Review

Luca Giovanella*, Ulla Feldt-Rasmussen, Frederik A. Verburg, Stephan K. Grebe, Mario Plebani and Penelope M. Clark

Thyroglobulin measurement by highly sensitive assays: focus on laboratory challenges

Abstract: Differentiated thyroid cancer (DTC) is the most common endocrine cancer and its incidence has increased in recent decades. The initial treatment consists of total thyroidectomy followed by ablation of thyroid remnants by radioactive iodine in most cases. As thyroid cells are the only source of thyroglobulin (Tg), circulating Tg serves as a biochemical marker of persistent or recurrent disease in the follow-up of DTC. Due to the suboptimal clinical detection rate of older Tg assays endogenous or exogenous thyrotropin (TSH) stimulations are recommended for unmasking occult disease. However, the development of new Tg assays with improved analytical sensitivity and precision at low concentrations now allows detection of very low Tg concentrations, reflecting minimal amounts of thyroid tissue, even without the need for TSH stimulation. Even if the use of these assays still has not found its way in current clinical guidelines, such assays are now increasingly used in clinical practice. As serum Tg measurement is a technically challenging assay and criteria to define a ‘highly sensitive’ assay may be different, a good knowledge of the technical difficulties and interpretation criteria is of paramount importance for both clinical thyroidologists, laboratory physicians and scientists involved in the care of DTC patients.

Keywords: differentiated thyroid carcinoma; functional sensitivity; immunoassay; limit of quantification; thyroglobulin.

Introduction

Thyroglobulin (Tg) is a large, 660-kDa, glycoprotein which normally is stored in the follicular colloid of the thyroid gland where it acts as a substrate for thyroid hormone production. As it is only produced by benign or well-differentiated malignant thyroid cells, it is a very good tumor marker for patients with differentiated thyroid carcinoma (DTC) [1], especially after removal of all benign and malignant thyroid tissue through surgery and I-131 ablation [2].

Over the years, the sensitivity and precision of Tg assays have improved by multiple orders of magnitude [3]. Currently, Tg measurement in combination with ultrasound of the neck is considered standard of care in the follow-up of DTC [4, 5]. However, until recently, optimal sensitivity of Tg assays for the detection of smaller disease foci required stimulation of endogenous Tg production by high serum thyroid stimulating hormone (TSH) concentrations [4–7], requiring expensive injections with recombinant human TSH or withdrawal of levothyroxine (LT4) medication which is associated with profound hypothyroidism. Further improvements in assay technology have however recently resulted in new highly sensitive Tg (hsTg) assays [8, 9]. From an increasing number of studies in the literature it appears that these assays are in fact sufficiently sensitive to obviate the need for TSH stimulation in most DTC patients [10, 11]. Therefore, the increasing adoption of these assays in clinical practice has considerable implications, such as a reduction of costs of DTC...
follow-up or avoidance of hypothyroidism. However, hsTg measurement is technically challenging and, in addition, criteria adopted to define a ‘highly sensitive’ assay by different manufacturers, laboratories and clinicians are very likely to diverge considerably.

The aim of the present paper is to review advances and challenges in Tg measurement techniques and their impact on clinical management of DTC patients. We hope this could be useful in providing a ‘sense of direction’ for laboratories and clinicians that are considering the use of hsTg assays.

Tg assays

Analytical performance

Imunoassay has been the main analytical technique used for the measurement of serum Tg, at first by competitive immunoassay and later by immunometric (reagent excess) assays. More recently, mass spectrometric methods have been developed. Each new assay format has been developed to attempt to overcome the major analytical challenges in measuring, as a tumor marker, a heterogeneous analyte of large molecular weight in the presence of endogenous and other interfering antibodies. The key performance characteristics can be listed as:

1. The clinical requirement for a wide measuring range of five orders of magnitude with a linear response;
2. The absence of a high dose hook effect;
3. Comparability of results obtained by different analytical methods;
4. Accurate results with quantitative recovery;
5. Absence of interference by anti-thyroglobulin antibodies (TgAb) and heterophilic antibodies (HAb);
6. Commutability of Tg in different sample matrices.

Whilst substantial efforts have been made to overcome the problems of antibody interference (discussed elsewhere in this article), it is notable that there is no Reference Method System including a Reference Method Procedure available for Tg and use of the BCR® 457 Certified Reference Material (formerly CRM 457) has not completely eliminated the notable differences in results obtained by different methods. Issues of commutability of the BCR® 457 material and the need for harmonization have yet to be addressed (discussed elsewhere in this article). Due to the lack of agreement between different immunoassays it has been recommended that patients should be followed up using the same assay [5, 6, 11, 12]. If an assay change is required then dual reporting of patient results for a period of time to allow comparison of results by re-baselining serum Tg quantitation by the new assay is suggested. In addition, it is suggested that clinical decision limits should be assay specific [11]. Laboratories will be familiar with the necessary experiments to be performed to verify assay performance [13, 14], namely assessment of linearity, measuring range, trueness (measurement bias), comparability through patient comparison studies, limit of detection/limit of quantitation/functional sensitivity. It is worth considering some particular points with regard to serum Tg [15]:

Assay diluents

Commercial assays may be provided with an assay diluent and specify a dilution value (e.g., 1 in 10 v/v). Laboratories may wish to consider whether this covers the range of concentrations that is required from a clinical perspective, such as the monitoring of metastatic disease where serum Tg may reach concentrations >1000 μg/L. The concentration at which the high dose hook effect has been excluded should also be determined. The feasibility of using an in-house human serum pool as diluent (with undetectable TgAb as measured by a suitable assay and Tg concentration <0.1 μg/L) may need to be investigated and the linearity over a wider concentration range determined.

Recovery experiments

Estimation of recovery of added Tg has been proposed as a method of assessing whether there is interference by endogenous antibodies (i.e., TgAb and HAb) though this is not currently advocated by current guidelines (see section on Interferences). Nevertheless, determination that an assay shows quantitative recovery should be performed as part of the method validation. Studies have shown that the measured recovery is dependent on the protocol used – in particular the Tg concentration, source of Tg (degree of iodination) and incubation time [16, 17]. Assessment of recovery using a source of Tg independent of the kit calibrators is suggested, though manufacturers may recommend a protocol if recovery is being determined in the context of assessing assay interference.

Analytical sensitivity

Recent years have seen growing recognition of the clinical need for improved precision of assays at low Tg concentration.
concentrations and this has been paralleled by improvements in assay performance. This inter-relationship has raised a number of issues: what constitutes a significant change (rise or fall) in serum Tg from a clinical perspective and from an analytical perspective? What is the lowest serum Tg concentrations that is clinically significant? In the first case, serial measurements may be necessary in order to assess the significance of any change, rather than acting on a single result [12, 18, 19]. In the second case, there has been a gradual improvement in assay sensitivity with the original radioimmunoassays (RIAs) reporting down to 2–5 μg/L, the first immunometric assays (IMA) down to 1 μg/L and more recently IMAs with limits of 0.1 μg/L. In comparison, reported mass spectrometric methods have sensitivities of 0.5–2 μg/L. However, these broad comparisons are limited because of differences in bias between different assays and different experimental and statistical methods were used to determine the sensitivity of the assays.

In the first instance analytical sensitivity has often been determined by repeat analysis of the zero calibrator and determination of the apparent concentration equal to the zero plus (for immunometric assays) two or three standard deviations of the signal (minus for competitive assays), when it is known as the limit of the blank (LOB). There are significant limitations to this approach. In the majority of cases the measured sensitivity will be below the concentration of the lowest concentration calibrator. The assumption is often made that the fitted standard curve is close to the measured dose-response curve. This may not be the case and can be difficult to assess for automated immunoassays leading to a misleading estimate of the LOB [13]. Although of limited use in understanding the precision of low concentration samples, the LOB can be useful when optimizing assay conditions during assay development. The limit of detection (LOD) is defined as the lowest analyte concentration that can be distinguished from the LOB using replicate analysis of a sample of known low concentration. LOD has similar limitations to those of LOB.

Thus functional sensitivity (FS) was introduced as a measure of analytical sensitivity and was originally described for assessing the sensitivity of TSH assays. The National Academy of Clinical Biochemistry (NACB) published guidelines which included advice on how to determine FS though it should be noted that these have now been archived ‘per NACB and National Guideline Clearinghouse policy’ (http://www.aacc.org/SiteCollectionDocuments/Archived%20and%20Historical/ThyroidArchived2010.pdf#page=1) [20]. The NACB protocol indicated that FS may be determined from between batch precision of measurement of patient pools, in the same test mode (singleton or duplicate) as patient samples over the clinically relevant concentration range over two different lots of reagents and calibrators and over a period of 6 months. The patient pools should be TgAb negative and should cover the clinically relevant concentration range. The protocol specifies three different concentration ranges for the patient pools. Whilst it may be difficult to obtain individual patient samples to determine a precision profile and thus FS, this approach can have advantages in terms of covering the concentration range and is recommended by the British Thyroid Association [12]. From the calculated precision profile a cut-off value corresponding to a CV of 20% (somewhat arbitrarily) is taken as the FS. The difference in FS between Tg assays has created a ‘generational’ nomenclature system with each subsequent generation exhibiting a substantial improvement (i.e., 10-fold).

It should be recognized, however, that there are also limitations to this approach when determining the sensitivity of an assay. For example from an analytical perspective the sample type may influence results – control sera should not be used due to potential matrix effects. There are also differences in the statistical approach to the calculation of the precision profile, principally to do with the identification of outliers and the confidence intervals of the profile [13]. The latter authors emphasize the need to determine confidence intervals of the precision profile – these may vary significantly over the concentration range and indicate the confidence with which the calculated FS can be viewed. In addition, the published literature may not be helpful by not providing details of the procedure used to generate the precision profile (e.g., whether within or between batch precision was used, how many samples were analyzed and whether samples were analyzed in singleton or duplicate). With sparse data the relationship between imprecision and concentration can be poorly estimated. Finally, assays may be adapted over time (e.g., through reagent changes or recalibrations) even though they keep the same brand name. A good example of this is provided by the ‘high sensitivity’ troponin assays where changes in reagent lots and calibration algorithms have led to changes in FS. Lastly the limit of quantitation (LOQ) is similar to the FS but does have an additional requirement for predefined goals for bias and imprecision [14, 21] and is increasingly used as a measure of sensitivity for both immunoassay and mass spectrometric assays. Manufacturers may quote LOB, LOD and LOQ as determined by regulatory authorities and national guidelines (e.g., those of the Clinical and Laboratory Standards Institute EPI7-A2) [22]. The relationship between these estimates of sensitivity is: LOB<LOD≤LOQ.

Given that assay performance can vary with time, operator, reagent lot, calibration, equipment maintenance
and other factors monitoring of sensitivity whether as FS or LOQ should be ongoing and laboratories should determine their own FS/LOQ rather than just quoting manufacturers’ data.

Thus, it becomes clear that in comparing the performance of different Tg assays and in order to provide clinicians with realistic interpretations of Tg results, it is necessary to know exactly how ‘sensitivity’ has been determined such that like can be compared with like. Some examples of these approaches are illustrated in Table 1.

**Development of mass spectrometric assays**

Protein identification and quantification by mass spectrometry (MS) has a long tradition in laboratory medicine. More recently, efforts have been directed towards the development of MS-based Tg assays. The MS work-flow for protein measurements involves a digestion of the sample with trypsin. Trypsin cleaves protein in a predictable fashion into peptides, which are measured by MS and identified by protein database matching. In the case of Tg, the tryptic digest will cleave all proteins in the sample. One can then specifically look for (one or several) tryptic peptides that are proteotypic for Tg based on predicted cleavage [32]. Quantification is obtained by including a standard curve based on synthetic version of the measured peptides and adding a non-radioactive isotopic version of the peptides in defined concentrations as internal standards to the samples before analysis [33]. In practice, however, many problems may affect Tg measurement by MS. One faces formidable signal to noise problems in identifying the quite low concentration Tg peptides in a trillion-fold higher background abundance of all the other peptides from all the other proteins. To overcome this problem, some form of sample enrichment is required. Separation of proteins based on size can be used before trypsin digestion. Even more effective is immune affinity purification of the desired Tg target peptide(s) from all the other peptides. A good antibody against the target peptide

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Values for Tg assays for LOD, LOB, LOQ or FS as reported in the literature or quoted by manufacturers.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assay</strong></td>
<td><strong>Manufacturer</strong></td>
</tr>
</tbody>
</table>
| Dynotest Tg Plus | BRAHMS Thermofisher, D  | IRMA               | LOB 0.16  
|               |                             |                   |  FS 0.4 [23]  
|               |                             |                   | M quoted values but corrected for difference in manufacturers standardization to CRM 457 (i.e., multiplied by 2) |
| Tg-Access    | Beckmann, US               | ICMA              | LOB 0.1 (M)*  
|               |                             |                   |  FS 0.1 [25] Precision profile determined from 6 pools, TgAb status not stated.  
|               |                             |                   |  FS 0.1 [26] Precision profile from 10 serum pools at a CV=20%.  
|               |                             |                   |  FS 0.11 [27] Method of determination not stated.  
|               |                             |                   |  FS 0.05 [28] According to NACB protocol.  |
| e-Iason TgCa | Iason GmbH (A)             | ELISA             | FS 0.02 (M)  
| ELSA-hTg     | CIS Bio International (F)  | IMA               | FS 0.8 [29] Method not stated.  
| TgII Roche Elecsys | Roche Diagnostics AG (CH) | ICMA          | LOB 0.02 (M)  
|               |                             |                   |  LOD 0.04 (M)  
|               |                             |                   |  LOQ 0.1 (M) Total allowable error of ≤30%  
|               |                             |                   |  All determined according to CLSI EP17-A requirements |
| Us-Tg        | BRAHMS Thermofisher (D)   | TRACE             | LOB=0.02 M  
|               |                             |                   |  LOD=0.04 M  
|               |                             |                   |  LOQ=0.1 M, total allowable error of ≤40%  
|               |                             |                   |  All determined according to CLSI EP17-A requirements |
| Immulite 2000| Siemens Healthcare (D)    | ICMA              | LOB=0.2 (M)  
|               |                             |                   |  FS=0.9 (M) Method of determination not stated  
|               |                             |                   |  FS=0.36 [30] according to NACB protocol  
|               |                             |                   |  FS=0.6 [31] based on in-house (non-manufacturer) calibration allowing for calculation of ‘negative’ concentrations |

ELISA enzyme-linked immunosorbent assay; FS, functional sensitivity; ICMA immunochromiluminometric assay; IECMA immunoelectrochemiluminometric assay; IRMA, immunoradiometric assay; LOB, limit of the blank; LOD, limit of detection; LOQ, limit of quantification; M, manufacturer’s quoted details; NA, not available; NQ, not quoted; TRACE, time-resolved amplified cryptate emission. ‘*’Insert package reports the term ‘analytical sensitivity’. 

Unauthenticated
Download Date | 2/18/18 4:46 PM
can achieve dramatic enrichment. Using one or both of these enrichment techniques, Tg measurement by MS becomes possible [34]. By limiting oneself to one or two peptide targets that ionize well, optimizing chromatography, and using selective reaction monitoring on high sensitivity mass filtering tandem MSs, rather than the higher resolving, but (typically) less sensitive, high resolution instruments used in proteomics research, analytical sensitivities can be achieved that are slightly higher than modern immunometric Tg immunoassays [35] (Figure 1).

**Standardization and harmonization of Tg assays**

Tg is a large (660 kDa), highly glycosylated dimeric molecule that is heterogeneous in serum due to differential splicing of Tg mRNA as well as carbohydrate and iodide heterogeneity. In addition, biosynthesis of the mature Tg molecule may become deregulated in thyroid tumor cells resulting in differences in the structure of circulating Tg protein. These changes can lead to exposure or masking of epitopes and hence differences in Tg immunoreactivity [36]. Different Tg assays employ a number of antibodies against Tg with varying specificity for different epitopes. Potentially this can result in variability in the measurement of different Tg isoforms in the patient’s specimen and ultimately to differences in Tg concentration reported by the assays [20, 37]. Early international collaborative studies showed that serum Tg concentrations varied by as much as 40%–60% between methods [38]. The introduction and use of the BCR® 457 has significantly reduced inter-method variability to about 30%, but has not eliminated it completely [39, 40]. Consequently, any change in Tg assay has the potential to disrupt serial monitoring and prompt inappropriate clinical vulnerability.

![Figure 1](https://example.com/figure1.png)  
**Figure 1** Schematic depiction of the workflow for current (high-pressure) liquid chromatography tandem mass spectrometry (LC-MS/MS; HPLC-MS/MS) measurement of Tg.
decisions. For longitudinal consistency of clinical care, consecutive measurements of Tg concentrations should be performed in the same laboratory using the same assay each time. If an assay change is unavoidable, a new baseline of the individual patient’s serum Tg concentrations should be established through parallel Tg measurements using both the old and the new assay [11]. Furthermore, internal and external quality control programs, including samples at low and very low Tg concentrations, are of pivotal importance for checking the precision, reproducibility (internal quality control) and accuracy (e.g., lack of bias of analytical results) of assays to ensure optimal patient care. Laboratories providing Tg measurement are required to participate in a certified national or international program of quality assurance [41].

Finally, it is worth bringing to attention that BCR® 457 was produced around 20 years ago, and although high temperature accelerated stability was investigated as part of the project [39] it cannot be excluded that long-term storage might have changed the epitope composition by degradation or unfolding. By using highly sensitive Tg methods a very precise calibrations could be more important than using the previous less sensitive ones, without a requirement for the low concentrations. It may thus be important to consider investigating consequences of the current assay calibration and comparability by the use of newer Tg methods, and perhaps it is time to consider producing a new calibrator.

Interferences

Depending on the population studied, the assay used and the definition of a positive test result, up to 25%–30% of patients with DTC have a positive test for TgAb at the time of diagnosis [16, 42]. Some authors proposed to measure serum Tg by RIA and IMA on the same sample as discordant result suggest TgAb interference [16, 43]. However, this method is technically demanding and time-consuming and, consequently, it is rarely employed in clinical practice. There are two approaches to detecting TgAb: recovery of added exogenous Tg or measurement of TgAb by immunnoassay [17, 36, 44, 45]. Whereas the first approach has a relatively wide reference range and is very dependent on experimental conditions (e.g., concentration of added Tg, two measurements of Tg that increases overall imprecision), the second approach will, depending on the assay, detect most TgAb-positive patients. In this context, a positive result is often defined as a result above the upper limit of the reference range. However, for thyroidectomized patients, lacking antigen for TgAb generation, due to limitations, the cut-off value for detection of analytical interference is probably much lower [45, 46].

In addition, not all patients with positive TgAb will show interference by measurement of recovery or discordance between Tg results using different assays [45]. Any TgAb assay should be standardized against the First International Reference Preparation 65/93, but, due to the heterogeneous nature of TgAb, no single assay can predict with absolute certainty whether TgAb in a given sample will interfere with Tg measurement [47]. Theoretically, a simple relationship exists between Tg and TgAb and the higher the TgAb concentration, the higher the Tg concentration that can be concealed by TgAb. Sometimes, however, seemingly low concentrations of TgAb may be associated with strong interference and conversely patients with high concentrations of TgAb show no evidence of interference with the Tg measurement [45]. Recently, new recovery tests using a low Tg concentration (i.e., 5–10 μg/L) were introduced [48]. While the performance of these ‘low concentration recovery tests’ (e.g., mini-recovery tests) has yet to be investigated extensively in patients with DTC, they may prove to serve as an addition to TgAb measurement [45]. All in all, any sample with a positive TgAb result and/or abnormal recovery test/discordant Tg result by different analytical methods (e.g., immunometric assay vs. radioimmunoassay) should be considered as unreliable for measuring serum Tg concentrations in patients with DTC. In addition, a small percentage of patients (1%–3% in the literature) show interference with Tg measurement due to HAb [49, 50]. These can bind animal antigens and form a bridge between capture and detection antibody leading to a falsely elevated (or, rarely, falsely decreased) Tg measurement in immunometric assays [51]. HAb interference may be detected either by recovery measurement or measurement of Tg in serially diluted sera (providing Tg concentrations are sufficiently high). In addition, as HAb interference is generally assay specific, the use of an alternate assay may both identify a false-positive sample and provide the correct test value. However, the method of choice, which is more specifically geared towards heterophile antibody interference, is to pre-treat a serum aliquot with proprietary blocking agents then compare the Tg result with an aliquot that was not pre-treated [45].

The majority of patients with DTC show an elevated pre-operative serum Tg, but the predictive role of this measurement is debatable as immunoassays cannot detect a difference between Tg from normal thyroid tissue and that secreted by thyroid cancer tissue. In clinical practice, however, patients with clear disease foci or significant thyroid remnants but undetectable serum Tg
levels despite a negative TgAb test results are occasionally encountered [52, 53]. This may occur when: the spatial conformation of Tg is changed leading to decreased immunoreactivity; the ability to secrete Tg is lost; the TgAb result is a false-negative or inappropriate cut-off values were used. To overcome this problem, an in vivo recovery test has been proposed: measurement of Tg and TgAb in any patient referred for thyroidectomy due to a suspicion of DTC [11, 45]. This strategy could provide ‘baseline’ Tg and TgAb concentrations, which, as stated by NACB guidelines [20], could theoretically allow assessment of the reliability of Tg and TgAb measurements after thyroidectomy, although the clinical value of this in vivo test has never been verified.

Several analytes have been tested as alternatives to Tg measurement in TgAb positive patients, such as Tg mRNA, thyrotropin receptor mRNA or thyroid peroxidase mRNA. However, in most studies it proved impossible to achieve clinically useful sensitivity and specificity levels [54].

Recently MS emerged as a suitable method to overcome interferences in Tg measurement; in fact trypsin digests all proteins in a sample, including Tg and any TgAb or HAb, by cleaving them at predictable sites. One can then specifically look for (one or several) tryptic peptides that are proteotypic for Tg (based on predicted cleavage), without any interference by TgAb. Kato et al. published a proof of principle for the tryptic digestion approach to overcome antibody interferences in Tg assays almost 15 years ago [32]. However, this group used an in-house ELISA, rather than MS, to detect Tg proteotypic peptides after digestion. Unfortunately, this system only achieved a LOD of approximately 10 μg/L, insufficient for modern clinical use. In 2008 Hoofnagle et al. revisited this approach, using tryptic digestion, followed by immunocapture of proteotypic peptides and detection of the captured peptides by MS [33]. Several other groups have since replicated this work [34, 35, 55]. The LOQ of these Tg MS assays is 0.5–1 μg/L, and all have shown excellent comparability (R²>0.95, slope 0.8–1.2) to immunoassays in samples that are TgAb negative. In TgAb positive samples, which have detectable Tg by hsTg immunoassay (FS/LOQ ~0.1 μg/L), all published MS Tg assays still correlated with a similar R² with the immunoassay, but demonstrated a slope of ~1.5, consistent with systematic under-recovery of Tg (of about 50%–60%) in the immunoassays. Finally, in samples that are TgAb positive, but have an undetectable Tg by sensitive immunoassay the Tg MS assays can detect Tg in 20%–25% of cases (Figure 2) [55].

The principle drawbacks of MS-based Tg assays are currently the complex and manual workflows, the long turn-around-time, largely owing to the several hours of tryptic digest, and the currently slightly higher LOD and LOQ when compared to hsTg immunoassays. Due to these limitations, Tg MS assays should not replace Tg immunoassays and MS testing should be reserved for TgAb positive patients. If a patient is known to be TgAb positive, Tg MS testing may be performed instead of immunoassay testing. For patients with uncertain TgAb status, TgAb should be measured before Tg testing. For patients who prove TgAb negative, Tg testing should then be reflexed to a Tg immunoassay, while samples from TgAb positive patients should be routed to Tg measurement by MS [35, 55].

Serum levels of TgAb are not correlated with the tumor load of the patient but rather indicate the activity of the immune system. Furthermore, the mere presence of TgAb in serum has thus far not conclusively been shown to correlate with a worse or better overall prognosis. However, TgAb can be used as ‘surrogate tumor marker’ as disease-free patients with high TgAb concentrations typically display a slow progressive TgAb decline over years (but may not achieve full TgAb-negativity, possibly because of the long-lived memory of plasma cells). Although the most common change in TgAb status is positive to negative in response to successful treatment the persistence of TgAb, a rising TgAb trend, the failure of TgAb to fall or a de-novo TgAb appearance is highly suspicious for active disease [11, 12, 45].

Clinical impact of new highly sensitive Tg assays

There is a wide consensus among various guidelines that Tg is an important, sensitive method for monitoring DTC
patients for the presence of residual or recurrent disease after total thyroidectomy and 131I remnant ablation. A hindrance for the preeminent position of Tg measurement thus far consisted of the need, recommended in all guidelines, for TSH stimulation in order to achieve optimal sensitivity. A ‘negative’ TSH-stimulated Tg measurement in combination with a negative clinical examination, negative neck US and, when indicated, negative additional imaging procedures, predicts a very low risk of recurrence in both low- and high-risk patients [56, 57]. Once these findings have been established, routine DTC follow-up should therefore consist of periodic clinical examination combined with neck ultrasound and Tg measurement on LT4 medication [2, 3]. An important question that needs to be answered is whether the latter is still the case for hsTg assays. Previous studies have already suggested that the additional yield of DTC recurrence through rhTSH stimulation was only 0.8% using Tg assays with a FS of about 1 μg/L [58, 59]. In a study by Giovanella et al. [60] the clinical sensitivity for the detection of DTC recurrence could be increased to 92% by employing a Tg assay with a FS of 0.4 μg/L while specificity did not change. In a further study by the same group [61], unstimulated serum Tg measurements were associated with a 96% negative predictive value (NPV) in 117 low-risk DTC, which increased to 99% when combined with neck US. In this study, rhTSH stimulated Tg measurement only detected one additional recurrence in 104 patients with an undetectable unstimulated Tg. Therefore, even using less sensitive assay technology, the additional value of an rhTSH stimulated Tg measurement is likely more modest than guidelines led us to believe, especially in low-risk patients. More recently, a number of studies were performed to investigate the diagnostic performance of basal serum hsTg measurement in the follow-up of DTC patients [7–10, 18, 24, 25, 29, 62, 63]. Smallridge et al. [64] evaluated 194 patients who also underwent rhTSH stimulation by using the Beckmann Coulter Access Tg assay (FS 0.1 μg/L). Of the 80 patients with a basal Tg below 0.1 μg/L, two (2.5%) had an rhTSH-stimulated Tg concentration above 2 μg/L. No patient had concurrent imaging suggestive of local recurrence or distant metastasis. However, if unstimulated hsTg was 0.1–0.5 or 0.6–2.0 μg/L, rhTSH-stimulated Tg was above 2 μg/L in 24.2% and 82.4% of cases, respectively. Recently, Giovanella et al. [10] reviewed and meta-analyzed data from nine studies including 3178 DTC patients and confirmed the very high NPV (98%–100%) of an undetectable basal hsTg (e.g., <0.1 μg/L). Nonetheless, these assays also have an adequate sensitivity for detection of recurrent disease (88%–98%). Therefore, hsTg assays seem to be able to obviate the need for measurement of TSH-stimulated Tg concentrations. Unfortunately, the improved sensitivity is associated with an unsatisfactory clinical specificity and positive predictive value (PPV). Patients with a low detectable basal hsTg (i.e., between 0.1 and 1 μg/L) can likely be considered as ‘disease-free’ after a negative TSH-stimulated Tg measurement. This means that the latter group of patients (who fortunately are a minority of the total DTC population) should be submitted to a TSH-stimulated Tg measurement [10–12] (Figure 3). However, although the low frequency of DTC recurrences impacts the ability to study PPVs, the PPV of an rhTSH-stimulated Tg above 1–2 μg/L appears comparable to a basal hsTg above 0.10–0.20 μg/L [19]. In addition, the trend in basal hsTg, measured when TSH is lowered/suppressed at constant level, should reflect changes in thyroid tissue mass and thus provide a sensitive parameter for disease [19, 24, 61, 65]. This is also supported by a growing number of studies showing the prognostic utility of monitoring the basal hsTg trend and Tg doubling time [66–69]. A further limitation to the available evidence is that most patients who were enrolled in the studies were affected by low-risk DTC. Data on patients with intermediate- and high-risk tumors are less robust, indicating that the above described approach could be restricted, at least for the moment, to low-risk DTC patients [11]; further studies should be performed in a broader spectrum of high-risk patients.

**FNAC-Tg from lymph nodes: clinical role, validation and quality assessment**

Despite the excellent prognosis of most DTC patients, a minority of them will develop a recurrence (or multiple recurrences in the long term). Whereas follicular thyroid carcinoma metastatizes to distant organs as lung or bone, the papillary thyroid carcinoma primarily disseminates to regional lymph nodes via the lymphatic vessels. Cervical recurrences may be detected clinically but are most frequently discovered on neck ultrasound (US); additional imaging procedures (i.e., radiiodine scan, PET/CT scan) are performed when distant metastases are expected. In case of suspicious US findings, fine-needle aspiration cytology (FNAC) is generally required to confirm or exclude recurrence by cytology [2, 4, 5]. Unfortunately, FNAC samples may be inadequate or even false-negative, especially in lymph nodes with small metastases, partial involvement or cystic changes. To improve the diagnostic yield of FNAC, several authors proposed measurement of Tg in aspirates (FNAC-Tg) [70–73]. In fact, FNAC-Tg
measurement using a cut-off level of 1 μg/L provides higher sensitivity and specificity than cytology in athyrotic patients with papillary thyroid carcinoma and concurrent suspicious US and increased serum Tg [23, 26].

The measurement of Tg in non-serum/plasma samples is problematic from an analytical perspective, not least because of the lack of experimental data to support the validity of results and the absence of formal support for this application by commercial manufacturers of Tg assays. The onus is thus on laboratories to determine whether analysis of non-serum/plasma samples is listed as an ‘Intended use’ in the Information for Users provided by manufacturers. In its absence, full analytical validation to regulatory standards by laboratories is required. Several potential confounding factors should be considered [27].

There may be differences in antigen structure and immunoreactivity. Furthermore, there may be matrix effects associated with the washout fluid not found in serum/plasma. Also the high dose hook effect requires exclusion. The TgAb can potentially be present in the washout fluid and affect the immunoassay. As a consequence laboratory investigation might include the following actions.

### Pre-analytical

Samples are collected by fine needle aspiration for cytology and after dispensation of the sample onto the appropriate slides, the needle is washed out. Published protocols vary, often giving minimal detail, such as volume of washout.
and fluid as ‘normal saline’ or phosphate-buffered saline. Where a serous fluid has been aspirated, dilution with saline may be omitted. Samples are collected from cervical lymph nodes or space occupying lesions in the thyroid bed [72].

1. Is the sample representative of the lymph node/ thyroid bed? Is the first, second, etc needle washout? How many lymph nodes were aspirated? (Action – Document).

2. Is the sample homogenous or does it contain cellular/particulate matter? (Action – Mix, centrifuge sample).

3. Does the sample contain hemoglobin? (Action – Discard, Tg contamination from peripheral blood during the collection procedure may invalidate the result).


**Analytical**

1. Are there any matrix effects from the fluid used to wash-out the collection needle? This may be assay specific. (Action – Determine the optimal fluid).

2. Concentrations of Tg reported in FNAC-washout fluid range from 0.1–>1000 μg/L. Is there a high dose hook effect? (Action – Establish linearity, assay all samples before and after dilution in an appropriate diluent).

3. Assay recovery from washout fluid of added Tg. Over-recovery of 100%–140% has been reported [6].

4. Endogenous Tg antibodies have been reported to affect and also to have no effect on the measurement of Tg in FNAC samples [28, 65, 71]. However, some of these studies are based on TgAb measurements in the patient’s serum or by inappropriate analytical methods for TgAb, such as agglutination methods. The analytically relevant question is whether there are TgAb in the sample being analyzed and whether these interfere in the immunoassay [30]. Measurement of TgAb themselves in washout fluid may equally be problematic, but attempts should be made to determine/validate the assays before use [26].

5. Quality assessment. Serum quality control materials may not adequately reflect the performance of Tg assays in washout fluids unless a serum-based diluent has been used as the washout fluid. Laboratories may wish to exchange washout samples with other laboratories to compare results, having established the stability of the samples.

**Post-analytical factors**

1. Consider whether results are reported per volume of washout fluid or per aspirate.


Close collaboration between laboratory and clinician is required in establishing protocols for the sampling and collection of samples and their accurate analysis.

**Conclusions**

The post-surgical follow-up of DTC is aimed at early identification of the small proportion of patients who have residual disease or will develop recurrence. In the absence of TgAb and heterophile antibodies, Tg measurements are nowadays the reference standard for clinical management of patients previously treated for DTC. Current clinical guidelines are still based on study results obtained using Tg assays with a FS of about 1 μg/L. Using such assays a TSH-stimulated Tg measurement was mandatory to achieve sufficient clinical sensitivity for the detection of persistent and/or recurrent disease. Even though the introduction of hsTg assays is not without challenges, especially where it comes to the sensitivity of current assays for the detection of interfering autoantibodies against Tg, there is an increasing body of evidence that an undetectable highly sensitive Tg during LT4 treatment is sufficient with a high NPV to forgo TSH stimulation in low-risk DTC patients. Patients with slightly increased basal hsTg (e.g., 0.1–1 μg/L) concentration should undergo a TSH-stimulated Tg measurement.

**Author contributions:** All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

**Financial support:** LG, FAV and PC served on an advisory board of Roche AG (September 2012–September 2013); LG received research support and accepted speaker’s fees from Roche AG; UFR serves on an advisory board of Genzyme; UFR has received research support from Genzyme; LG, FAV and UFR have accepted speaker’s fees from Genzyme.

**Employment or leadership:** None declared.

**Honorarium:** None declared.

**Competing interests:** The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.
References

29. Rosario PW, Purisch S. Does a highly sensitive thyroglobulin (Tg) assay change the clinical management of low-risk patients with thyroid cancer with Tg on T4 <1 ng/ml determined by traditional assays? Clin Endocrinol 2008;68:338–42.
44. Persson AC, Links TP, Wilde J, Sluijer WJ, Wolffenbuttel BH, van den Ouweland JM. Thyroglobulin (Tg) recovery testing with quantitative Tg antibody measurement for determining interference in serum Tg assays in differentiated thyroid carcinoma. Clin Chem 2006;52:1196–9.
46. 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;96:1283–91.


70. Boi F, Baghino G, Atzeni F, Lai ML, Faas G, Mariotti S. The diagnostic value for differentiated thyroid carcinoma metastases of thyroglobulin (Tg) measurement in washout fluid from fine-needle aspiration biopsy of neck lymph nodes is maintained in the presence of circulating anti-Tg antibodies. J Clin Endocrinol Metab 2006;91:1364–9.


**Bionotes**

Luca Giovanella MD, PhD is the Chief of the Department of Nuclear Medicine-PET/CT and Thyroid Centre at Oncology Institute of Southern Switzerland (IOSI) and the Scientific Advisor of the Department of Clinical Chemistry and Laboratory Medicine at Ente Ospedaliero Cantonale in Bellinzona and Lugano (Switzerland). He serves as Privat-Dozent (PD) at University of Zürich (Switzerland). His clinical and scientific activities are mainly focused on the diagnosis and treatment of thyroid diseases.

Ulla Feldt-Rasmussen MD, is Professor at University of Copenhagen and Chief Physician of Medical Endocrinology at National University Hospital. Her research interests involve the thyroid gland, cancer and autoimmunity. She has published more than 320 papers in peer-reviewed journals, organized international meetings and postgraduate courses, and led several European thyroid-related projects. Dr. Feldt-Rasmussen is reviewer and editorial board member of many international journals, belongs to numerous international professional organizations and has served as Secretary-Treasurer of European Thyroid Association, and as President of ETA Cancer Research Network. She has received several prestigious prizes and was Haines lecturer at Mayo Clinic in 2011, and Associate Professor at University of Messina, Italy.

Frederik A. Verburg
Department of Nuclear Medicine, RWTH University Hospital Aachen, Aachen, Germany

Frederik A. Verburg MD, PhD studied medicine at the universities of Leuven (Belgium) and Utrecht (The Netherlands) from 1997 to 2004. In 2008 he obtained his PhD in medicine at the University of Utrecht. He trained as a nuclear medicine physician at the University Medical Center Utrecht, The Netherlands, and the University Hospital Würzburg, Germany, from 2005 to 2010. From 2011 to the
present he is working as a consultant nuclear medicine physician and Assistant Professor of Nuclear Medicine at the RWTH University Hospital Aachen in Aachen, Germany.

Stephan K. Grebe  
Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, NY, USA

Stefan K. G. Grebe MD, PhD is Professor of Laboratory Medicine and Pathology (DLMP), Mayo Clinic College of Medicine. He trained in Internal Medicine and Endocrinology and jointed the DLMP of the Mayo Clinic as a Co-Director of the Endocrine Laboratory in 2002. He served as Chair of the Division of Clinical Biochemistry and Immunology from 2006 to 2014. His main clinical and research interests are currently focused on thyroid cancer tumor markers and clinical mass spectrometry.

Mario Plebani  
Department of Laboratory Medicine, University Hospital Padua, Padua, Italy

Mario Plebani obtained his medical degree summa cum laude from the Medical School of the University of Padova in 1975. He completed residency training and specialization in Laboratory Medicine (1978), and subsequently in Gastroenterology (1983), at the same University. He is full Professor of Clinical Biochemistry and Clinical Molecular Biology at the University of Padova, School of Medicine, Chief of the Department of Laboratory Medicine at the University-Hospital of Padova, and Chief of the Center of Biomedical Research (a specialized Center for quality in laboratory medicine for the Veneto Region). Currently, he is also Director of the Postgraduate School in Clinical Biochemistry at the Medical School of the Padova University and President of the Course for Medical Technologists at the same Medical School. He served as President of the International Society of Enzymology (ISE) for 4 years (2004–2008) and as President of the Italian Society of Clinical Biochemistry and Molecular Clinical Biology for 5 years (in 2003 and from 2007 to 2009). He is the Chairman of the IFCC Working group on ‘Laboratory errors and patient safety’ (WG LEPS) and in 2008 received the AACC Award for Outstanding Clinical Laboratory Contributions to Improving Patient Safety. Dr. Plebani is Editor in Chief of CCLM (Clinical Chemistry and Laboratory Medicine), and Associate editor of CRC Clinical Laboratory Sciences, and International Journal of Biological Markers. His main areas of research are quality in laboratory medicine, biomarkers in cancer and cardiovascular diseases, and in vitro allergy diagnostics.

Penelope M. Clark  
Clinical Laboratory Services, Queen Elizabeth Hospital Birmingham, Birmingham, UK

Penelope M. Clark PhD, is an Honorary Consultant Clinical Scientist and Senior Lecturer at The Queen Elizabeth Hospital Birmingham and the University of Birmingham. She headed the Regional Endocrine Laboratory with responsibility for national services for thyroid disease and pituitary disease including thyroid cancer. She is a Fellow of The Royal College of Pathologists and The Association for Clinical Biochemistry and Laboratory Medicine.