Differentiated Thyroid Cancer

Reexamination of Risk Groups and Outcome of Treatment

Laura E. Sanders, MD; Blake Cady, MD

Objective: To reexamine the age, metastases, extent, and size (AMES) risk criteria for well-differentiated thyroid cancer with the effect of therapy on outcome.

Design: Review of patient medical records and direct-contact follow-up.

Setting: Two tertiary referral centers.

Main Outcome Measures: Recurrence or death.

Patients: One thousand nineteen patients with well-differentiated thyroid cancer treated between 1940 and 1990.

Results: One thousand nineteen patients with well-differentiated thyroid cancer were treated between 1940 and 1990, with a mean follow-up of 13 years, including a recent group of 264 patients treated from 1980 to 1990 at 2 different institutions with a mean follow-up of 8 years. The AMES criteria were used to designate high- and low-risk patients. The entire group had 229 high- and 790 low-risk patients; the percentage of high-risk patients decreased slightly after 1960. From 1940 to 1960, 1960 to 1979, and 1980 to 1990, the high-risk groups had survival rates of 48%, 62%, and 47%, respectively. For the low-risk patients, survival rates were 96%, 98%, and 98%, respectively. Recurrences occurred in 5% of low-risk patients and were usually curable; in high-risk patients, recurrence was associated with a 75% mortality. In low-risk patients, there was no significant difference in recurrence or death according to type of operation (unilateral or bilateral) or use of radioactive iodine. In high-risk patients, there were trends toward but no significant improvement in survival with bilateral surgery and radioactive iodine therapy; thyroid replacement was associated with a significant improvement in survival.

Conclusions: The AMES risk criteria remain highly valid predictors of risk. They define most low-risk patients for whom radical treatment may add excess morbidity but not improve already excellent prognoses.

Arch Surg. 1998;133:419-425

Well-differentiated thyroid cancer, including papillary and follicular cancers, is a heterogeneous disease: most patients do extremely well, with few having recurrences and even fewer dying of the disease. The optimum treatment of patients with these tumors continues to be debated. Complications from surgery, although rarely fatal, can be a lifelong annoyance and are more common after more aggressive surgery. Tailoring surgery to the patient’s risk of recurrence or death, therefore, is essential to improving patient outcome not only from disease but also from morbidity associated with treatment. By examining the patient’s risk factors for prognosis, the surgeon can guide his or her practice to perform as much treatment as is useful and cause as little harm as possible.

Our objective was to reexamine the age, metastases, extent, and size (AMES) risk factors, described in 1985 as a refinement of earlier age-based risk groupings, with an additional 264 patients treated from 1980 to 1990 at 2 different institutions, the Lahey Hitchcock Medical Center in Burlington, Mass, and the New England Deaconess Hospital in Boston, Mass. We examined the continued validity of the risk factors with this expanded cohort of now 1019 patients treated at the 2 institutions.

RESULTS

There were 1019 patients treated between 1940 and 1990, with a median follow-up of 13 years (range, 0-47 years) overall. The median follow-up for patients treated before 1960 was 19 years; for 1960 to 1979, 15 years; and for 1980 to
MATERIALS AND METHODS

Medical records of 1019 patients initially treated for papillary and follicular thyroid cancer at the Lahey Hitchcock Medical Center and the New England Deaconess Hospital were reviewed. An existing database going back to 1940 was entered into a computer database and supplemented with material from tumor registries, patient medical records, and direct patient contact data from both hospitals. We examined the patients for age and sex; pathologic type, size, and extent of primary cancer (major capsular or blood vessel invasion in patients with follicular cancer and extension outside the thyroid in patients with papillary cancer); and presence of distant metastases. We also examined nodal involvement and its degree. We looked at the characteristics of the tumors during 3 periods: before 1960, 1960 to 1979, and 1980 to 1990. Twenty-year adjusted survival was calculated for all of these factors except for the 1980 to 1990 group, for whom 13-year adjusted survival was used.

We looked at treatment options for the whole group and for each group. Bilateral total thyroidectomy was infrequently performed until recently, bilateral surgery having generally been ipsilateral total with contralateral subtotal, or bilateral subtotal in earlier years. We compared outcomes and complications of bilateral and unilateral surgery. We also looked at node dissection, use of radioactive iodine, and use of thyroid replacement as postoperative treatments. Because this information was not available for patients at the New England Deaconess Hospital, we analyzed only those patients for whom it was known. External beam radiation was used as adjuvant treatment in the early years of this series but was abandoned and therefore was not examined in this analysis.

For purposes of analysis, we grouped the patients into high- and low-risk groups, as described in 1985.1 The low-risk group included women 50 years of age and younger and men 40 years of age and younger without evidence of distant metastases. Older patients were also included in the low-risk group if they had primary tumors less than 5 cm and papillary cancer without gross extrathyroidal invasion or follicular cancer without major capsular invasion and without blood vessel invasion. The high-risk group included all patients with distant metastases and older patients with either tumors greater than 5 cm or extrathyroidal extension of tumor.

Recurrence was defined as reappearance of a tumor thought to have been completely treated. Patients who were initially treated at other than the 2 primary institutions were not included because of incompleteness of data and possible bias toward poorer outcome.

Actuarial curves were generated using life-table analysis. Differences in survival distributions between groups were analyzed using a log-rank test. Multivariate analysis was performed using the Cox proportional hazards regression model.

1990, 8 years. There were 730 women (72%) and 289 men (28%), and the median age of the men (47 years) was older than that of the women (41 years).

Papillary (or mixed papillary follicular) cancers made up 76% of the cases; 20% were follicular and 4% were Hurthle cell. The distribution of follicular cancers did not change during the 3 study periods—before 1960, 1960 to 1979, and after 1980. The median size of the tumors was between 2.0 and 2.9 cm, and although this did not change over time, the percent of tumors greater than 5 cm dropped significantly after 1960 from 15% to 5% (P<.001).

In the entire group, there were 78% low-risk and 22% high-risk patients by the AMES criteria. The group treated before 1960 included the most high-risk patients (27%); there may have been even more high-risk patients in this group who were not detected because tumor size was not recorded in 166 of 432 of the patients treated earliest. Therefore, tumor size could not be used as a marker for higher risk. The 1960 to 1980 group included 18% high-risk patients and the 1980 to 1990 group contained 19% high-risk patients.

The AMES criteria remain highly discriminatory in predicting rates of recurrence and death. In the entire group of patients from 1940 through 1990, the adjusted survival rate at 20 years was 96% for low-risk and 50% for high-risk patients (P=.001). Figure 1, Figure 2, and Figure 3 indicate that this predictive value has held over time. For the group treated before 1960, 20-year survival was 96.5% for low-risk and 48.0% for high-risk patients (P=.001; Figure 1); for the group treated between 1960 and 1979, 20-year survival was 98.4% for low-risk and 62.3% for high-risk patients (P=.001; Figure 2); and for the group treated between 1980 and 1990, 15-year survival was 98.0% for low-risk and 47.0% for high-risk patients (P=.001; Figure 3).

Prediction of recurrence, and of the curability of recurrences, is similarly consistent. In all low-risk patients, risk of recurrence was 5%; among the high-risk patients, 31% of the tumors recurred. Of the 44 low-risk patients who had recurrence, 12 (27%) died of (10 patients) or with (2 patients) disease. Sixteen recurrences were local only and were treated surgically (12 patients), with radioactive iodine (5 patients) or with ex-
ternal beam radiation (5 patients); some patients re-
ceived more than 1 modality. Only 2 (13%) of these
patients died of disease. Sixteen patients had recur-
rences in the local lymph nodes only, and only 1 (6%)
of these patients died of disease. Even among the 12 pa-
tients with distant metastases (mainly in the lungs), 5 were
treated successfully and had long-term survival, leaving
58% dead of disease. Since 7 of the 10 deaths were the
result of lung metastases and only 3 of 32 local or node
recurrences led to deaths, more extensive neck surgery
is therefore unlikely to have improved the survival of low-
risk patients. Two of these local recurrences were in pa-
tients treated in the 1950s with external beam adjuvant
therapy, and their recurrences were 16 and 32 years later,
rising the possibility that the recurrences were a com-
plication of this aggressive treatment.

Among high-risk patients, however, recurrence was
a grave sign. Of the 77 high-risk patients who had re-
currence, 58 (75%) died of disease. Twenty-three had lo-
cal disease only; 16 (70%) died of disease. Twelve pa-
tients had recurrence in nodes only, with 6 (50%) dying
of disease. Of 42 patients with distant metastases, 36 (86%)
died of disease. Risk factors strongly predict tumor be-
havior, but does type of treatment affect outcome? We,
like others,²-⁴ found the effect of treatment to be much
more subtle.

Bilateral surgery did not significantly affect sur-
vival in any group. In comparing all unilateral vs bilat-
eral surgery, at first we found a trend toward better sur-
vival at 20 years. The 23-year adjusted survival rate was
86.8% vs 83.5% for bilateral thyroidectomy compared
with unilateral surgery (P=.58). This figure achieved
significance in the high-risk group at 52.7% vs 44.3%
(P=.037). As Byar et al⁵ suggested, we expected to see
the effect of treatment magnified in high-risk patients
who have a measurable incidence of recurrence and
death. However, when we removed from the unilateral
group patients who underwent biopsy only (evidently a
marker for disease thought to be untreatable), the sig-
nificance disappeared, although there remained a trend
toward better survival up to 15 years (Figure 4). Low-
risk patients did not show any difference at all, with the
survival curves of unilateral vs bilateral surgery super-
imposed (Figure 5).

Bilateral surgery was not associated with a de-
crease in recurrence. In the high-risk group, recurred
ences were 34% with unilateral vs 29% with bilateral sur-
gery (P=.51). For the low-risk group, recurrences were

---

Figure 2. Adjusted survival at 20 years for patients treated from 1960 to
1979.

Figure 3. Adjusted survival at 17 years for patients treated from 1980 to
1990.

Figure 4. Percent survival considering extent of surgery for high-risk
patients.

Figure 5. Percent survival considering extent of surgery for low-risk patients.
identical (5%) with unilateral and bilateral surgery ($P = 1.00$).

In analyzing lymph node involvement, we found that not all patients who underwent lymph node dissections had positive nodes because in the early years lymphadenectomy was performed prophylactically; conversely, not all patients with positive nodes had lymphadenectomy because of the occasional presence of microscopically involved nodes in the surgical specimen. Therefore, we examined lymphadenectomies separately from positive nodes.

Lymph node involvement has previously been shown not to affect survival adversely. $3,5,6$ This remained true in our group, with high-risk patients having a 20-year adjusted survival rate of 58.5% with involved nodes and 45.5% without involved nodes. The 20-year survival rate for low-risk patients was 94.6% with and 96.5% without involved nodes. Neither of these differences was significant.

Node dissection, however, seemed to be associated with improved survival in the high-risk patients alone, and this was highly significant. Thyroidectomy with any type of node dissection was associated with an adjusted 20-year survival rate of 67.0% vs 41.3% in the high-risk group ($P = .001$; Figure 6). Patients who had node dissections vs those who did not were comparable in age, size, and percentage with patients who had all gross disease removed. This difference appeared only in high-risk patients who had papillary cancer. Node dissection made no significant difference in the survival of high-risk patients with follicular cancer.

Node dissection did not improve survival in low-risk patients, who had adjusted 20-year survival rates of 96.5% and 96.0% with and without node dissections, respectively ($P = .57$).

Radioactive iodine, often cited and used as innocuous, was not clearly effective in our patients. High-risk patients had a 20-year adjusted survival rate of 59.0% with vs 52.2% without radioactive iodine therapy, which is a trend, but with a $P$ value of only .86. Low-risk patients had an adjusted survival rate of 99.2% vs 97.2% with and without radioactive iodine therapy, again a trend without statistical significance. With even larger numbers, these trends might achieve significance (although these 2 groups combined still did not have a significant difference in outcome).

Use of thyroid hormones was associated with a small but apparently real improvement in survival: 89.6% vs 85.0% adjusted 20-year survival ($P = .01$). High- and low-risk groups each showed a small trend toward better survival, which combined became significant. Data are not available on the incidence of osteoporosis caused by over-aggressive suppression of thyrotropin.

Complication rates were available only for patients from the Lahey Hitchcock Medical Center. In the 3 study periods, risk of permanent recurrent laryngeal nerve injury was 3.5% before 1960, 1.8% from 1961 to 1980, and 1.3% from 1980 to 1990. These figures do not differentiate deliberate sacrifice from inadvertent injury. Permanent hypoparathyroidism occurred in 3.3%, 2.2%, and 0% of patients in the 3 periods, respectively. Instances of temporary nerve palsy and temporary hypoparathyroidism were not recorded consistently.

Figure 6. Percent survival considering thyroidectomy and node dissection in high-risk patients.

Comment

We know much about predicting the prognosis of differentiated thyroid cancer and much less about how treatment affects outcome. Prospective randomized studies of the effect of therapy on outcome of differentiated thyroid cancer have been thought to be largely impractical because this cancer is associated with a good prognosis in most instances, is relatively uncommon, and can have a long interval to recurrence or death. $2$ Retrospective reviews have provided what information we have to guide treatment.

The recognition that the behavior of well-differentiated thyroid cancer varies enormously, but to a large extent predictably, is a concept that has become widely accepted in the last 20 years. From initial recognition that age was a critical factor, for reasons that could only be speculated on, $1,7-9$ further expansion and refinement of risk groups has taken place. At least 10 scoring systems $10$ now exist, which have been developed to predict patient outcome based on characteristics of the patient and the tumor: AMES from earlier articles $1$; AGES (age, tumor G, extent, and size) and MACIS (metastasis, age, completeness of resection, invasion, and size) from the Mayo Clinic $3$; EORTC from the European Thyroid Cancer Cooperative Group $3$; the TNM classification from the American Joint Committee on Cancer $11$; the Ohio classification of Mazzaferrari and Jhiang $2$ from Ohio and US Air Force patients; Clinical Class by DeGroot et al. $12$; and classification systems of Noguchi et al, $13$ Shaha et al, $14$ Akslen, $15$ and Brieryel et al. $10$ With any scoring system, the question arises how well it applies to a cohort of patients different from that on which it was originally developed. $16$ Recently, Brieryel et al. $10$ evaluated 382 patients with papillary and follicular cancer using all 10 of these scoring systems and found no significant difference among AGES, TNM, EORTC, MACIS, and AMES systems, which were most accurate in predicting the prognosis of patients with papillary cancer. Other groups $3,11,16-19$ have independently validated the useful-
The low-risk patient who is likely to do well does not need to be exposed to additional risk of even nonfatal surgical complications. Low-risk patients by the AMES criteria were found in this study, with 790 patients, to have identical rates of survival at 10 and 20 years whether they had unilateral or bilateral procedures. Even those who argue for more aggressive treatment have not demonstrated an improvement in survival in good-risk patients with more extensive surgery.2 This finding has been duplicated by the Mayo Clinic4 and many other groups.17,18,22,24,26,28-30

Recurrences in our low-risk patients were uncommon and even more uncommonly fatal; local recurrences, which in high-risk patients might conceivably be preventable by more aggressive local surgery, were associated with death due to disease in only 13% of patients with local recurrences in our low-risk group. Some researchers2,31 have found that there is a decreased risk of recurrence—as opposed to death—in low-risk patients with more aggressive surgery, but others have not duplicated this finding.3,5,29,30 Recurrences in low-risk patients are uncommon (5% in our series) and usually can be treated successfully. Of our 44 low-risk patients who had recurrences, only 10 (23%) died of disease, 2 of these possibly related to postoperative external beam radiation.

High-risk patients have a disease that behaves differently, and the therapeutic approach needs to be different. With a risk of death of more than 50%, aggressive treatment is more readily justified. Bilateral surgery showed a trend, although not statistically significant, toward decreasing both deaths due to disease and local recurrences, which were themselves associated with death in most high-risk patients. Lymph node dissections of any type were associated with improved survival in this group, which underscores the importance of thorough local control.

We can also draw conclusions about postoperative medical treatment. Thyroid hormone replacement to suppress thyrotropin seemed to be effective in our patients. It should be used judiciously to avoid osteoporosis, which can be associated with serious health effects. Radioactive iodine therapy showed a nonsignificant trend toward better survival, even in this large group. In high-risk patients, it should be used; in low-risk patients, which will include most women with childbearing potential, it should be used with awareness of its lack of minimal demonstrated usefulness.

The AMES criteria remain simple, powerful predictors of risk that easily can be applied during and after surgery to help choose for the individual patient the most treatment that will be helpful with the least troublesome complications.

For the low-risk patient, bilateral surgery is not demonstrably better than unilateral surgery, and unless the surgeon is extremely experienced and has a personal miserable complication rate, he or she should not hesitate to stop at unilateral surgery. Radioactive iodine, the use of which often demands bilateral surgery, offers nothing additional to the low-risk patient. Thyroid hormone to suppress thyrotropin may be helpful.

The high-risk patient, however, should have bilateral surgery, either ipsilateral total and contralateral sub-
total or—if this can be done with assurance of safety—bilateral total. Lymphadenectomy should be performed if there is any gross disease. Radioactive iodine and thyroid hormone should be given postoperatively, and recurrences should be sought and treated aggressively.


We thank Gerald Heatley, MS, for his assistance with the statistical analyses of the data.

Reprints: Laura E. Sanders, MD, Department of General Surgery, Lahey Hitchcock Medical Center, 41 Mall Rd, Burlington, MA 01805.

REFERENCES


DISCUSSION

Thomas Colacchio, MD, New Lebanon, NH: As we’ve heard, differentiated thyroid cancer is a relatively uncommon malignancy that generally carries an uncommonly favorable prognosis. Both of these characteristics have led to an ongoing debate regarding the best form of overall therapy, with a particular emphasis on the extent of surgery necessary to maintain the anticipated excellent prognosis. As we heard from Dr Wells yesterday, and reinforced by our authors today, there has not been and likely will not be a prospectively randomized trial conducted to answer this question. In addition, the realities of care delivery make a strategy of referral of these patients to centers of excellence an improbability. As a result, for now we must rely on the large, retrospective series to attempt to answer these questions.

There clearly is a subgroup of patients with well-differentiated thyroid cancer who do not enjoy this favorable prognosis and for whom the disease-related mortality approaches 50%. Effectively identifying this population and determining the optimal treatment strategy are the goals of the authors as they have presented the further application of their AMES risk assessment categories. They have reviewed the records of more than 1000 patients treated at 2 institutions during 5 decades. They demonstrated that their AMES risk categories continue to be effective in distinguishing high- from low-risk patients in this most recent cohort of 264 patients from 1980 to 1990. It is noteworthy and yet somewhat discouraging that their survival curves from these high- vs low-risk groups from the 3 study periods reported are nearly identical despite the fact that the percentage of tumors greater than 3 cm has decreased since 1960.

I would ask them to postulate if they would why these risk factors of age, metastasis, extent, and size appear to have such a dramatic effect on survival in this particular disease, especially since they seem to believe that no therapeutic decisions have any significant effect in the outcome in the low-risk group. Is this merely the result of disease stage of diagnosis, or are there other factors that differentiate the tumor biology in these 2 groups?

Although I would not choose to debate their assertion that in the setting of infrequent encounters with this disease the decision to perform less than a total thyroidectomy in a low-risk patient is reasonable to minimize morbidity, I have several questions of clarification that will help in evaluating their assess-
ment of the impact of treatment alternatives on the outcomes of these patients.

First, does the definition you use of recurrence mean ipsilateral disease recurrence following unilateral surgery, or does it also include occurrence of disease in the contralateral remnant thyroid?

Does your suggestion in the manuscript that the external beam radiation resulted in recurrence refer to the possibility of radiation-reduced disease in the remnant thyroid?

In your patients, was there evidence for dedifferentiation of the residual disease into a more anaplastic form in those who died with recurrence?

How many patients did you have in the unilateral biopsy only group who were eliminated from your data analysis?

What number of your patients underwent reoperation to remove the contralateral thyroid, and was this morbidity in the group analyzed separately?

Was the radioiodine therapy you described used to ablate a normal thyroid remnant or to treat recurrent disease?

Was the complication rate analysis stratified for unilateral vs bilateral thyroidectomy?

Finally, I congratulate the authors on the verification of the AMES classification because it has not been replicated in a second population. This report should put that criticism to rest. In addition, in a series from the University of Vermont that is unpublished of 100 patients with thyroid carcinoma, many of whom were treated by unilateral thyroidectomy and very few of whom had prophylactic radioactive iodine, we’ve had no deaths in the AMES low-risk population.

My question: what happens if you separate out the encapsulated follicular carcinoma? Is AMES not really a risk factor analysis for papillary carcinoma? I would imagine that histological identification of a true encapsulated follicular carcinoma would be a single defining factor for a low-risk group no matter what the patient age. I wonder if you have data to confirm that?

Francis Moore, MD, Boston, Mass: I’m a little concerned about the analysis of treatment because you haven’t stated what your criteria were for giving radioactive iodine to the low-risk group or extending surgery to the second side in the low-risk group or using thyroid hormone. To come to the conclusions about the effect of therapy, you would have to say that those choices had been made totally at random, which I sincerely doubt. I would like to know what the criteria were and what assurance we have that the patients who were selected in the low-risk group for more extensive therapy did not in fact have more extensive disease.

Dr Sanders: Thank you all for your thoughtful questions.

The first one raised by Dr Colacchio is, “Is there a difference in tumor biology?” We all think that there is. I am not aware of any substantive proposal for why this might be. Dr Cady suggested to me just before my presentation that perhaps Dr Wells will be able to apply his formidable mind to answering some of these questions. There were in some of the early papers some suggestions raised that maybe this was some hormonal thing because it appeared that women who were premenopausal and younger men had a better prognosis, but as far as I know this is purely speculation, the short answer being we don’t know.

How did we classify recurrences? Recurrence was any appearance of disease after the disease was thought to be eradicated, so that would include ipsilateral recurrence, contralateral recurrence, distant recurrence, any recurrence.

The point you mention of external beam radiation therapy, which I didn’t get to in my talk but which is in the paper, is an interesting one. Early in the experience, and because this series goes back to 1940, we have some very early patients; they were often given prophylactic external beam radiation even after some early tumors were treated in patients that we now know were low-risk group. There were a couple of these patients who 15 or 30 years later developed a recurrent disease, both of which appeared to have some anaplastic dedifferentiation. One could speculate that this was radiation induced. It’s hard to know.

Did we have any patients who dedifferentiated? There were a couple, of whom these were the most memorable.

How many patients had unilateral biopsy only? There were about 20 patients, so a relatively small number of the total, and these were mostly patients who presented with already widely metastatic disease and fairly early in the series.

We did not separate out the patients who had reoperation to remove the contralateral thyroid. Our practice has generally been to perform an ipsilateral total and a contralateral subtotal in patients who have a “follicular lesion” at frozen section. We would not then need to do completion thyroidectomy so there would not be many patients in this group.

Radioactive iodine. The nature of our database doesn’t allow us to differentiate between those who had treatments for ablation and treatments given prophylactically or patients who had that treatment to treat recurrent disease, so those patients are not broken out.

Did we stratify patients who had unilateral vs bilateral complications? No. As you see, there were few patients who had recurrent nerve injuries. Parenthetically, it is probably hard to get a real handle on what your recurrent nerve injury is. There are patients who perceive themselves to be hoarse who won’t let you look down their throats, and patients as we all know who have had a previous thyroidectomy (all, of course, by people other than anyone sitting in this room) who come for another side and when you do a laryngoscopy you discover that they have an asymptomatic paralyzed vocal cord.

Dr Monchik points out and makes a good argument for those who would pursue a more aggressive disease. It’s true that papillary cancer is often multicentric. There are autopsy studies that have shown in Finnish patients as high as 30% of people who have small occult papillary carcinomas. It has never been clear what the clinical significance of those occult papillary carcinomas is, and because, as has been acknowledged, we have only retrospective studies, it’s hard to say how much use radioactive iodine is and how useful it is to be able to follow up the patients with thyroglobulin. There is no question that in the hands of someone as capable as Dr Monchik, a bilateral procedure would be safe, but as Colin Thomas says, we would argue that if something is not worth doing, it is not worth doing well.

Finally, I congratulate Dr Foster on his successful series, and it agrees with what we think—that the people with low-risk disease do better.

Dr Moore, you have pointed out one of the flaws of any retrospective study. All of us who write about this are lighting our admittedly imperfect candles rather than cursing the lack of prospective studies. We were not able to separate out the criteria for selecting treatment with radioactive iodine or bilateral thyroidectomy or thyroid hormone. During 50 years, even in these patients restricted to 2 institutions, there have been a variety of practitioners doing this surgery. I wouldn’t say it has been random, but the criteria for choosing a given therapy have probably not been completely consistent.