Title Page:

Full Title: Cost-Effectiveness of FDG-PET/CT for Cytologically Indeterminate Thyroid Nodules: a Decision Analytic Approach.

Authors: Vriens D, MD¹; Adang EMM, PhD²; Netea-Maier RT, MD, PhD³; Smit JWA, MD, PhD⁴; de Wilt JHW, MD, PhD⁵; Oyen WJG, MD, PhD¹; de Geus-Oei LF, MD, PhD¹.

¹dept. of Radiology and Nuclear Medicine, ²Health Evidence, ³Internal Medicine section Endocrinology, ⁴Internal Medicine and ⁵Surgery; Radboudumc, Nijmegen, the Netherlands

Abbreviated Title: Cost-Effectiveness of FDG-PET/CT in thyroid nodules

Key Terms: Thyroid Nodules; Fine-Needle Aspiration Cytology; FDG-PET/CT; Surgery; Thyroid Cancer; Cost-Effectiveness; Economical Evaluation.

Word Count: 3,983

Manuscript Category: Original Article

Corresponding Author: Vriens D, MD, dept. of Radiology and Nuclear Medicine (756) (also reprint)

PO Box 9101, 6500 HB, Nijmegen, the Netherlands

Dennis.Vriens@radboudumc.nl

T +31-24-361 40 48 / F +31-24-799 91 05

Grants/Fellowships: D. Vriens was funded in part by The Netherlands Organisation for Health Research and Development (ZonMW)

Disclosure summary: DV: has nothing further to declare; Eemma: has nothing to declare; RTNM: has nothing to declare; JWAS: has nothing to declare; Wjgo: has nothing to declare; LFdGO: has nothing to declare.

Clinical Trial Registration Number: N/A
Abstract:

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 unbefitting as complications may occur. Preoperative 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 USA, two molecular tests of cytology material are routinely used for this purpose.

Objective: 5-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 analysed. 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 (QALY) 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 to 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 €3,684 compared to surgery in all. Also, compared to a gene expression classifier test and a molecular marker panel, the mean iNMB of FDG-PET/CT was €1,030 and €3,851, 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 modest increase of health-related quality of life. Compared to 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.
Introduction:

Thyroid nodules are common 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, the majority of which have no clinical relevance. Once a nodule is established, screening for cancer is warranted as most of thyroid carcinomas present as thyroid nodules (2). Especially in localized (~68%) and regional (~25%) stage at diagnosis, prognosis of differentiated thyroid carcinoma is favourable as 5-year relative survival in these patients is over 97% (2).

In case of unsuppressed thyrotropin (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 in 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, either 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 non-diagnostic or unsatisfactory specimens. Without further classification, in 69-88% of these patients the nodule is found to be benign at diagnostic hemithyroidectomy (lobectomy) (6). In most malignant nodules, secondary surgery with adjuvant treatment including radio-active iodine-131 thyroid remnant ablation (RRA) and TSH-suppression therapy is recommended. Only in case of subcentimetre (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 USA. One, a 167 gene-expression classifier (GEC), is used to minimize unnecessary diagnostic thyroid surgery and another one, a mutation marker panel (MMP), is used to select patients for initial total thyroidectomy thereby saving on two-step surgery. The GEC (7-9), showed a positive and negative predictive value 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/PTC and PAX8/PPARγ. It showed a positive and negative predictive value of 87% and 90%, respectively. Its limited negative predictive value 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 an FDG-PET was performed on previous-generation PET-scanners (most without CT capabilities and none with time-of-flight technology) prior to scheduled surgery and therefore Gold Standard histology was available. We described a positive and negative predictive value 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 out 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 negative predictive value 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 less symptoms and cosmetic complaints of a neck scar. Also, less patients would need lifelong daily thyroid hormone suppletion, as up to one third of lobectomized patients have functional insufficiency of the remaining thyroid tissue (16). Although rare, surgical complications may be severe (haemorrhage, infection, permanent hoarseness) (16-18) and could be decreased using the proposed strategy.

As 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 healthcare costs and patients’ health-related quality of life (HRQoL). We determined the cost-
effectiveness of an FDG-PET/CT driven approach compared to either a surgical approach (being standard of care in Europe/Asia) or one of both molecular tests (USA standard).
Material and Methods:

Decision model

An 8-(health)state Markov decision model, with yearly cycle length, was developed in accordance with the 2009 American Thyroid Association (ATA) guidelines for 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 cat. III and IV) was simulated being either driven by diagnostic thyroid surgery (surgery), a molecular test aiming at prevention of unnecessary surgery (GEC), a molecular test aiming at 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 8 potential 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 summarised 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.

As 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% (i.e. the weighted mean of all three study populations (6, 10, 12)) and the test sensitivities and specificities as stated in the original references.
Risk and probability estimation

The duration of each Markov cycle was considered to be one year, therefore transition between healthstates 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. 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 timeframe of one 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). These prices include reimbursement tariffs for the molecular test, FDG-PET/CT, to physicians, anaesthesia, pathology, laboratory investigations, US procedures, thyroid surgery, RRA, medication, hospital facilities and all other costs incurred during inpatient and outpatient treatment. Costs of both transient and permanent complications were based on estimates from literature; its wide distribution reflects the variety of severity of these complications.

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

\[1 \text{ January } 2013: € 1.00000 = \text{US$ 1.31139 = CAN$ 1.32909.}\]
All 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 microsimulation of 100,000 first-order trials (patients) for 10,000 second-order parameter samples over five cycles. A five year evaluation period was chosen as most costs (and HRQoL losses) are made in the first years, the ATA guidelines (3) have difficulty in providing recommendations after the first five years, over 50% of recurrences occur in the first two years (4, 25) and limited data to the fact of probabilities, costs and effects after the first five years are different between scenarios. Half-cycle correction was applied.

Results are displayed in a cost-effectiveness plane (26).

One-way sensitivity analyses were performed to explore the variation of base-case model parameters on their range of extremes (10,000 hypothetical patients, one thousand 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 five-year period for each scenario were used to compute the incremental Net Monetary Benefit (iNMB in €):

\[ iNMB = \lambda \cdot (E_2 - E_1) - (C_2 - C_1) \] Eq. 1

Where \( \lambda \) is the willingness-to-pay threshold, \( E \) are the effects (utilities) and \( C \) are the costs of both scenarios under comparison. The subscript 1 denotes the comparator (either 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: \( iNMB > 0 \).
The Dutch Council for Public Health and Health Care recommends a willingness-to-pay threshold of €
80,000/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>0 for a
wide willingness-to-pay range, is displayed.

Modelling and Monte-Carlo analysis was performed using TreeAge Pro Suite (version 2011, TreeAge
Software Inc., Williamstown, MA, USA). Data analyses were performed using Matlab (version
R2013a, MathWorks, Natick, MA, USA).
Results:

Base case cost-effectiveness analysis:

After 5 years of treatment for and follow-up after an FNAC-indeterminate thyroid nodule, mean discounted costs were €8,804 (surgery), €9,341 (GEC), €8,913 (MMP) and €7,983 (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 to surgery in all patients. Compared to GEC and MMP, it was €1,358 and €930 less expensive with slight differences in HRQoL over 5 year. The mean iNMB was €3,684 compared to surgery, €1,030 compared to GEC and €3,851 compared to MMP (table 3). The robustness of these findings is displayed in the cost-effectiveness plane in figure 3: all of the 10,000 projections actually show a reduction of costs of FDG-PET/CT compared to the other three strategies. None of these 10,000 simulations indicated that FDG-PET/CT would be more costly and less effective, less costly and less effective or more costly and more effective except in comparison with the GEC, where PET showed a lower HRQoL of 0.0040 QALY (i.e. 1.5 Quality-Adjusted Life Day). 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 versus any of the other 3 strategies (supplemental figure 1).

The fraction of futile surgery of histologically benign, FNAC-indeterminate thyroid nodules, for 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 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), i.e. almost halving unbeneficial surgery and surgery-related complications. Mean five-year overall and recurrence free survival in this population were 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 versus either thyroid surgery in all patients or MMP (in both mean incremental utility: -0.10 QALY) leading to a mean iNMB of € -7,418 (versus surgery) and €-7,264 (versus MMP). At a value for the utility attributed to watchful surveillance of 0.953 (versus surgery) or 0.952 (versus MMP), the mean iNMB equals € 0. For comparison, the utility attributed to the healthstate 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 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 work-up 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 figures 2a-c).
Discussion:

We presented an economic decision analytical model, forecasting that implementation of FDG-PET/CT in the work-up of FNAC-indeterminate thyroid nodules could lead to substantial reduction in direct medical costs and, compared to two of the three alternatives, modestly improvement of patients HRQoL over the duration of 5 years.

Avoidance of (complications of) unnecessary thyroid surgery to provide a definite histopathological diagnosis is the principal cause cost-reduction. The fraction of surgeries performed for a benign thyroid nodule could almost be halved when fully implementing FDG-PET/CT compared to thyroid surgery in all FNAC-indeterminate thyroid nodules (40.3% and 75.0%, respectively). As it is estimated that 60,220 men and women are diagnosed with DTC in the USA in 2013 (2) and about half are found after surgery for FNAC-indeterminate nodules (29, 30), it can be roughly estimated that 120,000 patients undergo thyroid surgery for a FNAC-indeterminate thyroid nodule, of whom 90,000 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 USA only, assuming FNAC-indeterminacy was the sole reason for thyroid surgery. Compared to the 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 draw-back, the somewhat lower specificity of FDG-PET/CT compared to 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).

We found a higher mean 5-year discounted costs of €8,913 (MMP) compared to €8,804 (surgery). This is similar to the published economic analysis (12), which describes an additional US$104 to the overall cost of nodule evaluation only. The numerical difference can be explained by the fact that Yip et al. allowed a second FNAC in case of a negative MMP, which is able to revoke FNAC-indeterminacy and thus futile surgery.

Compared to the economic analysis of the GEC (10), we found a higher mean 5-year discounted costs of €9,341 (GEC) compared to €8,804 (surgery), while these authors describe a lower economic burden when adopting the GEC (US$ 10,719 compared to US$ 12,172). The main reason explaining
our contrary conclusion is that we attribute lower values for surgery and surgery-related costs than
they do. E.g. in our model uncomplicated hemithyroidectomy plus 5-year follow-up would cost €
5,499, but in their model this would be US$ 10,319 (€ 8,311, indexed to January 2013). As 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 2 diagnostic tests only, while in there model it
saves enough to pay for over 3 GEC’s. 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 are not known for delayed
treatment therefore no additional costs or detrimental effects are incorporated in 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 upon the
transition from localized to regional disease (5-year 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 FWHM for the PET-scanners used in the meta-analysis). With state-of-art time-of-flight technology (3-4 mm FWHM) it is likely that sensitivity, and thus negative predictive value, 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, as 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 as 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 non-uniform discount rate for costs and utilities. As non-uniform 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%/year 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 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 (e.g. sick-leave days, decreased productivity, and money spent on care outside of 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. As these relations are complex and as it is impossible to accurately substantiate any assumption as to the quantitative relation 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). 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, as other reason 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 only 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 as 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 work-up of adult patients with thyroid nodules scheduled for surgery for FNAC-indeterminacy (i.e. cellular atypia, follicular neoplasia) could lead to a decrease in costs and a moderate increase in HRQoL compared to diagnostic surgery in all patients according to current European practice and is competitive to the current USA 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 robustness of these data. Prospective studies are needed to further support cost-effectiveness, implementability and to gain insight in false-positivity of FDG-PET/CT. Prospective head-to-head comparison to alternative strategies or combinations of strategies should be considered.
Acknowledgments:

The PhD project of D. Vriens was funded in part by The Netherlands Organisation for Health Research and Development (ZonMW).
References:


4. Oncoline 2007 Evidence Based Nation-wide Guideline Thyroid Carcinoma version 1.1. In: Integraal Kankercentrum Nederland IKNL

5. Cibas ES, Ali SZ 2009 The Bethesda System for Reporting Thyroid Cytopathology. Thyroid 19:1159-1165


unfavorable prognosis: an immunohistochemical study. Thyroid 12:747-754


51. Lee J, Park JH, Lee CR, Chung WY, Park CS 2013 Long-term outcomes of total thyroidectomy versus thyroid lobectomy for papillary thyroid microcarcinoma: Comparative analysis after propensity score matching. Thyroid

52. Mazzaferri EL, Jhiang SM 1994 Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer. The American journal of medicine 97:418-428

54. van Roosmalen J, van Hemel B, Suurmeijer A, Groen H, Ruitenbeek T, Links TP, Plukker JT 2010 Diagnostic value and cost considerations of routine fine-needle aspirations in the follow-up of thyroid nodules with benign readings. Thyroid 20:1359-1365


57. Mernagh P, Campbell S, Dietlein M, Luster M, Mazzaferri E, Weston AR 2006 Cost-effectiveness of using recombinant human TSH prior to radioiodine ablation for thyroid cancer, compared with treating patients in a hypothyroid state: the German perspective. European journal of endocrinology / European Federation of Endocrine Societies 155:405-414


**Figure Legends:**

**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 healthstates based on this decision tree (figure 2). Diamonds are decision nodes, decision are based on probabilities. Boxes are interventions and cost money. (c)TT: (completion) Total Thyroidectomy; FDG-PET/CT: FluoroDeoxyGlucose Positron Emission Tomography / Computed Tomography; FNAC: Fine-Needle Aspiration Cytology; GEC: Gene-Expression Classifier; HT: HemiThyroidectomy; MMP: Molecular Marker Panel; MT: Molecular Test; PA: histoPathology; RRA: Radioactive iodine-131 thyroid Remnant Ablation; TN: Thyroid Nodule; UPM: Unifocal Papillary Microcarcinoma

**Figure 2:** Markov tree. Simulated patients with FNAC-indeterminate TNs will end up in any of these eight healthstates (ellipses), based on the decision tree (figure 1). After each cycle duration (1 year), transitions to other healthstates may occur (arrows, transition probabilities). In case of recurrence after HT intervention (box) can take place, which has a certain decision (diamond). During surveillance after a negative FDG-PET/CT at some point suspicion for malignancy might arise and patient will undergo (diagnostic) surgery after all. Decision nodes are based on probabilities, interventions cost money, healthstates cost money and have a certain HRQoL. (c)TT: (completion) Total Thyroidectomy; FNAC: Fine-Needle Aspiration Cytology; HRQoL: Health Related Quality of Life; HT: HemiThyroidectomy; RRA: Radioactive iodine-131 thyroid Remnant Ablation; TN: Thyroid Nodule.
**Figure 3:** Incremental costs – incremental utility plot (cost-effectiveness plane) comparing FDG-PET/CT-driven treatment to current practice. Each of the 10,000 dots represents the mean value of 100,000 simulated patients. The left upper quadrant represents situations where the novel strategy is less effective but more costly than the conventional treatment, the right upper quadrant: more effective and more costly, left lower quadrant: less effective and less costly and the right lower quadrant: more effective but less costly (100%). The oblique lines represent a willingness-to-pay threshold of €20,000/QALY and €80,000/QALY, respectively. 50%-5, 75%-5, 95%-5 and 99% confidence ellipses are drawn. QALY: Quality-adjusted life year.

**Figure 4:** Tornado plots showing the results of one-way sensitivity analyses of top-10 inputs of the model on the iNMB of FDG-PET/CT versus one of the other three strategies (upper panel: surgery, middle panel: GEC, lower panel: MMP), for a willingness-to-pay threshold (λ) 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.
Table Legends and Tables:

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 (SA). CI: Confidence Interval; (c)TT: (completion) Total Thyroidectomy; Dist.: parameter stochastic distribution; DTC: Differentiated Thyroid Carcinoma; EO: Expert Opinion; FDG-PET/CT: FluoroDeoxyGlucose Positron Emission Tomography / Computed Tomography; GEC: Gene-Expression Classifier; HT: HemiThyroidectomy; MMP: Molecular Marker Panel; NPV: Negative Predictive Value; UPM: Unifocal Papillary Microcarcinoma.

Table 2: Base case parameter values and distributions for costs and utilities, including range used for one-way sensitivity analyses (SA). All cost parameters were assumed to be of gamma distributions and all utility parameters beta distributions. CI: Confidence Interval; (c)TT: (completion) Total Thyroidectomy; DOT: the system of imbursement of the Dutch Healthcare Authority; EO: Expert Opinion; GEC: Gene-Expression Classifier; FDG-PET/CT: FluoroDeoxyGlucose Positron Emission Tomography / Computed Tomography; HT: HemiThyroidectomy; MMP: Molecular Marker Panel; NZa: Dutch Healthcare Authority; QALY: Quality-Adjusted Life Year; RRA: Radioactive iodine-131 thyroid Remnant Ablation; SA: one-way Sensitivity Analysis.

Table 3: Base case main results. CI: confidence interval; GEC: Gene-Expression Classifier; FDG-PET/CT: FluoroDeoxyGlucose Positron Emission Tomography / Computed Tomography; iNMB: incremental Net Monetary Benefit (using a Willingness-to-Pay threshold ($\lambda$) of €80,000/QALY); MMP: Molecular Marker Panel; N/A: not applicable; QALY: quality-adjusted life year;
Table 1:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Distr.</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</td>
<td>Beta</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 [year]</td>
<td>Gamma</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 [year]</td>
<td>Gamma</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</td>
<td>Beta</td>
<td>0.00000031 (0.0000021 - 0.0000043)</td>
<td>(4)</td>
<td>-</td>
</tr>
<tr>
<td>Incidence of DTC in healthy males</td>
<td>Beta</td>
<td>0.0000013 (0.00000069 - 0.0000013)</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>Age/sex dependent</td>
<td>(42)</td>
<td>-</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</td>
<td>Beta</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</td>
<td>Beta</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>Diagnostic test characteristics:</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>(6, 13-15, 37-41)</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>(7, 10)</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>(7, 10)</td>
<td>0.65 - 0.99</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>(6, 10-15, 37-41)</td>
<td>0.75 - 0.99</td>
</tr>
<tr>
<td>Yearly probability of surgery after surveillance:</td>
<td>Beta</td>
<td>0.0070 (0.0014 - 0.021)</td>
<td>Computed (=1-NPV$^{0.2}$)</td>
<td>0.00 - 0.05</td>
</tr>
<tr>
<td>After negative FDG-PET/CT</td>
<td>Beta</td>
<td>0.0100 (0.0035 - 0.023)</td>
<td>Computed (=1-NPV$^{0.2}$)</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</td>
<td>Beta</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</td>
<td>Beta</td>
<td>0.088 (0.069 - 0.11)</td>
<td>(10, 16, 43-48)</td>
<td>0.01 - 0.26</td>
</tr>
<tr>
<td>Fraction of transient complications due to (c)TT</td>
<td>Beta</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</td>
<td>Beta</td>
<td>0.038 (0.023 - 0.056)</td>
<td>(10, 16, 43-48)</td>
<td>0.01 - 0.25</td>
</tr>
<tr>
<td>Fraction of death due to any type of surgery</td>
<td>Beta</td>
<td>0.0019 (0.00091 - 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</td>
<td>Beta</td>
<td>0.0047 (0.00020 - 0.016)</td>
<td>(51)</td>
<td>0.001 - 0.025</td>
</tr>
<tr>
<td>Yearly probability of cTT after recurrence after HT</td>
<td>Beta</td>
<td>0.917 (0.889 - 0.940)</td>
<td>(10)</td>
<td>0.90 - 1.00</td>
</tr>
<tr>
<td>Yearly probability of recurrence after (c)TT</td>
<td>Beta</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</td>
<td>Beta</td>
<td>0.0051 (0.0020 - 0.0095)</td>
<td>(10, 52)</td>
<td>0.00 - 0.01</td>
</tr>
<tr>
<td>Variable</td>
<td>Costs [€] Expected Value (95%-CI)</td>
<td>Source</td>
<td>Range for SA: Expected Value (95%-CI)</td>
<td>Source</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-----------------------------------</td>
<td>--------</td>
<td>---------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td><strong>Procedures:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDG-PET/CT</td>
<td>€ 1,002 (€ 816 - € 1,208)</td>
<td>N Za</td>
<td>€ 800 - € 1,200</td>
<td></td>
</tr>
<tr>
<td>GEC</td>
<td>€ 2,577 (€ 2,097 - € 3,106)</td>
<td>(10)</td>
<td>€ 1,611 - € 4,026</td>
<td></td>
</tr>
<tr>
<td>MMP</td>
<td>€ 523 (€ 426 - € 631)</td>
<td>(12)</td>
<td>€ 400 - € 650</td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>€ 4,419 (€ 3,595 - € 5,326)</td>
<td>(10, 12, 44, 53-56), DOT</td>
<td>€ 2,994 - € 16,878</td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>€ 6,238 (€ 5,075 - € 7,518)</td>
<td>(10, 12, 44, 53-57), DOT</td>
<td>€ 3,433 - € 20,796</td>
<td></td>
</tr>
<tr>
<td>cTT</td>
<td>€ 6,618 (€ 5,385 - € 7,977)</td>
<td>(10, 12, 53-55), DOT</td>
<td>€ 3,952 - € 16,878</td>
<td></td>
</tr>
<tr>
<td>RRA</td>
<td>€ 2,479 (€ 2,017 - € 2,987)</td>
<td>(44, 57-59), DOT</td>
<td>€ 1,277 - € 2,692</td>
<td></td>
</tr>
<tr>
<td><strong>Healthstates:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surveillance after FDG-PET/C or GEC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st year</td>
<td>€ 488 (€ 397 - € 589)</td>
<td>(10), DOT</td>
<td>€ 228 - € 889</td>
<td></td>
</tr>
<tr>
<td>2nd-5th year</td>
<td>€ 314 (€ 256 - € 379)</td>
<td>(10), DOT</td>
<td>€ 0 - € 493</td>
<td></td>
</tr>
<tr>
<td>Surveillance after HT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st year</td>
<td>€ 1,080 (€ 879 - € 1,077)</td>
<td>(10, 44), DOT</td>
<td>€ 317 - € 1,208</td>
<td></td>
</tr>
<tr>
<td>2nd-5th year</td>
<td>€ 0 (€ 0 - € 0)</td>
<td>EO, (10, 44), DOT</td>
<td>€ 0 - € 725</td>
<td></td>
</tr>
<tr>
<td>Transient complication due to HT</td>
<td>€ 645 (€ 525 - € 778)</td>
<td>(10, 45)</td>
<td>€ 188 - € 5,280</td>
<td></td>
</tr>
<tr>
<td>Permanent complication due to HT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st year</td>
<td>€ 4,441 (€ 3,613 - € 5,533)</td>
<td>(10), DOT</td>
<td>€ 3,123 - € 4,993</td>
<td></td>
</tr>
<tr>
<td>2nd-5th year</td>
<td>€ 772 (€ 628 - € 931)</td>
<td>(10), DOT</td>
<td>€ 55 - € 886</td>
<td></td>
</tr>
<tr>
<td>Recurrence after HT</td>
<td>€ 1,630 (€ 1,326 - € 1,964)</td>
<td>(10), DOT</td>
<td>€ 326 - € 2,013</td>
<td></td>
</tr>
<tr>
<td>Surveillance after (c)TT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st year</td>
<td>€ 1,321 (€ 1,075 - € 1,592)</td>
<td>(10, 44), DOT</td>
<td>€ 274 - € 1,772</td>
<td></td>
</tr>
<tr>
<td>2nd-5th year</td>
<td>€ 699 (€ 569 - € 842)</td>
<td>(10, 44), DOT</td>
<td>€ 180 - € 954</td>
<td></td>
</tr>
<tr>
<td>Transient complication due to (c)TT</td>
<td>€ 645 (€ 525 - € 778)</td>
<td>(10), DOT</td>
<td>€ 188 - € 5,154</td>
<td></td>
</tr>
<tr>
<td>Permanent complication due to (c)TT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st year</td>
<td>€ 5,282 (€ 4,298 - € 6,367)</td>
<td>(10), DOT</td>
<td>€ 3,724 - € 9,825</td>
<td></td>
</tr>
<tr>
<td>2nd-5th year</td>
<td>€ 899 (€ 732 - € 1,084)</td>
<td>(10), DOT</td>
<td>€ 317 - € 1,773</td>
<td></td>
</tr>
<tr>
<td>Recurrence after (c)TT</td>
<td>€ 1,347 (€ 1,096 - € 1,623)</td>
<td>(10), DOT</td>
<td>€ 326 - € 2,184</td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>€ 0 (€ 0 - € 0)</td>
<td>Convention</td>
<td>0 (0 - 0)</td>
<td></td>
</tr>
</tbody>
</table>
Table 3:

<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>€ 8,804 (€ 8,774 - € 8,835)</td>
<td>€ 9,341 (€ 9,300 - € 9,383)</td>
<td>€ 8,913 (€ 8,884 - € 8,942)</td>
<td>€ 7,983 (€ 7,941 - € 8,025)</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>
</tbody>
</table>

Incremental values of FDG-PET/CT compared to alternative strategy:

| Incremental costs [€] | € -822 (€ -871 - € -772) | € -1,358 (€ -1,377 - € -1,340) | € -930 (€ -970 - € -890) | N/A |
| Incremental utilities [QALY] | 0.036 (0.031 - 0.041) | -0.0040 (-0.0050 - -0.0030) | 0.037 (0.033 - 0.041) | N/A |
| iNMB [€] | € 3,684 (€ 3,278 - € 4,094) | € 1,030 (€ 916 - € 1,142) | € 3,851 (€ 3,528 - € 4,170) | N/A |
| Incremental futile surgery [%] | -34.7% (-34.3% - -35.2%) | +2.1% (+2.0% - +2.3%) | -34.7% (-34.3% - -35.2%) | N/A |
Supplemental Figure Legends

**Supplemental Figure 1:** Cost-effectiveness acceptability curves, plotting the probability of a positive iNMB (P(iNMB>€ 0) for a range of values for the willingness-to-pay threshold (λ). 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 (λ) 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.