PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/137703

Please be advised that this information was generated on 2019-04-18 and may be subject to change.
Cost-Effectiveness of FDG-PET/CT for Cytologically Indeterminate Thyroid Nodules: A Decision Analytic Approach


Departments of Radiology and Nuclear Medicine (D.V., W.J.G.O., L.F.d.G.-O.), Health Evidence (E.M.M.A.), Internal Medicine Section of Endocrinology (R.T.N.-M.), Internal Medicine (J.W.A.S.), and Surgery (J.H.W.d.W.), Radboudumc, 6500 HB Nijmegen, The Netherlands

Context: Patients with thyroid nodules of indeterminate cytology undergo diagnostic surgery according to current guidelines. In 75% of patients, the nodule is benign. In these patients, surgery was unnecessary and unbeneificial because complications may occur. Preoperative fluorodeoxyglucose-positron emission tomography/computed tomography (FDG-PET/CT) was found to have a very high negative predictive value (96%) and might therefore avoid futile surgery, complications, and costs. In the United States, two molecular tests of cytology material are routinely used for this purpose.

Objective: Five-year cost-effectiveness for routine implementation of FDG-PET/CT was evaluated in adult patients with indeterminate fine-needle aspiration cytology and compared with surgery in all patients and both molecular tests.

Design: A Markov decision model was developed to synthesize the evidence on cost-effectiveness about the four alternative strategies. The model was probabilistically analyzed. One-way sensitivity analyses of deterministic input variables likely to influence outcome were performed.

Setting and Subjects: The model was representative for adult patients with cytologically indeterminate thyroid nodules.

Main Outcome Measures: The discounted incremental net monetary benefit (iNMB), the efficiency decision rule containing outcomes as quality-adjusted life-years and (direct) medical cost, of implementation of FDG-PET/CT is displayed.

Results: Full implementation of FDG-PET/CT resulted in 40% surgery for benign nodules, compared with 75% in the conventional approach, without a difference in recurrence free and overall survival. The FDG-PET/CT modality is the more efficient technology, with a mean iNMB of €3684 compared with surgery in all. Also, compared with a gene expression classifier test and a molecular marker panel, the mean iNMB of FDG-PET/CT was €1030 and €3851, respectively, and consequently the more efficient alternative.

Conclusion: Full implementation of preoperative FDG-PET/CT in patients with indeterminate thyroid nodules could prevent up to 47% of current unnecessary surgery leading to lower costs and a modest increase of health-related quality of life. Compared with an approach with diagnostic surgery in all patients and both molecular tests, it is the least expensive alternative with similar effectiveness as the gene-expression classifier. (J Clin Endocrinol Metab 99: 3263–3274, 2014)
Thyroid nodules are common, and as many as 3%–8% of European adults have palpable nodules, but the risk of differentiated thyroid carcinoma in these nodules is less than 5%. In healthy adults, a screening ultrasound (US) can detect asymptomatic thyroid nodules in up to 68% of volunteers (1). Due to the increasing use of US and other imaging techniques, more and more asymptomatic thyroid nodules are discovered, most of which have no clinical relevance. Once a nodule is established, screening for cancer is warranted because most of the thyroid carcinomas present as thyroid nodules (2). In particular, in localized (<68%) and regional (<25%) stage at diagnosis, the prognosis of differentiated thyroid carcinoma is favorable because the 5-year relative survival in these patients is greater than 97% (2).

In case of unsuppressed TSH, the recommended initial diagnostic test of a thyroid nodule according to current guidelines is US-guided fine-needle aspiration cytology (FNAC) (3, 4). Aspirates are classified into six diagnostic categories according to the Bethesda System for Reporting Thyroid Cytopathology (5). In approximately 75% of patients, this will lead to a definite diagnosis and treatment, for benign, suspicious for malignancy, or definite malignant disease. However, in the remaining cases, repetitive FNAC cannot determine whether the lesion is benign or malignant due to cellular atypia, follicular neoplasia, or repetitive nondiagnostic or unsatisfactory specimens. Without further classification, in 69%–88% of these patients, the nodule is found to be benign at diagnostic hemi-thyroidectomy (lobectomy) (6). In most malignant nodules, secondary surgery with adjuvant treatment including radioactive iodine-131 thyroid remnant ablation (RRA) and TSH-suppression therapy is recommended. Only in the case of subcentimeter (pT1a), indolent, unifocal papillary microcarcinoma, additional treatment is considered unnecessary (3).

The use of one of two molecular tests as an adjunct to diagnosis in FNAC-indeterminate thyroid nodules is standard of care in the United States. One, a 167-gene expression classifier (GEC), is used to minimize unnecessary diagnostically relevant thyroid surgery, and another one, a mutation marker panel (MMP), is used to select patients for initial total thyroidectomy, thereby saving on the two-stage surgery. The GEC (7–9) showed a positive and negative predictive value (NPV) of 47% and 93%, respectively, and was found to be cost effective (10). Another molecular test (11) includes a MMP for mutations in BRAF and RAS and rearrangements in RET/papillary thyroid carcinoma and paired box transcription factor-8/peroxisomal proliferator-activated receptor-γ. It showed a positive and NPV of 87% and 90%, respectively. Its limited NPV made the authors suggest an up-front total thyroidectomy after a positive test result and lobectomy otherwise. By saving on two-stage surgery, they showed a moderate increase in costs of nodule evaluation (+18% or US$104 per patient overall costs) (12). Currently both these tests are unavailable in Europe or Asia.

Recently we summarized the data of 225 individual patients with indeterminate thyroid nodules from our own series (13) and five other published prospective studies (6). In all patients a fluorodeoxyglucose (FDG)-positron emission tomography (PET) was performed on previous-generation PET scanners (most without computed tomography (CT) capabilities and none with time of flight technology) prior to the scheduled surgery, and therefore, gold standard histology was available. We described a positive predictive value and NPV of 39% and 96%, respectively. These data were recently confirmed by two prospective series of 55 and 46 patients, respectively (14, 15), concluding that FDG-PET/CT could reduce the number of diagnostic (hemi)thyroidectomies by 13%–25% (15). Even though none of the studies summarized in our published meta-analysis (6) adopted the Bethesda criteria (five of six were published before its establishment (5)], confirmation of the performance of FDG-PET/CT in a Bethesda classified population (14, 15) supports its predictive value in this population.

Based on the high NPV of FDG-PET/CT to exclude malignancy in case of cellular atypia or follicular neoplasia in asymptomatic thyroid nodules, we hypothesize that its incorporation could reduce futile surgery from 74% to 39%. This would lead to fewer symptoms and cosmetic complaints of a neck scar. Also, fewer patients would need life-long daily thyroid hormone supplementation because up to one third of lobectomized patients have functional insufficiency of the remaining thyroid tissue (16). Although rare, surgical complications may be severe (hemorrhage, infection, permanent hoarseness) (16–18) and could be decreased using the proposed strategy.

Because surgery, hospitalization, follow-up, FDG-PET/CT, and both molecular tests entail significant costs, current health economic evaluation was undertaken to model the potential impact of implementation of each one of these tests separately in the work-up of FNAC-indeterminate thyroid nodules on direct health care costs and patients’ health-related quality of life (HRQoL). We determined the cost-effectiveness of an FDG-PET/CT driven approach compared with either a surgical approach (being the standard of care in Europe/Asia) or one of both molecular tests (US standard).

Materials and Methods

Decision model

An eight-(health) state Markov decision model, with yearly cycle length, was developed in accordance with the 2009 Amer-
American Thyroid Association (ATA) guidelines for the management of patients with thyroid nodules (3) and the strategies proposed by the developers of both molecular tests (10, 12). Treatment for adult patients with thyroid nodules that are scheduled for surgery based on indeterminate FNAC (Bethesda categories III and IV) was simulated, being driven by diagnostic thyroid surgery (surgery), a molecular test aiming at the prevention of unnecessary surgery (GEC), a molecular test aiming at the prevention of two-step surgery (MMP), and routine FDG-PET/CT. Branches were developed to represent patient care after an indeterminate FNAC result (decision tree, Figure 1), leading to one of eight health states. These health states include surveillance (after a negative FDG-PET/CT or GEC), surveillance after thyroid surgery, permanent complications due to thyroid surgery, recurrence after thyroid surgery, or death.

### FDG-PET/CT and molecular tests

Diagnostic performance of FDG-PET/CT is based on the six studies summarized in our meta-analysis (6). Diagnostic performance of the GEC is based on Li et al (10) and for the MMP on Yip et al (12). In contrast with Yip et al (12), we chose not to incorporate a repeated FNAC in any of the four study arms to homogenize the simulated clinical course in all patients.

Because the different tests were originally benchmarked on different populations, with individual study cancer prevalence ranging from 20% (12) to 32% (10), we computed positive and negative predictive values based on an uniform a priori risk of malignancy of 25% [ie, the weighted mean of all three study populations (6, 10, 12)] and the test sensitivities and specificities as stated in the original references.

![Decision Tree](image)

**Figure 1.** Decision tree. Simulated patients with FNAC-indeterminate TNs will either be treated based on diagnostic thyroid surgery, based on one of two molecular test or based on the result of FDG-PET/CT. They will enter the Markov model in one of eight health states based on this decision tree (see Figure 2). Diamonds are decision nodes, and decision are based on probabilities. Boxes are interventions and cost money. (c)TT, (completion) total thyroidectomy; HT, hemithyroidectomy; MT, molecular test; PA, histopathology; TN, thyroid nodule; UPM, unifocal papillary microcarcinoma.
Risk and probability estimation

The duration of each Markov cycle was considered to be 1 year; therefore, the transition between health states reflect annual probabilities governed by factors such as a priori probability of malignancy, surgical complication rates, recurrence rates, and age- and sex-specific mortality rates (Figure 2). Stochastic transition probabilities were collected from a variety of international literature sources including several other decision analyses on the diagnostic approach of an FNAC-indeterminate thyroid nodule (Table 1). Missing parameter values or those that varied highly among literature were elicited from a panel consisting of six medical, surgical, and imaging thyroid experts from the Radboudumc in Nijmegen and one health economist.

Cost and utility estimation

The Markov state information contained costs and utilities with a time frame of 1 year.

The model considers stochastic direct medical costs data (Table 2). These were derived from 2012 reimbursement rates of the Dutch system of Diagnosis-Treatment Combinations and published in the international literature. All prices were indexed to January 2013 euros, using country-specific consumer price indexes (19-22) and up-to-date exchange rates (23) (January 2013: €1.00000 = US$1.31139 = CAN$1.32909). These prices include reimbursement tariffs for the molecular test, FDG-PET/CT, to physicians, anesthesia, pathology, laboratory investigations, US procedures, thyroid surgery, RRA, medication, hospital facilities, and all other costs incurred during inpatient and outpatient treatment. The costs of both transient and permanent complications were based on estimates from literature; its wide distribution reflects the variety of the severity of these complications.

Utilities for each cycle in a particular health state were derived from literature (Table 2). Quality-adjusted life-years (QALYs) were calculated by the discounted sum of utilities over the 5-year evaluation period. Utility values from the literature were used where available or elicited from a previously mentioned expert panel based on time-trade-off weighting.

One-way sensitivity analyses were performed to explore the variation of base-case model parameters on their range of extremes (10,000 hypothetical patients, 1000 second order parameter samples). One-way sensitivity analyses for transition probabilities, costs, and utilities were performed over a wide range of values identified from the literature (Tables 1 and 2). Among the variables examined are parameters connected to the procedure and follow-up after hemithyroidectomy, the procedure and the follow-up after FDG-PET/CT, the performance of molecular tests and FDG-PET/CT (sensitivity, specificity), and the demographics of the population under review (prevalence of malignancy in thyroid nodules).

The mean costs and utilities acquired during this 5-year period for each scenario were used to compute the incremental net monetary benefit (iNMB in euros):

\[ \text{iNMB} = \lambda \cdot (E_2 - E_1) - (C_2 - C_1) \]

where \( \lambda \) is the willingness-to-pay threshold, \( E \) is the effects (utilities), and \( C \) is the costs of both scenarios under comparison. The subscript 1 denotes the comparator (surgery, GEC, or MMP) and 2 denotes FDG-PET/CT-driven treatment (27). From the iNMB, the decision rule for cost-effectiveness can be inferred: \( \text{iNMB} > 0 \).

The Dutch Council for Public Health and Health Care recommends a willingness-to-pay threshold of 680 000/QALY for conditions with a maximal disease burden (28), and this is used throughout this study. However the cost-effectiveness acceptability curve, defined as the probability of iNMB greater than 60, for a wide willingness-to-pay range, is displayed.

Modeling and Monte-Carlo analysis were performed using TreeAge Pro Suite (version 2011; TreeAge Software Inc). Data analyses were performed using Matlab (version R2013a; MathWorks).

Results

Base case cost-effectiveness analysis

After 5 years of treatment for and follow-up after an FNAC-indeterminate thyroid nodule, mean discounted costs and utilities were exponentially discounted at a constant rate of 4.0% and 1.5% per year, respectively (24).

Base case cost-effectiveness analysis and sensitivity analyses

For the base-case scenario, the model has been run in a probabilistic fashion, with micromutation of 100 000 first-order trials (patients) for 10 000 second-order parameter samples over five cycles. A 5-year evaluation period was chosen because most costs (and HRQoL losses) are made in the first years, the ATA guidelines (3) have difficulty in providing recommendations after the first 5 years, greater than 50% of recurrences occur in the first 2 years (4, 25), and only limited data exist supporting the fact that probabilities, costs and effects after the initial first 5 years differ between scenarios. Half-cycle correction was applied. Results are displayed in a cost-effectiveness plane (26).
Table 1. Accountability of Base-Case Parameter Values and Stochastic Distributions for Base Variables and Transition Probabilities, Including Range Used for One-Way Sensitivity Analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>Distribution</th>
<th>Expected Value (95% CI)</th>
<th>Source</th>
<th>Range for SA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discount rate of costs</td>
<td>Fixed</td>
<td>0.040</td>
<td>(10, 24)</td>
<td>0.030–0.050</td>
</tr>
<tr>
<td>Discount rate of utilities</td>
<td>Fixed</td>
<td>0.015</td>
<td>(10, 24)</td>
<td>0.010–0.050</td>
</tr>
<tr>
<td><strong>Population description</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction of female patients (\beta)</td>
<td>Dirichlet</td>
<td>0.86 (0.81–0.90)</td>
<td>(6, 10, 13, 37–41)</td>
<td>0.78–0.93</td>
</tr>
<tr>
<td>Age of female patient when diagnosed, y (\gamma)</td>
<td>Dirichlet</td>
<td>47.3 (21.0–73.6)</td>
<td>(6, 10, 13, 37–41)</td>
<td></td>
</tr>
<tr>
<td>Age of male patient when diagnosed, y (\gamma)</td>
<td>Dirichlet</td>
<td>55.6 (26.1–85.0)</td>
<td>(6, 10, 13, 37–41)</td>
<td></td>
</tr>
<tr>
<td>Incidence of DTC in healthy females (\beta)</td>
<td>Dirichlet</td>
<td>0.000 0031 (0.000 0021–0.000 0043)</td>
<td>(4)</td>
<td></td>
</tr>
<tr>
<td>Incidence of DTC in healthy males (\beta)</td>
<td>Dirichlet</td>
<td>0.000 0013 (0.000 00069–0.000 0013)</td>
<td>(4)</td>
<td></td>
</tr>
<tr>
<td>Yearly probability of death of any cause (not cancer related)</td>
<td>Life table</td>
<td></td>
<td>Age/sex dependent</td>
<td>(42)</td>
</tr>
<tr>
<td><strong>General probabilities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction HT of all surgery (\beta)</td>
<td>Dirichlet</td>
<td>0.95 (0.90–0.98)</td>
<td>EO</td>
<td>0.50–0.99</td>
</tr>
<tr>
<td>Fraction of UPM in indeterminate nodules (\beta)</td>
<td>Dirichlet</td>
<td>0.023 (0.0076–0.047)</td>
<td>(6, 10, 13, 37–41)</td>
<td>0.01–0.10</td>
</tr>
<tr>
<td>Prevalence of cancer in indeterminate nodules</td>
<td>Dirichlet</td>
<td>0.25 (0.22–0.28)</td>
<td>(6, 7, 10–15, 37–41)</td>
<td>0.15–0.35</td>
</tr>
<tr>
<td><strong>Diagnostic test characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDG-PET/CT sensitivity</td>
<td>Dirichlet</td>
<td>0.95 (0.88–0.99)</td>
<td>(6, 13–15, 37–41)</td>
<td>0.70–0.99</td>
</tr>
<tr>
<td>FDG-PET/CT specificity</td>
<td>Dirichlet</td>
<td>0.48 (0.40–0.55)</td>
<td>(7, 10)</td>
<td>0.35–0.70</td>
</tr>
<tr>
<td>GEC sensitivity</td>
<td>Dirichlet</td>
<td>0.92 (0.85–0.97)</td>
<td>(11, 12)</td>
<td>0.65–0.99</td>
</tr>
<tr>
<td>GEC specificity</td>
<td>Dirichlet</td>
<td>0.52 (0.44–0.59)</td>
<td>(11, 12)</td>
<td>0.40–0.75</td>
</tr>
<tr>
<td>MMP sensitivity</td>
<td>Dirichlet</td>
<td>0.59 (0.49–0.69)</td>
<td>(11, 12)</td>
<td>0.35–0.70</td>
</tr>
<tr>
<td>MMP specificity</td>
<td>Dirichlet</td>
<td>0.98 (0.96–0.99)</td>
<td>(11, 12)</td>
<td>0.75–0.99</td>
</tr>
<tr>
<td><strong>Yearly probability of surgery after surveillance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After negative FDG-PET/CT (\beta)</td>
<td>Dirichlet</td>
<td>0.0070 (0.0014–0.021)</td>
<td>Computed ((1–\text{NPV}^{\alpha}))</td>
<td>0.00–0.05</td>
</tr>
<tr>
<td>After negative GEC (\beta)</td>
<td>Dirichlet</td>
<td>0.010 (0.0035–0.023)</td>
<td>Computed ((1–\text{NPV}^{\alpha}))</td>
<td>0.00–0.05</td>
</tr>
<tr>
<td><strong>Complications of surgery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction of transient complications due to HT (\beta)</td>
<td>Dirichlet</td>
<td>0.039 (0.0020–0.064)</td>
<td>(10, 16, 43–48)</td>
<td>0.01–0.60</td>
</tr>
<tr>
<td>Fraction of permanent complications due to HT (\beta)</td>
<td>Dirichlet</td>
<td>0.088 (0.069–0.11)</td>
<td>(10, 16, 43, 44, 47, 49)</td>
<td>0.01–0.26</td>
</tr>
<tr>
<td>Fraction of transient complications due to (c)TT (\beta)</td>
<td>Dirichlet</td>
<td>0.19 (0.10–0.30)</td>
<td>(10, 16, 43–48)</td>
<td>0.01–0.65</td>
</tr>
<tr>
<td>Fraction of permanent complications due to (c)TT (\beta)</td>
<td>Dirichlet</td>
<td>0.038 (0.023–0.056)</td>
<td>(10, 16, 43, 44, 47, 49)</td>
<td>0.01–0.25</td>
</tr>
<tr>
<td>Fraction of death due to any type of surgery (\beta)</td>
<td>Dirichlet</td>
<td>0.0019 (0.000 91–0.0032)</td>
<td>(10, 50)</td>
<td>0.00–0.01</td>
</tr>
<tr>
<td><strong>Recurrence/cancer-related death</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yearly probability of recurrence after HT for UPM (\beta)</td>
<td>Dirichlet</td>
<td>0.0047 (0.000 20–0.016)</td>
<td>(51)</td>
<td>0.001–0.025</td>
</tr>
<tr>
<td>Yearly probability of cTT after recurrence after HT (\beta)</td>
<td>Dirichlet</td>
<td>0.017 (0.089–0.940)</td>
<td>(10)</td>
<td>0.90–1.00</td>
</tr>
<tr>
<td>Yearly probability of recurrence after (c)TT (\beta)</td>
<td>Dirichlet</td>
<td>0.027 (0.019–0.037)</td>
<td>(10, 52)</td>
<td>0.01–0.07</td>
</tr>
<tr>
<td>Yearly probability of death due to cancer (\beta)</td>
<td>Dirichlet</td>
<td>0.0051 (0.0020–0.0095)</td>
<td>(10, 52)</td>
<td>0.00–0.01</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; (c)TT, (completion) total thyroidectomy; Distribution, parameter stochastic distribution; EO, expert opinion; HT, hemithyroidectomy; SA, sensitivity analyses; UPM, unifocal papillary microcarcinoma.
costs were €8804 (surgery), €9341 (GEC), €8913 (MMP), and €7983 (FDG-PET/CT).

Their mean discounted utilities were 4.52, 4.56, 4.52, and 4.55 QALY, respectively. Therefore, FDG-PET/CT-driven surgery proved to be the more efficient alternative, being on average €822 less expensive per patient with moderately higher HRQoL of 0.036 QALY over 5 years compared with surgery in all patients. Compared with GEC and MMP, it was €1358 and €930 less expensive except in comparison with the GEC, in which PET showed a lower HRQoL of 0.0040 QALY (ie, 1.5 quality-adjusted life days). This makes a convincing case that the FDG-PET/CT modality is the most efficient approach. For the willingness-to-pay range of €0–€80 000/QALY, the probability of a positive iNMB equals 1 for PET vs any of the other three strategies (Supplemental Figure 1).

The fraction of futile surgery of histologically benign, FNAC-indeterminate thyroid nodules was 75% (surgery), 38% (GEC), 75% (MMP), and 40% (FDG-PET/CT), respectively. Therefore, unbeneficial surgery could potentially be decreased by up to 37% and 35% by the full implementation of GEC and FDG-PET/CT, respectively. This would lead to a reduction of surgery-related (permanent) complications (including surgery related death) from 7.7% (surgery or MMP) to 4.4% (GEC) or 4.6% (FDG-PET/CT), ie, almost halving unbeneficial surgery
Table 3. Base Case Main Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Surgery (95% CI)</th>
<th>GEC (95% CI)</th>
<th>MMP (95% CI)</th>
<th>FDG-PET/CT (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costs, €</td>
<td>€8804 (€8774–€8835)</td>
<td>€9341 (€9300–€9383)</td>
<td>€8913 (€8884–€8942)</td>
<td>€7983 (€7941–€8025)</td>
</tr>
<tr>
<td>Futile surgery, %</td>
<td>75.0% (74.8%–75.3%)</td>
<td>38.2% (37.8%–38.5%)</td>
<td>75.0% (74.8%–75.3%)</td>
<td>40.3% (39.9%–40.7%)</td>
</tr>
<tr>
<td>Incremental values of FDG-PET/CT vs alternative strategy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incremental costs, €</td>
<td>€−822 (€−871–€−772)</td>
<td>€−1358 (€−1377–€−1340)</td>
<td>€−930 (€−970–€−890)</td>
<td>N/A</td>
</tr>
<tr>
<td>Incremental utilities, QALY</td>
<td>0.036 (0.031–0.041)</td>
<td>−0.0040 (−0.0050−0.0030)</td>
<td>0.037 (0.033–0.041)</td>
<td>N/A</td>
</tr>
<tr>
<td>iNMB, €</td>
<td>€3684 (€3278–€4094)</td>
<td>€1030 (€916–€1142)</td>
<td>€3851 (€3528–€4170)</td>
<td>N/A</td>
</tr>
<tr>
<td>Incremental futile surgery, %</td>
<td>−34.7% (−34.3% to −35.2%)</td>
<td>+2.1% (+2.0%−+2.3%)</td>
<td>−34.7% (−34.3% to −35.2%)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; N/A, not applicable.

and surgery-related complications. Mean 5-year overall and recurrence-free survival in this population was similar in all four strategies, being 96.5% and 97.2% respectively.

One-way sensitivity analyses

The most influential parameter (under assumptions of independency) was found to be the utility attributed to watchful surveillance (after a negative FDG-PET/CT scan or GEC). At the minimum evaluated value (0.90), a worse quality of life was found for FDG-PET/CT-driven treatment vs either thyroid surgery in all patients or MMP (in both mean incremental utility: −0.10 QALY), leading to a mean iNMB of €−7418 (vs surgery) and €−7264 (vs MMP). At a value for the utility attributed to watchful surveillance of 0.953 (vs surgery) or 0.952 (vs MMP), the mean iNMB equals €0. For comparison, the utility attributed to the health state after uncomplicated hemithyroidectomy is set at 0.99.

Other parameters that proved influential in affecting cost-effectiveness included the utility of surveillance and permanent complications after hemithyroidectomy, the probability of hemithyroidectomy-induced (transient and permanent) complications, the probability of performing hemithyroidectomy as the primary method for thyroid surgery, and surgical mortality as well as the costs of a hemithyroidectomy procedure. In comparison with the GEC, which has a similar place in the workup as FDG-PET/CT, the crucial parameters leading to a preference of GEC over FDG-PET/CT were the test specificity of both (sensitivity and specificity), the cost-price of the GEC, the test sensitivity of the GEC, and the yearly probability that surgery has to be performed after a (false negative) FDG-PET/CT. For the range of the prevalence of thyroid carcinoma tested (15%–35%), FDG-PET/CT was the preferred modality over the GEC (see Figure 4 and Supplemental Figure 2, A–C).

Discussion

We presented an economic decision analytical model, forecasting that implementation of FDG-PET/CT in the workup of FNAC-indeterminate thyroid nodules could lead to a substantial reduction in direct medical costs and, compared with two of the three alternatives, modestly im-
Avoidance of (complications of) unnecessary thyroid surgery to provide a definite histopathological diagnosis is the principal cause of cost-reduction. The fraction of surgeries performed for a benign thyroid nodule could almost be halved when fully implementing FDG-PET/CT compared with thyroid surgery in all FNAC-indeterminate thyroid nodules (40.3% and 75.0%, respectively). Because it is estimated that 60 220 men and women are diagnosed with differentiated thyroid carcinoma (DTC) in the United States in 2013 (2) and about half are found after surgery for FNAC-indeterminate nodules (29, 30), it can be approximately estimated that 120 000 patients undergo thyroid surgery for a FNAC-indeterminate thyroid nodule, of whom 90 000 undergo surgery for a benign disease. Full implementation of FDG-PET/CT could save up to 42 000 unnecessary surgeries annually, €99 million and 4.3 thousand QALYs in the United States only, assuming FNAC indeterminacy was the sole reason for thyroid surgery. Compared with those in the US current practice of GEC, a change from full implementation of GEC to FDG-PET/CT could potentially result in an annual cost-reduction of €164 million. On the drawback, the somewhat lower specificity of FDG-PET/CT compared with the GEC might lead to a modestly higher fraction of surgery for benign nodules of 2.1%, responsible for a negligible (but negative) effect on HRQoL (Table 3 and Figure 3).
attribute lower values for surgery and surgery-related costs than they do. For example, in our model uncomplicated hemithyroidectomy plus 5-year follow-up would cost €5499, but in their model this would be US$10 319 (€8311, indexed to January 2013). Because we adopted the same cost-price of the GEC, this example shows that in our model the prevention of one uncomplicated surgery by the GEC equals the costs of two diagnostic tests only, whereas in their model it saves enough to pay for more than 3 GECs. This is further supported by the fact that the costs attributed to the GEC was one of the most influential determinants in one-way sensitivity analysis (Figure 4, middle panel).

Modest improvement of HRQoL was found as long as estimated HRQoL of surveillance after a negative FDG-PET/CT was higher than 0.95, this parameter was found to be the sole variable that could lead to a situation in which an FDG-PET/CT-driven approach did not dominate current European practice of surgery in all patients and even a decremental net monetary benefit. To the best of our knowledge, currently there have been no prospective studies published that investigate the HRQoL of a wait-and-see policy in benign thyroid nodules. To further substantiate this parameter and our results, a prospective study should be undertaken to investigate the consequences of implementation in daily practice with respect to (in)direct costs, measured HRQoL, and other measures of effectiveness.

The HRQoL attributed to surveillance after a negative FDG-PET/CT could be depreciated due to factors related to the thyroid nodule itself or to the fear of a false-negative FDG-PET/CT result [1.3% of all FDG-PET/CT scans performed were false negative (6)]. The former can be prevented by not offering FDG-PET/CT in case thyroid surgery is considered for other than mere diagnostic purposes only. A false-negative FDG-PET/CT scan could delay treatment for thyroid malignancy. Our model assumes that on average most of these are treated during a 5-year follow-up period. Outcome with respect to progression-free and overall survival, costs, and HRQoL is not known for delayed treatment; therefore, no additional costs or detrimental effects are incorporated into the model. However, the oncological-, economical-, and HRQoL-related consequences are considered to be minimal due to the relative indolent course of this disease. Furthermore, there is limited impact on survival on the transition from localized to regional disease [5 y relative overall survival: 99.9% and 97.4%, respectively (2)], all with good treatment options. Finally, the false-negative ratio is based on the sensitivity of FDG-PET/CT, which was found to be highly dependent on the scanners’ resolution (5–8 mm Full Width at Half Maximum for the PET scanners used in the meta-analysis). With state-of-the-art time-of-flight technology (3–4 mm Full Width at Half Maximum), it is likely that sensitivity, and thus NPV, are higher and that the false-negative cases that occur are the smallest DTCs.

General weaknesses of any model are oversimplification of daily practice and the accuracy of the definition of each parameter. However, the current model was designed closely adhering to the ATA guidelines. By using data from a variety of sources including international literature, government publications, guidelines, and expert estimates and allowing a stochastic uncertainty in these estimates, we substantiated the generalizability of the model.

When the available literature showed heterogeneous parameter values, we elicited these from an expert panel because we expected that this variation was both based on study heterogeneity and threshold effect due to unclear definitions. For example, parameter values for probability, costs, and utility of complications highly depend on what the authors define as complication: if a minor bleeding is included in the definition of a transient complication, the probability of having a transient complication will increase, the average costs will decrease, and the average utility will probably increase. By adopting a higher scale parameter, determining the statistical dispersion of the distribution, we tried to cover these higher uncertainties.

It is likely that the value of QALYs rise over time, and because this rise is not taken into account by other means in an economic evaluation, it is suggested to discount utilities with a lower rate than costs (24). Therefore, we adopted a nonuniform discount rate for costs and utilities. Because nonuniform discounting is still uncommon in the international literature (24, 31), we repeated the analyses of the base-case scenario with a uniform discount rate of 3%/y for both costs and utilities, showing no different conclusion.

One-way sensitivity analyses over a plausible but wide range of parameter estimates showed that the outcome of the simulations were most critically influenced by the utility of surveillance after a negative FDG-PET/CT or hemithyroidectomy, costs of hemithyroidectomy, fractions and utilities of hemithyroidectomy-induced complications (including death), distribution of the initial type of surgery, and FDG-PET/CT sensitivity and specificity. Furthermore, only direct costs for a 5-year duration were computed. One could argue that indirect costs (eg, sick leave days, decreased productivity, and money spent on care outside the medical setting), would further support the inclusion of FDG-PET/CT in the diagnostic algorithm. A limitation of the sensitivity analyses is the assumption of independency. The parameters in the model are clearly related due to threshold effects. Because these relationships are complex and because it is impossible to accu-
rately substantiate any assumption as to the quantitative relationship between these parameters, this was not attempted, and a wide range value for the sensitivity analyses was chosen.

Due to the limited specificity and positive predictive value, still 40% of patients undergo thyroid surgery for a benign thyroid nodule. The only independent predictive factor for FDG-uptake in literature was cellular atypia (present in both benign and malignant nodules). The current literature mainly focuses on FDG uptake in known thyroid carcinoma (32–35) or (in vitro) in thyroid cells (36); therefore, the limited specificity of FDG-PET/CT for (FNAC indeterminate) thyroid nodules is still poorly understood.

Test characteristics of FDG-PET/CT are based on populations with a heterogeneous fraction of people suffering from multinodular disease [15%–71% (6, 14, 15)], which might influence results for two reasons: 1) from a methodological point of view, the nodule under investigation by FNAC, FDG-PET/CT, and histopathology might not be the same and 2) the result of a negative FDG-PET/CT might not modify surgical treatment decision because reasons other than merely indeterminate FNAC might be the reason for surgery. In practice the former issue is being by most studies by only including patients with a clear, dominant nodule. The latter can be overcome by offering FDG-PET/CT to patients that are scheduled for surgery only for reason of indeterminate FNAC. Although this further selected population is different from that we obtained the negative and positive predictive value of FDG-PET/CT, we believe that the robustness of our main conclusions shown by one-way sensitivity analysis is still valid for a wide range of values. The global impact might be overestimated because not all patients with a FNAC indeterminate thyroid nodule and a negative FDG-PET/CT might wish to refrain from surgery.

In conclusion, our cost-utility analysis demonstrates that full implementation of FDG-PET/CT in the workup of adult patients with thyroid nodules scheduled for surgery for FNAC-indeterminacy (ie, cellular atypia and follicular neoplasia) could lead to a decrease in costs and a moderate increase in HRQoL compared with diagnostic surgery in all patients according to current European practice and is competitive to the current US standard of the GEC. These results are primarily based on a decrease in costs and complications of surgery in patients with benign thyroid nodules that are not resected for being symptomatic. Sensitivity analyses showed the robustness of these data. Prospective studies are needed to further support cost-effectiveness and implementability and to gain insight in the false positivity of FDG-PET/CT. A prospective head-to-head comparison with alternative strategies or combinations of strategies should be considered.

Acknowledgments

The PhD project of D.V. was funded in part by The Netherlands Organisation for Health Research and Development.

Address all correspondence and requests for reprints to: D. Vriens, MD, Department of Radiology and Nuclear Medicine (756), PO Box 9101, 6500 HB, Nijmegen, The Netherlands. E-mail: dennis.vriens@radboudumc.nl.

D.V. was supported in part by The Netherlands Organisation for Health Research and Development.

Disclosure Summary: The authors have nothing to declare.

References

13. de Geus-Oei LF, Pieters GF, Bonenkamp JJ, et al. 18F-FDG PET


42. Central Bureau for Statistics. Central Bureau for Statistics: Life-tables. http://statline.cbs.nl/StatWeb/publication/DM%253D%252FNL%2526PA%253D%252F3736%2526ned%2526D1%253D%252Fa%2526D2%253D%252Fa%2526D4%253D%252F1%2526HDR%3D%252FG%2526STB%253D%252G2%2526G%2526VW%253D%252AT. Accessed April 14, 2013.


51. Lee J, Park JH, Lee CR, Chung WY, Park CS. Long-term outcomes of total thyroidectomy versus thyroid lobectomy for papillary thy-


**Supplemental Figure Legends**

**Supplemental Figure 1:** Cost-effectiveness acceptability curves, plotting the probability of a positive iNMB ($P(\text{iNMB}>€ 0)$) for a range of values for the willingness-to-pay threshold ($\lambda$). The dotted line is at a willingness-to-pay threshold of € 80,000/QALY. FDG-PET/CT: FluoroDeoxyGlucose Positron Emission Tomography / Computed Tomography; GEC: Gene-Expression Classifier; iNMB: incremental Net Monetary Benefit; MMP: Molecular Marker Panel; QALY: Quality-Adjusted Life Year.

**Supplemental Figure 2a-c:** Tornado plots showing the results of one-way sensitivity analysis of all inputs of the model on the iNMB versus one of the other three strategies (a: surgery, b: GEC, c: MMP), for a willingness-to-pay threshold ($\lambda$) of € 80,000/QALY, the whiskers represent the limits of the 95%-confidence interval; the ranges of tested values tested are between parentheses. The vertical dotted line is set at the mean iNMB of the base-case scenario. The vertical line at € 0 represents the break-even situation at a willingness-to-pay threshold of € 80,000/QALY. (c)TT: (completion) Total Thyroidectomy; FDG-PET/CT: FluoroDeoxyGlucose Positron Emission Tomography / Computed Tomography; GEC: Gene-Expression Classifier; HT: HemiThyroidectomy; iNMB: incremental Net Monetary Benefit; MMP: Molecular Marker Panel; QALY: Quality-Adjusted Life Year.