Serum Thyroglobulin (Tg) Monitoring of Patients with Differentiated Thyroid Cancer Using Sensitive (Second-Generation) Immunometric Assays Can Be Disrupted by False-Negative and False-Positive Serum Thyroglobulin Autoantibody Misclassifications

Carole Spencer, Ivana Petrovic, Shireen Fatemi, and Jonathan LoPresti

University of Southern California (C.S., I.P., J.L.), Los Angeles, California 90089; and Kaiser Permanente (S.F.), Panorama City, California 91402

Context: Reliable thyroglobulin (Tg) autoantibody (TgAb) detection before Tg testing for differentiated thyroid cancer (DTC) is critical when TgAb status (positive/negative) is used to authenticate sensitive second-generation immunometric assay (2GIMA) measurements as free from TgAb interference and when reflexing “TgAb-positive” sera to TgAb-resistant, but less sensitive, Tg methodologies (radioimmunoassay [RIA] or liquid chromatography-tandem mass spectrometry [LC-MS/MS]).

Objective: The purpose of this study was to assess how different Kronus (K) vs Roche (R) TgAb method cutoffs for “positivity” influence false-negative vs false-positive serum TgAb misclassifications that may reduce the clinical utility of reflex Tg testing.

Methods: Serum Tg2GIMA, TgRIA, and TgLc-MS/MS measurements for 52 TgAb-positive and 37 TgAb-negative patients with persistent/recurrent DTC were compared. A total of 1426 DTC sera with TgRIA of ≥1.0 μg/L had false-negative and false-positive TgAb frequencies determined using low Tg2GIMA/TgRIA ratios (<75%) to indicate TgAb interference.

Results: TgAb-negative patients with disease displayed Tg2GIMA, TgRIA, and TgLc-MS/MS serum discordances (% coefficient of variation = 24 ± 20%, range, 0%–100%). Of the TgAb-positive patients with disease, 98% had undetectable/lower Tg2GIMA vs either TgRIA or TgLc-MS/MS (P < .01), whereas 8 of 52 (15%) had undetectable Tg2GIMA + TgLc-MS/MS associated with TgRIA of ≥1.0 μg/L. Receiver operating characteristic curve analysis reported more sensitivity for TgAb method K vs R (81.9% vs 69.1%, P < .001), but receiver operating characteristic curve cutoffs (≥0.6 kIU/L [K] vs > 40 kIU/L [R]) had unacceptably high false-negative frequencies (22%–32%), whereas false positives approximated 12%. Functional sensitivity cutoffs minimized false negatives (13.5% [K] vs 21.3% [R], P < .01) and severe interferences (Tg2GIMA, <0.10 μg/L; 0.7% [K] vs 2.4% [R], P < .05) but false positives approximated 23%.

Conclusions: Reliable detection of interfering TgAbs is method and cutoff dependent. No cutoff eliminated both false-negative and false-positive TgAb misclassifications. Functional sensitivity cutoffs were optimal for minimizing false negatives but have inherent imprecision (20% coefficient of variation) that, exacerbated by TgAb biologic variability during DTC monitoring, could cause TgAb status to fluctuate for patients with low TgAb concentrations, prompting unnecessary Tg method changes and disrupting Tg monitoring. Laboratories using reflexing should limit Tg method changes by considering a patient’s Tg + TgAb testing history in addition to current TgAb status before Tg method selection. (J Clin Endocrinol Metab 99: 4589–4599, 2014)

Abbreviations: CV, coefficient of variation; DTC, differentiated thyroid cancer; FS, functional sensitivity; 2GIMA, second-generation immunometric assay; K, Kronus; LC-MS/MS, liquid chromatography-tandem mass spectrometry; LOQ, limit of quantitation; MCO, manufacturer-recommended cutoff; NS, not significant; R, Roche; ROC, receiver operating characteristic; RIA, radioimmunoassay; Tg, thyroglobulin, TgAb, thyroglobulin autoantibody.
Serum thyroglobulin (Tg) is the primary biochemical tumor marker used to detect recurrence in patients with differentiated thyroid cancers (DTCs) (1). Unfortunately, the thyroglobulin autoantibodies (TgAbs) present in 25% to 30% of patients with DTCs can interfere with Tg measurement (2–11). Automated (second-generation) immunometric assays (2GIMAs) are rapidly becoming the standard of care because they have superior functional sensitivity (FS) (0.05–0.10 μg/L) for detecting basal Tg without recombinant human TSH stimulation (9, 12–18). However, TgAb interference with Tg2GIMA causes underestimation (false low/undetectable) Tg2GIMA (10, 11, 19). In contrast, the radioimmunoassay (RIA) and liquid chromatography-tandem mass spectrometry (LC-MS/MS) classes of Tg methods resist TgAb interference but have an order of magnitude inferior FS (0.5–1.0 μg/L), lack automation, and are not universally available (4, 6, 7, 9, 11, 12, 17, 19, 20).

Laboratories often reflex Tg measurement to RIA or LC-MS/MS when the serum Tg concentration is above a fixed cutoff set to define TgAb “positivity.” This strategy is designed to maximize clinical sensitivity by restricting Tg2GIMA measurement to “TgAb-negative” sera, while minimizing interference by reflexing “TgAb-positive” sera to a TgAb-resistant Tg methodology (RIA or LC-MS/MS) (6, 11, 19, 20). Clearly, the sensitivity and specificity of the TgAb method has a critical impact on the reliability of this reflex strategy, because false-negative TgAb tests can lead to inappropriately low/undetectable Tg2GIMA that can mask disease, whereas false-positive TgAb tests may prompt unnecessary reflexing to a less sensitive methodology that may fail to detect low Tg disease (21–27). Guidelines caution against unnecessarily changing Tg methods because of wide disparities in numeric Tg values reported by different methods for the same serum (9, 12, 17, 18, 28–30).

Studies use concordance between TgAb methods to assess the reliability of TgAb detection (2, 10, 31–35). This study directly evaluated the effects of interfering TgAbs on Tg measurement in terms of a low ratio (<75%) between values reported by a TgAb-sensitive Tg2GIMA and a TgAb-resistant TgRIA (2, 10, 11, 17, 36, 37). The Kronus TgAb method was selected for testing because this semi-automated radioassay predates current automated TgAb tests and has provided stable TgAb values for more than 2 decades (4). The Roche TgAb method was selected because laboratories adopted this method (38) after our previous study (10) found it to be more sensitive than 2 other automated TgAb tests (Beckman and Siemens) compared with Kronus as the reference.

Sensitivity differences between TgAb methods reflect the assay design, the specificity of the TgAb test reagents, and the cutoff selected to define a “positive” TgAb result. Previously, we reported that manufacturer-recommended cutoffs (MCOs) for TgAbs were set too high to reliably detect interfering TgAbs and were more appropriate for diagnosing thyroid autoimmunity (10). The goals of the current study were to assess whether lower cutoffs could reduce false-negative and minimize false-positive TgAb misclassifications that could have a negative impact on DTC monitoring when a fixed TgAb cutoff value was used to reflex Tg testing to different methods.

## Materials and Methods

### Tg methods

Both Tg methods were standardized against the International Reference Preparation CRM-457.

#### TgRIA

This TgRIA, developed by the USC Endocrine Laboratory, University of Southern California, Los Angeles (4, 10, 28, 39) had first-generation FS (0.5 μg/L) established by guidelines (Tg assay FS being the lowest Tg concentration measured with 20% between-run coefficient of variation [CV] in human serum over 6–12 months [clinical interval typical for DTC monitoring] employing ≥2 different reagent lots and calibrators [30]). Between-run precisions for TgAb-positive human sera measured over 1 year were 14.6%, 7.6%, 6.5%, 9.3%, and 7.8% for Tg concentrations of 0.75, 3.2, 12.0, 20.3, and 34.8 μg/L, respectively. Within-run precisions were 8.4%, 7.1%, 1.5%, and 5.3% at Tg concentrations of 0.94, 2.0, 16.1, and 31.6 μg/L, respectively.

#### Tg2GIMA

This Tg2GIMA method was the Access immunochemiluminometric method (Beckman Coulter). The FS established for TgAb-negative human serum was 0.05 μg/L. However, this study considered values of <0.10 μg/L “undetectable” in accord with other reports (15). Between-run precisions for TgAb-negative human sera measured over a 14-month period were 13.0%, 6.5%, 4.2%, and 4.0% for Tg concentrations of 0.13, 0.59, 7.0, and 54 μg/L, respectively. Within-run precisions were 4.1%, 3.2%, 1.7%, and 1.9% at Tg concentrations of 0.15, 0.76, 7.0, and 106 μg/L, respectively.

#### TgLC-MS/MS

TgLC-MS/MS measurements were made by the Mayo Medical Laboratories (40). The limit of quantitation (LOQ) (20% CV between runs made over 1 month) was 0.5 μg/L (6, 9).

#### Tg2GIMA/TgRIA ratios: used to indicate TgAb interference

A low Tg2GIMA/TgRIA ratio (<75%) was considered to indicate TgAb interference as previously established (2, 10, 36, 37). Sera with TgAb concentrations below any specific cutoff displaying a low Tg2GIMA/TgRIA ratio were considered false negative. Conversely, sera with TgAb values more than or equal to the cutoff without a low Tg2GIMA/TgRIA ratio were considered false positive. Sera with severe interference had unequivo-
cally detectable TgRIA (≥1.0 μg/L) and undetectable (<0.10 μg/L) Tg²⁵⁴IMA.

**TgAb assays**

Both TgAb assays claimed standardization against the World Health Organization International Reference Preparation IRP 65/93 and were performed according to the manufacturers’ protocols.

**Kronus/RSR TgAb (K)**

Method K was a semiautomated radioassay (Kronus also known as RSR) that uses ¹²⁵I-labeled Tg to bind TgAb in a diluted (1:21) serum specimen. The TgAb-Tg ¹²⁵I-labeled complex is precipitated with protein A. Between-run precisions for human sera measured over 1 year were 14.3%, 9.5%, 8.7%, and 11.1% for TgAb concentrations of 0.7, 2.6, 5.2, and 11.0 kIU/L, respectively. TgAb concentrations of >17 kIU/L were measured at appropriate dilutions. Within-run precisions of TgAb measured in human sera were 5.7%, 3.1%, 1.9%, and 3.1% at concentrations of 0.8, 3.0, 6.2, and 14.6 kIU/L, respectively. The LOD (within-run precision of zero matrix [41]) was 0.3 kIU/L.

**Roche Elecsys TgAb (R)**

Method R was the automated electrochemiluminescent competitive immunometric assay method (Roche Diagnostics) whereby serum TgAb competes for biotinylated human Tg with ruthenium-labeled TgAb. The Tg-TgAb complexes form and bind streptavidin-coated microparticles and are magnetically captured onto the surface of an electrode. Between-run precisions for human sera measured over 1 year were 22.5%, 20.1%, 13.1%, 9.3%, 14.5%, 8.9%, and 16.7% for concentrations of 13, 17, 27, 35, 87, 146, and 1329 kIU/L, respectively. Within-run precisions were 4.9%, 5.1%, and 5.6% at concentrations of 63, 115, and 2894 kIU/L, respectively. The LOD (within-run precision of zero matrix) was 0.3 kIU/L.

**Serum specimen groups**

Group A comprised 89 sera from patients with DTC (88 papillary and 1 Hurtle cell) drawn a median of 8 days (range, 0–167 days) before detection of persistent/recurrent disease by biopsy or anatomic imaging. Group A was used to compare the different classes of the Tg methods (Tg²⁵⁴IMA, TgRIA, and TgLC-MS/MS). Of the patients, 37 were TgAb negative (below method K and R FSs), and 52 were classified as TgAb positive by both methods. Group B comprised 1426 sera from 1110 patients with DTC selected for sufficient volume for Tg²⁵⁴IMA and TgAb (both K and R) with unequivocally detectable TgRIA (range, 1.0–40 μg/L), allowing Tg²⁵⁴IMA/TgRIA ratio calculations. Group B sera were selected to cover a range of TgAb values from LOD to very high TgAb concentrations. Group C comprised 607 sequential DTC sera received for routine Tg + TgAb testing that had K and R TgAb measurements and were used to establish the range and frequency of TgAb values typical of clinical practice.

**Serum TgAb analyses using different cutoffs for TgAb positivity**

Methods K and R TgAb data for groups B and C were analyzed as subgroups covering the measurement range. Cutoffs for TgAb positivity represented those frequently used by laboratories: LOD, FS, receiver operating characteristic (ROC) curve, and MCO. For method K, these cutoffs were 0.3 (LOD), 0.4 (FS), 0.5, 0.6, 0.7 (ROC curve), 0.8, 0.9, 1.0 (MCO), 1.2, 1.5, 2, 3, 5, 10, 25, 50, 100, and ≥250 kIU/L, and for method R the cutoffs were 10 (LOD), 15, 18, 20, 22 (FS), 25, 30, 35, 41 (ROC curve), 50, 60, 80, 115 (MCO), 200, 400, 600, and ≥1000 kIU/L. ROC curve analysis on group B subgroups used low Tg²⁵⁴IMA/TgRIA ratios (≤75%) to detect interfering TgAb.

**Statistical analyses**

ROC curve analysis was performed using MedCalc 12.3.0 (Mariakerke) software. A true-positive serum had TgAb values more than or equal to the selected cutoff with an abnormally low Tg²⁵⁴IMA/TgRIA ratio (≤75%) indicating TgAb interference (2, 10, 17, 36, 37). A false-positive serum had TgAb values more than or equal to the selected cutoff without a low Tg²⁵⁴IMA/TgRIA ratio (≥75%) indicating the absence of TgAb interference. A true-negative serum had TgAb below the cutoff without a low Tg²⁵⁴IMA/TgRIA ratio. A false-negative serum had TgAb below the cutoff with a low Tg²⁵⁴IMA/TgRIA ratio. Statistical analyses were performed with XLSTAT and Student t tests using SPSS software (version 13.0). Statistical significance was set at a value of P < .05.

**Results**

Figure 1 shows comparative serum Tg²⁵⁴IMA, TgLC-MS/MS, and TgRIA measurements for the group A patients with documented disease. Of these patients, 37 had no TgAb detected (below the FS of both methods K and R) (Figure 1A), and 52 had TgAb detected by both methods (Figure 1B). TgAb-negative patients had comparable Tg²⁵⁴IMA/TgRIA vs Tg²⁵⁴IMA/TgLC-MS/MS ratios (median [range]: 81 [1%–182%] vs 91 [1%–221%], respectively, not significant [NS]). Despite correlations between the methods (Tg²⁵⁴IMA = 0.94TgRIA − 3.1, r = 0.98; Tg²⁵⁴IMA = 1.05TgLC-MS/MS − 2.1, r = 0.98; and TgLC-MS/MS = 0.88TgRIA − 0.4, r = 0.98), between-method % CVs for the 3 measurements made on individual sera were high (mean ± SD, 24 ± 20, range, 0%–100%). Of the patients, 17 of 37 (46%) displayed >20% between-method discordances, and in 5 of 37 (14%) discordances exceeded 30% (indicated in Figure 1A as solid red circles). Serum Tg values for the highly discordant patients were <0.10/3.5/2.9, 16.6/39.4/25.4, 2.0/6.1/1.8, 14.1/6.4/39.6, and 0.30/0.9/0.8 μg/L for Tg²⁵⁴IMA, TgLC-MS/MS, and TgRIA, respectively. These between-method discordances exceeded the <10% CV expected for repetitive measurements made with a single method. One TgAb-negative patient with disease had no Tg detected by biopsy or anatomic imaging.
any method. The ratios for the TgAb-positive patients with disease were comparable irrespective of whether TgRIA or TgLC-MS/MS was used as the denominator (mean ± SD [range] 41 ± 29 [3%-99%] vs 50 ± 26 [5%-115%] for Tg2GIMA/TgRIA vs Tg2GIMA/TgLC-MS/MS ratios, respectively, NS). Both the Tg2GIMA/TgRIA and Tg2GIMA/TgLC-MS/MS ratios for group A TgAb-positive patients were significantly lower than those for group A TgAb-negative patients (P < .001). The group A TgAb-positive sera also displayed correlations between the methods: Tg2GIMA = 0.58TgRIA − 0.99, r = 0.95; Tg2GIMA = 0.60TgLC-MS/MS − 0.35, r = 0.97; and TgLC-MS/MS = 0.96TgRIA − 1.10, r = 0.98. A total of 51 of 52 (98%) Tg2GIMA values for TgAb-positive patients were either undetectable (20 of 52, 38%) or lower than either TgLC-MS/MS (P < .02) or TgRIA (P < .01). TgLC-MS/MS was below the LOQ in 12 patients (23%) (indicated by red squares in Figure 1B). Six of 12 sera (solid squares) had unequivocally undetectable TgLC-MS/MS values (no peak = solid red squares) and sera with marginally detectable TgLC-MS/MS values in the 0.3 to 0.5 μg/L range (open red squares) (6, 9, 40).

any method. The ratios for the TgAb-positive patients with disease were comparable irrespective of whether TgRIA or TgLC-MS/MS was used as the denominator (mean ± SD [range] 41 ± 29 [3%-99%] vs 50 ± 26 [5%-115%] for Tg2GIMA/TgRIA vs Tg2GIMA/TgLC-MS/MS ratios, respectively, NS). Both the Tg2GIMA/TgRIA and Tg2GIMA/TgLC-MS/MS ratios for group A TgAb-positive patients were significantly lower than those for group A TgAb-negative patients (P < .001). The group A TgAb-positive sera also displayed correlations between the methods: Tg2GIMA = 0.58TgRIA − 0.99, r = 0.95; Tg2GIMA = 0.60TgLC-MS/MS − 0.35, r = 0.97; and TgLC-MS/MS = 0.96TgRIA − 1.10, r = 0.98. A total of 51 of 52 (98%) Tg2GIMA values for TgAb-positive patients were either undetectable (20 of 52, 38%) or lower than either TgLC-MS/MS (P < .02) or TgRIA (P < .01). TgLC-MS/MS was below the LOQ in 12 patients (23%) (indicated by red squares in Figure 1B). Six of 12 sera (solid squares) had unequivocally undetectable TgLC-MS/MS (no peak detected), whereas 6 of 12 sera had marginal TgLC-MS/MS in the 0.3 to 0.5 μg/L range. Two patients with undetectable TgLC-MS/MS also had undetectable Tg2GIMA and TgRIA, and 2 with marginal TgLC-MS/MS had some Tg detected by both Tg2GIMA (0.16 and 0.19 μg/L) and TgRIA (0.9 and 2.8 μg/L), whereas the remain-

The Endocrine Society. Downloaded from press.endocrine.org by [individual/User displayNames] on 18 December 2014 at 13:52. For personal use only. No other uses without permission. All rights reserved.
ing TgAb. Table 1 summarizes subgroup characteristics, showing the performance at the most commonly used cutoffs (LOD, FS, ROC curve, and MCO) by bold type. Table 1 and Figure 3 show that as the TgAb cutoff increased, the number of false-negative misclassifications rose and the number of false-positive misclassifications declined. Comparable median TgRIA and Tg2GIMA were seen for the cutoffs close to the FS, but as the cutoff increased median Tg2GIMA progressively decreased to 0.10 μg/L at the highest cutoff, whereas median TgRIA remained relatively stable across the TgAb range. The preferential influence of rising TgAb on Tg2GIMA but not TgRIA produced a progressive decline in median Tg2GIMA/TgRIA ratios. The frequency of severe interference steadily rose with the increasing cutoff to peak at ~15% at high TgAb concentrations. ROC curve analysis (Figure 4) reported optimal cutoffs of >0.6 kIU/L for method K vs >40 kIU/L for method R and showed that method K had higher sensitivity for TgAb detection (81.9% vs 69.1%, K vs R, respectively, *P < .001*) and more area under the curve (0.89 vs 0.85, respectively, *P < .001*). With ROC curve cutoffs of 0.7 kIU/L (K) vs 41 kIU/L (R), method K classified fewer sera as TgAb negative (46.3% vs 55.1%, K vs R, respectively, *P < .001*) of which fewer were *false negatives* (21.7% vs 31.6%, K vs R, respectively, *P < .001*). Furthermore, fewer false-negative sera displayed severe interference using method K (2.1% vs 6.2%, K vs R, respectively, *P < .001*). When the FS cutoff recommended by guidelines (30) was used, method K also displayed superior sensitivity. Specifically, although the percentage of TgAb negatives was comparable using FS cutoffs (32.1% vs 34.9%, K vs R, respectively, NS), fewer method K TgAb negatives were *false negatives* (13.5% vs 21.3%, K vs R, respectively, *P < .01*) and fewer displayed severe interference (0.7% vs 2.4%, K vs R, respectively, *P < .05*). In addition, using method K, fewer false-negative sera had inappropriately low Tg2GIMA (0.10 μg/L) (0.7% vs 4.4%, K vs R, respectively, *P < .001*). Consistent with superior method K sensitivity, fewer method K false negatives were judged positive by method R compared with method R false negatives judged positive by method K (16.1% vs 50.9% K vs R, respectively, *P < .001*). A chart review revealed that a significant number of false-negative sera were from patients with DTC with a prior (method K) history of TgAb positivity (32.2% vs 20.8%, K vs R, respectively). Taken together, these data suggested that a low Tg2GIMA/Tg RIA ratio detected interfering TgAb more sensitively than direct TgAb measurement using either method, as reported previously (37).

Methods K and R had comparable false-positive frequencies using FS cutoffs (22.4% vs 23.8%, K vs R, respectively, NS). However, method R had a wider subfunctional sensitivity range (10–21 kIU/L) vs that for method K (0.1–0.3 kIU/L), suggesting a higher method R potential for reporting false positives using cutoffs of <22 kIU/L.

Figure 2. A, Relationship between TgRIA (abscissa) and Tg2GIMA (ordinate) for the 367 group B sera that were classified as TgAb negative according to the FS cutoff of both methods K and R. B, Relationship between TgRIA (abscissa) and Tg2GIMA (ordinate) for the 834 group B sera that were classified as TgAb-positive according to the FS cutoff of both methods K and R. The table below shows the median TgAb for methods K and R, and the percentage of severe interferences (Tg2GIMA of <0.10 μg/L) seen when TgAb-positive group B sera were analyzed according to TgRIA values (1.0–2.5, 2.6–5.0, 5.1–10, or >10 μg/L).

<table>
<thead>
<tr>
<th>TgRIA (μg/L) (n)</th>
<th>1.0–2.5</th>
<th>2.6–5.0</th>
<th>5.1–10</th>
<th>&gt;10</th>
</tr>
</thead>
<tbody>
<tr>
<td>% with Tg2GIMA &lt;0.10 μg/L</td>
<td>39</td>
<td>27</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>median TgAb (K) kIU/L</td>
<td>2.0</td>
<td>6.0</td>
<td>3.5</td>
<td>2.0</td>
</tr>
<tr>
<td>median TgAb (R) kIU/L</td>
<td>55</td>
<td>188</td>
<td>154</td>
<td>134</td>
</tr>
</tbody>
</table>

The Endocrine Society. Downloaded from press.endocrine.org by [individualUser.displayName] on 18 December 2014 at 13:52 For personal use only. No other uses without permission. All rights reserved.
Discussion

Reliable TgAb detection before Tg testing is critical when serum Tg$^{2G}$IMA measurements are used as a DTC tumor marker. A “negative” TgAb test is used to authenticate the absence of TgAb interference, whereas a “positive” test suggests that the Tg$^{2G}$IMA may be unreliable and report falsely low/undetectable serum Tg values that could mask disease—the most serious problem that compromises the clinical utility of Tg$^{2G}$IMA testing. Laboratories often maximize the clinical sensitivity of Tg$^{2G}$IMA measurement while minimizing the TgAb interference problem by reflexing TgAb-positive sera to a TgAb-resistant, but less sensitive, class of Tg method (RIA or LC-MS/MS) (5, 17). This study has clinical implications for laboratories that perform reflex Tg testing. The study confirmed that false-negative TgAb misclassifications were unacceptably high (30%–40%) using the MCOs for TgAb (10), whereas the FS cutoff minimized both false negatives and severe interferences associated with falsely low/undetectable Tg$^{2G}$IMA. However, the FS cutoffs were associated with an approximate 20% false-positive frequency that could prompt unnecessarily reflexing of many sera to less sensitive RIA or LC-MS/MS methodology, which could fail to detect disease associated with low Tg concentrations (21–27). The FS cutoff also has an inherent 20% between-run imprecision (30) that could lead to false fluctuations in TgAb status (positive to negative or vice versa) while monitoring patients with low Tg concentrations and, as a result, prompt unnecessary changes in the Tg method used. Guidelines caution about the need for Tg method continuity, because different numeric Tg values are reported when the same serum sample is measured by different methods (4, 9, 12, 17, 28–30). Although there were correlations between the methods in the absence of TgAb
overall, the discordance between the Tg\(^{2G}\)IMA, TgRIA, and TgLC-MS/MS measurements made for the individual TgAb-negative patients with disease was higher (mean 24\% CV, range 0\%–100\%) than would be typical when Tg\(^{2G}\)IMA is used consistently to monitor an individual patient at 6- to 12-month intervals (9). The 5 patients (14\%) who displayed severe (>30\% CV) between-method discordances (red circles in Figure 1A) emphasize how serum Tg monitoring could be disrupted by unnecessarily Tg method changes. Laboratories that reflex Tg testing to different methods based on the TgAb FS cutoff could guard against unnecessarily changing the Tg method by considering the patient’s Tg and TgAb testing history, in addition to the TgAb status of the current specimen, before selecting an appropriate Tg method.

This study evaluated how different Kronus/RSR vs Roche TgAb method cutoffs for positivity influenced TgAb false-negative and false-positive frequencies, using a low Tg\(^{2G}\)IMA/TgRIA ratio (<75\%) to indicate the presence of interfering TgAb (2, 10, 36, 37). Table 1 and Figure 3 show that false positives declined whereas false negatives rose with a rising cutoff, but that no cutoff used for either method eliminated all false-negative and false-positive TgAb misclassifications. ROC curve analysis reported higher Kronus sensitivity vs Roche sensitivity, although both the ROC curve–determined and manufacturer-recommended cutoffs of both methods had unacceptably high false-negative frequencies (20\%–40\%) that included many cases of severe interference (Tg\(^{2G}\)IMA of <0.10 µg/L) (10, 33, 34). This is the first study of TgLC-MS/MS

<table>
<thead>
<tr>
<th>Table 1. Group B Sera Subgroup Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serina Group, n (%)</td>
</tr>
<tr>
<td>TgAb, Median, kIU/L</td>
</tr>
<tr>
<td>Tg RIA, Median, µg/L</td>
</tr>
<tr>
<td>Tg(^{2G})IMA, Median, µg/L</td>
</tr>
<tr>
<td>Tg(^{2G})IMA/Tg RIA Ratio, Median, %</td>
</tr>
<tr>
<td>False Positive, %</td>
</tr>
<tr>
<td>False Negative, %</td>
</tr>
<tr>
<td>Tg(^{2G})IMA, %</td>
</tr>
<tr>
<td>&lt;0.30 µg/L</td>
</tr>
<tr>
<td>&lt;0.10 µg/L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cutoff, kIU/L</th>
<th>Serina Group, n (%)</th>
<th>TgAb, Median, kIU/L</th>
<th>Tg RIA, Median, µg/L</th>
<th>Tg(^{2G})IMA, Median, µg/L</th>
<th>Tg(^{2G})IMA/Tg RIA Ratio, Median, %</th>
<th>False Positive, %</th>
<th>False Negative, %</th>
<th>Tg(^{2G})IMA, %</th>
<th>&lt;0.30 µg/L</th>
<th>&lt;0.10 µg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3 (LOD)</td>
<td>458 (32.1)</td>
<td>0.3</td>
<td>15</td>
<td>6.6</td>
<td>5.5</td>
<td>93.8</td>
<td>22.4</td>
<td>13.5</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>0.4 (FS)</td>
<td>62 (4.3)</td>
<td>0.4</td>
<td>23</td>
<td>3.2</td>
<td>3.2</td>
<td>88.2</td>
<td>22.4</td>
<td>13.5</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>0.5</td>
<td>93 (6.5)</td>
<td>0.5</td>
<td>21</td>
<td>3.5</td>
<td>3.5</td>
<td>79.3</td>
<td>19.1</td>
<td>15.4</td>
<td>1.5</td>
<td>1.2</td>
</tr>
<tr>
<td>0.6</td>
<td>47 (3.3)</td>
<td>0.6</td>
<td>28</td>
<td>3.2</td>
<td>2.7</td>
<td>74.3</td>
<td>14.6</td>
<td>19.4</td>
<td>3.8</td>
<td>1.8</td>
</tr>
<tr>
<td>0.7 (ROC curve)</td>
<td>39 (2.7)</td>
<td>0.7</td>
<td>30</td>
<td>2.4</td>
<td>2.0</td>
<td>68.4</td>
<td>12.5</td>
<td>21.7</td>
<td>5.0</td>
<td>2.1</td>
</tr>
<tr>
<td>0.8</td>
<td>35 (2.5)</td>
<td>0.8</td>
<td>34</td>
<td>1.9</td>
<td>0.8</td>
<td>43.0</td>
<td>11.0</td>
<td>23.7</td>
<td>7.7</td>
<td>2.4</td>
</tr>
<tr>
<td>0.9</td>
<td>39 (2.7)</td>
<td>0.9</td>
<td>39</td>
<td>1.8</td>
<td>0.6</td>
<td>49.9</td>
<td>10.3</td>
<td>26.2</td>
<td>8.6</td>
<td>3.7</td>
</tr>
<tr>
<td>1.0 (MCO)</td>
<td>58 (4.1)</td>
<td>1.0</td>
<td>40</td>
<td>2.1</td>
<td>1.3</td>
<td>47.0</td>
<td>9.3</td>
<td>28.6</td>
<td>10.3</td>
<td>3.8</td>
</tr>
<tr>
<td>1.2</td>
<td>64 (4.5)</td>
<td>1.3</td>
<td>45</td>
<td>2.4</td>
<td>1.5</td>
<td>55.0</td>
<td>8.7</td>
<td>32.5</td>
<td>11.4</td>
<td>4.7</td>
</tr>
<tr>
<td>1.5</td>
<td>75 (5.3)</td>
<td>1.7</td>
<td>55</td>
<td>2.1</td>
<td>0.45</td>
<td>24.7</td>
<td>7.0</td>
<td>35.6</td>
<td>14.3</td>
<td>5.0</td>
</tr>
<tr>
<td>2</td>
<td>78 (5.5)</td>
<td>2.5</td>
<td>75</td>
<td>1.9</td>
<td>0.34</td>
<td>16.1</td>
<td>5.3</td>
<td>39.3</td>
<td>16.8</td>
<td>6.7</td>
</tr>
<tr>
<td>3</td>
<td>97 (6.8)</td>
<td>3.8</td>
<td>134</td>
<td>2.1</td>
<td>0.17</td>
<td>9.1</td>
<td>4.8</td>
<td>43.1</td>
<td>20.4</td>
<td>8.7</td>
</tr>
<tr>
<td>5</td>
<td>91 (6.4)</td>
<td>6.8</td>
<td>289</td>
<td>2.4</td>
<td>0.52</td>
<td>6.8</td>
<td>4.6</td>
<td>47.6</td>
<td>22.5</td>
<td>11.4</td>
</tr>
<tr>
<td>10</td>
<td>170 (11.6)</td>
<td>15.0</td>
<td>447</td>
<td>3.4</td>
<td>0.10</td>
<td>3.7</td>
<td>3.2</td>
<td>50.9</td>
<td>24.2</td>
<td>12.7</td>
</tr>
<tr>
<td>25</td>
<td>29 (2.0)</td>
<td>33</td>
<td>661</td>
<td>3.5</td>
<td>0.03</td>
<td>1.3</td>
<td>1.7</td>
<td>53.2</td>
<td>24.7</td>
<td>14.2</td>
</tr>
<tr>
<td>50</td>
<td>29 (2.0)</td>
<td>71</td>
<td>1312</td>
<td>4.7</td>
<td>0.10</td>
<td>2.6</td>
<td>1.1</td>
<td>54.2</td>
<td>25.9</td>
<td>15.2</td>
</tr>
<tr>
<td>100</td>
<td>30 (2.1)</td>
<td>157</td>
<td>1347</td>
<td>4.7</td>
<td>0.10</td>
<td>2.0</td>
<td>0.0</td>
<td>55.1</td>
<td>27.2</td>
<td>15.4</td>
</tr>
<tr>
<td>≥250</td>
<td>32 (2.2)</td>
<td>738</td>
<td>3621</td>
<td>7.7</td>
<td>0.10</td>
<td>2.4</td>
<td>0.0</td>
<td>59.5</td>
<td>26.6</td>
<td>15.8</td>
</tr>
</tbody>
</table>

TgAb cutoffs frequently used by laboratories to define a “positive” TgAb (LOD, FS, ROC curve, and MCO) are shown in bold.
measurement of clinically defined sera, contrasting with previous studies that were limited to methodologic correlations of TgLC-MS/MS with Tg2GIMA or TgRIA (11, 19). Severe TgAb interference causing Tg2GIMA underestimation was evident from undetectable Tg2GIMA seen for 38% of the TgAb-positive patients with disease (Figure 1B) and lower Tg2GIMA vs either TgLC-MS/MS or TgRIA seen for an additional 60% of these patients. The paradoxically undetectable TgLC-MS/MS seen for 12 of 52 (23%) TgAb-positive patients despite the presence of disease warrants further study. Two of these patients had no Tg detected by any method, and 8 had undetectable Tg2GIMA and TgLC-MS/MS but unequivocally detectable (≥1.0 μg/L) TgRIA. It was more common to see TgAb-positive sera with TgRIA higher (>50%) than TgLC-MS/MS than sera with higher TgLC-MS/MS than TgRIA (56% vs 11%, respectively, P < .01), consistent with either TgAb interference causing TgRIA overestimation or false-positive TgRIA (42, 43) or, alternatively, the presence of a polymorphic tumor Tg that failed to generate the target Tg peptide necessary for LC-MS/MS detection (20). It was striking that both TgLC-MS/MS and TgRIA values were significantly lower (P < .001) when patients with disease had TgAb detected (Figure 1, B vs A). This observation lends support to past studies, suggesting that increased metabolic clearance of Tg-TgAb complexes may be responsible (5, 44–47). If the metabolic clearance of Tg complexed with TgAb is faster than the clearance of free Tg, high Tg assay FS would be especially critical for detecting disease in TgAb-positive patients (22, 26).

Guidelines mandate that each DTC specimen have a TgAb status determined directly by immunoassay and not a Tg recovery test (1, 5, 6, 17, 28, 30). Current TgAb methods vary in sensitivity, specificity, and the numeric values they report, despite claiming to use the same primary calibrator (Medical Research Council [MRC] 65/93) (4, 10, 12, 31–35). This study of 2 well-established TgAb methods (Kronus/RSR vs Roche) found that both the intrinsic sensitivity of the method and the cutoff selected for TgAb positivity influenced the reliability of TgAb detection. In addition, Tg concentrations of >1000 μg/L interfered with the Roche method, as described previously (33). Some sera were classified as TgAb positive by one method but TgAb negative by the other, and some low TgAb concentrations caused profound interference whereas high TgAb in other sera appeared nonreactive (2, 5, 10, 17, 48). These qualitative TgAb differences reflect a complex matrix of factors that include the numeric cutoff for positivity (10, 33, 34) as well as the epitope specificity of the individual patient’s serum Tg antibodies for binding the thyroglobulin reagent used in the TgAb test (10, 33, 35, 48–50). Comparative TgAb method studies fail to assess how Tg antibodies in individual sera affect Tg measurement (10, 31–35). Either a positive TgAb test and/or discordant Tg results between different analytical methods (eg, Tg2GIMA vs TgRIA) are recognized indicators for possible Tg interference (18). The current study used an abnormally low Tg2GIMA/TgRIA ratio (<75%) to assess the influence of interfering TgAb on Tg measurement, as reported previously (2, 5, 10, 18, 36, 37, 51). This Tg2GIMA/TgRIA ratio parameter had the advantages of being independent of the measured TgAb concentration as well as the potential to amplify TgAb influences by distorting Tg2GIMA and TgRIA values in opposite directions. Thus, the unidirection (underestimation) typical of TgAb interference with IMA, coupled with either an unaffected or overestimated TgRIA, would produce a lower Tg2GIMA/TgRIA ratio (2, 4, 9, 10, 36, 37). Although specimen availability and cost constraints prevented the use of Tg2GIMA/TgLC-MS/MS ratios for the entire study, the utility of the Tg2GIMA/TgRIA ratio parameter is supported by the concordance found between ratios calculated for patients with disease (Figure 1) using either TgRIA or TgLC-MS/MS as the denominator. Specifically, when Tg was detectable, making ratio calculations possible, the Tg2GIMA/TgRIA vs Tg2GIMA/TgLC-MS/MS ratios were comparable both in the presence and absence of TgAb. It should be noted that the group A TgAb-negative patients with disease displayed a wider range of Tg2GIMA/TgRIA ratios than that seen for group B TgAb-negative results that comprised a larger number of sera but probably included fewer patients with disease (with no clinical information available). The wider range of Tg2GIMA/TgRIA ratios seen for the group A TgAb-negative patients reflected the discordances seen
between the 3 methods that characterized these patients (Figure 1A) and probably reflected different method specificities for detecting heterogeneous, tumor-derived, Tg molecules (51).

When FS cutoffs were used, the Roche method reported significantly more false negatives with severe interferences than the Kronus method (2.4% vs 0.7%, R vs K, respectively, P < .05), and more false negatives with inappropriately low (<0.30 μL) Tg2GIMA values (4.4% vs 0.7%, R vs K, respectively, P < .001). Underestimation causing falsely low (but detectable) Tg2GIMA values has as much potential to disrupt clinical management as severe interference, because a Tg2GIMA of <0.30 μL predicts the absence of disease (14–16, 18, 24, 25, 27, 29). Now that radioiodine remnant ablation is no longer considered a necessary treatment for low-risk DTC (52–56), an increasing number of disease-free patients will undergo lifelong monitoring of low basal (non-TSH–stimulated) Tg2GIMA concentrations (0.10–1.0 μg/L range) arising from normal remnant tissue (9, 24, 25, 27, 57, 58). TgAb positivity is associated with a higher DTC recurrence risk (59–62) and changes in a patient’s TgAb status have clinical implications: a TgAb decline or disappearance is considered a good prognostic sign, whereas a rise in or de novo appearance of TgAb suggests active disease (2–7, 59, 60, 63–68). When sera are misclassified as TgAb negative and have inappropriately low Tg2GIMA values due to interference, any recurrence that causes a rising TgAb could further suppress Tg2GIMA and be misinterpreted as a good prognostic sign.

This study has a number of clinical implications. It is the first to show that a detectable or higher TgLC-MS/MS is frequently seen for TgAb-positive patients with persistent/recurrent DTC and undetectable/low Tg2GIMA (11, 19, 20), confirming that TgAb interference causing Tg2GIMA underestimation can mask disease (Figure 1B). The study emphasizes how concordance between Tg2GIMA and Tg measured by a different class of method (RIA or LC-MS/MS) may help prove that a Tg2GIMA result is free from TgAb interference (17, 18). The primary focus of the study was the influence of TgAb method cutoffs on the reliability of TgAb detection. No cutoff could be identified for either the Kronus or Roche TgAb method that eliminated both false-negative and false-positive serum TgAb misclassifications. Because TgAb interference causing Tg2GIMA underestimation is considered the most serious clinical problem, it is optimal to use the FS cutoff to define TgAb positivity because this cutoff minimizes false negatives. However, a significant percentage of sera (~20%) would be misclassified as TgAb-positive using FS and could be unnecessarily reflexed to the less sensitive TgRIA or TgLC-MS/MS/MS methods that may fail to detect disease associated with low Tg concentrations (21–27). In addition, use of the FS cutoff can cause the TgAb status of patients with low TgAb concentrations to fluctuate between “negative” and “positive” over time, as a result of assay imprecision (15%–20%) exacerbated by the TgAb biologic variability (~15%) expected during DTC monitoring (69). Changes in a patient’s TgAb status that were solely methodologic could prompt unnecessary switching between Tg methodologies (Tg2GIMA to TgRIA or TgLC-MS/MS, or vice versa) and potentially disrupt serial Tg monitoring, thereby negatively affecting clinical management. The potential for disrupting Tg monitoring is clearly evident from the significant serum Tg2GIMA, TgRIA, and TgLC-MS/MS differences seen in Figure 1A that appear unrelated to TgAb and possibly reflect method specificity differences for tumor Tg detection. Laboratories that reflex Tg testing to different methods should minimize unnecessary Tg method changes by considering the individual patient’s Tg and TgAb measurement history in addition to the TgAb status of the current specimen before selecting an appropriate Tg method and establishing a new baseline for patients when a change in the Tg or TgAb method becomes necessary. Physicians should recognize the technical limitations of the current Tg and TgAb measurements illustrated by this study and interpret serum Tg and TgAb values relative to the clinical status of the patient.

**Acknowledgments**

We thank Dr Stefan Grebe and Brian Netzel for collaborating to provide the TgLC-MS/MS analyses at the Mayo Medical Laboratories.

Address all correspondence and requests for reprints to: Carole A. Spencer, MT, PhD, USC Endocrine Laboratory, 126 West Del Mar Boulevard, Pasadena, CA 91105-2805. E-mail: cs Spencer@usc.edu.

Results from this work were presented in part at the 82nd Annual Meeting of the American Thyroid Association, Quebec City, Canada, 2012.

Disclosure Summary: The authors have nothing to disclose.

**References**


10. Spencer C, Petrovic I, Fatemi S. Current thyroglobulin autoantibody (TgAb) assays often fail to detect interfering TgAb that can result in the reporting of falsely low/undetectable serum Tg IMA values for patients with differentiated thyroid cancer. *J Clin Endocrinol Metab*. 2011;96:1283–1291.


15. Chindris AM, Diehl NN, Crook JE, Fatourechi V, Smallridge RC. Undetectable sensitive serum thyroglobulin (<0.1 ng/mL) in 163 patients with follicular cell-derived thyroid cancer: results of rTSH stimulation and neck ultrasonography and long-term biochemical and clinical follow-up. *J Clin Endocrinol Metab*. 2012;97:2714–2723.


38. Algeciras-Schimnich A, Lasho MA, Ness KM, Cheryk LA, Grebe SK. The Roche Elecsys and Siemens-Centaur thyroglobulin autoantibody assays show comparable clinical performance to the recently unavailable Beckman-Coulter access thyroglobulin autoantibody
45. Feldt-Rasmussen U, Petersen PH, Date J, Madsen CM. Sequential changes in serum thyroglobulin (Tg) and its autoantibodies (TgAb) following subtotal thyroidectomy of patients with preoperatively detectable TgAb. Clin Endocrinol (Oxf). 1980;12:29–38.
49. Spencer CA. Clinical review: Clinical utility of thyroglobulin antibody (TgAb) measurements for patients with differentiated thyroid cancers (DTC). J Clin Endocrinol Metab. 2011;96:3615–3627.
52. Sabra MM, Grewal RK, Ghosein RA, Tuttle RM. Higher administered activities of radioactive iodine are associated with less structural persistent response in older, but not younger, papillary thyroid cancer patients with lateral neck lymph node metastases. Thyroid. 2014;24:1088–1095.
62. Soyłuk O, Boztepe H, Aral F, Alagol F, Öz bey NC. Papillary thyroid carcinoma patients assessed to be at low or intermediary risk after primary treatment are at greater risk of long term recurrence if they are thyroglobulin antibody positive or do not have distinctly low thyroglobulin at initial assessment. Thyroid. 2011;21:1301–1308.