Comparison of Immunoradiometric Assays for Determination of Thyroglobulin: A Validation Study

L.A. Tortajada-Genaro,^{1*} M.P. Cózar,² J.L. Moreno Frigols,³ and C. Romero de Avila²

¹Departamento Química, Universidad Politécnica de Valencia, Valencia, Spain

²Servicio de Medicina Nuclear, Hospital Clínico de Valencia, Valencia, Spain

³Unidad de Radiofarmacia, Hospital Clínico de Valencia, Valencia, Spain

In this study we compared and validated commercial immunoradiometric assays (IRMA) to determine thyroglobulin (Tg) levels in serum. From a set of 440 samples, 68 were selected to calculate the validation parameters and the clinical performance of the assays. The commercial kits evaluated were the Tg-CTK (DiaSorin), IRMAZenco Tg (ZenTech), and SELco-Tg (Medipan). We found that 21% of the collected samples were in the critical range of concentration. Detection limits were calculated as being below 3 µg/L. Intra- and inter-reproducibility were lower than 3.1% and 9.2%, respectively. Dilution and recovery studies provided quantitative determinations. Correlation regression coefficients

from the results of the methods were obtained. The determined concentrations were compared with the clinical evidence of disease. Variation in the 125-iodine-labeled antibody concentration and control charts showed the robustness of the methods. Analysis time and the simplicity of the methods were also evaluated. Reliable Tg determination is important for monitoring patients with differentiated thyroid cancer (DTC), controlling other thyroid diseases, and assessing the quality of imaging techniques. A strategy for verification and comparison based on analytical parameters and clinical performance is proposed. *J. Clin. Lab. Anal.* 21:147–153, 2007.

© 2007 Wiley-Liss, Inc.

Key words: immunoradiometric assay; IRMA; thyroglobulin; analytical validation

INTRODUCTION

Thyroglobulin (Tg) is a glycoprotein that consists of two subunits of 330 kDa. It is the precursor of thyroid hormones T₃ and T₄. Low concentrations of Tg are excreted into the blood, and elevated Tg levels occur in various thyroid diseases (1).

Because it is highly specific, measurement of serum Tg levels is used to confirm differentiated thyroid cancer (DTC). Since this measure is a reliable and sensitive tumor marker, it is useful for the diagnosis and detection of residual carcinoma after surgery and 131-iodine treatment, and monitoring for recurrent carcinoma (1–3).

Additionally, recent studies have shown that the performance of [¹⁸F]-2-fluoro-2-deoxyglucose positron emission tomography (PET-FDG) depends on Tg values (>10–15 µg/L) (4,5). Finally, the importance of this determination is emphasized because serum Tg levels are the critical factor in DTC protocols (2).

Although other methods have been proposed (6,7), the immunoradiometric assay (IRMA) is an important

method for the routine detection of Tg in human serum (1–3). IRMA assays are based on two antigen-specific antibodies that recognize different binding sites on the antigen (Tg). The first antibody is 125-iodine-labeled (Ab*), and the second is immobilized on the inner surface of the tube (Ab). A sandwich-type complex is formed (Ab-Tg-Ab*). The radioactivity measured is related to the Tg concentration. The sample concentration is calculated using a calibration curve from radioactive signal data and the concentration of Tg standards.

However, the determination of Tg is hampered by various methodological problems, including discrepancies among methods, high detection limits, intra-individual biological variability, thyroglobulin autoanti-

*Correspondence to: Luis Antonio Tortajada Genaro, Departamento Química, Universidad Politécnica de Valencia, Camí de Vera s/n, 46022 Valencia, Valencia, Spain. E-mail: luitorge@qim.upv.es

Received 26 July 2006; Accepted 22 November 2006

DOI 10.1002/jcla.20156

Published online in Wiley InterScience (www.interscience.wiley.com).

body interference (TgAb), and “hook” problems (low values in samples with high Tg concentrations due to alterations in antibody-antigen union) (8–11).

Recently developed methods avoid the interference of TgAb and thus reduce discrepancies among the results (12–14). Moreover, new methods have been standardized with the use of a human thyroglobulin reference material (CRM 457) (15). Nevertheless, the availability of different commercial kits makes it necessary to establish comparison criteria in order to select a suitable method for routine analysis.

In this study we compared three commercial IRMA assays for the determination of Tg: Tg-CTK (DiaSorin, Saluggia, Italy), IRMAZenco Tg (ZenTech, Angleur, Belgium), and SELco-Tg (Medipan, Selchow, Germany). For evaluation of reliability, different analytical parameters (sensitivity, application range, detection limit, repeatability, reproducibility, selectivity, accuracy, and robustness) were calculated under different experimental conditions. We also assessed the clinical performance of the methods by measuring serum Tg levels in different patients. Our objective was to establish guidelines for clinical laboratories to select the optimum IRMA method.

MATERIALS AND METHODS

Patients

Samples from 440 patients were analyzed in the Radioimmunoanalysis Laboratory of Hospital Clínico of Valencia. The patients (278 females and 162 males) were between 14 and 82 years of age (median 48). A total of 384 patients were being seen for routine DTC follow-up. The rest of the patients had healthy thyroids or were being analyzed for other thyroid diseases.

The blood samples were aseptically collected over a 1-year period. Aliquots were frozen (-20°C) after centrifugation (15 min, 760g). To avoid degradation, the samples were stored for only 10–15 days. A total of 28 analysis series (8–29 samples/series) were obtained.

A subset of serum samples was randomly chosen to calculate the prediction abilities of the calibration models. A dilution study was performed using free thyroglobulin serum provided by the manufacturers and applying the following dilution factors: 1, 0.5, 0.1, and 0.05. A recovery study was performed by adding Tg standard (50 $\mu\text{g/L}$) to the samples.

Tg concentration was determined in serum samples from randomly selected patients ($N = 68$). The main characteristics of the patients are summarized in Table 1.

Twenty percent of the patients were classified as normal or as having various benign thyroid diseases, such as hyper- or hypothyroidism. The rest of the patients were diagnosed as having DTC of the papillar

TABLE 1. Characteristics of selected patients

	Men	Women
Number	21	47
Age ^a	43.3 (15–74)	47.8 (27–73)
TSH (mU/L) ^a	63.8 (0.01–225)	28.5 (0.01–269)
AntiTg (U/mL) ^a	37.5 (0–256)	19.8 (0–100)
% Patients CDT	72.2	84.2

^aValues are average (minimum–maximum).

type (92%) or follicular type (8%). Levels of AntiTg ranged between 0–256 U/mL. Since biosynthesis of Tg depends on thyroid stimulating hormone (TSH), the patients were classified according to the treatment step. Half of the selected patients were analyzed with $\text{TSH} \leq 0.2 \text{ mU/L}$ (suppressive treatment). Most of the patients (60%) were also monitored by radioisotope imaging.

We performed a dilution study with labeled antibodies by applying the following dilution factors: 1, 0.9, 0.5, 0.1, and 0.01.

Instrumentation

For semiautomatic procedures we employed a shaker (bioMerieux, Marey L' Étoile, France), a washing system (Nichols Institute Diagnostics, San Juan Capistrano, CA), and a LKB-Wallac 1277 Gamma naster automa gamma counter (Wallac, Turku, Finland).

For automatic procedures we used a STRATEC SR300 (Birkenfeld, Germany), which incorporates a pipetting module for automatic addition of tracer reagent, standards, controls, and samples; a shaking and incubating module; a washing module; and a gamma counter.

Procedure

The kits evaluated in this study were the Tg-CTK (DiaSorin), IRMAZenco Tg (ZenTech) and SELco-Tg (Medipan). We selected these kits because the manufacturers have tested them using a reference material (CRM 457, Community Bureau of References). This reference material is composed of lyophilized human thyroglobulin, the properties of which were previously validated in an interlaboratory comparative study (15).

These methods use human Tg standards and mouse monoclonal antibody 125-iodine label. Assay tubes are coated with rabbit polyclonal antibody for the Tg-CTK method, and mouse monoclonal for the IRMAZenco and SELco-Tg methods.

For the Tg-CTK method, 50 μL of standards or samples and 250 μL of radiolabeled antibody were added in plastic tubes coated with antibody, while for SELco-Tg method are 100 μL of standards or samples

and 200 μ L of radiolabeled antibody. The tubes were incubated for 16–24 and 15–20 hr, respectively.

For the IRMAZenco method, 100 μ L of standards or samples and 100 μ L of diluent were added in plastic tubes coated with antibody, incubated for 2 hr and washed with 2 mL of washing solution. Then 250 μ L of labeled antibody were added and the tubes were incubated for 18–24 hr.

For the three methods, after sandwich complex formation (Ab-Tg-Ab*), each tube was washed (three cycles) to remove any unbound labeled antibody. Radioactivity was measured with a gamma counter.

Validation of the detection limits and reproducibility was performed with the use of semiautomatic and automatic procedures.

Statistical Analysis

A calibration curve was obtained for each series of analysis and calculated as the double-logarithm relationship between counts per minute (cpm) corrected by background signal and the concentration of the standard. Statistical tests were calculated with a 95% confidence level.

RESULTS

The values of 440 samples were classified according to regions defined according to the clinical implications of the Tg levels (Fig. 1). The interval for patients with healthy thyroids fluctuated between 2 and 40 ng/mL. For patients who were free of metastasis after a complete ablation of thyroid tissue, the serum level of Tg was lower than 1 ng/mL when TSH treatment was completely suppressed (63% of studied patients). The patients with recurrence showed levels of about 2 ng/mL. Levels higher than 1 ng/mL and 2 ng/mL (10%) are considered as the alert limit for screening for DTC and starting diagnostic tests and aggressive treatment, respectively. Levels higher than 40 ng/mL (6%) were found in the serum of patients with benign thyroid disease.

Fig. 1. Percentage of serum samples according to thyroglobulin concentration intervals (μ g/L). Number of analyzed samples = 440.

In this study, 21% of the serum Tg levels were in the critical range or concentration interval (1–4 ng/mL) where the Tg value is more critical for making decisions regarding patient management. The results show the relevance of an accurate and precise determination of Tg.

The characteristics of thyroglobulin assays for three compared kits are summarized in Table 2. The verified working intervals of the assays include the interval of clinical utility; however, these intervals can be considered too wide for routine analysis applications. Tg concentrations without a hook effect are suitable for clinical application. There are important differences between the three methods in maximum B/T % (percentage binding rates related to the total radioactivity), which is a measurement of the proportion of Tg that is linked to the total labeled antibody. The SELco-Tg method revealed a higher excess of free 125-iodine-labeled antibody.

The detection limit or minimal detectable concentration ($<0.3 \mu$ g/L) calculated from the calibration curve confirmed the methods' applicability for monitoring patients with ablation of the thyroid gland. We verified these values by calculating the detection limit from the reproducibility of the background signal ($N=10$) measured using two detectors (gamma counters of LKB and STRATEC SR300 instruments). The results were statistically similar to the values shown in Table 2.

We evaluated the assay precision by measuring the relative standard deviation ($N=3$). Intra-assay reproducibility was calculated from replicates of samples in the same analysis, and interassay reproducibility was calculated from replicates of samples in different analyses. The F-test demonstrated that precision was similar for the three kits. The results indicate that the methods are suitable for patient monitoring.

We tested for accuracy by performing a dilution study, and obtained quantitative determinations (91–119% and 97–102%). We also performed a recovery

TABLE 2. Characteristics of thyroglobulin assays

	Tg-CTK	IRMAZENco	SELco-Tg
Number standards	6	7	6
Working range	0–500	0–600	0.3–250
Control samples	2.8 ± 1.0	28 ± 6	2.2 ± 0.9
Concentrations (ng/mL)	59 ± 8		55 ± 10
High dose hook effect	>2000	>100000	>2000
Maxim B/T %	60	70	40
Detection limit (ng/mL)	0.3	0.1	0.3
Intraassay precision (%)	2.9	3.1	3.0
Interassay precision (%)	5.0	9.2	7.2
Incubation steps	1	2	1
Assay duration (hr)	16	2+18	15

using human Tg standards, and obtained quantitative values (95–102%).

Calibration curves calculated from the radioactive response of the Tg standards were studied (Fig. 2). Response and concentration were modeled according to a double logarithmic function. A deviation for low concentrations was observed for the Tg-CTK and IRMAZENco Tg kits. This finding alerted us to possible limitations and errors in the determinations from the samples that contained low Tg concentrations using these kits.

Serum Tg concentrations of selected patients ($N = 68$) with different TSH levels and AntiTg concentrations were compared (Table 1).

The Tg-CTK kit provided higher Tg levels than the detection limit for 30% of patients without evidence of DTC. Although the values were less than $1 \mu\text{g/L}$, the samples can be considered as false positives. The Tg values of other kits were not statistically comparable with values obtained using the Tg-CTK kit for low Tg concentrations. These results indicate the presence of a constant systematic error in this method, and this error was not observed in the other kits.

Figure 3 shows the relation between concentrations of Tg measured with the SELco-Tg and IRMAZENco Tg kits. A significant correlation between results using a linear regression model ($R^2 = 0.986$) was obtained. A t-test indicated that the ordinate was verified statistically to be equal to zero (0.3 ± 0.4), demonstrating the absence of constant error between the methods. The slope of the linear regression model was statistically equal to unity (1.05 ± 0.06), demonstrating the absence of proportional discrepancies between the two methods.

We studied the robustness of the methods by modifying some of the assay conditions. The concentra-

Fig. 2. Calibration curves obtained using the Tg-CTK, IRMAZENco Tg, and SELco Tg assays.

tion of 125-iodine-labeled antibody was modified in order to alter the relationship between radiolabeled antibody (Ab^*) and Tg needed for immunoreaction. Figure 4 illustrates the response radioactive curve of the sandwich complex for different Tg concentrations and different Ab^* dilutions using the SELco-Tg kit. The reaction yield was not affected for small modifications of the antibody/Tg quotient. Only variations of this quotient higher than 50% (dilution factor = 0.5) modified the response as a function of the Tg concentration, and changed the calibration curve slope.

We verified this observation by determining the concentration of control samples. Figure 5 shows that

Fig. 3. Linear correlation analysis between thyroglobulin concentrations obtained using the IRMAZENco Tg and SELco Tg assays. Number of samples = 68.

Fig. 4. Radioactive response curve of Ab-Tg-Ab^* for different thyroglobulin concentrations and different dilutions of antibody using the SELco Tg assay.

Fig. 5. Concentration of thyroglobulin in control samples for different dilutions of antibody using the SELco Tg assay.

the Tg concentrations in these samples had a significant error when the variation of the quotient Ab^*/Tg was higher than 50% using the SELco-Tg kit. Similar results were obtained using the other tested kits. In conclusion, the assays were robust for a wide range of antibody-Tg reaction conditions.

We also evaluated robustness in control charts by measuring the number of out-of-limit values. Concentrations of control samples out of admitted critical intervals were observed in less than 8% of the analysis series.

Finally, we compared the methods in terms of ease of operation and analysis time. The Tg-CTK method required a lower sample volume (50 µL), which reduced the number of determinations in the same serum sample and the volume of extracted blood required. With respect to analysis duration, the SELco-Tg assay yielded a significant reduction (up to 25%). The IRMAZENco Tg assay requires two steps for the formation of the $Ab-Tg-Ab^*$ complex, whereas the other assays use only one step.

DISCUSSION

Although Tg determination is used for diagnosis and control of different diseases, the main application is monitoring of DTC. The presence of recurrence or metastasis is detected when Tg is $>10 \mu\text{g/L}$ in patients without hormonal treatment, and $>2-3 \mu\text{g/L}$ in patients with treatment. A remission is considered when Tg values are undetectable with high TSH. An elevation of Tg levels in a patient after total thyroidectomy necessitates a complete study, including different imaging techniques, ablative $^{131}\text{-iodine}$, or surgery (2,3).

Therefore, as Fig. 1 shows, clinical laboratories must deal with samples from patients with a wide range of

levels, and treatment can be modified as a function of the analysis results. Given the variety of newly-developed methods, a rigorous strategy of comparison should be standardized to allow selection of the appropriate assay according to the criteria of each laboratory. In the same way, an internal validation to determine adequate assays should be realized. This protocol should be applied even though manufacturers have validated the method using CRM457 (15). The selected method for Tg determination should be verified in order to ensure reliable results.

Comparison and validation involve the determination of different quality parameters. In this study we assessed the sensitivity, applicability range, detection limit, repeatability, reproducibility, selectivity, accuracy, and robustness of the three tested methods.

To check the sensitivity, a representation of radioactive response and concentration of $Ab-Tg-Ab^*$ is recommended. This kind of representation informs about possibilities for sample interpolation in the application range (16). The tested methods allowed us to calculate serum Tg concentrations in the useful clinical range. However, the Tg-CTK method showed irregularities for low concentrations. The SELco-Tg method presented a higher excess of free $^{125}\text{-iodine}$ -labeled antibody, although no unspecific interactions were detected.

A low detection limit can inform about the apparition of a recessive in a patient after a complete ablation of thyroids gland, since methods must provide undetectable Tg levels in samples of free metastasis patients. The detection limits were lower than $1 \mu\text{g/L}$ (the clinical alert limit) for the three methods. Nevertheless, the Tg-CTK method levels were higher than the detection limit for 30% of patients without evidence of DTC (false positive).

For Tg determination it is generally recommended that patients should be monitored by one method and one laboratory. This emphasizes the fact that each laboratory must estimate its own precision values by performing repeatability and reproducibility studies in order to control uncertainty associated with the Tg results (12). For the three methods examined here, we observed no significant difference, and the precision values were suitable for monitoring DTC patients.

A high selectivity of Tg determination is an important criterion in the selection of a method due to hook effect problems or TgAb interferences (8–11). The use of highly concentrated samples informs about the hook effect. The methods must provide high limits without negative error due to alterations of the antibody-Tg union. Dilution studies also inform about the absence of these effects because quantitative recovery percentages must be calculated.

In internal validation studies of Tg determination, recovery assays should be included to detect errors due to the presence of AntiTg (8,14). Quantitative recoveries in patients with high AntiTg values have demonstrated the selectivity of tested methods.

A recovery study, by addition and/or by dilution, allows one to evaluate the accuracy against specific perturbing factors (AntiTg) and unspecific factors, such as typical experimental errors associated with an immunodetermination (8,14). Nonquantitative recoveries indicate the presence of proportional systematic errors due to handling and/or measuring, or alterations in patient serum.

Determination of concentration in the same set of representative patients and the calculation of linear regression allow the quantitative comparison of several methods (16). Correlation coefficient and hypothesis tests for independent term and slope validate Tg levels. It is possible to decide whether the methods provide comparable values or systematic errors.

The study of robustness provides valuable information about a checked method. Variations of experimental conditions should not alter Tg determination. Consequently, an IRMA method is selected when the Tg values are robust to different factors, such as labeled antibody concentration or detector oscillations. The tested methods provided suitable results regarding analytical robustness.

Guidelines about reliable results of a clinical laboratory demand quality for each measured series. One can achieve this objective by including the determination of control samples with human Tg standards and checking the tolerance limits. In addition, inter-day analyses of patient samples can be used as a quality control to detect perturbing effects of interference associated with other components of serum sample.

The number of steps required and analysis duration are parameters that the clinical laboratory should take into account. The studied methods have a long incubation time (overnight), and the IRMAZENco Tg method involves a two-step procedure. Worse features are obtained from longer-duration methods because they require more handling of samples and more instrument and personnel time.

CONCLUSIONS

Checking of validation quality parameters is needed for the selection of a suitable IRMA method for Tg determination in serum. The Tg values obtained can be used as decisive information regarding diseases. A laboratory must establish selection criteria, including analytical, clinical, operative, and economical considerations.

In this study we have proposed a selection strategy that considers general and specific methodological problems and critical concentration levels. The first objective of this strategy is to test the accuracy and precision associated with Tg concentrations. The selected method must provide the real concentration of serum Tg with the lowest dispersion. The second objective is to verify that these quality characteristics are maintained during different analyses to allow the reliable monitoring of thyroid diseases.

The proposed validation includes dilution and recovery studies, verification of detection limits by comparing different procedures, checking of calibration models, reproducibility and repeatability assays, variation of experimental conditions, and control charts.

This validation protocol was applied for the evaluation of three commercial methods. According to the results, the SELco-Tg method shows better global features, including a shorter analysis time and the absence of systematic errors.

REFERENCES

1. Dunn JT. Clinical usefulness of serum thyroglobulin assays. In: Hamburger JI, editor. *Diagnostic methods in clinical thyroidology*. New York: Springer Verlag; 1989. p 127–157.
2. Schlumberger M, Pacini F, Wiersinga WM, et al. Follow-up and management of differentiated thyroid carcinoma: a European perspective in clinical practice. *Eur J Endocrinol* 2004;151: 539–548.
3. Spencer CA, LoPresti JS, Fatemi S, Nicoloff JT. Detection of residual and recurrent differentiated thyroid carcinoma by serum thyroglobulin measurement. *Thyroid*. 1999;9:435–441.
4. Giammarile F, Hafdi Z, Bornaud C, et al. Is [¹⁸F]-2-fluoro-2-deoxy-d-glucose (FDG) scintigraphy with non-dedicated positron emission tomography useful in the diagnostic management of suspected metastatic thyroid carcinoma in patients with no detectable radioiodine uptake? *Eur J Endocrinol* 2003;149: 293–300.
5. Ruiz Franco-Baux JV, Borrego Dorado I, Gomez Camarero P, et al. F-18-fluorodeoxyglucose positron emission tomography on patients with differentiated thyroid cancer who present elevated human serum thyroglobulin levels and negative I-131 whole body scan. *Rev Esp Med Nucl* 2005;24:5–13.
6. Morgenthaler NG, Froehlich J, Rendl J, et al. Technical evaluation of a new immunoradiometric and a new immunoluminometric assay for thyroglobulin. *Clin Chem* 2002;48:1077–1083.
7. Erali M, Bigelow RB, Meikle AW. ELISA for thyroglobulin in serum: recovery studies to evaluate autoantibody interference and reliability of thyroglobulin values. *Clin Chem* 1996;42:766–770.
8. Okosieme OE, Evans C, Moss L, Parkes AB, Premawardhana LD, Lazarus JH. Thyroglobulin antibodies in serum of patients with differentiated thyroid cancer: relationship between epitope specificities and thyroglobulin recovery. *Clin Chem* 2005;51:729–734.
9. Spencer CA, Bergoglio LM, Kazarosyan M, Fatemi S, LoPresti JS. Clinical impact of thyroglobulin (Tg) and Tg autoantibody method differences on the management of patients with differentiated thyroid carcinomas. *J Clin Endocrinol Metab* 2005;90:5566–5575.

10. Marquet PY, Daver A, Sapin R, et al. Highly sensitive immunoradiometric assay for serum thyroglobulin with minimal interference from autoantibodies. *Clin Chem* 1996;42: 258–262.
11. Massart C, Maugendre D. Importance of the detection method for thyroglobulin antibodies for the validity of thyroglobulin measurements in sera from patients with Graves disease. *Clin Chem* 2002;48:102–107.
12. Ferrari L, Seregni E, Aliberti G, et al. Comparative evaluation of two methods to assay thyroglobulin serum concentrations in patients with differentiated thyroid carcinomas. *Q J Nucl Med Mol Imaging* 2004;48:237–242.
13. Giovanella L, Ceriani L, Garancini S. High-sensitive 2nd generation thyroglobulin immunoradiometric assay. Clinical application in differentiated thyroid cancer management. *Q J Nucl Med* 2002;46:319–322.
14. Bourrel F, Hoff M, Regis H, Courriere P, Caron P. Immunoradiometric assay of thyroglobulin in patients with differentiated thyroid carcinomas: need for thyroglobulin recovery tests. *Clin Chem Lab Med* 1998;36:725–730.
15. Feldt-Rasmussen U, Profilis C, Colinet E, et al. Human thyroglobulin reference material (CRM 457). I: assessment of homogeneity, stability and immunoreactivity. *Ann Biol Clin (Paris)* 1996;54:337–342.
16. Campins P, Tortajada-Genaro LA, Messeguer S, Blasco F, Sevillano A, Molins C. Creatinine determination in urine samples by batchwise kinetic procedure and flow injection analysis using the Jaffé reaction: chemometric study. *Talanta* 2001;55:1079–1089.