In Thyroidectomized Patients with Thyroid Cancer, a Serum Thyrotropin of 30 μ U/mL After Thyroxine Withdrawal Is Not Always Adequate for Detecting an Elevated Stimulated Serum Thyroglobulin

Laticia A. Valle,^{1-3,*} Revital L. Gorodeski Baskin,^{1-3,*} Kyle Porter,^{3,4} Jennifer A. Sipos,¹⁻³
Raheela Khawaja,¹⁻³ Matthew D. Ringel,¹⁻³ and Richard T. Kloos^{1-3,5,†}

Background: The thyrotropin (TSH) level or duration of thyroid hormone withdrawal (THW) required to detect stimulated thyroglobulin (Tg) in differentiated thyroid cancer (DTC) monitoring is unknown. The objective of this study was to evaluate the TSH cutoff of $>30 \mu\text{U/mL}$ as a means to detect stimulated $\text{Tg} \geq 2 \text{ ng/mL}$ after THW ($\text{THW-Tg} \geq 2$), and sensitivity of the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) questionnaire for detecting hypothyroid symptoms.

Methods: This was a prospective longitudinal cohort study done at a tertiary academic medical center. Forty-seven patients with DTC undergoing their first Tg stimulation or after previously abnormal Tg stimulation had weekly measurements of TSH and Tg during the 4 weeks THW, and repeated questionnaire assessments.

Results: TSH did not reach a plateau in any patient, and in those whose Tg did not remain undetectable, Tg continued to rise. Seventy-five percent of patients had an undetectable $\text{Tg} < 0.2 \text{ ng/mL}$ at baseline (95% were $< 0.5 \text{ ng/mL}$) with 16% remaining undetectable throughout THW. The majority of patients (72.7% and 97.8%) achieved $\text{TSH} > 30 \mu\text{U/mL}$ by 3 and 4 weeks THW, respectively. Of the 15 patients with maximum stimulated $\text{THW-Tg} \geq 2$, 38% were detected before the minimal $\text{TSH} > 30 \mu\text{U/mL}$ cutoff. At 2 weeks THW, 3 had a $\text{TSH} > 30 \mu\text{U/mL}$, and none of them had $\text{Tg} \geq 2 \text{ ng/mL}$. At 3 weeks THW, 11 had a $\text{TSH} > 30 \mu\text{U/mL}$, and 64% of them had $\text{Tg} \geq 2 \text{ ng/mL}$. Only 60% were detected at 3-week THW regardless of their TSH level. Eighty-six percent were detected by $\text{TSH } 60\text{--}80 \mu\text{U/mL}$. Conversely, all patients whose serum Tg was $< 0.2 \text{ ng/mL}$ when their serum TSH was $> 20 \mu\text{U/mL}$ did not achieve a $\text{THW-Tg} \geq 2$.

Conclusion: The minimal TSH cutoff of $> 30 \mu\text{U/mL}$ was inadequate to detect many patients with final stimulated $\text{THW-Tg} \geq 2$ during complete THW. $\text{TSH} > 80\text{--}100 \mu\text{U/mL}$ was a better cutoff, achieved in only 53% after 4-week THW. Conversely, we propose a preliminary THW-stopping rule for ending THW early in selected patients. In patients with a $\text{Tg} < 0.2 \text{ ng/mL}$ when $\text{TSH} > 20 \mu\text{U/mL}$, all had a final stimulated $\text{Tg} \leq 2 \text{ ng/mL}$, potentially saving qualifying patients 40% of THW duration compared to 4-week THW. FACIT-F correlated with TSH, but was not sensitive to detect mild hypothyroidism.

Introduction

DIFFERENTIATED THYROID CANCER (DTC) surveillance relies upon serum thyroglobulin (Tg) and neck ultrasonography (US) due to their high sensitivities to detect persistent or recurrent thyroid cancer (1–3). To achieve maximal sensitivity, Tg levels are measured after thyrotropin (TSH) stimulation, particularly when patients have low or unde-

tectable Tg levels with low TSH levels. The most accurate stimulated Tg cutoff to detect or predict persistent disease is $> 1\text{--}2.5 \text{ ng/mL}$ (3–9).

The TSH level necessary to achieve adequate Tg stimulation after thyroid hormone withdrawal (THW) has not been determined, and the commonly used cutoff is derived from the level thought necessary for radioactive iodine (RAI) imaging. Edmonds *et al.* reported a study of seven patients

¹Department of Internal Medicine; ²Division of Endocrinology, Diabetes, and Metabolism; ³The Ohio State University Medical Center; ⁴Center for Biostatistics; ⁵Division of Nuclear Medicine; The Ohio State University, Columbus, Ohio.

*These two authors are co-first authors.

†Present affiliation: Veracyte, Inc., South San Francisco, California.

where a TSH level $>30 \mu\text{U/mL}$ was necessary for adequate uptake on an RAI whole-body scan (10). Others have postulated that a TSH level $>25\text{--}30 \mu\text{U/mL}$ or after 2–3 weeks of THW is adequate (7,11–14). The average number of THW days to achieve a TSH $>30 \mu\text{U/mL}$ in a pediatric DTC population is about 14 days (11,15–17), and about 19 days in adults (15,18,19). Still, it is unclear if and when TSH or Tg levels plateau, and if Tg continues to rise once the minimum TSH cutoff has been achieved. One THW study over 21–23 days demonstrated an increase in the mean Tg in 31 patients, from 2.36 to 3.94 to 5.72 ng/mL as the TSH rose from a mean of 36.9 to 49.9 to 58.5 $\mu\text{U/mL}$, respectively (20).

Various questionnaires have assessed quality of life (QOL) during THW (21–24). QOL scores are lower during THW at the time of ablation compared to euthyroid recombinant human TSH (rhTSH)-prepared patients or controls (21,22,25–32). Finding the minimal duration and/or degree of hypothyroidism for adequate Tg stimulation testing may minimize the impact on QOL for patients undergoing THW. However, questionnaires assessing hypothyroidism have not been applied on a weekly basis during THW. Thus, the diminution of QOL to potentially spare by shortening THW has not been quantified.

The primary aims of this study were to examine the relationship between Tg and TSH during 4 weeks of THW, to evaluate the adequacy of TSH cutoffs and the duration of THW, and to determine if there is a THW-stopping rule where one could effectively rule out a subsequent significant rise in Tg without completing the full THW. The secondary aims were to determine if Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-F) is a sensitive tool to detect various degrees of hypothyroidism, if there is recall bias when applying the questionnaire weekly versus every other week, and determine if other factors in patients with DTC significantly affect their FACIT-F scores or the rise in TSH during THW.

Materials and Methods

Study patients

Forty consecutive patients with DTC were enrolled between September 2009 and September 2011 from the Ohio

State University (OSU). Inclusion criteria were age >18 year, pre-enrollment Tg level $<0.5 \text{ ng/mL}$ on thyroid hormone (TH), and a previously abnormal TSH stimulated Tg ($\geq 0.5 \text{ ng/mL}$) or one year post-thyroidectomy and RAI remnant ablation undergoing their first stimulated Tg test. Patients were presumed by medical history and examination to have no pituitary/hypothalamic disease to invalidate measurements of TSH and stimulated Tg. Subjects were excluded if they were not capable of informed consent, unable to comply with the study protocol, pregnant or planning pregnancy, unable to tolerate the study protocol due to comorbid medical conditions, minors or dependents, or if they had positive Tg antibodies. Subjects provided informed consent, and signed a Health Insurance Portability and Accountability Act (HIPAA) authorization policy. The OSU Internal Review Board (IRB) approved the study.

Thirty-eight of 40 enrolled subjects had sufficient biochemical data for analysis, and 38 had sufficient FACIT-F data for analysis. An additional 7 off-study patients had a previous abnormal stimulated Tg ($\geq 0.5 \text{ ng/mL}$) and met inclusion criteria, and their demographic and biochemical data were included utilizing a separate IRB-approved protocol, but FACIT-F data were not obtained.

This was the first stimulated Tg for 8 of the combined 45 patients with analyzable biochemical data (Table 1). Medical records were reviewed for neck US and CT neck and chest within 14 months of enrollment utilizing results nearest study visit 1, and all laboratory, pathology, surgery, and nuclear medicine encounters since initial diagnosis. Subjects were staged according to the American Joint Committee on Cancer's Tumor Node Metastases (TNM) classification scale (33).

Tg measurement

The Tg assay used was Immulite 2000 (Siemens, Deerfield, IL). The manufacturer-reported functional sensitivity is 0.9 ng/mL , and analytical sensitivity is 0.2 ng/mL . The interassay variability, defined as the coefficient of variation (CV), at controls of 0.34, 1.4, and 51.0 ng/mL is 14.7%, 7.1%, and 6.1%, respectively. A detectable Tg level for this study was $\geq 0.2 \text{ ng/mL}$.

TABLE 1. PATIENT CHARACTERISTICS

Patient characteristics	Naïve (n = 8)	Previous abnormal stimulated Tg (n = 32)	Off-study (previous abnormal) (n = 7)
Age (years)	44.1 (33–54)	43 (19–75)	43.5 (25–65)
Sex (F:M)	1:1	4.2:1	2:1
Pathology, PTC:FTC:FVPTC (n)	6:1:1	26:2:4	7:0:0
TNM 7th staging, I:II:III:IV (n)	4:2:1:1	26:2:4:0	6:1:0:0
Ablations (n)	1.13 (1–2)	1.3 (1–4)	1.5 (1–2)
Total RAI activity (mCi)	142 (75–350) ^{a*}	175 (32–435) ^b	169 (29–376)
Duration of disease (years)	1.125 (1–2)	8.56 (2–34)	5.8 (2–14)
Reoperations (mean)	0.125 (0–1)	0.6 (0–2)	0.4 (0–1)
THW (mean)	2	2.84 (2–4)	3 (2–4)
rhTSH stimulations (mean)	0	2.13 (1–6)	1.5 (0–3)
Basal TSH ($\mu\text{U/mL}$)	0.51 (0.04–1.62)	0.10 (0.01–3.06)	0.12 (0.16–0.02)

Data are presented as mean (range) or number of patients. THW total includes the current study.

^aExcluded 2 patients with unknown activity.

^bExcluded 1 patient with unknown activity.

Tg, thyroglobulin; THW, thyroid hormone withdrawal; TSH, thyrotropin; rhTSH, recombinant human TSH.

Anti-Tg antibody measurement

The anti-Tg antibody (TgAb) assay used was the DPC 2000 Immulite anti-TgAb kit (Siemens, Deerfield, IL). The sensitivity is 2.2 IU/mL, and range is <20–3000 IU/mL. A negative TgAb test is <20 IU/mL.

Study protocol

Enrolled patients underwent baseline (visit 1) TSH, Tg battery (Tg and TgAb), and questionnaire, and then THW to achieve a TSH >30 μ U/mL and at least 4 weeks of THW (5 visits). Subjects were randomized to complete the FACIT-F questionnaire weekly (visits 1–5) or every other week (visits 1, 3, and 5). Subjects stopped TH therapy, and once a week (every 6–10 days) underwent Tg battery testing.

The TSH level necessary for Tg cutoffs of ≥ 0.5 , 1, and 2 ng/mL was considered the TSH level associated with the first Tg level to meet or exceed the Tg threshold (Table 2). For example, patient #21 had a Tg level of 1.2 ng/mL at a TSH level of 38.0 μ U/mL, and 1 week later, Tg was 3.1 ng/mL at a TSH of 62.1 μ U/mL. We assumed that they did not reach the Tg cutoff of ≥ 2 ng/mL until TSH was 62.1 μ U/mL. This probably creates a systematic bias toward concluding that higher TSH levels are required for a given Tg cutoff.

QOL questionnaire

FACIT-F is comprised of the Functional Assessment of Cancer Therapy-General (FACT-G) questions and the Fatigue Scale (FS). The former consists of 4 subgroups of questions: physical well-being (PWB), social/family well-being (SWB), emotional well-being (EWB), and functional well-being (FWB). The FS consists of additional questions focused on the effects of fatigue on activities of daily living. The FACIT-F was administered at the start of each visit, depending on ran-

domization, by the same clinician each time. Patients were left alone to complete the questionnaire before any blood draw to minimize bias or interfering variables. Patients were blinded to all results until they finished the study. Scoring the QOL questionnaire was as follows: a FACT-G score (0–108) was the sum of the 4 subgroups (PWB, SWB, EWB, FWB); FACIT-F score (0–160) was the sum of FACT-G + FS; and the FACIT-F Trial Outcome Index (TOI) was the sum of PWB, FWB, and FS (0–108). Higher scores imply higher QOL.

Statistical analysis

We attempted to predict Tg levels at THW week 4 using Tg and TSH measures from earlier weeks. Equations to predict Tg values were obtained by extending the observed slope in Tg relative to TSH between the first and second, first and third, and second and third Tg measures obtained before week 4. Another equation was obtained by estimating the least-squares regression slope for all Tg measures obtained before week 4. The observed TSH level at week 4 was entered into each equation to generate the predicted Tg at this observed TSH level. The predicted Tg was compared to the observed Tg, and the percentage of predictions within 20% of the observed value was calculated.

FACIT QOL questionnaires were analyzed using linear mixed models. Overall FACIT scores and subscales were the outcome variables. Models were fitted with either TSH or week of THW as the primary independent variable. Age at THW, sex, stage, histology, years since diagnosis, number of ablations, number of surgeries, number of withdrawals were included in the model as covariates. A random subject effect was included to account for correlation between QOL scores from the same patient over time. Recall bias was tested by including a randomization group effect to FACIT scores collected at weeks 0, 2, and 4. The capacity of FACIT scores to detect hypothyroidism and changes in TSH was evaluated by generalized linear models with a logit link. The outcomes were TSH ≥ 10 μ U/mL or an increase of 10 μ U/mL or more in TSH, and the FACIT score or subscale was the independent variable. The models accounted for correlation within the same patient. These models were fit both to all time points and to a restricted sample consisting only of TSH measures ≤ 40 μ U/mL, to limit the outcome to mild-moderate hypothyroidism.

An optimal THW-stopping rule was developed from observed Tg and TSH values. More complex stopping rules based on TSH area under the curve, or TSH in conjunction with THW week, were also considered. The negative predictive value (NPV) of the stopping rule was calculated for maximum stimulated Tg ≥ 0.5 , 1, and 2 ng/mL. Exact binomial confidence intervals for these NPVs were calculated. A multivariate logistic regression model was fit to the outcome TSH > 80 μ U/mL at 4-week THW and TSH > 30 μ U/mL at 3-week THW, with age at THW, sex, basal TSH, and mean duration of disease included as independent variables. The percentage of patients reaching Tg ≥ 0.5 , 1, and 2 ng/mL was compared between week-3 THW and TSH > 80 using Fisher's Exact tests. All hypothesis tests were evaluated at the $\alpha = 0.05$ significance level. For summary statistics and analyses, undetectable Tg and TSH results were considered zero, and TSH values were rounded to 2 decimal places, or to the nearest integer for simplicity.

TABLE 2. ADEQUATE THYROTROPIN LEVELS FOR STIMULATED THYROGLOBULIN LEVELS OF MORE THAN 0.5, 1, AND 2 NG/mL

TSH (μ U/mL)	Tg ≥ 0.5 ng/mL (n = 35)	Tg ≥ 1 ng/mL (n = 26)	Tg ≥ 2 ng/mL (n = 15)
<20	8/35 (23%)	3/26 (12%)	1/15 (7%)
20–<25	11/16 (69%)	6/13 (46%)	3/13 (23%)
25–<30	15/19 (79%)	10/17 (59%)	5/13 (38%)
30–<40	19/22 (86%)	13/18 (72%)	5/11 (45%) ^a
40–<60	26/28 (93%)	18/23 (78%)	9/12 (75%)
60–<80	33/34 (97%)	22/25 (88%)	12/14 (86%)
80–<100	33/33 (100%) ^a	22/22 (100%) ^a	12/12 (100%) ^a
100–<120	34/34 (100%)	24/24 (100%)	15/15 (100%)
120–<150	35/35 (100%)	26/26 (100%)	15/15 (100%)

Patients are added to the numerator and denominator at the TSH interval when their respective Tg level crosses the Tg threshold of interest (e.g., ≥ 0.5 , 1.0, and 2.0 ng/mL), and to all higher TSH intervals. Patients contribute only to the denominator of the highest TSH interval when its respective Tg level is below the designated Tg threshold, and to all lower TSH intervals. Patients do not contribute to the numerator or the denominator of any intervening TSH intervals. This method accounts for the changing denominators in each designated Tg column. All patients, except two (baseline Tg 0.7 and 0.8 ng/mL), had Tg < 0.5 ng/mL before THW.

^aNo new patient crosses the Tg threshold in that interval, but the percentage rises due to the lower denominator.

Results

Relationship between Tg and TSH

Data were sufficient to examine the relationship between Tg and TSH in 44 patients. The rate of rise in Tg as TSH increased was approximately linear for most patients, but it was different for each patient (Fig. 1). To determine if THW could be shortened, we tried to predict each patient's final Tg using the individual slope between two of the first three Tg values relative to TSH or by the regression slope for the first 3 Tg values. However, the best model could predict the week-4 THW Tg to within 20% of the actual value in only 40% of patients.

For most patients, the TSH and Tg both continued to rise throughout the 4-week THW. One patient was excluded from all Tg analyses due to inconsistent Tg values suggesting interference. Five patients (11%) had a plateau in Tg levels between the last 2 weeks of THW (Tg values did not change beyond the CV for the Tg assay).

Thirty-three patients had an undetectable baseline Tg level <0.2 ng/mL (42/44 patients had a baseline Tg level <0.5 ng/mL) with 16% remaining undetectable throughout the THW. Two patients met inclusion criteria but demonstrated study basal Tg levels of 0.7 and 0.8 ng/mL. The baseline TSH min, max, mean, and median were <0.04 , 3.01, 0.37, and 0.07 μ U/mL, respectively. Seventy-six percent of patients had a baseline TSH below the assay normal range, and 64% had a baseline TSH suppressed to <0.10 μ U/mL.

Most patients (97.8%) demonstrated a TSH value >30 μ U/mL within 4 weeks of THW with one exception who required 5-week THW to achieve this level (Fig. 2a). Seven patients had an undetectable Tg level <0.2 ng/mL throughout the THW. Their mean TSH was 105.19 μ U/mL at THW completion (range 49.97–148.36 μ U/mL). At the end of THW the mean TSH for all 45 patients was 88.87 μ U/mL (range 33.73–170.50 μ U/mL), and 72.7% of patients achieved a TSH

FIG. 1. The relationship of thyroglobulin (Tg) to thyrotropin (TSH) in the 15 patients who had a final Tg stimulation to >2.0 ng/mL during thyroid hormone withdrawal (THW). *Inset:* A zoomed view that emphasizes the cutoffs of Tg 2.0 ng/mL and TSH 30 μ U/mL. Patients that cross the vertical line at TSH = 30 μ U/mL fail to demonstrate their Tg >2.0 ng/mL until the TSH rises well above TSH >30 μ U/mL.

FIG. 2. (a) Percent of patients achieving a TSH level >20 , 30 , and 80 μ U/mL by week of THW. (b) Percent of patients with detectable Tg levels ≥ 0.5 , 1.0 , and 2.0 ng/mL by week of THW. Patients were categorized by their week-4 Tg grouping of $0.5-1.0$, $1.0-2.0$, or ≥ 2.0 ng/mL and the earlier weeks reflect when they cross the threshold for their particular grouping (e.g., those with a week-4 Tg ≥ 2.0 ng/mL are only represented in that category at earlier weeks). (c) Percent of patients with a final Tg ≥ 2.0 ng/mL that become detectable at 0.5 , 1.0 , and ≥ 2.0 ng/mL by week of THW.

>30 μ U/mL by 3 weeks of THW (Fig. 2a). The mean baseline TSH of patients that took longer than 3 weeks to achieve a TSH value >30 μ U/mL was 0.11 μ U/mL compared to the mean basal TSH value of 0.48 μ U/mL for the others ($p=0.09$). Only 53.3% of patients achieved a TSH >80 μ U/mL by 4 weeks of THW. The mean baseline TSH of patients with TSH ≤ 80 μ U/mL after 4-week THW was 0.31 μ U/mL compared to the mean basal TSH of 0.43 μ U/mL for the others ($p=0.54$). On multivariate analysis, age at THW, sex, basal TSH, and mean duration of disease were not significant predictors of TSH ≤ 80 μ U/mL after 4-week THW ($p=0.06$ for sex, with female patients having higher odds of TSH ≤ 80 μ U/mL after 4-week THW).

TSH level adequate for a stimulated Tg ≥ 0.5 , 1, and 2 ng/mL

Table 2 and the Figure 1 insert demonstrate that the minimal TSH cutoff of $25 < 30 \mu\text{U/mL}$ was inadequate to detect all of the patients who eventually demonstrated Tg levels ≥ 0.5 , 1, or 2 ng/mL during 4-week THW. The highest TSH values needed to detect a stimulated Tg level ≥ 0.5 , 1, or 2 ng/mL were 130, 146, and $115 \mu\text{U/mL}$, respectively. Of patients with maximum stimulated THW-Tg ≥ 2 (34% of all patients), only 38% were detected before the minimal TSH was $25 < 30 \mu\text{U/mL}$. Eighty-six percent of patients with eventual stimulated THW-Tg ≥ 2 were detected when the TSH was $60 < 80 \mu\text{U/mL}$. Overall, a TSH > 80 or $> 100 \mu\text{U/mL}$ was needed to detect nearly all patients who ultimately achieved Tg levels ≥ 0.5 , 1, or 2 ng/mL (Table 2).

TSH vs. THW weeks to detect stimulated Tg cutoffs

As Tg continued to rise in most patients who achieved Tg levels ≥ 0.5 , 1, or 2 ng/mL during 4-week THW, Figure 2b demonstrates that measurement of Tg after 2 or 3 weeks of THW failed to correctly categorize many patients who eventually achieved a final Tg level of $\geq 0.5 < 1$, $\geq 1 < 2$, or ≥ 2 ng/mL during the study. Of patients who achieved a final THW-Tg ≥ 2 , only 20% and 60% of patients were identified after 2 and 3 weeks of THW, respectively (Fig. 2b, c). At 2-week THW, 3 had a TSH $> 30 \mu\text{U/mL}$, and none of them had a Tg ≥ 2 ng/mL. At 3-week THW, 11 had a TSH $> 30 \mu\text{U/mL}$, and 64% of them had Tg ≥ 2 ng/mL. All patients with THW-Tg ≥ 2 ng/mL within the protocol that ended after 4-week THW plus a TSH $> 30 \mu\text{U/mL}$ achieved this endpoint at 4-week THW. It is unknown if more patients would have achieved THW-Tg ≥ 2 with a longer duration of THW.

We analyzed whether the TSH level or the number of THW weeks was more important to identify patients who achieved final Tg levels ≥ 0.5 , 1, or 2 ng/mL. TSH $> 80 \mu\text{U/mL}$ identified 100% of patients at each threshold (Table 2), a significantly higher percentage than were identified at week 3 (Fig. 2b; Tg $0.5 < 1$ [100% vs. 38%, $p=0.03$], Tg $1 < 2$ [100% vs. 42%, $p=0.01$], and Tg ≥ 2 ng/mL [100% vs. 60%, $p=0.02$]). Failure to reach a TSH of > 80 by THW week 4 occurred in 47% (21/45) of patients (Fig. 2a).

Correlation of stimulated Tg result with anatomic imaging

Anatomic imaging identified no significant thyroid remnant in any patient. In patients with a final Tg stimulation < 2 ng/mL, 1 patient had a neck CT that was negative and 52% had a neck US performed; it was negative or the most significant lymph nodes were ≤ 5 mm in 95% and no biopsies were performed. In 1 patient a $5 \text{ mm} \times 6 \text{ mm}$ benign-appearing lymph node was reported. In patients with a final THW-Tg ≥ 2 , 86% had an US; it was negative or the most significant lymph nodes were ≤ 5 mm in 25%. No biopsies were performed. The remaining 75% had some lymph nodes > 5 mm with benign features and maximum size 10 mm. The 2/15 patients with a Tg > 2.0 ng/mL without an US had biochemically stable disease. One patient had an US-FNA that was negative. Eight patients had additional neck CT imaging (all negative) and 4 had additional chest CT imaging with nonspecific findings.

THW-stopping rule

In the nine patients with an undetectable basal Tg (< 0.2 ng/mL) and final stimulated THW-Tg ≥ 2 , 100% had a detectable Tg level before a TSH cutoff of $20 \mu\text{U/mL}$. A Tg of < 0.2 ng/mL was observed in 11 patients at a TSH $> 20 \mu\text{U/mL}$, and none had a final stimulated THW-Tg ≥ 2 . Of those with an undetectable Tg and a TSH $> 20 \mu\text{U/mL}$, 91% did not manifest a final stimulated Tg ≥ 1 ng/mL. Only 1 patient, #1033, with undetectable Tg at TSH of $26.80 \mu\text{U/mL}$ (THW week 2) had a TSH-stimulated Tg > 1.0 (1.1 ng/mL) at THW week 4 when the TSH was $129.87 \mu\text{U/mL}$. The NPV was 100% for patients with an undetectable Tg at a TSH cutoff $> 20 \mu\text{U/mL}$ for stimulating Tg ≥ 2 ng/mL [95% confidence interval (CI) 72%–100%], and 91% for stimulating to a Tg ≥ 1 ng/mL [CI 55%–99%] (Fig. 3a, b), preliminarily suggesting that individuals with an undetectable Tg level with a TSH $> 20 \mu\text{U/mL}$ have a low likelihood of residual thyroid cancer.

FACIT-F scores

Nineteen patients were randomized to complete the FACIT-F weekly, and 19 took it every other week (visits 1, 3, and 5). Total FACIT-F scores were significantly correlated to TSH ($p < 0.001$) and THW week, as were TOI, and all subscale scores (including the FS), except EWB ($p=0.34$). The TSH level and THW week were highly correlated ($r=0.85$). FACIT-F scores were not correlated with any factors listed in Table 1. There was no significant difference or recall bias in FACIT-F or subscale scores between patients taking the questionnaire weekly versus every other week. FACIT-F was unable to detect TSH $> 10 \mu\text{U/mL}$ ($p=0.73$) compared to baseline. FACIT-F was unable to detect a $10 \mu\text{U/mL}$ change in TSH when restricted for a TSH $< 40 \mu\text{U/mL}$ except in the subscale of social well-being ($p=0.01$). There was a significant

FIG. 3. (a) Stopping rule. Patients who have an undetectable Tg at a TSH $> 20 \mu\text{U/mL}$ will not have a final stimulated Tg level ≥ 2 ng/mL at the end of 4 weeks of THW as depicted by the symbols below representing all data points from all 11 patients that met the stopping rule criteria (SRC). *Asterisk denotes undetectable Tg level. (b) Negative predictive values of the Stopping rule to achieve different peak Tg cutoff values.

FIG. 4. Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) total, physical well-being, functional well-being, and fatigue scale scores shown with box plots for all patients by week of THW. The shaded boxes are the interquartile range between the 75th and 25th percentiles. The hash marks are the minimum and maximum values, excluding outliers (○). Mean (◇) and median (horizontal line within the box) scores are shown. Significant differences in scores between weeks of THW shown with the following symbols: ‡, § correspond to weeks 3 and 4, respectively (all p -values ≤ 0.02). For example, FACIT-F total scores at week 0 were significantly different compared to week 3 (‡), and 4 (§).

difference in FACIT-F scores between weeks 2–3 ($p=0.02$), 3–4 ($p=0.01$), and weeks 2–4 ($p<0.001$) as shown in Figure 4.

Discussion

In our study, a TSH level of 145 and 115 $\mu\text{U}/\text{mL}$ was necessary to capture all patients with Tg stimulations ≥ 1 and 2 ng/mL, respectively. In our study, only 73% of patients achieved a TSH $>30 \mu\text{U}/\text{mL}$ by 3 weeks, which is less than other studies that report between 90% and 97% of patients at 3 weeks (15, 18, 32). This may be, in part, due to the difference in the frequency of TSH laboratory draws (weekly versus every 2–4 days) or the level of baseline TSH suppression, which is difficult to compare given differing functional sensitivities between studies. Our study is also limited by its small sample size, frequency of only once-weekly laboratory draws, and THW duration of only 4 weeks. If our cohort continued the THW beyond 4 weeks, it is likely that more patients would have achieved a final Tg level ≥ 1 or 2 ng/mL. How this would have affected our conclusions is unknown.

When the Tg became detectable during THW, it increased in most patients with increasing TSH levels as suggested by

prior investigators (20). Still, the rate of rise (slope) was different for each patient, and we were unable to accurately predict the final Tg using the patient's data from early time points during THW. How long the TSH and/or Tg would continue to rise is unknown. Our study demonstrates that stimulated Tg levels during THW must be interpreted in the context of the corresponding TSH level, or else false conclusions about the disease status may be drawn. This caution extends to the use of fixed Tg cutoffs to indicate residual disease, or to prompt additional testing such as neck US, chest CT, FDG-PET imaging, or RAI scanning, or alter treatment such as RAI therapy, or the degree of prescribed TSH suppression. For example, would different interpretations or management happen if patient #1002 stopped THW after 1 week when their Tg was $<0.2 \text{ ng/mL}$ with a TSH of 9.82 $\mu\text{U}/\text{mL}$? What about if they stopped THW after 2 weeks instead of 4 weeks when their Tg was 1.2 ng/mL with a TSH of 29.92 $\mu\text{U}/\text{mL}$ versus a Tg of 7.4 ng/mL with a TSH of 56.65 $\mu\text{U}/\text{mL}$, respectively? What would their Tg be if their TSH rose to >80 – $100 \mu\text{U}/\text{mL}$? This example demonstrates the complexities of using fixed cutoff values during THW when both TSH and Tg change over time. This emphasizes the need

for consistency in the method and manner in which stimulated Tg is performed, and to acknowledge differences when consistency is not possible. In practice, that should mean that instead of considering a patient as having a stimulated Tg of X, that one should recognize the patient as having a stimulated Tg of X with a TSH of Y after Z weeks of THW. As a result, the clinician is prompted to recognize the dynamic variables of TSH, weeks of THW, and method of Tg stimulation when they compare this Tg to a future or past result.

Nearly all patients in our study with a stimulated Tg ≥ 0.5 , 1, and 2 ng/mL were detected by 4 weeks of THW (by study design), but the minimal TSH cutoff of 25–30 $\mu\text{U/mL}$ was inadequate to detect many of these patients, including 62% whose final Tg was >2 ng/mL. A TSH cutoff of >80 or >100 $\mu\text{U/mL}$ was a more sensitive cutoff (Table 2) and was more sensitive than 3-week THW (Fig. 2b). As only 53% of patients achieved a TSH >80 $\mu\text{U/mL}$ after 4-week THW (Fig. 2a), this implies that many patients would require >4 -week THW to more sensitively detect the Tg cut-points evaluated in this study. One may question the significance of finding a stimulated THW-Tg ≥ 2 only when the TSH is very high. A Tg ≥ 2 ng/mL was an important cutoff level in the second phase-III comparison trial of rhTSH and THW to detect thyroid remnant or cancer (7). In that study, the patients achieved TSH levels that approached 80 mU/L during THW with their mean TSH levels of 71 and 69 $\mu\text{U/mL}$ for arms I and II, respectively. Additionally, other studies have correlated anatomic findings with stimulated Tg levels comparable to ours during 3–5-week THW, suggesting that they too achieved similarly high TSH levels (8,9). Our patients lacked significant structural disease. However, our cohort is likely enriched for patients with persistent biochemical disease, yet lacking significant anatomic disease, because it is dominated by patients previously screened for structurally persistent disease (which led some to reoperation with curative intent before inclusion in this study). Thus, whether or not our lack of anatomic findings is applicable to other patients is unknown.

It is important for clinicians to recognize that higher Tg levels are typically achieved during THW than after rhTSH stimulation (7,34,35). Schlumberger *et al.* showed a 2-fold difference in THW-stimulated Tg compared to those after rhTSH, but with similar recurrence rates in both groups (9). The reproducibility of stimulated Tg measurements during THW and during rhTSH has not been well studied, while variability of Tg measurements between assays is well recognized (9,34). Data from Spencer *et al.* suggest that the reproducibility of rhTSH-Tg stimulation testing over time is variable (34).

The role of stimulated Tg testing in patient follow-up is diminishing as Tg assays with improved functional sensitivities emerge. Still, the data of Spencer *et al.* suggest that even with 2nd-generation Tg assays that there may still be a role for TSH stimulation when the basal Tg is between 0.1 and 0.5 ng/mL to clarify the disease status (34). Others have demonstrated impaired specificity of low functional sensitivity Tg assays, and that Tg testing accuracy was improved by TSH stimulation testing (8,9).

The second aim of our study was to evaluate FACIT-F for QOL during THW. Here, FACIT-F was able to detect QOL changes associated with hypothyroidism after 4 weeks of THW when compared to baseline. This ability was previously

shown by Taieb *et al.* (21, 22) and by other questionnaires [FACT-G, Billewicz score, SF-36, and QOL-Thyroid scale (4,21,22,30,32,36,37)]. We detected no recall bias in patients completing the questionnaire weekly when compared to those completing it every other week. However, neither FACIT-F nor FACT-G was found to be a sensitive tool for detecting mild hypothyroidism, similar to other questionnaires (38–40). Overall, FACIT-F was highly correlated to TSH, and QOL scores significantly differed over the last two weeks of THW (between weeks 2–3, 3–4, and 2–4) calculated in total score, TOI scores, and all subscales scores, except social and EWB. In our study a TSH >20 $\mu\text{U/mL}$ was reached in 44% and 86% of our patients by THW week 2 or 3, respectively. Had we stopped THW with the first TSH >20 $\mu\text{U/mL}$ with an undetectable Tg, consistent with our preliminary stopping rule, then 40% of the THW duration and its hypothyroid symptoms could have been avoided in these patients.

In conclusion, the majority of patients with detectable Tg levels manifest a continued rise of both TSH and Tg throughout 4 weeks of THW. The minimal TSH cutoff of >30 $\mu\text{U/mL}$ is inadequate to detect many patients that eventually demonstrated a stimulated THW-Tg ≥ 2 . A TSH cutoff of >80 or >100 $\mu\text{U/mL}$ was more reliable to detect these patients, suggesting that consistent methods and intensity of stimulation are necessary for adequate comparisons when monitoring patients with DTC. However, when the Tg was undetectable (<0.2 ng/mL) at a TSH >20 $\mu\text{U/mL}$, then with 91% and 100% certainty, their final Tg did not stimulate to ≥ 1 or 2 ng/mL, respectively, during 4 weeks of THW. Such a stopping rule may be appropriate for patients with a low pretest probability of having residual thyroid cancer who are selected for stimulated Tg testing.

Disclosure Statement

L.A.V., R.L.G., K.P., J.A.S., and R.K. having nothing to disclose. J.A.S. and R.T.K. are recent speakers for Genzyme Corporation. M.D.R. was previously on an advisory board for Veracetyte, Inc. R.T.K. was a consultant to Genzyme Corporation. This research and manuscript were generated while R.T.K. was an OSU faculty member.

References

- Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Pacini F, Schlumberger M, Sherman SI, Steward DL, Tuttle RM 2009 Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 19:1167–1214.
- Pacini F, Molinaro E, Castagna MG, Agate L, Elisei R, Ceccarelli C, Lippi F, Taddei D, Grasso L, Pinchera A 2003 Recombinant human thyrotropin-stimulated serum thyroglobulin combined with neck ultrasonography has the highest sensitivity in monitoring differentiated thyroid carcinoma. *J Clin Endocrinol Metab* 88:3668–3673.
- Mazzaferri EL, Kloos RT 2002 Is diagnostic iodine-131 scanning with recombinant human TSH useful in the follow-up of differentiated thyroid cancer after thyroid ablation? *J Clin Endocrinol Metab* 87:1490–1498.
- Wartofsky L 2002 Management of low-risk well-differentiated thyroid cancer based only on thyroglobulin measurement after recombinant human thyrotropin. *Thyroid* 12:583–590.

5. Kloos RT, Mazzaferri EL 2005 A single recombinant human thyrotropin-stimulated serum thyroglobulin measurement predicts differentiated thyroid carcinoma metastases three to five years later. *J Clin Endocrinol Metab* **90**:5047–5057.
6. Kloos RT 2010 Thyroid cancer recurrence in patients clinically free of disease with undetectable or very low serum thyroglobulin values. *J Clin Endocrinol Metab* **95**:5241–5248.
7. Haugen BR, Pacini F, Reiners C, Schlumberger M, Ladenson PW, Sherman SI, Cooper DS, Graham KE, Braverman LE, Skarulis MC, Davies TF, DeGroot LJ, Mazzaferri EL, Daniels GH, Ross DS, Luster M, Samuels MH, Becker DV, Maxon HR, 3rd, Cavalieri RR, Spencer CA, McEllin K, Weintraub BD, Ridgway EC 1999 A comparison of recombinant human thyrotropin and thyroid hormone withdrawal for the detection of thyroid remnant or cancer. *J Clin Endocrinol Metab* **84**:3877–3885.
8. Brassard M, Borget I, Edet-Sanson A, Giraudet AL, Mundler O, Toubreau M, Bonichon F, Borson-Chazot F, Leenhardt L, Schvartz C, Dejax C, Brenot-Rossi I, Toubert ME, Torlontano M, Benhamou E, Schlumberger M 2011 Long-term follow-up of patients with papillary and follicular thyroid cancer: a prospective study on 715 patients. *J Clin Endocrinol Metab* **96**:1352–1359.
9. Schlumberger M, Hitzel A, Toubert ME, Corone C, Troalen F, Schlageter MH, Claustat F, Koscielny S, Taieb D, Toubreau M, Bonichon F, Borson-Chazot F, Leenhardt L, Schvartz C, Dejax C, Brenot-Rossi I, Torlontano M, Tenenbaum F, Bardet S, Bussiere F, Girard JJ, Morel O, Schneegans O, Schlienger JL, Prost A, So D, Archambeaud F, Ricard M, Benhamou E 2007 Comparison of seven serum thyroglobulin assays in the follow-up of papillary and follicular thyroid cancer patients. *J Clin Endocrinol Metab* **92**:2487–2495.
10. Edmonds CJ, Hayes S, Kermode JC, Thompson BD 1977 Measurement of serum TSH and thyroid hormones in the management of treatment of thyroid carcinoma with radioiodine. *Br J Radiol* **50**:799–807.
11. Tamai H, Suemastu H, Kurokawa N, Esaki M, Ikemi T, Matsuzuka F, Kuma K, Nagataki S 1979 Alterations in circulating thyroid hormones and thyrotropin after complete thyroidectomy. *J Clin Endocrinol Metab* **48**:54–58.
12. Hilts SV, Hellman D, Anderson J, Woolfenden J, Van Antwerp J, Patton D 1979 Serial TSH determination after T3 withdrawal or thyroidectomy in the therapy of thyroid carcinoma. *J Nucl Med* **20**:928–932.
13. Hershman JM, Edwards CL 1972 Serum thyrotropin (TSH) levels after thyroid ablation compared with TSH levels after exogenous bovine TSH: implications for 131-I treatment of thyroid carcinoma. *J Clin Endocrinol Metab* **34**:814–818.
14. Goldman JM, Line BR, Aamodt RL, Robbins J 1980 Influence of triiodothyronine withdrawal time on 131I uptake post-thyroidectomy for thyroid cancer. *J Clin Endocrinol Metab* **50**:734–739.
15. Serhal DI, Nasrallah MP, Arafah BM 2004 Rapid rise in serum thyrotropin concentrations after thyroidectomy or withdrawal of suppressive thyroxine therapy in preparation for radioactive iodine administration to patients with differentiated thyroid cancer. *J Clin Endocrinol Metab* **89**:3285–3289.
16. Kuijt WJ, Huang SA 2005 Children with differentiated thyroid cancer achieve adequate hyperthyrotropinemia within 14 days of levothyroxine withdrawal. *J Clin Endocrinol Metab* **90**:6123–6125.
17. Grigsby PW, Siegel BA, Bekker S, Clutter WE, Moley JF 2004 Preparation of patients with thyroid cancer for 131I scintigraphy or therapy by 1–3 weeks of thyroxine discontinuation. *J Nucl Med* **45**:567–570.
18. Sanchez R, Espinosa-de-los-Monteros AL, Mendoza V, Brea E, Hernandez I, Sosa E, Mercado M 2002 Adequate thyroid-stimulating hormone levels after levothyroxine discontinuation in the follow-up of patients with well-differentiated thyroid carcinoma. *Arch Med Res* **33**:478–481.
19. Liel Y 2002 Preparation for radioactive iodine administration in differentiated thyroid cancer patients. *Clin Endocrinol (Oxf)* **57**:523–527.
20. Tahboub RM, Nyalakonda K, Arafah BM 2009 Change in serum TSH and thyroglobulin levels in patients with differentiated thyroid cancer after short-term thyroid hormone withdrawal: an effective approach in monitoring patients and disease surveillance. The Endocrine Society, *ENDO 09 91st Annual Meeting*. Abstract no. OR12-05.
21. Taieb D, Sebag F, Cherenko M, Baumstarck-Barrau K, Fortanier C, Farman-Ara B, De Micco C, Vaillant J, Thomas S, Conte-Devolx B, Loundou A, Auquier P, Henry JF, Mundler O 2009 Quality of life changes and clinical outcomes in thyroid cancer patients undergoing radioiodine remnant ablation (RRA) with recombinant human TSH (rhTSH): a randomized controlled study. *Clin Endocrinol (Oxf)* **71**:115–123.
22. Taieb D, Baumstarck-Barrau K, Sebag F, Fortanier C, De Micco C, Loundou A, Auquier P, Palazzo FF, Henry JF, Mundler O 2011 Health-related quality of life in thyroid cancer patients following radioiodine ablation. *Health Qual Life Outcomes* **9**:33.
23. Billewicz WZ, Chapman RS, Crooks J, Day ME, Gossage J, Wayne E, Young JA 1969 Statistical methods applied to the diagnosis of hypothyroidism. *Q J Med* **38**:255–266.
24. Shacham S 1983 A shortened version of the profile of mood states. *J Pers Assess* **47**:305–306.
25. Schroeder PR, Haugen BR, Pacini F, Reiners C, Schlumberger M, Sherman SI, Cooper DS, Schuff KG, Braverman LE, Skarulis MC, Davies TF, Mazzaferri EL, Daniels GH, Ross DS, Luster M, Samuels MH, Weintraub BD, Ridgway EC, Ladenson PW 2006 A comparison of short-term changes in health-related quality of life in thyroid carcinoma patients undergoing diagnostic evaluation with recombinant human thyrotropin compared with thyroid hormone withdrawal. *The Journal of clinical endocrinology and metabolism* **91**:878–884.
26. Pacini F, Ladenson PW, Schlumberger M, Driedger A, Luster M, Kloos RT, Sherman S, Haugen B, Corone C, Molinaro E, Elisei R, Ceccarelli C, Pinchera A, Wahl RL, Lebouilleux S, Ricard M, Yoo J, Busaidy NL, Delpassand E, Hanscheid H, Felbinger R, Lassmann M, Reiners C 2006 Radioiodine ablation of thyroid remnants after preparation with recombinant human thyrotropin in differentiated thyroid carcinoma: results of an international, randomized, controlled study. *J Clin Endocrinol Metab* **91**:926–932.
27. Luster M, Felbinger R, Dietlein M, Reiners C 2005 Thyroid hormone withdrawal in patients with differentiated thyroid carcinoma: a one hundred thirty-patient pilot survey on consequences of hypothyroidism and a pharmacoeconomic comparison to recombinant thyrotropin administration. *Thyroid* **15**:1147–1155.
28. Lee J, Yun MJ, Nam KH, Chung WY, Soh EY, Park CS 2010 Quality of life and effectiveness comparisons of thyroxine withdrawal, triiodothyronine withdrawal, and recombinant thyroid-stimulating hormone administration for low-dose radioiodine remnant ablation of differentiated thyroid carcinoma. *Thyroid* **20**:173–179.

29. Dueren C, Dietlein M, Luster M, Plenzig F, Steinke R, Grimm J, Groth P, Eichhorn W, Reiners C 2010 The use of thyrogen in the treatment of differentiated thyroid carcinoma: an intraindividual comparison of clinical effects and implications of daily life. *Exp Clin Endocrinol Diabetes* **118**: 513–519.
30. Dow KH, Ferrell BR, Anello C 1997 Quality-of-life changes in patients with thyroid cancer after withdrawal of thyroid hormone therapy. *Thyroid* **7**:613–619.
31. Davids T, Witterick IJ, Eski S, Walfish PG, Freeman JL 2006 Three-week thyroxine withdrawal: a thyroid-specific quality of life study. *Laryngoscope* **116**:250–253.
32. Chow SM, Au KH, Choy TS, Lee SH, Yeung NY, Leung A, Leung A, Leung HL, Shek CC, Law SC 2006 Health-related quality-of-life study in patients with carcinoma of the thyroid after thyroxine withdrawal for whole body scanning. *Laryngoscope* **116**:2060–2066.
33. Sobin L, Gospodaraowicz M, Wittekind C 2009 TNM Classification of Malignant Tumours. Seventh edition. Wiley-Blackwell, Hoboken, NJ, pp 58–62.
34. Spencer C, Fatemi S, Singer P, Nicoloff J, Lopresti J 2010 Serum basal thyroglobulin measured by a second-generation assay correlates with the recombinant human thyrotropin-stimulated thyroglobulin response in patients treated for differentiated thyroid cancer. *Thyroid* **20**: 587–595.
35. Pacini F, Molinaro E, Lippi F, Castagna MG, Agate L, Ceccarelli C, Taddei D, Elisei R, Capezzone M, Pinchera A 2001 Prediction of disease status by recombinant human TSH-stimulated serum Tg in the postsurgical follow-up of differentiated thyroid carcinoma. *J Clin Endocrinol Metab* **86**:5686–5690.
36. Leboeuf R, Perron P, Carpentier AC, Verreault J, Langlois MF 2007 L-T3 preparation for whole-body scintigraphy: a randomized-controlled trial. *Clin Endocrinol (Oxf)* **67**:839–844.
37. Husson O, Haak HR, Oranje WA, Mols F, Reemst PH, van de Poll-Franse LV 2011 Health-related quality of life among thyroid cancer survivors: a systematic review. *Clin Endocrinol (Oxf)* **75**:544–554.
38. Helfand M 2004 Screening for subclinical thyroid dysfunction in nonpregnant adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* **140**:128–141.
39. Canaris GJ, Manowitz NR, Mayor G, Ridgway EC 2000 The Colorado thyroid disease prevalence study. *Arch Intern Med* **160**:526–534.
40. Lindeman RD, Schade DS, LaRue A, Romero LJ, Liang HC, Baumgartner RN, Koehler KM, Garry PJ 1999 Subclinical hypothyroidism in a biethnic, urban community. *J Am Geriatr Soc* **47**:703–709.

Address correspondence to:

Richard T. Kloos, MD

Veracyte, Inc.

7000 Shoreline Court, Suite 250

South San Francisco, CA 94080

E-mail: richard.kloos@veracyte.com