear regression analysis gave an overall correlation (groups 1 and 2, n = 120) of CIS = (1.25 × Stanford) + 2.2, r = 0.964, or, for sera of group 1 (exclusively thyroid-cancer patients, n = 50) CIS = (1.25 × Stanford) − 0.61, r = 0.986. A similar correlation was obtained when we analyzed the CIS standards by the Stanford assay (n = 6): CIS = (1.20 × Stanford) − 9.01, r = 0.999, indicating that the two standards were not quite equivalent. Corresponding standard solutions with the same measurable Tg concentrations are defined by 1.2 times higher Tg values in the CIS assay as
Table 1. Comparison of Three Assays for Thyroglobulin

<table>
<thead>
<tr>
<th>Measurement of anti-Tg positive sera</th>
<th>NMS</th>
<th>CIS</th>
<th>Stanford</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tg standards</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Tg concn assigned to kit standards relative to Stanford standard = 1.0</td>
<td>2.28</td>
<td>1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>Specificity of antiserum</td>
<td>Cross reacts with antigen other than intact Tg</td>
<td>No signif. cross reaction found</td>
<td>No signif. cross reaction found</td>
</tr>
<tr>
<td>Precision (CV), %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-assay</td>
<td>3.9-7.9 (5.1-8.2)*</td>
<td>4.7-12.0 (3.9-5.0)</td>
<td>4.1-11.3</td>
</tr>
<tr>
<td>Inter-assay</td>
<td>4.5-11.5 (6.6-13.0)</td>
<td>6.2-14.8 (6.2-8.9)</td>
<td>5.9-15.6</td>
</tr>
<tr>
<td>Assay sensitivity, μg/L</td>
<td>Not determined* (5)</td>
<td>3 (3)</td>
<td>2.5</td>
</tr>
<tr>
<td>Assay range, μg/L</td>
<td>Not determined* (5-500)</td>
<td>3-1000 (3-1000)</td>
<td>2.5-1000</td>
</tr>
<tr>
<td>Normal range, μg/L (intact thyroid)</td>
<td>(≤60)</td>
<td>(≤60)</td>
<td>≤40</td>
</tr>
<tr>
<td>Reference values (μg/L) for thyroid-cancer patients*</td>
<td>Undetd, ≤15</td>
<td>Undetd, ≤10</td>
<td>Undetd, ≤10</td>
</tr>
</tbody>
</table>

*Data in parentheses: those provided by the manufacturers.

**Cannot be determined definitively, given the characteristics of the present NMS standard and antiserum.

*See Methods.

---

Fig. 1. Dose–response curves obtained with the standard solutions provided in the CIS kit (●●●●) and with serial dilutions of two representative sera from patients with metastatic thyroid cancer (▲▲▲▲ and ▲▲▲▲), of the highest CIS Tg standard = 1000 μg/L (○○○○), or of the highest NMS Tg standard = 500 μg/L (□□□□), all in human Tg-free serum
compared with the Stanford assay (= 1.0).

In contrast, Tg results generated by the NMS assay correlated less well with either the CIS or the Stanford assay. For all sera in groups 1 and 2 which had no detectable anti-Tg and could thus be analyzed by the NMS assay (n = 101), we found NMS = (0.23 × CIS) + 9.11, r = 0.855, or NMS = (0.31 × Stanford) + 8.8, r = 0.888, respectively. The corresponding equations for group 1 (n = 31) were, NMS = (0.21 × CIS) + 8.59, r = 0.841, or NMS = (0.27 × Stanford) + 7.77, r = 0.877, respectively.

Moreover, the NMS standards, when measured by the two IRMAs, produced Tg results with a totally different correlation between methods than that obtained for patients' sera. The regression equations for the NMS standards were (n = 7), NMS = (2.34 × CIS) + 14.24, r = 0.994, and NMS = (2.28 × Stanford) + 0.93, r = 1, with a slope approximately 10-fold that observed for patients' sera or for the CIS or Stanford standards.

Clinical Evaluation

Figure 4 shows results of Tg measurements in 50 patients...
results of all her $^{131}$I scans were still negative up to that date. All three assays were also in agreement in producing positive results (or a positive and a borderline result by the NMS assay) for two other patients who as yet have no positive clinical findings, and gave a negative or borderline Tg result (CIS: < 3; NMS: 5; Stanford: 10 $\mu$g/L) for one patient with a radioiodine uptake of 0.5%, which was confined to the thyroid bed and which could be due to either a thyroid remnant or recurrent disease.

The major difference between methods was the percentage of undetectable Tg results, as opposed to low but measurable Tg results, in patients thought to be free of disease.

Using the CIS assay, 34 patients (85%) had undetectable Tg and only three patients with total thyroidectomy and $^{131}$I ablation had measurable Tg of 5–10 $\mu$g/L. The Stanford assay gave undetectable Tg in 30 patients (75%) and Tg of 5–10 $\mu$g/L in seven patients (two with total thyroidectomy and $^{131}$I treatment and five with total or partial thyroidectomy but no $^{131}$I). By contrast, the NMS RIA gave undetectable Tg in only 12 (30%) patients without disease, while eight patients, including five with total thyroidectomy and $^{131}$I, had Tg results of 10 to 15 $\mu$g/L.

**Discussion**

We found the three assays for Tg in serum to differ significantly in methodology, analytical performance, and clinical applicability and usefulness.

Results for Tg by the CIS and Stanford assays correlated well ($r = 0.964$). The parallelism observed between the Tg standards and serial dilutions of clinical specimens, or analytical recoveries close to those expected theoretically, indicate the high specificity of the antisera used in these two assays (Figures 1 and 3). The minor differences between these two IRMAs are primarily related to non-uniformity of the standard concentrations, rendering the CIS results somewhat higher than the Stanford results. The sensitivities of the CIS (3 $\mu$g/L) and the Stanford assays (2.5 $\mu$g/L) were also comparable, when one takes into account the differences in the standard concentrations.

On the other hand, the NMS assay gave a weaker correlation with the CIS ($r = 0.855$) and the Stanford assay ($r = 0.888$), and the fact that we obtained totally different correlations for the kit standards and patients' sera raises the question of whether these NMS standards are suitable for the interpolation of Tg concentrations in sera. Also, the distinctly different slopes of the various dose-response curves produced by this assay suggest that the NMS antiserum is less specific.

Tg is a large, non-homogeneous protein, which contains many antigenic binding sites. Several forms of Tg with varying iodine content have been isolated. Tg from malignant glands is not necessarily identical with that used to raise antibodies, and repeated freezing and thawing degrades the molecule. Thus, it is plausible that antisera raised in different laboratories against Tg preparations of various degrees of purity or integrity recognize different antigenic determinants and bind different forms of Tg. For instance, an antiserum with high immunoreactivity towards a minor component, a fragment, or other degradation product could explain the several-fold higher Tg values for the NMS standards, as opposed to those for clinical specimens or for other standard preparations.

Differences between the matrices of samples and standards—e.g., the NMS standards contain animal proteins of undisclosed composition—could potentially also cause analytical inaccuracies by interfering with the second-antibody precipitation. Major matrix-associated problems were ruled
out by the fact that dilutions of the highest NMS standard in human Tg-free serum from athyreotic patients produced largely the same dose–response curve as did the kit standards. However, minor matrix effects may account for the deviations in the low range when the CIS or NMS standards were analyzed by the Stanford assay (Figure 3).

Several studies (7, 8, 16–26) on thyroid-cancer patients, including one (21) from this laboratory, indicate that measurement of Tg is a useful diagnostic adjunct to the whole-body 131I scan, the standard technique for determining the presence or absence of disease. Most investigators found that undetectable serum Tg “by a sensitive assay” is highly predictive of a negative whole-body 131I scan. Ashcraft and Van Herle (16) proposed measurements of Tg as the first-line screening test, to be performed at six months to yearly intervals during long-term management of patients treated for differentiated thyroid cancer. Using this strategy, all patients with unmeasurable Tg can be spared repeated radioisotope scans, and only those with Tg above the detection limit or above a certain second reference point, respectively (depending on the patient’s history and treatment), require further evaluation such as a second measurement of Tg and an 131I scan after T4 has been stopped for four to six weeks.

Based on our previous experience with the Stanford assay, and in agreement with Ashcraft and Van Herle (16), we have determined two reference values for each assay. Two cutoff points are necessary because concentrations of Tg in thyroid-cancer patients depend on the extent of thyroid surgery, and (or) prior radioiodine ablation(s) (131I-dose and timing), and on whether or not the patient is taking T4 as replacement at the time. The reference values were selected to provide maximum clinical sensitivity rather than a compromise between optimal clinical sensitivity and specificity. In dealing with a first-line screening test for a serious, life-threatening disease it is more important not to miss any patients with disease than to exclude false positives as much as possible.

Both IRMA evaluated correctly all 10 patients with clinically documented disease; the NMS assay identified only eight. Aside from that, the two IRMA were also more efficient than the competitive binding RIA in distinguishing patients who required no further work-up. Eighty-five percent, or 75%, of disease-free patients had undetectable Tg by the CIS or Stanford assay. By comparison, only 12 patients, 30% of those disease free, or 24% of group 1 had no anti-Tg and undetectable Tg by the NMS RIA, while results for a larger proportion were inconclusive (5–15 μg/L).

About 30% (SD 10%) of all thyroid-cancer patients have anti-Tg antibodies (13). IRMA-type assays permit quantification of Tg in the presence of anti-Tg, because Tg is stripped from interfering human anti-Tg with an excess of solid-phase rabbit anti-Tg antibody. In spite of this, when anti-Tg concentrations are very high, falsely negative results can be produced by IRMA (13). Thirteen of the patients we studied had low (anti-Tg ≤ 50 units/mL) and three patients had moderately elevated antibody levels (anti-Tg = 80–380 units/mL), so it is possible that some of their Tg results by IRMA are false negatives. Of more than 100 patients seen at our clinic during the past years, we have not encountered a single case with low to moderate anti-Tg concentrations where Tg was undetectable but metastatic lesions were large enough to be seen on a whole-body 131I scan. Thus, we assume that patients with Tg undetectable by a sensitive IRMA, even in conjunction with low anti-Tg, may not require the annual isotope scans.

Differences in the specificity of the antisera may have contributed to some controversy in the literature as to the clinical value of Tg measurements. Two recent studies by Echenique et al. (27) and Schwartz et al. (28), involving the NMS kit, in which concentrations of Tg in serum were correlated with the patients’ clinical status and 131I scan, demonstrated considerable overlap in Tg results from patients with active disease and those with no evidence of metastatic disease. One of the reports suggested caution in the use of results for Tg (by the NMS assay) as a replacement for the scan (27).

In conclusion, our data suggest that the first two commercial assays are not equivalent, and that the clinical utility of Tg measurements will, at least to a certain degree, depend on the assay used. The CIS assay appeared to be more nearly accurate, is more generally applicable for clinical use, and thus will probably have a more beneficial impact on patient care.

References
18. Botach H, Schulz E, Lockner B. Serum thyroglobulin estima-


