Follow up Approaches in Thyroid Cancer: A Risk Adapted Paradigm

R. Michael Tuttle, MD\textsuperscript{a,b,*}, Rebecca Leboeuf, MD\textsuperscript{a,b}

\textsuperscript{a}Joan and Sanford I. Weill Medical College of Cornell University, Department of Medicine, 1300 York Avenue, New York, NY 10021, USA
\textsuperscript{b}Memorial Sloan-Kettering Cancer Center, Zuckerman Research Building, 1275 York Avenue, New York, NY 10021, USA

Data from the Surveillance, Epidemiology, and End Results Program predicts that one out of every 136 people born in the United States today will be diagnosed with thyroid cancer at some point during their lifetime [1]. An estimated 33,550 cases of thyroid cancer will be diagnosed in 2007 (8,070 men and 25,480 women). While 30-year disease specific survival rates can exceed 95\% [2,3], the 5-year survival rate in older patients presenting with distant metastatic disease can be as low as 56\% [4]. Because most patients with thyroid cancer respond very well to treatment, it is not surprising that an estimated 366,466 thyroid cancer survivors are living in the United States as of January 2004 [5].

The primary goal of follow-up in these thyroid cancer survivors is to identify and appropriately treat the nearly 30\% of patients who may experience a clinically significant recurrence (which may occur as long as 20 to 30 years after initial therapy), based on the assumption that early detection and treatment of recurrent disease lowers morbidity and prolongs life [3]. As the paradigm for disease detection in low-risk patients moves away from routine diagnostic whole-body radioactive iodine (RAI) scanning and toward a greater reliance on neck ultrasonography and serum thyroglobulin (Tg) determinations as primary tools for disease detection, it is critical to understand both the strengths and limitations of this new follow-up paradigm.

Recently, several international authorities on thyroid cancer have published guidelines outlining generally similar approaches to the treatment and follow-up of thyroid cancer patients [6–11]. In each of these guidelines,
estimates of risk for recurrence and risk of disease-specific death are used to guide both initial treatment and follow-up recommendations.

By combining an understanding of the biology of the disease (response to initial therapy, likelihood of recurrence, common sites of recurrence, time to recurrence) with an improved understanding of the clinical utility of the specific test (sensitivity, specificity, and negative predictive value), we as health care providers can tailor the extent and intensity of our follow-up recommendations to the specific risk of disease-specific death and clinically evident recurrence in an individual patient. Therefore, the goal of this article is to describe a risk-adapted framework that will guide the practicing clinician in the selection and timing of appropriate follow-up tests for individual patients. The article will focus on the follow-up of patients with follicular-derived thyroid cancers (papillary and follicular thyroid cancers), leaving other rarer forms of thyroid cancer (medullary, anaplastic, lymphoma) for other investigators and articles.

Risk stratification

To begin to determine which tests would be appropriate for detection of disease-specific death or recurrent disease, it is critical to be able to appropriately risk stratify patients shortly after initial diagnosis, based on the likelihood of death or recurrence of thyroid cancer. Thereafter, just as importantly, one has to define likely sites of recurrence, as well as the likelihood that the recurrence will be RAI avid (or fluorodeoxyglucose [FDG] avid) and produce sufficient quantities of Tg to allow detection of recurrent disease. Fortunately, several different risk stratification systems have been published, each being very efficient at defining patients at high or low risk for disease-specific mortality (reviewed in Tuttle and colleagues[12]). Unfortunately, even in the best of these systems, there continues to be a small risk of death from disease in patients classified as low risk, warranting long-term follow-up. Several tumor-related and patient-related factors are commonly found in multivariate analysis to be predictors of disease-specific death in thyroid cancer: age at diagnosis, size of the primary tumor, histology of the tumor, locally invasive thyroid cancer (completeness of resection), and presence of distant metastases[12]. Lymph node involvement has been reported in some studies to increase disease recurrence without having an impact on disease-specific survival[13], while other studies demonstrate an increase in disease specific mortality in patients with lymph node involvement[3].

Each of these staging systems provides good risk stratification based on data available shortly after initial therapy. The authors refer to this as “initial risk stratification” and it certainly is a good starting point for initial decision making. However, in the follow-up, a patient’s additional data is accumulated that may either increase or decrease this initial risk estimate. The authors refer to the integration of the predictive values of these
additional data into the initial risk estimates as “ongoing risk stratification.” Over the years, the ongoing risk stratification provides important additional information that may significantly alter the initial assessment of the risk of recurrence, risk of death, and likely sites of recurrences.

**Initial risk stratification**

*Initial estimates of the risk of death from thyroid cancer*

The American Joint Committee on Cancer (AJCC) staging system, while commonly used throughout oncology, is designed to risk stratify for death, but not necessarily for recurrence [14]. So while AJCC staging is useful in identifying patients likely to die of thyroid cancer, it is not particularly helpful in predicting recurrence. Therefore, the addition of a postoperative clinicopathologic staging system should be used in conjunction with the AJCC staging system to improve prediction of risk for recurrence and thereby inform the authors’ follow-up testing paradigm [7]. While a wide variety of staging systems are useful, the MACIS (Metastasis, Age, Completeness of resection, Invasion, Size of primary tumor) scoring system has proven to be predictive of disease-specific death and is readily applied based on the usual clinical information available to the clinician, and therefore is most likely to be informative to clinicians caring for patients with thyroid cancer [15].

In the authors’ experience, even without a mathematical scoring system, several readily available clinical factors quickly stratify patients into one of four groups, based on the likelihood of dying from thyroid cancer (very low, low, intermediate, high) (Table 1). This table is based upon the variables repeatedly shown to be significant in multivariate analysis predicting death from thyroid cancer. The intermediate risk group comprises patients in whom the patient factors and tumor factors present a combination of high-risk and low-risk features. While it is clear that these intermediate risk patients are neither truly high risk nor low risk, the precise risk of death is hard to quantify for an individual patient. Therefore, the authors classify them as an intermediate group that deserves concern greater than the low-risk group, but certainly does not have the death rate associated with patients with all of the high-risk features. So while no risk stratification system can identify a cohort of patients with a 0% risk of death from disease, it can separate patients into high, intermediate, low, and very low risk of death from disease, and therefore should have a major impact on both the extent and intensity of the authors’ proposed follow-up and therapies.

*Initial estimates of the risk of clinically evident recurrence*

Unlike most other solid tumors, the risk of recurrence and the risk of death from thyroid cancer are not always concordant [3]. For example, young children with thyroid cancer have a very high recurrence rate (often exceeding 30%–40%) but 30-year disease specific death rates that are less
<table>
<thead>
<tr>
<th></th>
<th>Very low risk</th>
<th>Low risk</th>
<th>Intermediate risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis</td>
<td>&lt;45 years</td>
<td>&lt;45 years</td>
<td>Young patients (&lt;45 yrs)</td>
<td>&gt;45 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Classic PTC &gt; 4 cm</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Or vascular invasion</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Or microscopic extrathyroidal extension</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Or worrisome histology of any size&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Primary tumor size</td>
<td>&lt;1 cm</td>
<td>1 cm–4 cm</td>
<td>Older patients (&gt;45 yrs)</td>
<td>&gt;4-cm classic PTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Classic PTC &lt; 4 cm</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Or worrisome histology</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;1 cm–2 cm confined to the thyroid&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Histology</td>
<td>Classic PTC, confined to the thyroid gland&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Classic PTC, confined to the thyroid gland&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Histology in conjunction with age as above</td>
<td>Worrisome histology &gt; 1 cm–2 cm&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Completeness of resection</td>
<td>Complete resection</td>
<td>Complete resection</td>
<td>Complete resection</td>
<td>Incomplete tumor resection</td>
</tr>
<tr>
<td>Lymph node involvement</td>
<td>None apparent</td>
<td>Present or absent&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Present or absent&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Present or absent&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Distant metastasis</td>
<td>None apparent</td>
<td>None apparent</td>
<td>None apparent</td>
<td>Present</td>
</tr>
</tbody>
</table>

Only those patients meeting all criteria within the respect column would be classified as very low risk, or low risk. Older patients with either incomplete tumor resection or presence of distant metastasis are considered high risk, irrespective of tumor size and specific histology. Patients with a combination of risk factors (age, histology, and tumor size) crossing over between columns are classified as intermediate risk patients.

<sup>a</sup> Worrisome histologies includes histologic subtypes of PTC, such as tall cell variant, columnar variant, insular variant, and poorly differentiated thyroid cancers.

<sup>b</sup> Confined to the thyroid gland with no evidence of vascular invasion or extrathyroidal extension.

<sup>c</sup> Cervical lymph node metastases in older patients, but probably not in younger patients, may confer an increased risk of death from disease.
than 1% to 2%. However, older patients (more than 60 years old at diagnosis) have a high recurrence rate that is associated with an increase in disease-specific death. These differences likely reflect both the different biologies of papillary thyroid cancer (PTC) in young patients (classic PTC, RAI avid) compared with older patients (often more poorly differentiated tumors, RAI refractory), but also a difference in response to therapy (older patients more unlikely to be made disease free with additional surgery or RAI compared with younger patients). While the overall goal of follow-up in thyroid cancer survivors is prevention of disease-specific death, it is critical that clinically significant recurrence be identified and appropriately treated. It is hoped that early detection and therapy of disease recurrence will prevent disease-specific death.

While the risk factors that predict recurrence are often the same as those that predict an increased risk of death from thyroid cancer, factors—such as presence of metastatic disease at diagnosis or incomplete surgical resection—are more difficult to apply to “disease recurrence” strategies, as these patients usually have persistent disease (despite initial therapy) and are dealt with as patients at very high risk of persistent disease, progressive disease, or new metastatic lesions. Therefore, these patients are excluded from the following follow-up paradigm because the intensity and tools for follow-up need to be formulated on an individual patient basis in patients with incomplete surgical resection or distant metastatic disease present at diagnosis.

The risk factors for development of recurrent disease in those patients who have had complete resection of the initial tumor and no evidence of distant metastases at initial evaluation are presented in Table 2. Age at diagnosis continues to be an important predictive factor, with the highest risk of recurrence in the very young and the elderly. Similarly, size of the primary tumor is also linked to risk of recurrence in most studies, as are the histology of the primary tumor and the presence of gross extrathyroidal extension at the time of initial surgery. Most, but not all, studies demonstrate that the presence of identifiable cervical lymph node metastases at diagnosis carries a higher risk of subsequent identification of recurrent disease.

Over the last several years, pathology reports are providing more detailed information regarding the subtypes of thyroid cancer, the presence of microscopic multifocal disease, and the presence of microscopic extrathyroidal extension. While these features likely are indicative of a more aggressive thyroid cancer, it remains unclear whether the presence or absence of these risk factors should convey a significant risk of recurrence independent of size of the primary tumor, the age of the patient, and the completeness of resection. To be conservative, the authors have used these factors to classify as moderate risk, but additional studies are needed to make sure that the authors are not unnecessarily up-staging these patients, particularly when the primary tumor is quite small.
Ongoing risk stratification

Ongoing estimates of the impact of therapy (treatment variable)

Though each of the published staging systems provides very good risk stratification based on information obtainable within the first few weeks (or months) after initial therapy, none of the staging systems include variables that reflect the effectiveness of initial therapy, the negative predictive values of follow-up evaluations, or the impact of the length of time a patient has been free of disease after initial therapy. In clinical practice, the authors incorporate all of these additional factors into an ongoing reassessment of risk for both recurrence and death from thyroid cancer as time passes. In a simplistic fashion, an individual patient response to therapy can be classified as excellent, acceptable, or incomplete (Table 3).

Table 2
Risk stratification for the likelihood of clinically evident recurrence from thyroid cancer following complete resection of primary tumor in patients with no evidence of distant metastases at initial evaluation

<table>
<thead>
<tr>
<th></th>
<th>Low risk</th>
<th>Intermediate risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at diagnosis</strong></td>
<td>Any age</td>
<td>20–60 years</td>
<td>&lt;20 or &gt;60 years</td>
</tr>
<tr>
<td><strong>Primary tumor size</strong></td>
<td>&lt;1 cm&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1 cm–4 cm</td>
<td>&gt;4 cm</td>
</tr>
<tr>
<td><strong>Histology</strong></td>
<td>Classic PTC,</td>
<td>Classic PTC,</td>
<td>Other than classic</td>
</tr>
<tr>
<td></td>
<td>confined to the</td>
<td>minor extrathyroidal</td>
<td>PTC, gross</td>
</tr>
<tr>
<td></td>
<td>thyroid gland</td>
<td>extension, or</td>
<td>extra-thyroidal extension or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>vascular invasion</td>
<td>vascular invasion</td>
</tr>
<tr>
<td><strong>Lymph node involvement</strong></td>
<td>None apparent</td>
<td>Present or absent</td>
<td>Present</td>
</tr>
</tbody>
</table>

Patients with incomplete tumor resection or distant metastasis at diagnosis are very likely to have persistent disease even after aggressive initial therapy, and therefore are dealt with differently than the more usual patient without evidence of distant metastasis, in which all gross evidence of disease has been resected and are therefore not included in this risk stratification scheme.

<sup>a</sup> Classic PTC less than 1 cm with microscopic extrathyroidal extension, vascular invasion, or lymph node metastases would be considered intermediate risk.

Ongoing risk stratification

Ongoing estimates of the impact of therapy (treatment variable)

Though each of the published staging systems provides very good risk stratification based on information obtainable within the first few weeks (or months) after initial therapy, none of the staging systems include variables that reflect the effectiveness of initial therapy, the negative predictive values of follow-up evaluations, or the impact of the length of time a patient has been free of disease after initial therapy. In clinical practice, the authors incorporate all of these additional factors into an ongoing reassessment of risk for both recurrence and death from thyroid cancer as time passes. In a simplistic fashion, an individual patient response to therapy can be classified as excellent, acceptable, or incomplete (Table 3).

Ideally, after initial therapy patients are rendered free of disease and, therefore, all follow-up studies will demonstrate no evidence of disease, in which case the patient would be classified as having an excellent response to initial therapy and be very unlikely to die from thyroid cancer or even develop clinically significant recurrence. However, an incomplete response to initial therapy that is manifest by early development of clinically significant recurrent disease or markedly elevated (or rising) serum Tg after initial therapy would place the patient at significant risk for continued disease progression.

In actual clinical practice, the authors’ follow-up studies often are not absolutely definitive as to the presence or absence of low level disease. For example, very low level Tg (on suppression or stimulation) could represent either residual normal thyroid tissue or low level thyroid cancer. Millimeter-sized abnormal lymph nodes in the neck found during the follow-up
Table 3
Response to therapy variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Excellent response&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Acceptable response</th>
<th>Incomplete response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suppressed Tg&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Undetectable</td>
<td>Detectable but &lt;1 ng/mL</td>
<td>&gt;1 ng/mL</td>
</tr>
<tr>
<td>Stimulated Tg&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Undetectable</td>
<td>&lt;10 ng/mL</td>
<td>&gt;10 ng/mL</td>
</tr>
<tr>
<td>Trend in suppressed Tg&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Remains undetectable</td>
<td>Declining</td>
<td>Stable or rising</td>
</tr>
<tr>
<td>Anti-Tg antibodies</td>
<td>Absent</td>
<td>Absent or declining</td>
<td>Persistent or rising</td>
</tr>
<tr>
<td>Neck ultrasonography</td>
<td>No evidence of disease</td>
<td>Non-specific changes in thyroid bed;</td>
<td>Evidence of structurally significant recurrent or persistent disease in the thyroid bed (&gt;1 cm);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probable inflammatory lymph nodes; Stable millimeter sized cervical lymph node even if abnormal by ultrasound criteria</td>
<td>Cervical lymph nodes (&gt;1 cm), or distant metastases, particularly if structurally progressive or FDG avid</td>
</tr>
<tr>
<td>Diagnostic RAI whole body scan&lt;sup&gt;d&lt;/sup&gt;</td>
<td>No evidence for RAI-avid disease</td>
<td>No evidence for RAI-avid disease; Very faint uptake in thyroid bed only</td>
<td>Persistent or recurrent RAI-avid disease present</td>
</tr>
<tr>
<td>Cross sectional imaging (MRI, CT)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>No evidence of disease</td>
<td>Nonspecific changes</td>
<td>Structural disease present</td>
</tr>
<tr>
<td>FDG positron emission tomography (PET) scanning&lt;sup&gt;d&lt;/sup&gt;</td>
<td>No evidence of disease</td>
<td>Nonspecific changes consistent with normal variants or inflammatory changes</td>
<td>FDG-avid disease present</td>
</tr>
</tbody>
</table>

<sup>a</sup> Patients deemed to have an excellent or acceptable response to therapy generally warrant observation without additional specific therapy, while patients with an incomplete response are likely to require additional evaluation and treatment.

<sup>b</sup> Stimulated and suppressed Tg value cut-offs optimized for patients treated with total thyroidectomy and RAI remnant ablation.

<sup>c</sup> While most sensitive and specific in patients with total thyroidectomy and RAI remnant ablation, a rising Tg over time should also prompt further evaluation in patients treated with less than total thyroidectomy, or with total thyroidectomy without RAI remnant ablation. This highlights the crucial importance of measuring serum Tg in the same laboratory, to ensure comparability among the samples over time.

<sup>d</sup> While these studies are not routinely recommended for all patients without additional high-risk features or clinical suspicion of persistent or recurrent disease, results from these studies can be used as additional response to therapy measures, if done.
of many thyroid cancer patients could represent persistent low level disease or simply benign inflammatory lymph nodes. Similarly, FDG-PET scanning often cannot differentiate small inflammatory lymph nodes from small volume thyroid cancer. Even though the clinician cannot be 100% certain these patients have had an excellent response to therapy, at the worst they have low-level persistent disease that may or may not ever become clinically evident. Therefore, in the absence of other clinical indications, these indeterminate results can be considered an acceptable outcome that warrants continued follow-up, reserving additional treatments for evidence of disease progression. Most patients in this acceptable category will be there by virtue of suppressed Tg values detectable but less than 1 ng/mL or stimulated Tg values less than 10 ng/mL, usually with a neck ultrasonography that is either negative or nonspecific. In this setting, serum Tg values will often decline over time with levothyroxine suppression and observation alone [16], further supporting a cautious observation approach rather than an aggressive therapy.

Even though the acceptable response category probably includes many patients who will never develop clinically significant disease recurrence, common sense demands a closer follow-up of these patients, with low level Tg values or nonspecific imaging studies to identify the small number of patients in this category that will develop rising serum Tg values or increasing size of small metastatic lymph nodes that may require additional therapy.

When estimating the response to therapy, it is critical that sufficient time be allowed to pass to see both the effectiveness of initial therapy and to identify early clinical recurrences that may become manifest in the first 2 to 3 years after initial therapy. For example, serum Tg levels often fall for several years after RAI ablation in the absence of identifiable structural disease, which would result in down-staging of the patient from an “acceptable” response to an “excellent” response over the years. Conversely, small lymph nodes in the neck that are destined to develop into clinically significant recurrences may not be recognized for several years after initial therapy. Development of new abnormal cervical lymph nodes would result in up-staging the patient to an incomplete response and call for more careful observation and possibly additional therapy. Therefore, it is important to consider both the disease-free survival interval and the initial risk stratification when considering when response to therapy classification is most likely to yield reliable long-term prognostic information.

Influence of disease-free survival on risk of recurrence

Even though clinically evident disease recurrence can develop as many as 30 to 40 years after initial therapy, large retrospective studies consistently demonstrate that the majority of recurrences were detected in the first 10 to 15 years of follow-up [2,3]. Although not yet proven, it is likely that the increased sensitivity of the follow-up testing paradigm that the authors are currently using will identify most recurrences (or persistent disease)
sooner than the techniques used in these older studies, resulting in higher rates of recurrence or persistent disease in the first 5 years after therapy, with low rates thereafter. Because thyroid cancer is a slow, progressive disease that plays out over 30 to 40 years, all patients require lifelong follow-up to identify those few patients who develop clinically significant recurrences years after their initial therapy. However, it seems likely that serum Tg on suppression will be an adequate screen for detection of late recurrences in patients who have had appropriate initial therapy with risk specific follow-up in the first few years after diagnosis.

Until long-term studies are completed, demonstrating the negative predictive values of undetectable Tg values and neck ultrasonography based on risk stratification and response to therapy at specific time points after initial therapy, the authors will be conservative in the approach to integrating disease-free survival into the ongoing re-risk stratification paradigm. It seems reasonable to use a 2 to 10 year disease-free survival period (depending on initial risk stratification) to identify those patients who are at lowest risk for development of clinically significant thyroid cancer recurrence. When considered as a continuous variable, the longer the disease-free survival period (with appropriate follow-up studies), the lower the risk of clinically significant recurrence.

It is important to emphasize that disease-free survival is defined not only by time since initial therapy, but also the use of appropriate follow-up testing (usually suppressed Tg with or without neck ultrasound). The authors are often referred patients 25 to 30 years after initial therapy for “recurrent disease” who have large volume metastatic disease with serum Tg values on suppression that exceed 1,000 ng/mL. In some retrospective studies, these patients would be considered to have recurrent disease 25 to 30 years after initial therapy, when in reality they likely have had persistent disease for several years. Therefore, the authors are only confident that patients have had prolonged disease-free survival and are at low risk of recurrence if there are adequate follow-up studies on the patient in addition to prolonged survival after initial therapy.

**Influence of follow-up testing on risk of recurrence or death**

In the follow-up of thyroid cancer patients, the focus is often on tests that have very high sensitivity for disease detection. And while this is certainly important, it is also important to understand the positive-predictive value and negative-predictive value of the various tests that are used to identify recurrent disease. A test result with a high negative-predictive value will lower the assessment of the risk for recurrence in an individual patient, while a test result with a high positive-predictive value would demand continued close follow-up and further evaluation.

One of the difficulties in assessing the predictive value of specific tests is that the test will perform differently based on the pretest probability of disease in an individual patient. For example, a normal neck ultrasound in
a high-risk patient with a suppressed serum Tg of 800 ng/mL would not significantly lower the estimate of risk for recurrence. However, a normal neck ultrasound in a low-risk patient with an undetectable Tg on suppression has a negative-predictive value that exceeds 95%, and is therefore very reassuring [17]. Each of the published guidelines is similar in regard to the follow-up studies to be done in the first few years after initial therapy [6–10]. These generally include serum Tg on suppression every 6 months and neck ultrasound at 6 months to 12 months, and then yearly for 3 to 5 years based on the risk of the patient. Stimulated Tg values and diagnostic whole body scanning are used more selectively in patients at a high rather than a low risk of recurrence treated with total thyroidectomy and RAI ablation. Additional cross-sectional imaging and FDG-PET scanning are reserved for higher risk patients or clinical indications of non-RAI persistent disease. The similarities and differences in the published guidelines have been previously reviewed [12].

Therefore, while precise predictive values of individual tests in specific patients are hard to define, negative test results are generally reassuring and should lower the risk for development of recurrent disease. Because each of these tests are also used as a response-to-therapy variable in Table 3, they play a critical role in the assessment of how effective the initial therapy was, and therefore should reflect the risk of disease recurrence and death.

The primary follow up test predicting disease specific death from thyroid cancer is FDG-PET scanning [18]. Because FDG-PET scanning identifies metabolically active, non-RAI avid thyroid cancers, it is not surprising that patients with markedly positive FDG-PET scans have a poor prognosis, both in terms of progression of disease but also in terms of disease specific mortality. FDG-avid lesions seldom respond to RAI therapy [19], and as such FDG-PET scanning can be used as an adverse response to therapy variable.

Influence of initial therapeutic choices on estimating the impact of therapy

Obviously, the Tg values presented as excellent, acceptable, and incomplete are most sensitive and specific when patients have undergone total thyroidectomy and radioactive iodine ablation. However, by definition, less than total thyroidectomy is only recommended for low-risk patients [20], and as such intensive follow-up and highly sensitive and specific disease detection techniques are not required. In patients treated with less than total thyroidectomy, serial determinations of Tg—without thyroid stimulating hormone (TSH) stimulation—and serial neck ultrasound evaluating both the contralateral residual thyroid tissue and cervical lymph nodes would routinely be recommended. Excellent response to therapy would be a stable Tg over time (with similar TSH levels) and no evidence of structural disease in the contralateral lobe or neck lymph nodes. An example of an acceptable response could be a millimeter-sized colloid nodule in the contralateral lobe, or nonspecific
lymph nodes on neck ultrasound. Obviously, the threshold for a complete thy-
roidectomy is low and would include development of significant abnormali-
ties in the contralateral thyroid lobe, a rising serum Tg, or evidence of
metastatic disease on neck ultrasound or other cross-sectional imaging.

Serum Tg values are also more difficult to interpret in patients treated
with total thyroidectomy without radioactive iodine remnant ablation.
Even though many of these patients will have Tg levels less than 1 ng/mL
on suppression, values as high as 5 ng/mL to 10 ng/mL are likely to repre-
sent residual normal thyroid tissue and not necessarily persistent disease.
Postoperative Tg values will vary with the completeness of the total thyroi-
dectomy and the level of TSH suppression. Precise cut-off values that dif-
ferentiate residual normal thyroid tissue from metastatic thyroid cancer are not
well defined. In the authors’ center, where the surgeons routinely leave less
than 1% to 2% residual normal thyroid, a suppressed serum Tg 1 month
after surgery that exceeds 10 ng/mL to 20 ng/mL raises concern for persis-
tent thyroid cancer and would warrant additional imaging and RAI rem-
nant ablation. Whether or not the patient underwent RAI ablation,
a rising serum Tg on suppression over time with similar levels of TSH sup-
pression would be suspicious for recurrent disease and defined as an incom-
plete response to therapy.

Secondary risk stratification

While continuous, ongoing risk stratification occurs at each follow-up
visit, it seems reasonable to define a point in time where secondary risk strat-
ification can be seen as a gateway to less intensive long-term follow-up con-
sisting of yearly physical examination and suppressed Tg values. While the
exact time point for secondary risk stratification remains to be defined, from
a clinical perspective it seems reasonable to re-evaluate the risk of the pa-
tient based primarily on response to treatment approximately 2 years after
initial therapy.

By 2 years after initial therapy, several additional pieces of data that were
not available at the time of initial risk assessment are now available for in-
corporation into an updated risk classification scheme that incorporates sev-
eral responses to therapy variables (eg, one or two neck ultrasounds, several
suppressed Tg values, sometimes stimulated Tg values, and diagnostic
whole-body follow-up scans). Additionally, serum Tg values often continue
to decline for at least 12 to 18 months after RAI ablation, so a 2 year time
point would allow a reasonable period of time to assess whether the Tg is
rising or falling in a given patient. While a time point this early in the course
of the disease may not accurately define which patients are “cured” of dis-
ease, it can certainly guide the recommendations regarding the types and in-
tensity of follow-up studies that are likely to be cost effective.

It must be emphasized that the secondary risk stratification at 2 years af-
ter initial therapy is designed simply to guide the follow-up paradigm and not
to accurately predict which patients are “cured” of disease. While all thyroid cancer patients are at risk of recurrence, the follow-up paradigm the authors recommend should be based both on the likelihood of recurrence and the likelihood that a minimal disease detection paradigm (suppressed Tg and physical examination) will identify recurrent disease at a time that the disease is still treatable with minimal morbidity and mortality.

From a practical standpoint, additional risk stratification time points could be considered at year 5 for moderate-risk patients, and at year 10 for high-risk patients before being entirely comfortable that a minimal follow-up paradigm of suppressed Tg and physical examination without additional studies is adequate for disease detection in these patients with more than low risk for recurrence. However, it is very likely that patients initially staged as moderate- and high-risk patients that have an excellent response to therapy as determined at the 2-year secondary risk stratification endpoint can be down staged to a lower risk category. Additional long-term studies incorporating both initial stage of disease and status of response to therapy variables at 2 years after diagnosis are needed before the authors can be comfortable down-staging these higher risk patients as early as 2 years after initial therapy.

Clearly, specific follow up recommendations will vary for individual patients, based on initial risk stratification, disease-free survival time, and response to therapy variables, but this method does provide an approach to secondary risk stratification several years after initial therapy and should guide the long-term follow-up approach.

**A risk-adapted paradigm**

From a clinical perspective, the intensity and methods of follow-up surveillance should be based on the individual patient’s risk of recurrence or death from thyroid cancer. This follow-up paradigm is at first based on initial estimates of risk based on data available shortly after diagnosis and initial treatment. The paradigm is then modified as new data becomes available over time, continually adjusting the risk estimates based on the new information obtained: a risk-adapted follow-up paradigm. Based on the length of disease-free survival and the estimate of response to therapy, the intensity and methods of follow-up are adjusted to match the new risk estimates.

Initial risk assessment for death from disease and disease recurrence is done using one of the widely available prognostic staging systems. Patients with gross disease remaining despite aggressive initial therapy, or with distant metastases at diagnosis, may require additional therapy with RAI or external beam radiation therapy and more intensive follow-up with RAI imaging, FDG-PET imaging, and cross-sectional imaging. Because of the wide variety of presentations of these patients, it is difficult to develop a meaningful generalizable follow-up paradigm in this context. In general, cross-sectional
imaging at 6-month intervals is appropriate with additional FDG-PET scanning and RAI scanning, as indicated by the specific clinical conditions.

However, a risk-adapted follow-up scheme is much more applicable to patients who had all evidence of gross disease resected and who have no evidence of distant metastases at diagnosis. In general, the authors expect the treatments to result in a very low disease-specific mortality and a low, but significant, risk of recurrent thyroid cancer. As described above, the follow-up scheme is begun by classifying patients based on the best estimate of risk of recurrence into low, moderate, or high risk (Table 4). In many aspects, the first 2 years of follow-up for all three risk groups is similar, with suppressed Tg values done every 6 months and neck ultrasonography on a yearly basis. Additional cross-sectional imaging, FDG-PET scanning, and routine diagnostic whole body scanning are reserved for high-risk patients. The occasional intermediate risk patient with low level Tg values and indeterminate ultrasonography may also undergo diagnostic whole body scanning looking for RAI-avid disease that may be amenable to a second dose of RAI.

Over this 2-year period, risk stratification is an ongoing process, incorporating additional data into the initial risk assessment and either lowering or raising the estimate of the risk of recurrence based on these follow-up tests. All patients then undergo a formal secondary risk assessment at 2 years after initial therapy. At that point, all data gathered over the last 2 years are reviewed and evaluated to determine the response to initial therapy. Patients with an excellent response to therapy should have a very low risk of recurrence and immediately enter into a long-term, minimal follow-up program of yearly physical examinations and suppressed serum Tg values. Additional neck ultrasonography, stimulated Tg, or other testing modalities are only used if there is some clinical suspicion of recurrent disease or the patient was at very high risk of recurrence in the initial staging evaluation.

At the time of secondary risk stratification, patients often have very low level Tg values, very small abnormal cervical lymph nodes, or nonspecific changes on cross-sectional imaging studies. While these findings would not be considered an excellent response to therapy, they can be considered an acceptable response to therapy that probably is best served with cautious observation. Additional neck ultrasonography is probably warranted in this group of patients on a yearly basis for up to 5 years. At that point, if there has been no structural disease progression and the Tg remains at a stable low level, it seems reasonable to transition to the long-term, minimal follow-up program of yearly physical examination, suppressed serum Tg, and occasional neck ultrasound every few years.

Patients with an incomplete response to therapy manifested by rising serum Tg values, structurally progressive disease, or persistent FDG-PET positive disease are at the highest risk for development of clinically evident recurrent disease and require continued intensive follow-up with neck ultrasonography, cross-sectional imaging, RAI imaging, and FDG-PET imaging.
Table 4
Risk-adapted follow-up paradigm for detecting recurrent disease in patients with complete tumor resection and no evidence of distant metastases at initial risk stratification

<table>
<thead>
<tr>
<th></th>
<th>Low risk</th>
<th>Intermediate risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial estimate of risk of recurrence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>First 2 years of follow-up</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suppressed Tg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Q 6 months</td>
<td>Q 6 months</td>
<td>Q 6 months</td>
</tr>
<tr>
<td>Stimulated Tg&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Not required</td>
<td>1–2 years</td>
<td>1–2 years</td>
</tr>
<tr>
<td>Neck ultrasonography&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Q year × 2</td>
<td>Q year × 2</td>
<td>Q year × 2</td>
</tr>
<tr>
<td>Diagnostic RAI whole body scan&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Not required</td>
<td>1–2 years</td>
<td>1–2 years</td>
</tr>
<tr>
<td>Cross sectional imaging (MRI, CT)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Not required</td>
<td>Not required</td>
<td>If Tg elevated or high clinical suspicion</td>
</tr>
<tr>
<td>FDG-PET scanning&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Not required</td>
<td>Not required</td>
<td>If Tg elevated, RAI scan negative</td>
</tr>
<tr>
<td><strong>Secondary risk stratification</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Response to therapy assessment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>Yearly physical examination, yearly suppressed Tg&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Yearly physical examination, yearly suppressed Tg, stimulated Tg to document undetectable Tg on suppression, continued observation/assessment of indeterminate structural abnormalities for at least another 2–3 years&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Consider additional cross-sectional imaging, possibly FDG-PET scan and the need for additional therapy.</td>
</tr>
<tr>
<td>Acceptable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incomplete</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Patients with stable low level Tg values and stable small structurally abnormal lymph nodes that have been stable for 5 years (and therefore still in the acceptable response category) can transition to yearly follow-up, suppressed Tg values, and occasional neck ultrasonography to document continued structurally stability.

<sup>b</sup> Intermediate or high-risk patients, even in the setting of an excellent treatment response at year 2, may still be periodically require neck ultrasonography, depending on the specifics of each individual case. Additionally, patients in whom the initial therapy was less than total thyroidectomy and RAI ablation may benefit from occasional neck ultrasounds over the next 5 to 10 years, because Tg on suppression is less sensitive for detection of recurrent disease in this setting.
The majority of patients with an incomplete response to therapy will require additional therapy, with appropriate use of surgical resection, RAI therapy, external beam irradiation, and systemic therapies.

While this risk-adapted approach will have better sensitivity and specificity in patients treated with total thyroidectomy and RAI ablation, it can also be readily applied to patients treated with either less than total thyroidectomy or total thyroidectomy without RAI ablation. At most centers, these will be patients initially classified as low or intermediate risk for recurrent disease. The authors follow these patients in a similar fashion to those patients treated with total thyroidectomy and RAI ablation, with a serum Tg on suppression every 6 months and with a neck ultrasound yearly for 2 years. By definition, these are rather low-risk patients, so whole body RAI scanning is not required, and without RAI ablation stimulated Tg values have little meaning and are therefore not done.

Even patients treated with less than total thyroidectomy and RAI ablation should undergo secondary risk stratification at 2 years. The definition of excellent and acceptable response to therapy must be modified because an undetectable Tg, while present in many of them, would not be a requirement for excellent response to therapy. Thus, the authors consider a stable Tg over time, in conjunction with an ultrasound that shows no evidence of recurrent disease, as an excellent response to therapy. Very few patients fall into an “acceptable response to therapy” category because concern by the patients (and clinicians) about even subtle changes in the contralateral lobe or cervical lymph node ultrasonography during the follow-up of these patients often leads to additional therapy. While the authors are comfortable following most of these low- and intermediate-risk patients without additional surgery or RAI, we recognize that many patients and clinicians are not, and would recommend additional diagnostic tests or treatment.

The other difference in the long-term follow-up recommendations in patients treated with less than total thyroidectomy and RAI ablation is that neck ultrasonography is recommended as part of their ongoing follow up for the next 5 to 10 years, usually done every 2 to 3 years, depending on the individual risks factors of each patient. The rationale for this additional testing is that Tg on suppression without total thyroidectomy and RAI ablation is less sensitive for detection of small volume recurrent disease in the contralateral lobe or in cervical lymph nodes in these patients. These recommendations reflect a decreased reliance in the sensitivity and specificity of serum Tg and an increase in the importance of neck ultrasonography in the follow-up of low- to intermediate-risk patients treated with less than total thyroidectomy and RAI ablation.

Summary

After initial therapy to remove all evidence of gross disease, the primary goal of follow-up of thyroid cancer patients is to detect and treat recurrent
disease to minimize morbidity and mortality. The authors’ recommendations regarding a follow-up paradigm should be informed by initial risk stratification, ongoing risk stratification, response to therapy variables, and disease-free survival interval. In this article, the authors provide a framework that begins to integrate each of these factors into the complex decision-making process that is clinical medicine. With appropriate selection of tests, based on a risk-adapted approach, physicians should be able to identify patients with clinically significant disease recurrence in a timely fashion, without subjecting every thyroid cancer survivor to needless, expensive, time-consuming, and worrisome tests that are unlikely to uncover clinically significant disease.

References


FOLLOW UP APPROACHES IN THYROID CANCER


