The role of thyroglobulin doubling-time in differentiated thyroid cancer: a meta-analysis

1,2Luca Giovanella, 3Maria Luisa Garo, 4Domenico Albano, 5Rainer Görges, 1,6Luca Ceriani

1Clinic for Nuclear Medicine, Imaging Institute of Southern Switzerland, Ente Ospedaliero Cantonale, Bellinzona (Switzerland)
2Clinic for Nuclear Medicine, University Hospital of Zürich, Zürich (Switzerland)
3Mathsly Research, Brescia and Vibo Valentia (Italy)
4Nuclear Medicine, University of Brescia and Spedali Civili Brescia, Brescia, Italy
5Department of Nuclear Medicine, University Hospital of Essen, Essen, Germany
6Faculty of Biomedical Sciences, Institute of Oncology Research, Università della Svizzera Italiana, Bellinzona, Switzerland

Corresponding author
Prof. Dr. med. Luca Giovanella MD PhD
Clinic for Nuclear Medicine - Imaging Institute of Southern Switzerland
Via A. Gallino 12, Bellinzona 6500, Switzerland
Phone: +41-91-811 72 12  Fax: +41-91-811 85 20
E-mail: luca.giovanella@eoc.ch

Short title: Thyroglobulin doubling-time
Keywords: thyroglobulin, doubling-time, differentiated thyroid cancer, prognosis
Word Count: 3070
ABSTRACT

Objective: In patients with differentiated thyroid cancer (DTC), recurrences may occur in up to 20% and may have a fatal outcome in 10% of cases. Thyroglobulin-Doubling Time (Tg-DT) values may contribute to predict response to treatment and disease recurrence in DTC patients. This study aimed to address the following questions: (1) Are Tg-DT values indicative of response to treatments in patients with DTC (i.e., “treatment monitoring?”) (2) Is Tg-DT predictive of 2-[18F]fluoro-2-deoxy-D-glucose (2-[18F]FDG) positron emission tomography/computed tomography (PET/CT) in patients with DTC? (3) Are Tg-DT values predictive of DTC prognosis (i.e., “prediction”)?

Design: Systematic review and meta-analysis

Methods: Methodology was registered in the PROSPERO database (CRD42021257947). A systematic search was carried out in PubMed, Web Of Science, and Scopus from June to August 2021 without time and language restrictions.

Results: Eleven studies were included for a total of 1421 patients. Positive association between Tg-DT < 1 year and recurrence or disease progression were observed. Tg-Dt was found to be (2-[18F]FDG) PET/CT results in patients with DTC. Area Under the Curve (AUC) = 0.86 (95%CI: 0.83-0.89). Sensitivity: 0.84 (0.64;0.94); Specificity: 0.71 (0.35; 0.92); DOR: 13.1 (3.1; 55.0); LR+: 2.9 (1.0; 8.1); LR-: 0.22 (0.1; 0.5). For patients with Tg-DT < 1 year (n = 247), the survival risk ratio was 2.09 [95%CI: 1.49; 2.94].

Conclusions: Tg-DT values are valuable in predicting response to treatment and disease recurrence in patients with DTC, as well as their overall survival. In addition, Tg-DT significantly increases the detection rate of 2-[18F]-FDG PET/CT.
INTRODUCTION

In many cases, differentiated thyroid carcinoma (DTC) has an indolent course and a generally favorable prognosis. However, recurrences may occur in up to 20% of DTC patients, with 10% of them having a fatal outcome (1). Serum thyroglobulin (Tg) is the pivotal DTC biomarker during follow-up of these patients: an undetectable Tg level (after excluding interfering anti-Tg autoantibodies) is associated with excellent treatment response and favorable prognosis. In contrast, detectable Tg levels may signal persistent or recurrent disease. Moreover, absolute Tg concentrations are correlated with tumor load and are widely employed to assess the extension of the disease and evaluate the response to treatments (2). Nevertheless, this single tumor marker measurement may not be exhaustive in the comprehension of disease status and treatment response because not intrinsically inclusive of previous measurements and the overall trend (3). Therefore, evaluating dynamic changes in Tg concentration over time is necessary to predict recurrence rates and overall survival. Tg doubling time (Tg-DT) has been demonstrated as a valuable biomarker to predict loco-regional recurrences, distant metastases, and survival independently from classical prognostic factors (e.g., TNM stage, age, and gender) (4). Recently, the Tg-DT was also proved useful to predict results of 2-[18F]fluoro-2-deoxy-D-glucose (2-[18F]FDG) positron emission tomography/computed tomography (PET/CT) in patients with detectable Tg levels who had a negative radioiodine whole-body scintigraphy (WBS) (5, 6). Despite relevant literature, however, recent management guidelines contain little specific advice on the use and interpretation criteria of Tg-DT and, consequently, Tg-DT is sparsely adopted in clinical practice. Therefore, the present study was prompted to provide a systematic review and meta-analysis of updated literature to obtain more robust evidence on Tg-DT performance in DTC assessment. Specifically, we aimed to address three clinical questions:

1. Are Tg-DT values indicative of response to treatments in patients with DTC (i.e., “treatment monitoring”)?

2. Is Tg-DT predictive of (2-[18F]FDG) PET/CT results in patients with DTC?
3. Are Tg-DT values predictive of DTC prognosis (i.e., “prediction”)?

**Materials and Methods**

*Protocols and Registration*

The systematic review and meta-analysis were performed according to the PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) guidelines (7). The methodology was registered in the PROSPERO (*International prospective register of systematic reviews*) database under the protocol number CRD42021257947.

*Eligibility criteria*

All original peer-reviewed research publications were considered. Inclusion criteria for eligible studies were: (1) At least two consecutive Tg measurements under the thyroid hormone replacement therapy to calculate Tg kinetics; (2) No interfering anti-Tg antibodies; (3) Prospective or retrospective studies.

*Information sources and search strategy*

A systematic search strategy was carried out on PubMed, Embase, Web of Science, and Scopus from May to July 2021 without time and language restrictions. The literature search strategy was based on the following keywords: (thyroglobulin) AND (doubling time OR doubling-time OR DT). Additionally, hand searches were performed to identify possible articles other than those found in the electronic databases. Finally, a further hand search of the citation lists of the included studies was performed. Two reviewers (L.G. and M.G.) performed the first (title/abstract screening) and second (full-text assessment) steps of the search process. Any disagreement was discussed and then solved by consensus.

*Study selection*
After removing duplicates and excluding not eligible articles, the potentially relevant articles were screened by reading titles and abstracts. Two reviewers selected the eligible studies (L.G. and M.G.) independently. Then, full texts of the eligible articles, i.e., those that met the inclusion and exclusion criteria, were retrieved. The final eligibility of each study was assessed, and the reasons for exclusion were recorded. Finally, two authors (L.G. and L.C.) executed the definitive article selection. In case of disagreement, it was resolved by discussion.

Data Extraction

Data extraction was organized in tables containing the following information:

1) Study characteristics: first author, year, country.
2) Study’s sample size and patients’ characteristics (sex, mean age, % of males).
3) Number of patients with papillary thyroid cancer and number of patients with follicular thyroid cancer.
4) Poorly differentiated or Hürthle or aggressive variants.
5) Sample size included in Tg-DT calculation.
6) Follow-up in years.
7) Number of consecutive Tg level measurements used to determine Tg-DT.
8) Rates of disease progression or recurrence of DTC.

No numerical information was extracted from the figures reported in the study publications.

Quality assessment

Two authors independently assessed the risk of bias of included studies using the NIH quality assessment tool for observational cohorts (https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools). The tool comprises fourteen criteria; the overall assessment is rated good, fair, or poor. Possible disagreements were resolved by discussion and consensus among all authors.
Statistical Analysis

The hierarchical summary receiver operating characteristics (HSROC) and bivariate methods were performed to determine the predictivity of (2-[18F]FDG) PET/CT results in patients with DTC. For cells containing zero, the continuity correction of 0.5 was used (8). Positive likelihood ratios greater than 2.0 or negative likelihood ratios lower than 0.5 with 95% CIs not including 1.0 were considered statistically significant (9). The risk ratio for survival was summarized through the DerSimonian–Laird random-effects model due to the nature of the studies. We had hypothesized that the treatment effect was influenced not only by the residual effect but also by unexpected factors as unmeasured comorbidity, age, or tumor stage. Heterogeneity was assessed using Cochrane Q-test and the I² statistic, where a p-value < 0.05 was taken to indicate statistically significant heterogeneity. According to the Cochrane Handbook for Systematic Reviews of Interventions, the ranges of interpretation for I² are as follows: 0–40% may be unimportant; 30–60% may represent moderate heterogeneity; 50–90% may represent substantial heterogeneity, and 75–100% may have considerable heterogeneity. Heterogeneity was also investigated through the L’Abbé plot: studies that deviated from the effect-size line considerably were reported as heterogeneity. Subgroup analyses were performed for studies mainly composed of patients with papillary thyroid cancer, studies carried out on more than 50 and 100 patients, respectively, and studies with at least four consecutive Tg measurements. Publication bias was not assessed because fewer than ten studies were included in the meta-analysis (10). A sensitivity analysis was performed to examine whether overall findings were robust to the chosen analysis method. Finally, the power of meta-analysis was calculated using the metapow command. Meta-analysis was carried out using STATA17 (StataCorp., College Station, TX, USA).

Results

Study Selection
The literature search identified 1830 studies. After removing duplicate records, 1685 studies were screened. Eleven studies between 2011 and 2021 met the inclusion and exclusion criteria (Figure 1). Sixteen reports were excluded, mainly because irrelevant to the main topic.

*Risk of bias*

Five studies were good as assessed by the NIH tool (2, 11-14). Sample size justification or power description were not reported in all included studies. In addition, the percentage of patients lost during the follow-up was not reported in several works (6, 11-16). Quality assessment of the included studies has been reported in Supplementary Table 1.

*Study characteristics*

One thousand four hundred and twenty-one (1,421) patients were included (Table 1). Patients’ mean age ranged between about 40 and 77 years. Samples of seven studies were prevalently composed by patients with papillary thyroid cancer for a total of 910 patients (4, 6, 12, 14-17), while the remaining four studies were prevalently composed by patients with follicular thyroid carcinoma for a total of 343 patients (5, 11, 13, 18). Six studies had a sample size lower than 100 patients (5, 13, 15-18), while five studies had a larger sample size that ranged between 102 (6) and 426 (4) patients. The percentage of males/females ranged between 18.1% (4) and 59.6%(15). Tg-DT calculation was reported for 918 patients. Six studies used at least four TG level measurements to determine Tg-DT (12, 14-16). Two studies used at least three Tg level measurements (5, 13), while the remaining three studies used respectively two (6, 17) and five (18) Tg level measurements. Three studies (13, 14, 18) were carried out in Germany, three in Japan (4, 12, 16), one in Switzerland (6), France (17), Turkey (5), USA (15), and Italy (11), respectively. The median follow-up was 6.1 years ranging between 157 days (16) to 11.1 years (15).
**Is Tg-DT indicative of response to treatment in patients with DTC?**

Five studies investigated the association between Tg-DT and DTC recurrence or disease progression (4, 12, 14, 16, 18) (Supplementary Table 2).

A positive association between Tg-DT < 1 year and recurrence or disease progression was observed in four studies (4, 14, 16, 18). Miyauchi et al. (2011) reported higher rates of loco-regional recurrence in patients with Tg-DT < 1 year (43.8% at five years and 78.6% at ten years) than in patients with Tg-DT ranging between 1 and 3 years (23.5% at five years, 72.6% at ten years), and in patients with Tg-DT ≥ 3 years (23.6% at five years and 42.5% at ten years) (4). Kelder et al. (2014) showed DTC progression in 8 patients out of nine with Tg-DT < 1 year (18). Verburg et al. (2017), in a sample of 174 patients, observed recurrence in 8 patients with Tg-DT < 1 year and no recurrence in patients with Tg-DT ≥ 1 year (14). Finally, Zhang et al. (2020) reported Tg-DT < 1 year in all patients with disease progression (16). Only one study reported no association between Tg-DT values and disease progression (12).

**Is Tg-DT predictive of (2-[18F]FDG) PET/CT results in patients with DTC?**

Four studies were used to determine hierarchical summary receiver-operating characteristics curve (HSROC) (5, 6, 11, 18). The Area Under the Curve (AUC) was 0.86 [95%CI: 0.83; 0.89] (Figure 2). The summary estimates of sensitivity and specificity was 0.84 [95%CI: 0.64; 0.94], 0.71 [95%CI: 0.35; 0.92] respectively. The pooled estimates of L+, L-, and DOR were 2.9 [95%CI: 1.0; 8.1], 0.22 [95%CI: 0.1; 0.5], and 13.1 [95%CI: 3.1; 55.0].

**Is Tg-DT predictive of survival outcomes in patients with DTC?**

Seven studies were included in the principal comparative analysis about the survival risk ratio between patients with Tg-DT < 1 year and those with Tg-DT ≥ 1 year (4, 11-14, 17).
For patients with Tg-DT < 1 year (n = 247), the risk ratio was 2.09 [95%CI: 1.49; 2.94] (test of $\theta$: $z = 4.25, p < 0.001$; test of homogeneity: $Q = \chi^2 (6) = 8.06, p = 0.23$; $I^2 = 25.6\%$) (Figure 3). L’Abbè Plot (Supplementary Figure 1) showed that Miyauchi et al. (2011) and Rössing et al. (2016) were outliers. After outliers exclusion, the risk ratio for patients (n = 159) with Tg-DT < 1 year was 1.93 [95%CI: 1.47; 2.54] (test of $\theta$: $z = 4.70, p < 0.001$; test of homogeneity: $Q = \chi^2 (6) = 4.03, p = 0.40$; $I^2 = 0.83\%$) (Figure 4). The risk ratio increased in patients with papillary thyroid cancer and Tg-DT < 1 year (n = 127) [2.38 (95%CI: 1.71; 3.30; test of $\theta$: $z = 5.16, p < 0.001$; test of homogeneity: $Q = \chi^2 (4) = 2.49, p = 0.65$; $I^2 = 0.00\%$] (Figure 5). Risk ratio of patients with Tg-DT < 1 year showed not significantly changes after subgroup analysis related to sample size (Supplementary Figures 2 and 3). The risk ratio for patients with Tg-DT < 1 year determined with at least four Tg measurements (n = 122) was 2.35 [95%CI: 1.66; 3.33] (test of $\theta$: $z = 4.82, p < 0.001$; test of homogeneity: $Q = \chi^2 (3) = 2.45, p = 0.48$; $I^2 = 0.00\%$) (Supplementary Figure 4). The sensitivity analysis showed no relevant difference in the case of different meta-analysis approaches (random vs. fixed) or change in effect size measurements (RR vs. OR) (Supplementary Table 3). Meta-analysis power was 82.9 (95%CI: 80.4; 85.2) (number of patients in each group: 54; level of significant = 0.05; number of simulations the power calculation is based on = 1000) (Supplementary Figure 5).

**Discussion**

Tumor marker doubling time (DT) has been a predictive marker for outcomes in patients with various types of cancer, as prostate cancer (DT of prostate-specific antigen) (19-23) or medullary thyroid cancer (DT of calcitonin and carcinoembryonic antigen) (24). The first analysis dealing with calcitonin-DT, published in the early 80s (25), showed that calcitonin-DT highly correlated with life expectancy and tumor recurrence.
As the main result of our study, a Tg-DT < 1 year robustly predicted the response to treatment ("treatment monitoring") and the patients’ outcome over time ("prediction"). In addition, Tg-DT values also emerged as reliable predictors of (2-[18F]FDG) PET/CT results. In particular, patients with Tg-DT < 1 year carry a higher risk of having an incomplete response to treatment and reduced survival than patients with longer Tg-DT. Among patients with negative radioiodine WBS, those with shorter Tg-DT values are more likely to have positive (2-[18F]FDG) PET/CT results that, in turn, confers a higher risk of reduced survival. The idea of a strict relationship between Tg kinetics and (2-[18F]FDG) PET/CT results is coherent and likely dependent on the aggressive behaviour and high growing tumor rate of (2-[18F]FDG)-avid DTC, usually associated with worse prognosis. Several potential sources of between-study heterogeneity and uncertainty should be considered in this study. First, patients’ inclusion criteria varied significantly among the included studies. Rössing et al. (2016) only included patients with proved structural persistent or relapsing disease after radioiodine therapy; Miyauchi et al. (2011) included a more heterogeneous series of DTC patients including patients without radioiodine therapy. In Rossing et al. (2016), the mortality risk of patients with a Tg-DT <5 months was more than twice as high as the mortality risk of patients with a Tg-DT of >14 months. Multivariate analysis, including the covariates of the prevalence of distant or locoregional metastases, Tg blood level, and degree of radioiodine accumulation, confirmed a doubling of mortality risk of patients with a Tg-DT <5 months compared to a Tg-DT of >14 months. However, it was not possible to confirm Tg-DT as an independent predictor for survival rate for all patients with progressive disease. However, when focusing on patients with a Tg level (geometric mean) >100 ng/mL representing a high tumor load, significant differences were found in survival rates when the Tg-DT was classified into the three groups < 3 months, 3–12 months, and >12 months (p < 0.05). This is not surprising as patients with lower Tg levels are likely those with reduced tumor burden and increased
radioiodine avidity, increasing the weight of such covariates. However, in line with other included studies, a Tg-DT of 14 months or more remained a robust favorable prognostic predictor even in patients with high tumor load.

Second, follow-up periods ranged from 157 days to 11.1 years in different studies. A unique feature of survival data is that not all patients experience the event (i.e., recurrence, death) by the end of the observation period (i.e., censoring phenomenon). This is more likely when a short follow-up period is available. Consequently, fewer events are expected in studies with shorter follow-up than those with more extended observation periods (26). On the other hand, a reduced Tg-DT predicted an increased risk of major events in patients only followed for short periods. Accordingly, a significant impact of follow-up length on our results seems unlikely.

Third, a different number of Tg measurements (from two to five) and mathematical methods were adopted to calculate Tg-DT values in included studies. The accuracy of the determined Tg-DT increases with the number of available Tg measurements. Rössing et al. (2016) showed in their study that the mean estimated error of Tg-DT was inversely correlated with the number of Tg measurements, being 21% for three measurements and decreasing to 8% for seven or more measurements. But even if different timing and measurement points may have influenced Tg kinetic evaluation, a Tg-DT < 1 year was homogeneously related to an increased risk of major events and positive (2-[18F]FDG) PET/CT results. Moreover, no major differences between mathematical methods to determine kinetic parameters were reported (3).

Fourth, different Tg and TgAb assays were employed, and inter-assays variability results should be considered in comparing studies. However, if the same assays are used in individual patients, the Tg-DT values remain comparable as opposed to the absolute Tg values (27, 28). Finally, it should be noted that an increase in Tg does not always reflect tumor progression:
pitfalls are, for example, therapeutically induced re-differentiation effects that can occur both under therapy with retinoids (29, 30) and under specific protein kinase inhibitors (31). In spite of a tumor regress proven by morphological and functional imaging, a Tg increase can occur here (as an expression of increasing differentiation of the tumor cells), and the aforementioned prognostic statements for the Tg-DT do not apply.

Fifth, the degree of TSH might affect Tg levels and, consequently, Tg-DT values. Interestingly, Angell and colleagues previously demonstrated that a significant influence on the Tg measurement is not detectable for TSH values below 0.5 mUI/L (32). As shown in Table 1, TSH below 0.3, 0.2, and 0.1 mUI/L were reported in 781, 682, and 360 patients, respectively. Indeed, TSH levels were not reported by Kelders and colleagues (9 patients) (18), while Iwasaki and colleagues reported non-elevated TSH levels with any additional information (128 patients), respectively (12). Both studies included patients with advanced metastatic disease for whom clinical guidelines recommend TSH suppression. Accordingly, it is reasonable to expect suppressed TSH values in most patients. All in all, a significant confounding effect of TSH levels is unlikely in our analyzed patients populations.

Conclusions

Our results indicate that Tg-DT values are valuable in predicting response to treatment and disease recurrence in patients with DTC, as well as their overall survival. In addition, Tg-DT significantly increases the detection rate of (2-[18F]FDG) PET/CT. According to our findings, Tg-DT values <1 year should alert the thyroid team and prompt more aggressive diagnostic and therapeutic approaches. However, standardization of Tg-DT evaluation (i.e., the timing of Tg measurements, kinetic model, and calculation method) and large clinical multicentre studies are required to define better the role of Tg-DT in this setting.
Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

The present study did not receive any specific grant from any funding agency in the public, commercial, or non-for-profit sector.
References


2. Giovanella L. Circulating biomarkers for the detection of tumor recurrence in the postsurgical follow-up of differentiated thyroid carcinoma Current opinion in oncology 2020 32 7-12.


11. Albano D, Tulchinsky M, Dondi F, Mazzoletti A, Bertagna F, Giubbini R. The role of Tg kinetics in predicting 2-[(18)F]-FDG PET/CT results and overall survival in patients affected by differentiated thyroid carcinoma with detectable Tg and negative 131I-scan Endocrine 2021 74 332-9.


18. Kelders A, Kennes LN, Krohn T, Behrendt FF, Mottaghy FM, Verburg FA. Relationship between positive thyroglobulin doubling time and 18F-FDG PET/CT-positive, 131I-negative lesions Nuclear medicine communications 2014 35 176-81.

20. Daskivich TJ, Regan MM, Oh WK. Prostate specific antigen doubling time calculation: not as easy as 1, 2, 4 The Journal of urology 2006 176 1927-37.


Differentiated Thyroid Cancer Patient with Dabrafenib and Trametinib Treatment

_Angell TE, Spencer CA, Rubino BD, Nicoloff JT, LoPresti JS. In search of an unstimulated thyroglobulin baseline value in low-risk papillary thyroid carcinoma patients not receiving radioactive iodine ablation._ 

_Thyroid_ 2014 **24** 1127-33.
Figure Legends

**Figure 1** – PRISMA Flow-Chart

**Figure 2** – Hierarchical summary receiver-operating characteristics (HSROC) plot for all included studies (n = 4). Area Under the Curve (AUC) = 0.86 (95%CI: 0.83-0.89). Sensitivity: 0.84 (0.64;0.94); Specificity: 0.71 (0.35; 0.92); DOR: 13.1 (3.1; 55.0); LR+: 2.9 (1.0; 8.1); LR-: 0.22 (0.1; 0.5).

**Figure 3** - Forest Plot for all studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year (n = 7). Risk ratio = 2.09 (95%CI 1.49; 2.94). Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.

**Figure 4** - Forest Plot for studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year without outliers (n = 5). Risk ratio = 1.93 (95%CI 1.47; 2.54). Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.

**Figure 5** – Subgroup analysis - Forest Plot for studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year composed mainly of patients with papillary thyroid carcinoma (n = 5). Risk ratio = 2.38 (95%CI 1.71; 3.39). Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.

**Supplementary - Figure Legends**

**Figure 1** – L’Abbè Plot (n = 7). Larger circles indicate more precise studies. Circles on the green line indicate that for those studies, the log risk-ratio is zero. Circles that deviate from the effect-size line greatly could be a sign of study heterogeneity. Studies (Miyauchi et al. 2011 and Rössing et al. 2016) numbers 1 and 2 were reported as outliers.

**Figure 2** – Subgroup analysis - Forest Plot for studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year with at least 50 patients (n = 6). Risk ratio = 2.07 (95%CI 1.40; 3.06).
Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.

Figure 3 – Subgroup analysis - Forest Plot for studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year in studies with at least 100 patients (n = 3). Risk ratio = 2.03 (95%CI 1.27; 3.25). Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.

Figure 4 – Subgroup analysis - Forest Plot for studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year with at least four consecutive Tg measurements. Risk ratio = 2.35 (95%CI 1.66; 3.33). Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.

Figure 5 – Power of the meta-analysis as a Function of the number of studies and heterogeneity. Number of patients for each group: 54 – Three between-studies dispersion: small (0.20), medium (0.30), and large (0.40). With seven studies, power would be around 57%, 89%, and 99% if dispersion was small, medium, or large.
## Table 1 - Characteristics of the included studies (n = 11)

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Country</th>
<th>Patients</th>
<th>Mean age, yr</th>
<th>% Males (male/female)</th>
<th>PTC</th>
<th>FTC</th>
<th>DTC type</th>
<th>Prevalence</th>
<th>Poorly differentiated or Hurthle or Aggressive variant</th>
<th>TSH</th>
<th>Patients included for Tg-DT</th>
<th>Follow-up, yr</th>
<th>Number of Tg measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miyauchi</td>
<td>2011</td>
<td>Japan</td>
<td>426</td>
<td>51.5</td>
<td>18.1 (77/349)</td>
<td>426</td>
<td>0</td>
<td>PTC</td>
<td>0</td>
<td>Suppressed to &lt; 0.1 mIU/L</td>
<td>137</td>
<td>7.3</td>
<td>≥ 4</td>
<td></td>
</tr>
<tr>
<td>Giovanella</td>
<td>2013</td>
<td>Switzerland</td>
<td>102</td>
<td>48.2</td>
<td>24.5 (25/99)</td>
<td>87</td>
<td>37</td>
<td>PTC</td>
<td>0</td>
<td>Suppressed (range &lt;0.01-0.02 mIU/L)</td>
<td>102</td>
<td>3.8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Kelders</td>
<td>2014</td>
<td>Germany</td>
<td>65</td>
<td>58</td>
<td>40% (26/39)</td>
<td>14</td>
<td>29</td>
<td>FTC</td>
<td>2</td>
<td>NR</td>
<td>9</td>
<td>NR</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Rössing</td>
<td>2016</td>
<td>Germany</td>
<td>99</td>
<td>16-77</td>
<td>37.3 (37/62)</td>
<td>33</td>
<td>66</td>
<td>FTC</td>
<td>0</td>
<td>TSH &lt; 0.3 mIU/L</td>
<td>99</td>
<td>4.7</td>
<td>≥ 3</td>
<td></td>
</tr>
<tr>
<td>Wassermann</td>
<td>2016</td>
<td>France</td>
<td>91</td>
<td>&lt; 45 yr: 14 (15%); ≥ 45 yr: 77 (85%)</td>
<td>38% (35/56)</td>
<td>47</td>
<td>29</td>
<td>PTC</td>
<td>15</td>
<td>TSH &lt; 0.2 mIU/L</td>
<td>31</td>
<td>8.7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Verburg</td>
<td>2017</td>
<td>Germany</td>
<td>174</td>
<td>NR</td>
<td>43% (75/99)</td>
<td>102</td>
<td>72</td>
<td>PTC</td>
<td>0</td>
<td>Suppressed</td>
<td>174</td>
<td>8.9</td>
<td>≥ 4</td>
<td></td>
</tr>
<tr>
<td>Manohar</td>
<td>2018</td>
<td>USA</td>
<td>62</td>
<td>63.2</td>
<td>59.6% (37/25)</td>
<td>44</td>
<td>18</td>
<td>PTC</td>
<td>0</td>
<td>Suppressed</td>
<td>50</td>
<td>11.1</td>
<td>≥ 4</td>
<td></td>
</tr>
<tr>
<td>Iwasaki</td>
<td>2019</td>
<td>Japan</td>
<td>147</td>
<td>70.3</td>
<td>41.6% (32/45)</td>
<td>65</td>
<td>12</td>
<td>PTC</td>
<td>0</td>
<td>TSH non-elevated without any further indication</td>
<td>128</td>
<td>5.1</td>
<td>≥ 4</td>
<td></td>
</tr>
<tr>
<td>Zhang</td>
<td>2020</td>
<td>Japan</td>
<td>21</td>
<td>62.5 ± 14.1</td>
<td>47.6 (10/11)</td>
<td>16</td>
<td>5</td>
<td>PTC</td>
<td>0</td>
<td>TSH &lt; 0.1 mIU/L</td>
<td>21</td>
<td>157 days</td>
<td>≥ 4</td>
<td></td>
</tr>
<tr>
<td>Albano</td>
<td>2021</td>
<td>Italy</td>
<td>139</td>
<td>56</td>
<td>54% (75/64)</td>
<td>72</td>
<td>53</td>
<td>FTC</td>
<td>14</td>
<td>Suppressed</td>
<td>139</td>
<td>3.7</td>
<td>≥ 2</td>
<td></td>
</tr>
<tr>
<td>Araz</td>
<td>2021</td>
<td>Turkey</td>
<td>95</td>
<td>52.6</td>
<td>54% (15/13)</td>
<td>4</td>
<td>22</td>
<td>FTC</td>
<td>2</td>
<td>TSH &lt; 0.1 mIU/L</td>
<td>28</td>
<td>7.1</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

PTC: Papillary Thyroid Carcinoma; FTC: Follicular Thyroid Carcinoma – NR: Not Reported; yr: years; Tg: Thyroglobulin
Figure 1 – PRISMA Flow-Chart

Identification of studies via databases and registers

Identification
- Records identified from: Databases (n = 1830)
- Records removed before screening: Duplicate records removed (n = 145)

Screening
- Records screened (n = 1685)
- Records excluded (n = 1656)
- Reports sought for retrieval (n = 29)
- Reports not retrieved (n = 2)
- Reports assessed for eligibility (n = 27)
- Reports excluded:
  - Irrelevant to the main topic (n = 13)
  - No response (n = 1)
  - Repetitive publication (n = 2)

Included
- Studies included in review (n = 11)
- Reports of included studies (n = 11)
Figure 2 – Hierarchical summary receiver-operating characteristics (HSROC) plot for all included studies (n = 4). Area Under the Curve (AUC) = 0.86 (95%CI: 0.83-0.89). Sensitivity: 0.84 (0.64;0.94); Specificity: 0.71 (0.35; 0.92); DOR: 13.1 (3.1; 55.0); LR+: 2.9 (1.0; 8.1); LR-: 0.22 (0.1; 0.5).
Figure 3 - Forest Plot for all studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year (n = 7). Risk ratio = 2.09 (95%CI 1.49; 2.94). Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.
Figure 4 - Forest Plot for studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year without outliers (n = 5). Risk ratio = 1.93 (95% CI 1.47; 2.54). Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.

<table>
<thead>
<tr>
<th>Study</th>
<th>Tg-DT &lt; 1 year</th>
<th></th>
<th>Tg-DT ≥ 1 year</th>
<th></th>
<th>Risk ratio with 95% CI</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Survivors</td>
<td>Survivors</td>
<td>Non-Survivors</td>
<td>Survivors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wasserman, 2016</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>20</td>
<td>2.60 [ 0.95, 7.00]</td>
<td>7.44</td>
</tr>
<tr>
<td>Verburg, 2017</td>
<td>12</td>
<td>11</td>
<td>30</td>
<td>121</td>
<td>2.63 [ 1.58, 4.35]</td>
<td>28.96</td>
</tr>
<tr>
<td>Manazar, 2018</td>
<td>21</td>
<td>13</td>
<td>5</td>
<td>11</td>
<td>1.98 [ 0.91, 4.28]</td>
<td>12.47</td>
</tr>
<tr>
<td>Iwasaki, 2019</td>
<td>15</td>
<td>30</td>
<td>14</td>
<td>69</td>
<td>1.98 [ 1.05, 3.72]</td>
<td>18.65</td>
</tr>
<tr>
<td>Albano, 2021</td>
<td>20</td>
<td>32</td>
<td>25</td>
<td>62</td>
<td>1.34 [ 0.83, 2.16]</td>
<td>32.47</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.93 [ 1.47, 2.54]</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: τ² = 0.00, I² = 0.83%, H² = 1.01

Test of θ = 0; Q(4) = 4.03, p = 0.40
Test of θ = 0; z = 4.70, p = 0.00
Figure 5 – Subgroup analysis - Forest Plot for studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year composed mainly of patients with papillary thyroid carcinoma (n = 5). Risk ratio = 2.38 (95% CI 1.71; 3.39). Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.
Supplementary Materials

Table 1 - Results of the individual components of the quality assessment of the included studies

<table>
<thead>
<tr>
<th>Author</th>
<th>1° criterion</th>
<th>2° criterion</th>
<th>3° criterion</th>
<th>4° criterion</th>
<th>5° criterion</th>
<th>6° criterion</th>
<th>7° criterion</th>
<th>8° criterion</th>
<th>9° criterion</th>
<th>10° criterion</th>
<th>11° criterion</th>
<th>12° criterion</th>
<th>13° criterion</th>
<th>14° criterion</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miyauchi et al., 2011</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>NA</td>
<td>Fair</td>
</tr>
<tr>
<td>Giovanella et al., 2013</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>NA</td>
</tr>
<tr>
<td>Kelders et al., 2014</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>NA</td>
</tr>
<tr>
<td>Rössing et al., 2016</td>
<td>Yes</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>NA</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Wassermann et al., 2016</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>NA</td>
</tr>
<tr>
<td>Verburg et al., 2017</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>NA</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Manohar et al., 2018</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>NA</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Iwasaki et al., 2019</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>NA</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Zhang et al. 2020</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>NA</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Albano et al., 2021</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>NA</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Araz et al., 2021</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>Yes</td>
<td>CD</td>
<td>No</td>
<td>NA</td>
</tr>
</tbody>
</table>

Legends: Criteria: 1) Was the research question or objective in this paper clearly stated? 2) Was the study population clearly specified and defined? 3) Was the participation rate of eligible persons at least 50%? 4) Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants? 5) Was a sample size justification, power description, or variance and effect estimates provided? 6) For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured? 7) Was the time frame sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed? 8) For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)? 9) Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants? 10) Was the exposure(s) assessed more than once over time? 11) Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants? 12) Were the outcome assessors blinded to the exposure status of participants? 13) Was loss to follow-up after baseline 20% or less? 14) Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)? – CD: Cannot Determine – NA: Not Applicable
**Supplementary Materials**

**Table 2 – Recurrence or disease progression (n = 5 studies)**

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Patients included for Tg-DT</th>
<th>Outcome</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miyauchi</td>
<td>2011</td>
<td>137</td>
<td>Recurrence</td>
<td>Tg-DT &lt; 1 year: at 5 year: 43.8%, at 10 year: 78.6%; 1 &lt; Tg-DT &lt; 3: at 5 year 23.5%, at 10 year: 72.6%; Tg-DT ≥ 3 years: at 5 year: 23.6%, at 10 year: 42.5%.</td>
</tr>
<tr>
<td>Kelders</td>
<td>2014</td>
<td>9</td>
<td>Disease progression</td>
<td>Tg-DT &lt; 1 year: 8/9 patients</td>
</tr>
<tr>
<td>Verburg</td>
<td>2017</td>
<td>174</td>
<td>Recurrence</td>
<td>Tg-DT &lt; 1 year: 8/174 patients with recurrence</td>
</tr>
<tr>
<td>Iwasaki</td>
<td>2019</td>
<td>128</td>
<td>Disease progression</td>
<td>Tg-DT is not indicative of disease progression</td>
</tr>
<tr>
<td>Zhang</td>
<td>2020</td>
<td>21</td>
<td>Disease progression</td>
<td>Disease progression in 11 patients with Tg-DT &lt; 1 year (RDT)</td>
</tr>
</tbody>
</table>
## Supplementary Material

### Table 3 – Sensitivity Analysis for Tg-DT < 1 year vs. Tg-DT ≥ 1 year in Non-Survivors and Survivors

<table>
<thead>
<tr>
<th></th>
<th>Random</th>
<th>Fixed</th>
<th>RR</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\theta}$</td>
<td>2.09 [1.49; 2.94]</td>
<td>2.09 [1.58; 2.76]</td>
<td>2.09 [1.49; 2.94]</td>
<td>2.93 [1.94; 4.42]</td>
</tr>
<tr>
<td>Test of $\hat{\theta}$: $z = 4.25$, p &lt; 0.001</td>
<td>Test of $\hat{\theta}$: $z = 5.19$, p &lt; 0.001</td>
<td>Test of $\hat{\theta}$: $z = 4.25$, p &lt; 0.001</td>
<td>Test of $\hat{\theta}$: $z = 5.12$, p &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>Test of Q: $\chi^2(6) = 8.06; p = 0.23$</td>
<td>Test of Q: $\chi^2(6) = 8.13; p = 0.23$</td>
<td>Test of Q: $\chi^2(6) = 8.06; p = 0.23$</td>
<td>Test of Q: $\chi^2(6) = 7.08; p = 0.31$</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1 – L’Abbé Plot (n = 7). Larger circles indicate more precise studies. Circles on the green line indicate that for those studies, the log risk-ratio is zero. Circles that deviate from the effect-size line greatly could be a sign of study heterogeneity. Studies (Miyauchi et al. 2011 and Rössing et al. 2016) numbers 1 and 2 were reported as outliers.
**Figure 2** – Subgroup analysis - Forest Plot for studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year with at least 50 patients (n = 6). Risk ratio = 2.07 (95% CI 1.40; 3.06). Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.

<table>
<thead>
<tr>
<th>Study</th>
<th>Tg-DT &lt; 1 year</th>
<th>Tg-DT ≥ 1 year</th>
<th>Risk ratio with 95% CI</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Survivors</td>
<td>Survivors</td>
<td>Non-Survivors</td>
<td>Survivors</td>
</tr>
<tr>
<td>Miyachi, 2011</td>
<td>2</td>
<td>18</td>
<td>1</td>
<td>116</td>
</tr>
<tr>
<td>Rössing, 2016</td>
<td>14</td>
<td>54</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>Verburg, 2017</td>
<td>12</td>
<td>11</td>
<td>30</td>
<td>121</td>
</tr>
<tr>
<td>Manohar, 2018</td>
<td>21</td>
<td>13</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Iwasaki, 2019</td>
<td>15</td>
<td>30</td>
<td>14</td>
<td>69</td>
</tr>
<tr>
<td>Albano, 2021</td>
<td>20</td>
<td>32</td>
<td>25</td>
<td>62</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $t^2 = 0.08, I^2 = 35.79\%, H^2 = 1.56$

Test of $\theta = 0; Q(5) = 7.79, p = 0.17$

Test of $\theta = 0; z = 3.96, p = 0.00$
Figure 3 – Subgroup analysis - Forest Plot for studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year in studies with at least 100 patients (n = 3). Risk ratio = 2.03 (95%CI 1.27; 3.25). Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.

<table>
<thead>
<tr>
<th>Study</th>
<th>Tg-DT &lt; 1 year</th>
<th>Tg-DT ≥ 1 year</th>
<th>Risk ratio with 95% CI</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Survivors</td>
<td>Survivors</td>
<td>Non-Survivors</td>
<td>Survivors</td>
</tr>
<tr>
<td>Miyachi, 2011</td>
<td>2</td>
<td>18</td>
<td>1</td>
<td>116</td>
</tr>
<tr>
<td>Verburg, 2017</td>
<td>12</td>
<td>11</td>
<td>30</td>
<td>121</td>
</tr>
<tr>
<td>Iwasaki, 2019</td>
<td>15</td>
<td>30</td>
<td>14</td>
<td>69</td>
</tr>
<tr>
<td>Albano, 2021</td>
<td>20</td>
<td>32</td>
<td>25</td>
<td>62</td>
</tr>
</tbody>
</table>

Overall: Risk ratio = 2.03 [1.27, 3.25]

Heterogeneity: τ² = 0.11, I² = 49.52%, H² = 1.98
Test of θ = 0: Q(3) = 5.64, p = 0.11
Test of θ = 0: z = 2.94, p = 0.00
Figure 4 – Subgroup analysis - Forest Plot for studies comparing survival for Tg-DT < 1 year vs. Tg-DT ≥ 1 year with at least four consecutive Tg measurements. Risk ratio = 2.35 (95%CI 1.66; 3.33).

Number of survived and non-survived patients according to the Tg-DT values. Random-effects DerSimonian–Laird model.
Figure 5 – Power of the meta-analysis as a Function of the number of studies and heterogeneity.

Number of patients for each group: 54 – Three between-studies dispersion: small (0.20), medium (0.30), and large (0.40). With seven studies, power would be around 57%, 89%, and 99% if dispersion was small, medium, or large.