Title Page:

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

Authors: Vriens D, MD[1]; Adang EMM, PhD[2]; Netea-Maier RT, MD, PhD[3]; Smit JWA, MD, PhD[4]; de Wilt JHW, MD, PhD[5]; Oyen WJG, MD, PhD[1]; de Geus-Oei LF, MD, PhD[1].

[1]dept. of Radiology and Nuclear Medicine, [2]Health Evidence, [3]Internal Medicine section Endocrinology, [4]Internal Medicine and [5]Surgery; 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; EMMA: 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 unbeneficial 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 healthstates. These healthstates 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)[1]. 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] January 2013: € 1.00000 = 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 micosimulation 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 €):

$$iNMB = \lambda \cdot (E_2 - E_1) - (C_2 - C_1) \qquad \text{Eq. 1}$$

Where λ 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 (5-8mm FWHM for the PET-scanners used in the meta-analysis). With state-of-art time-of-flight technology (3-4mm 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:

1. **Guth S, Theune U, Aberle J, Galach A, Bamberger CM** 2009 Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. European journal of clinical investigation 39:699-706
2. **Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA, (eds.)** 2013 SEER Cancer Statistics Review, 1975-2010,. In. http://seer.cancer.gov/csr/1975_2010/, based on November 2012 SEER data submission, posted to the SEER web site, 2013.: National Cancer Institute, Bethesda, MD.
3. **Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Pacini F, Schlumberger M, Sherman SI, Steward DL, Tuttle RM** 2009 Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer (American Thyroid Association Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer). Thyroid 19:1167-1214
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
6. **Vriens D, de Wilt JH, van der Wilt GJ, Netea-Maier RT, Oyen WJ, de Geus-Oei LF** 2011 The role of [18F]-2-fluoro-2-deoxy-d-glucose-positron emission tomography in thyroid nodules with indeterminate fine-needle aspiration biopsy: systematic review and meta-analysis of the literature. Cancer 117:4582-4594
7. **Alexander EK, Kennedy GC, Baloch ZW, Cibas ES, Chudova D, Diggans J, Friedman L, Kloos RT, LiVolsi VA, Mandel SJ, Raab SS, Rosai J, Steward DL, Walsh PS, Wilde JI, Zeiger MA, Lanman RB, Haugen BR** 2012 Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. N Engl J Med 367:705-715
8. **Jameson JL** 2012 Minimizing unnecessary surgery for thyroid nodules. N Engl J Med 367:765-767
9. **Walsh PS, Wilde JI, Tom EY, Reynolds JD, Chen DC, Chudova DI, Pagan M, Pankratz DG, Wong M, Veitch J, Friedman L, Monroe R, Steward DL, Lupo MA, Lanman RB, Kennedy GC** 2012 Analytical performance verification of a molecular diagnostic for cytology-indeterminate thyroid nodules. J Clin Endocrinol Metab 97:E2297-2306

10. **Li H, Robinson KA, Anton B, Saldanha IJ, Ladenson PW** 2011 Cost-effectiveness of a novel molecular test for cytologically indeterminate thyroid nodules. J Clin Endocrinol Metab 96:E1719-1726
11. **Nikiforov YE, Ohori NP, Hodak SP, Carty SE, LeBeau SO, Ferris RL, Yip L, Seethala RR, Tublin ME, Stang MT, Coyne C, Johnson JT, Stewart AF, Nikiforova MN** 2011 Impact of mutational testing on the diagnosis and management of patients with cytologically indeterminate thyroid nodules: a prospective analysis of 1056 FNA samples. J Clin Endocrinol Metab 96:3390-3397
12. **Yip L, Farris C, Kabaker AS, Hodak SP, Nikiforova MN, McCoy KL, Stang MT, Smith KJ, Nikiforov YE, Carty SE** 2012 Cost impact of molecular testing for indeterminate thyroid nodule fine-needle aspiration biopsies. J Clin Endocrinol Metab 97:1905-1912
13. **de Geus-Oei LF, Pieters GF, Bonenkamp JJ, Mudde AH, Bleeker-Rovers CP, Corstens FH, Oyen WJ** 2006 18F-FDG PET reduces unnecessary hemithyroidectomies for thyroid nodules with inconclusive cytologic results. J Nucl Med 47:770-775
14. **Deandreis D, Al Ghuzlan A, Auperin A, Vielh P, Caillou B, Chami L, Lumbroso J, Travagli JP, Hartl D, Baudin E, Schlumberger M, Leboulleux S** 2012 Is (18)F-fluorodeoxyglucose-PET/CT useful for the presurgical characterization of thyroid nodules with indeterminate fine needle aspiration cytology? Thyroid 22:165-172
15. **Munoz Perez N, Villar Del Moral JM, Muros Fuentes MA, Lopez de la Torre M, Arcelus Martinez JI, Becerra Massare P, Esteva Martinez D, Canadas Garre M, Coll Del Rey E, Bueno Larano P, Ferron Orihuela JA** 2013 Could F-FDG-PET/CT avoid unnecessary thyroidectomies in patients with cytological diagnosis of follicular neoplasm? Langenbeck's archives of surgery / Deutsche Gesellschaft fur Chirurgie
16. **McHenry CR, Slusarczyk SJ** 2000 Hypothyroidisim following hemithyroidectomy: incidence, risk factors, and management. Surgery 128:994-998
17. **Moon HG, Jung EJ, Park ST, Jung TS, Jeong CY, Ju YT, Lee YJ, Hong SC, Choi SK, Ha WS** 2008 Thyrotropin level and thyroid volume for prediction of hypothyroidism following hemithyroidectomy in an Asian patient cohort. World J Surg 32:2503-2508
18. **Shaha AR** 2007 TNM classification of thyroid carcinoma. World J Surg 31:879-887
19. Dutch Consumer Price Index. In: http://statlinecbsnl/StatWeb/publication/?VW=T&DM=SLNL&PA=71311ned, accessed 14/04/2013
20. US Consumer Price Index. In: ftp://ftpblsgov/pub/specialrequests/cpi/cpiaitxt, accessed 14/04/2013

21. Canadian Consumer Price Index. In: http://wwwstatcangcca/tables-tableaux/sum-som/l01/cst01/econ46a-enghtm, accessed 14/04/2013
22. German Consumer Price Index. In: http://wwwrateinflationcom/consumer-price-index/germany-historical-cpi, accessed 14/04/2013
23. Currency Exchange Rates. In: http://wwwxecom, accessed 14/04/2013
24. **Hakkaart-van Roijen L, Tan SS, Bouwmans CAM** 2010 Manual for Cost Research.Dutch manual, update appeared in 2004 and 2010. Published by the Dutch Government. Diemen: College voor Zorgverzekeringen
25. **Molinaro E, Giani C, Agate L, Biagini A, Pieruzzi L, Bianchi F, Brozzi F, Ceccarelli C, Viola D, Piaggi P, Vitti P, Pacini F, Elisei R** 2013 Patients with differentiated thyroid cancer who underwent radioiodine thyroid remnant ablation with low-activity (1)(3)(1)I after either recombinant human TSH or thyroid hormone therapy withdrawal showed the same outcome after a 10-year follow-up. J Clin Endocrinol Metab 98:2693-2700
26. **Nixon RM, Wonderling D, Grieve RD** 2010 Non-parametric methods for cost-effectiveness analysis: the central limit theorem and the bootstrap compared. Health economics 19:316-333
27. **Woertman WH, Van De Wetering G, Adang EM** 2013 Cost-Effectiveness on a Local Level: Whether and When to Adopt a New Technology. Medical decision making : an international journal of the Society for Medical Decision Making
28. **Pomp M, Brouwer W, Rutten F** 2007 QALY-time: New medical Technology, cost-effectiveness and guidelines ["QALY-tijd: Nieuwe medische technologie, kosteneffectiviteit en richtlijnen"] (CPB Document 152). In:
29. **Lew JI, Snyder RA, Sanchez YM, Solorzano CC** 2011 Fine needle aspiration of the thyroid: correlation with final histopathology in a surgical series of 797 patients. Journal of the American College of Surgeons 213:188-194; discussion 194-185
30. **Wang CC, Friedman L, Kennedy GC, Wang H, Kebebew E, Steward DL, Zeiger MA, Westra WH, Wang Y, Khanafshar E, Fellegara G, Rosai J, Livolsi V, Lanman RB** 2011 A large multicenter correlation study of thyroid nodule cytopathology and histopathology. Thyroid 21:243-251
31. **Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ** 2011 Discounting and decision making in the economic evaluation of health-care technologies. Health economics 20:2-15
32. **Schonberger J, Ruschoff J, Grimm D, Marienhagen J, Rummele P, Meyringer R, Kossmehl P, Hofstaedter F, Eilles C** 2002 Glucose transporter 1 gene expression is related to thyroid neoplasms with an

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

33. **Hooft L, van der Veldt AA, van Diest PJ, Hoekstra OS, Berkhof J, Teule GJ, Molthoff CF** 2005 [18F]fluorodeoxyglucose uptake in recurrent thyroid cancer is related to hexokinase i expression in the primary tumor. J Clin Endocrinol Metab 90:328-334
34. **Kim BH, Kim IJ, Kim SS, Kim SJ, Lee CH, Kim YK** 2010 Relationship between biological marker expression and fluorine-18 fluorodeoxyglucose uptake in incidentally detected thyroid cancer. Cancer biotherapy & radiopharmaceuticals 25:309-315
35. **Kaida H, Hiromatsu Y, Kurata S, Kawahara A, Hattori S, Taira T, Kobayashi M, Uchida M, Yamada K, Mihashi H, Umeno H, Kage M, Nakashima T, Hayabuchi N, Ishibashi M** 2011 Relationship between clinicopathological factors and fluorine-18-fluorodeoxyglucose uptake in patients with papillary thyroid cancer. Nucl Med Commun 32:690-698
36. **Deichen JT, Schmidt C, Prante O, Maschauer S, Papadopoulos T, Kuwert T** 2004 Influence of TSH on uptake of [18F]fluorodeoxyglucose in human thyroid cells in vitro. Eur J Nucl Med Mol Imaging 31:507-512
37. **Kresnik E, Gallowitsch HJ, Mikosch P, Stettner H, Igerc I, Gomez I, Kumnig G, Lind P** 2003 Fluorine-18-fluorodeoxyglucose positron emission tomography in the preoperative assessment of thyroid nodules in an endemic goiter area. Surgery 133:294-299
38. **Kim JM, Ryu JS, Kim TY, Kim WB, Kwon GY, Gong G, Moon DH, Kim SC, Hong SJ, Shong YK** 2007 18F-fluorodeoxyglucose positron emission tomography does not predict malignancy in thyroid nodules cytologically diagnosed as follicular neoplasm. J Clin Endocrinol Metab 92:1630-1634
39. **Sebastianes FM, Cerci JJ, Zanoni PH, Soares J, Jr., Chibana LK, Tomimori EK, de Camargo RY, Izaki M, Giorgi MC, Eluf-Neto J, Meneghetti JC, Pereira MA** 2007 Role of 18F-fluorodeoxyglucose positron emission tomography in preoperative assessment of cytologically indeterminate thyroid nodules. J Clin Endocrinol Metab 92:4485-4488
40. **Hales NW, Krempl GA, Medina JE** 2008 Is there a role for fluorodeoxyglucose positron emission tomography/computed tomography in cytologically indeterminate thyroid nodules? Am J Otolaryngol 29:113-118
41. **Traugott AL, Dehdashti F, Trinkaus K, Cohen M, Fialkowski E, Quayle F, Hussain H, Davila R, Ylagan L, Moley JF** 2010 Exclusion of malignancy in thyroid nodules with indeterminate fine-needle aspiration cytology after negative 18F-fluorodeoxyglucose positron emission tomography: interim analysis. World J Surg 34:1247-1253
42. Central Bureau for Statistics: Life-tables. In: http://statlinecbsnl/StatWeb/publication/?DM=SLNL&PA=37360ned&D

1=0&D2=a&D3=a&D4=l&HDR=G1,T&STB=G2,G3&VW=T, accessed 14/04/2013
43. **Rosato L, Avenia N, Bernante P, De Palma M, Gulino G, Nasi PG, Pelizzo MR, Pezzullo L** 2004 Complications of thyroid surgery: analysis of a multicentric study on 14,934 patients operated on in Italy over 5 years. World J Surg 28:271-276
44. **Shrime MG, Goldstein DP, Seaberg RM, Sawka AM, Rotstein L, Freeman JL, Gullane PJ** 2007 Cost-effective management of low-risk papillary thyroid carcinoma. Arch Otolaryngol Head Neck Surg 133:1245-1253
45. **Vidal-Trecan GM, Stahl JE, Eckman MH** 2004 Radioiodine or surgery for toxic thyroid adenoma: dissecting an important decision. A cost-effectiveness analysis. Thyroid 14:933-945
46. **Lee YS, Nam KH, Chung WY, Chang HS, Park CS** 2010 Postoperative complications of thyroid cancer in a single center experience. Journal of Korean medical science 25:541-545
47. **Spanknebel K, Chabot JA, DiGiorgi M, Cheung K, Curty J, Allendorf J, LoGerfo P** 2006 Thyroidectomy using monitored local or conventional general anesthesia: an analysis of outpatient surgery, outcome and cost in 1,194 consecutive cases. World J Surg 30:813-824
48. **Stoll SJ, Pitt SC, Liu J, Schaefer S, Sippel RS, Chen H** 2009 Thyroid hormone replacement after thyroid lobectomy. Surgery 146:554-558; discussion 558-560
49. **Esnaola NF, Cantor SB, Sherman SI, Lee JE, Evans DB** 2001 Optimal treatment strategy in patients with papillary thyroid cancer: a decision analysis. Surgery 130:921-930
50. **Hundahl SA, Cady B, Cunningham MP, Mazzaferri E, McKee RF, Rosai J, Shah JP, Fremgen AM, Stewart AK, Holzer S** 2000 Initial results from a prospective cohort study of 5583 cases of thyroid carcinoma treated in the united states during 1996. U.S. and German Thyroid Cancer Study Group. An American College of Surgeons Commission on Cancer Patient Care Evaluation study. Cancer 89:202-217
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
53. **Hooft L, Hoekstra OS, Boers M, Van Tulder MW, Van Diest P, Lips P** 2004 Practice, efficacy, and costs of thyroid nodule evaluation: a retrospective study in a Dutch university hospital. Thyroid 14:287-293

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
55. **Heller M, Zanocco K, Zydowicz S, Elaraj D, Nayar R, Sturgeon C** 2012 Cost-effectiveness analysis of repeat fine-needle aspiration for thyroid biopsies read as atypia of undetermined significance. Surgery 152:423-430
56. **Zanocco K, Heller M, Elaraj D, Sturgeon C** 2012 Is subtotal thyroidectomy a cost-effective treatment for Graves disease? A cost-effectiveness analysis of the medical and surgical treatment options. Surgery 152:164-172
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
58. **Wang TS, Cheung K, Mehta P, Roman SA, Walker HD, Sosa JA** 2010 To stimulate or withdraw? A cost-utility analysis of recombinant human thyrotropin versus thyroxine withdrawal for radioiodine ablation in patients with low-risk differentiated thyroid cancer in the United States. J Clin Endocrinol Metab 95:1672-1680
59. **Mernagh P, Suebwongpat A, Silverberg J, Weston A** 2010 Cost-effectiveness of using recombinant human thyroid-stimulating hormone before radioiodine ablation for thyroid cancer: the Canadian perspective. Value Health 13:180-187
60. **Kebebew E, Duh QY, Clark OH** 2000 Total thyroidectomy or thyroid lobectomy in patients with low-risk differentiated thyroid cancer: surgical decision analysis of a controversy using a mathematical model. World J Surg 24:1295-1302

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%-, 75%-, 95% - 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 (λ) of € 80,000/QALY); MMP: Molecular Marker Panel; N/A: not applicable; QALY: quality-adjusted life year;

Table 1:

Variable:	Distr.:	Expected Value (95%-CI)			Source:	Range for SA:
Base variables:						
Discount rate of costs	Fixed	0.040			(10, 24)	0.030 – 0.050
Discount rate of utilities	Fixed	0.015			(10, 24)	0.010 – 0.050
Population description:						
Fraction of female patients	Beta	0.86	(0.81	- 0.90)	(6, 10, 13, 37-41)	0.78 - 0.93
Age of female patient when diagnosed [year]	Gamma	47.3	(21.0	- 73.6)	(6, 10, 13, 37-41)	-
Age of male patient when diagnosed [year]	Gamma	55.6	(26.1	- 85.0)	(6, 10, 13, 37-41)	-
Incidence of DTC in healthy females	Beta	0.0000031	(0.0000021	- 0.0000043)	(4)	-
Incidence of DTC in healthy males	Beta	0.0000013	(0.00000069	- 0.0000013	(4)	-
Yearly probability of death of any cause (not cancer related)	Life-table	Age/sex dependent			(42)	-
General probabilities:						
Fraction HT of all surgery	Beta	0.95	(0.90	- 0.98)	EO	0.50 - 0.99
Fraction of UPM in indeterminate nodules	Beta	0.023	(0.0076	- 0.047)	(6, 10, 13, 37-41)	0.01 - 0.10
Prevalence of cancer in indeterminate nodules	Dirichlet	0.25	(0.22	- 0.28)	(6, 7, 10-15, 37-41)	0.15 - 0.35
Diagnostic test characteristics:						
FDG-PET/CT sensitivity	Dirichlet	0.95	(0.88	- 0.99)	(6, 13-15, 37-41)	0.70 - 0.99
FDG-PET/CT specificity	Dirichlet	0.48	(0.40	- 0.55)		0.35 - 0.70
GEC sensitivity	Dirichlet	0.92	(0.85	- 0.97)	(7, 10)	0.65 - 0.99
GEC specificity	Dirichlet	0.52	(0.44	- 0.59)		0.40 - 0.75
MMP sensitivity	Dirichlet	0.59	(0.49	- 0.69)	(11, 12)	0.35 - 0.70
MMP specificity	Dirichlet	0.98	(0.96	- 0.99)		0.75 - 0.99
Yearly probability of surgery after surveillance:						
After negative FDG-PET/CT	Beta	0.0070	(0.0014	- 0.021)	Computed *(=1-NPV$^{0.2}$)*	0.00 - 0.05
After negative GEC	Beta	0.010	(0.0035	- 0.023)	Computed *(=1-NPV$^{0.2}$)*	0.00 - 0.05
Complications of surgery:						
Fraction of transient complications due to HT	Beta	0.039	(0.0020	- 0.064)	(10, 16, 43-48)	0.01 - 0.60
Fraction of permanent complications due to HT	Beta	0.088	(0.069	- 0.11)	(10, 16, 43, 44, 47, 49)	0.01 - 0.26
Fraction of transient complications due to (c)TT	Beta	0.19	(0.10	- 0.30)	(10, 16, 43-48)	0.01 - 0.65
Fraction of permanent complications due to (c)TT	Beta	0.038	(0.023	- 0.056)	(10, 16, 43, 44, 47, 49)	0.01 - 0.25
Fraction of death due to any type of surgery	Beta	0.0019	(0.00091	- 0.0032)	(10, 50)	0.00 - 0.01
Recurrence/Cancer related death:						
Yearly probability of recurrence after HT for UPM	Beta	0.0047	(0.00020	- 0.016)	(51)	0.001 - 0.025
Yearly probability of cTT after recurrence after HT	Beta	0.917	(0.889	- 0.940)	(10)	0.90 - 1.00
Yearly probability of recurrence after (c)TT	Beta	0.027	(0.019	- 0.037)	(10, 52)	0.01 - 0.07
Yearly probability of death due to cancer	Beta	0.0051	(0.0020	- 0.0095)	(10, 52)	0.00 - 0.01

Table 2:

Variable:	Costs* [€] Expected Value (95%-CI)	Source	Range for SA:	Utility [QALY/yr] Expected Value (95%-CI)	Source	Range for SA:
Procedures:						
FDG-PET/CT	€ 1,002 (€ 816 - € 1,208)	NZa	€ 800 - € 1,200	-	-	
GEC	€ 2,577 (€ 2,097 - € 3,106)	(10)	€ 1,611 - € 4,026	-	-	
MMP	€ 523 (€ 426 - € 631)	(12)	€ 400 - € 650	-	-	
HT	€ 4,419 (€ 3,595 - € 5,326)	(10, 12, 44, 53-56), DOT	€ 2,994 - € 16,878	-	-	
TT	€ 6,238 (€ 5,075 - € 7,518)	(10, 12, 44, 53-57), DOT	€ 3,433 - € 20,796	-	-	
cTT	€ 6,618 (€ 5,385 - € 7,977)	(10, 12, 53-55), DOT	€ 3,952 - € 16,878	-	-	
RRA	€ 2,479 (€ 2,017 - € 2,987)	(44, 57-59), DOT	€ 1,277 - € 2,692	-	-	
Healthstates:						
Surveillance after FDG-PET/C or GEC				0.98 (0.95 - 1.00)	(10)	0.90 - 0.99
1st year	€ 488 (€ 397 - € 589)	(10), DOT	€ 228 - € 889			
2nd-5th year	€ 314 (€ 256 - € 379)	(10), DOT	€ 0 - € 493			
Surveillance after HT				0.99 (0.96 - 1.00)	(10, 49)	0.90 - 0.99
1st year	€ 1,080 (€ 879 - € 1,077)	(10, 44), DOT	€ 317 - € 1,208			
2nd-5th year	€ 0 (€ 0 - € 0)	EO, (10, 44), DOT	€ 0 - € 725			
Transient complication due to HT	€ 645 (€ 525 - € 778)t	(10, 45)	€ 188 - € 5,280	0.94 (0.89 - 0.98)	(10, 45)	0.90 - 0.99
Permanent complication due to HT				0.70 (0.61 - 0.79)	(10, 49, 60)	0.62 - 0.99
1st year	€ 4,441 (€ 3,613 - € 5,353)	(10)	€ 3,123 - € 4,993			
2nd-5th year	€ 772 (€ 628 - € 931)	(10)	€ 55 - € 886			
Recurrence after HT	€ 1,630 (€ 1,326 - € 1,964)	(10)	€ 326 - € 2,013	0.60 (0.50 - 0.69)	(10, 49, 60)	0.54 - 0.98
Surveillance after (c)TT				0.97 (0.93 - 0.99)	(10, 49)	0.90 - 0.99
1st year	€ 1,321 (€ 1,075 - € 1,592)	(10, 44), DOT	€ 274 - € 1,772			
2nd-5th year	€ 699 (€ 569 - € 842)	(10, 44), DOT	€ 180 - € 954			
Transient complication due to (c)TT	€ 645 (€ 525 - € 778)	(10)	€ 188 - € 5,154	0.94 (0.89 - 0.98)	(10, 45)	0.90 - 0.99
Permanent complication due to (c)TT				0.65 (0.55 - 0.74)	(10, 49, 60)	0.21 - 0.97
1st year	€ 5,282 (€ 4,298 - € 6,367)	(10)	€ 3,724 - € 9,825			
2nd-5th year	€ 899 (€ 732 - € 1,084)	(10)	€ 317 - € 1,773			
Recurrence after (c)TT	€ 1,347 (€ 1,096 - € 1,623)	(10)	€ 326 - € 2,184	0.60 (0.50 - 0.69)	(10, 49, 60)	0.54 - 0.98
Death	€ 0 (€ 0 - € 0)	Convention		0 (0 - 0)	Convention	

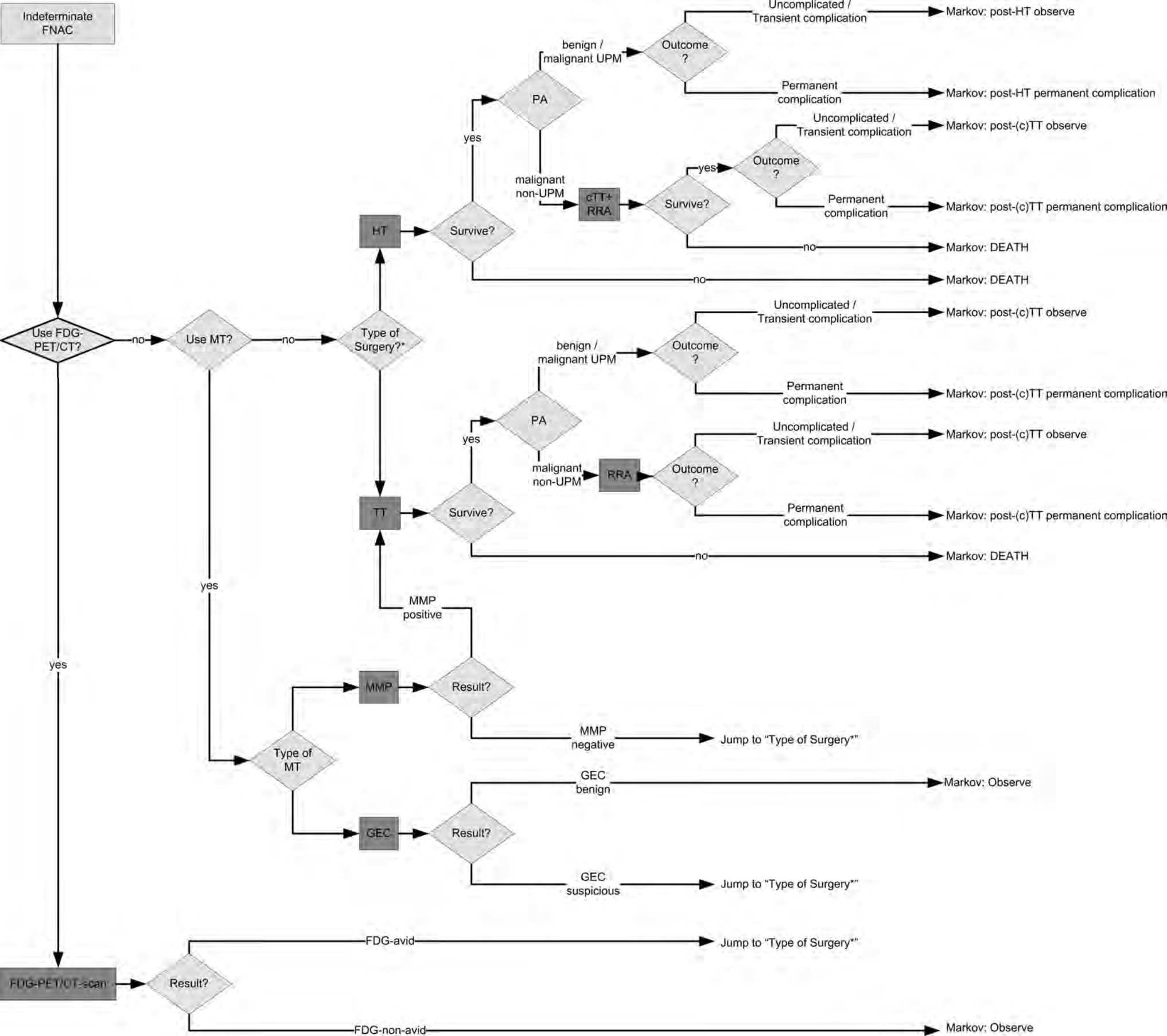
Table 3:

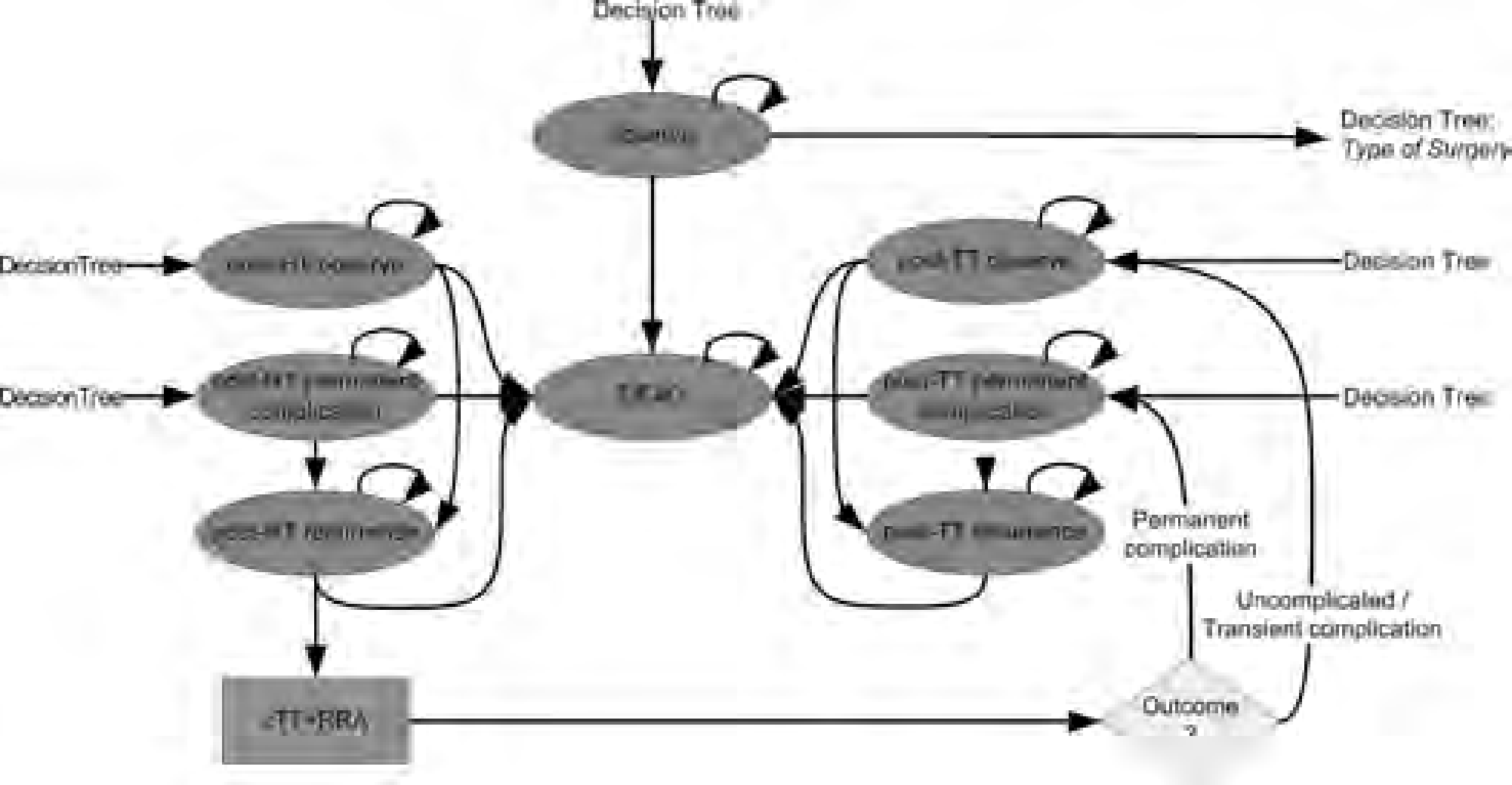
Variable:	**Surgery** (95%-CI):	**GEC** (95%-CI):	**MMP** (95%-CI):	**FDG-PET/CT** (95%-CI):
Absolute values:				
- *costs [€]*	€ 8,804 (€ 8,774 - € 8,835)	€ 9,341 (€ 9,300 - € 9,383)	€ 8,913 (€ 8,884 - € 8,942)	€ 7,983 (€ 7,941 - € 8,025)
- *utilities [QALY]*	4.516 (4.512 - 4.520)	4.556 (4.552 - 4.560)	4.515 (4.511 - 4.519)	4.552 (4.548 - 4.556)
- *Futile surgery [%]*	75.0% (74.8% - 75.3%)	38.2% (37.8% - 38.5%)	75.0% (74.8% - 75.3%)	40.3% (39.9% - 40.7%)
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

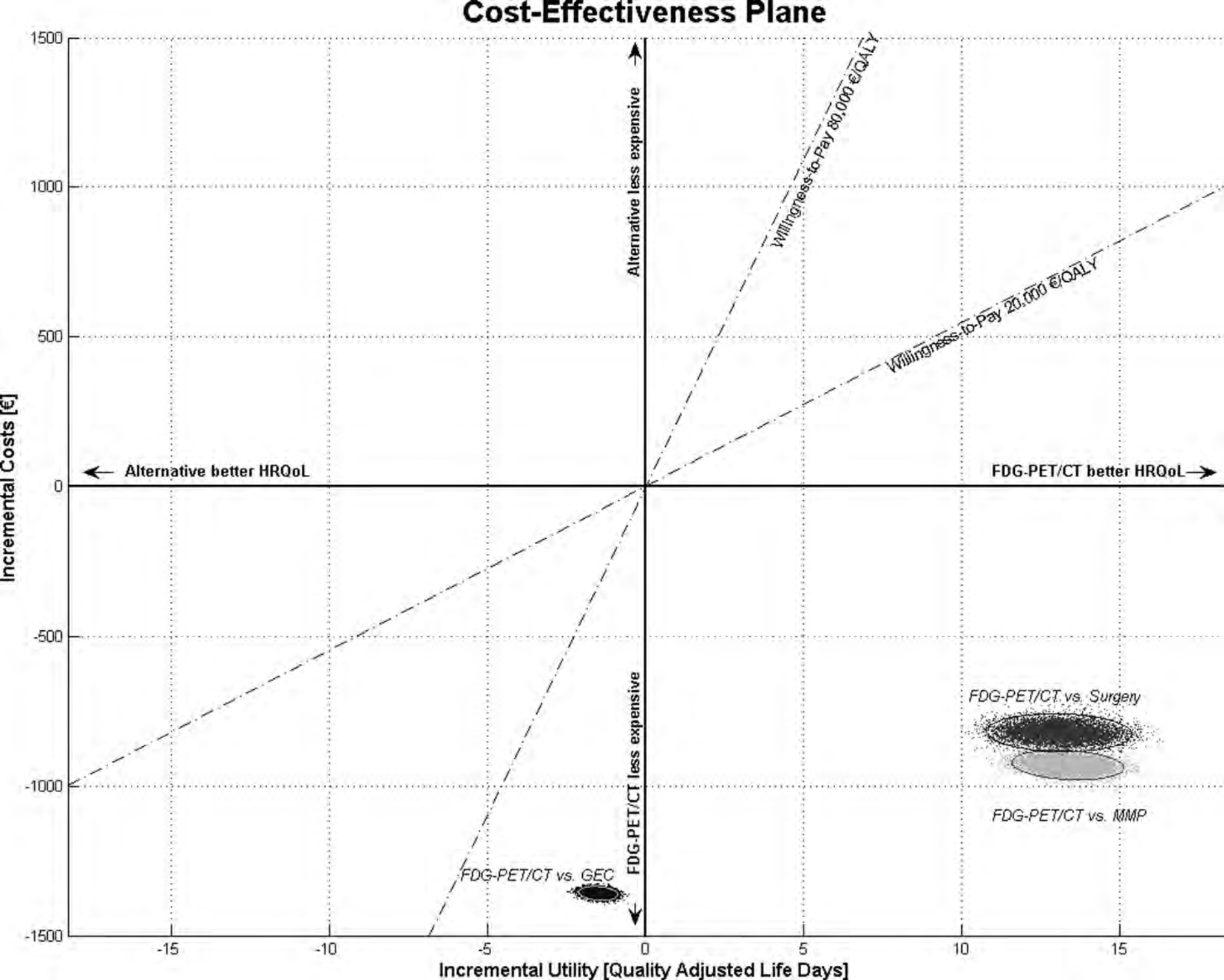
Supplemental Figure Legends

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.

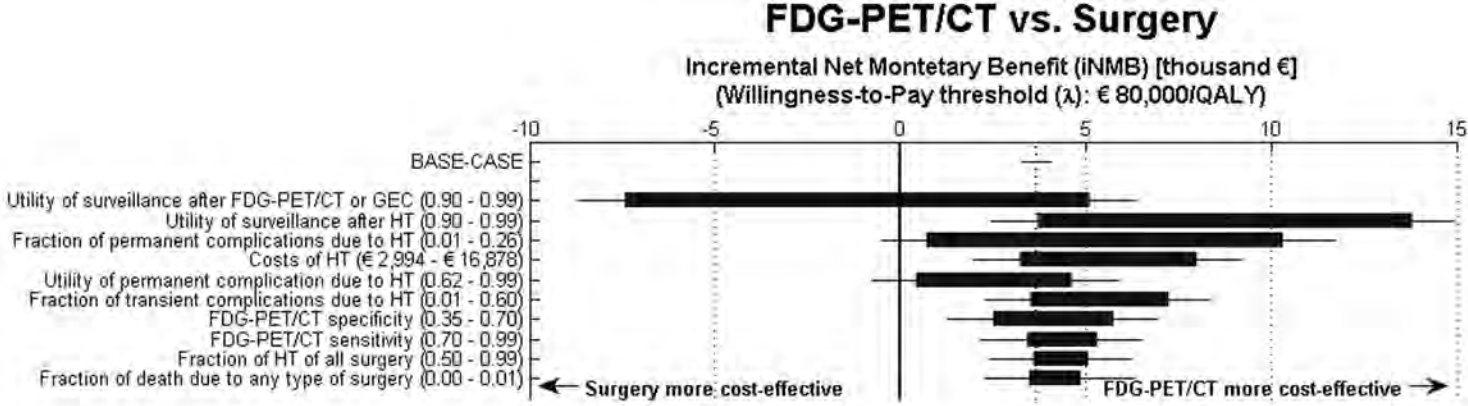
Supplemental Figure 2a-c: Tornado plots showing the results of one-way sensitivity analysis of all inputs of the model on the iNMB versus one of the other three strategies (a: surgery, b: GEC, c: MMP), for a willingness-to-pay threshold (λ) of € 80,000/QALY, the whiskers represent the limits of the 95%-confidence 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.


Indeterminate FNAC
Use FDG-PET/CT?
no
Use MT?
no
Type of Surgery?*
HT
Survive?
yes
PA
benign / malignant UPM
Outcome ?
Uncomplicated / Transient complication
Markov: post-HT observe
Permanent complication
Markov: post-HT permanent complication
malignant non-UPM
cTT+ RRA
Survive?
yes
Outcome ?
Uncomplicated / Transient complication
Markov: post-(c)TT observe
Permanent complication
Markov: post-(c)TT permanent complication
no
Markov: DEATH
no
Markov: DEATH
TT
Survive?
yes
PA
benign / malignant UPM
Outcome ?
Uncomplicated / Transient complication
Markov: post-(c)TT observe
Permanent complication
Markov: post-(c)TT permanent complication
malignant non-UPM
RRA
Outcome ?
Uncomplicated / Transient complication
Markov: post-(c)TT observe
Permanent complication
Markov: post-(c)TT permanent complication
no
Markov: DEATH
yes
MMP positive
yes
Type of MT
MMP
Result?
MMP negative
Jump to "Type of Surgery*"
GEC
Result?
GEC benign
Markov: Observe
GEC suspicious
Jump to "Type of Surgery*"
yes
FDG-PET/CT-scan
Result?
FDG-avid
Jump to "Type of Surgery*"
FDG-non-avid
Markov: Observe

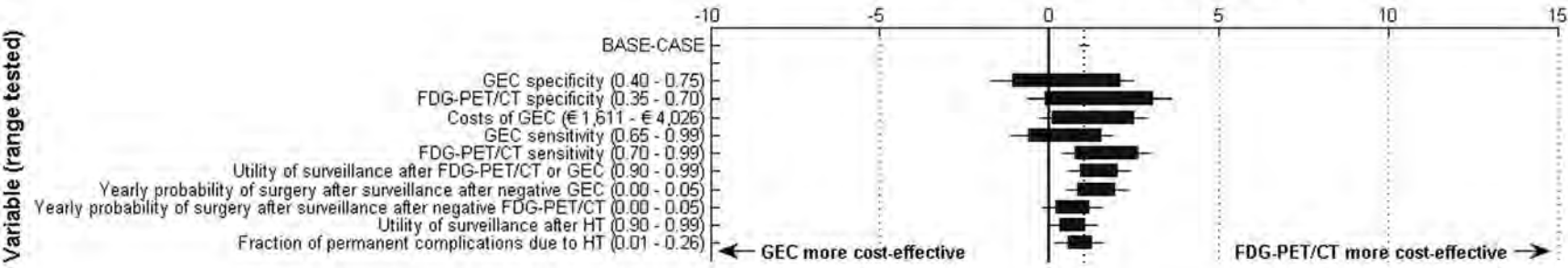

Decision Tree
Decision Tree:
Type of Surgery
DecisionTree
Decision Tree
DecisionTree
Decision Tree
Permanent
complication
Uncomplicated /
Transient complication
Outcome


Cost-Effectiveness Plane
Alternative less expensive
Willingness-to-Pay 80,000 €/QALY
Willingness-to-Pay 20,000 €/QALY
Alternative better HRQoL
FDG-PET/CT better HRQoL
FDG-PET/CT less expensive
FDG-PET/CT vs. Surgery
FDG-PET/CT vs. MMP
FDG-PET/CT vs. GEC
Incremental Costs [€]
Incremental Utility [Quality Adjusted Life Days]
1500
1000
500
0
-500
-1000
-1500
-15
-10
-5
0
5
10
15

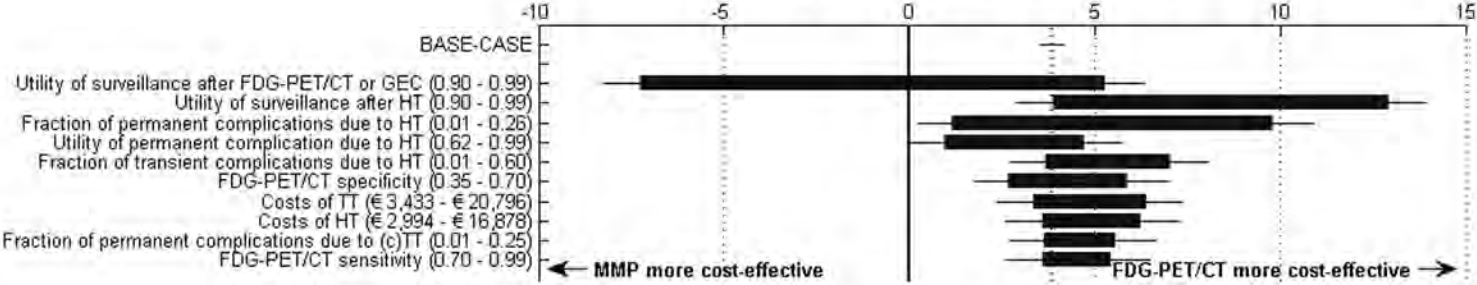
FDG-PET/CT vs. Surgery

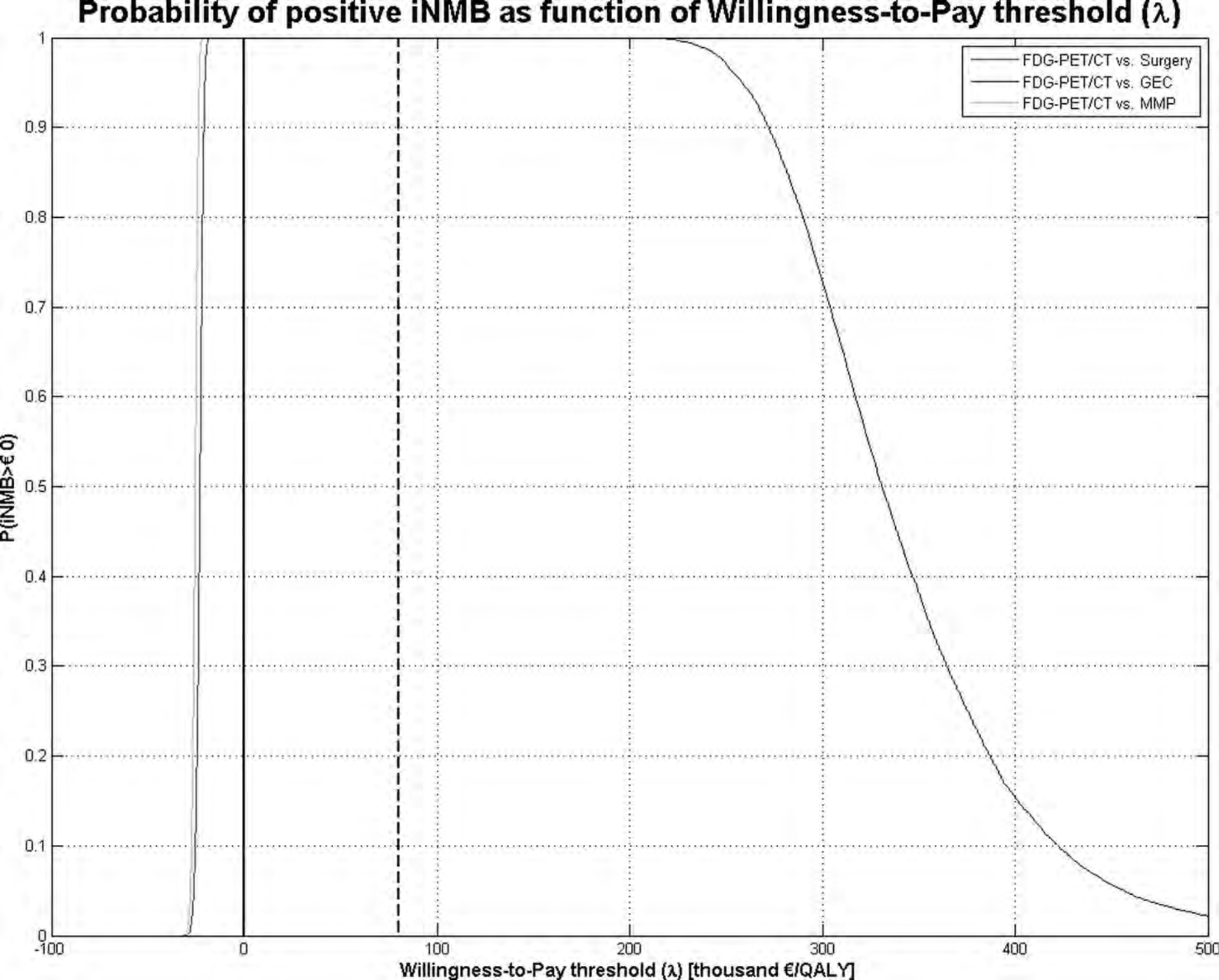


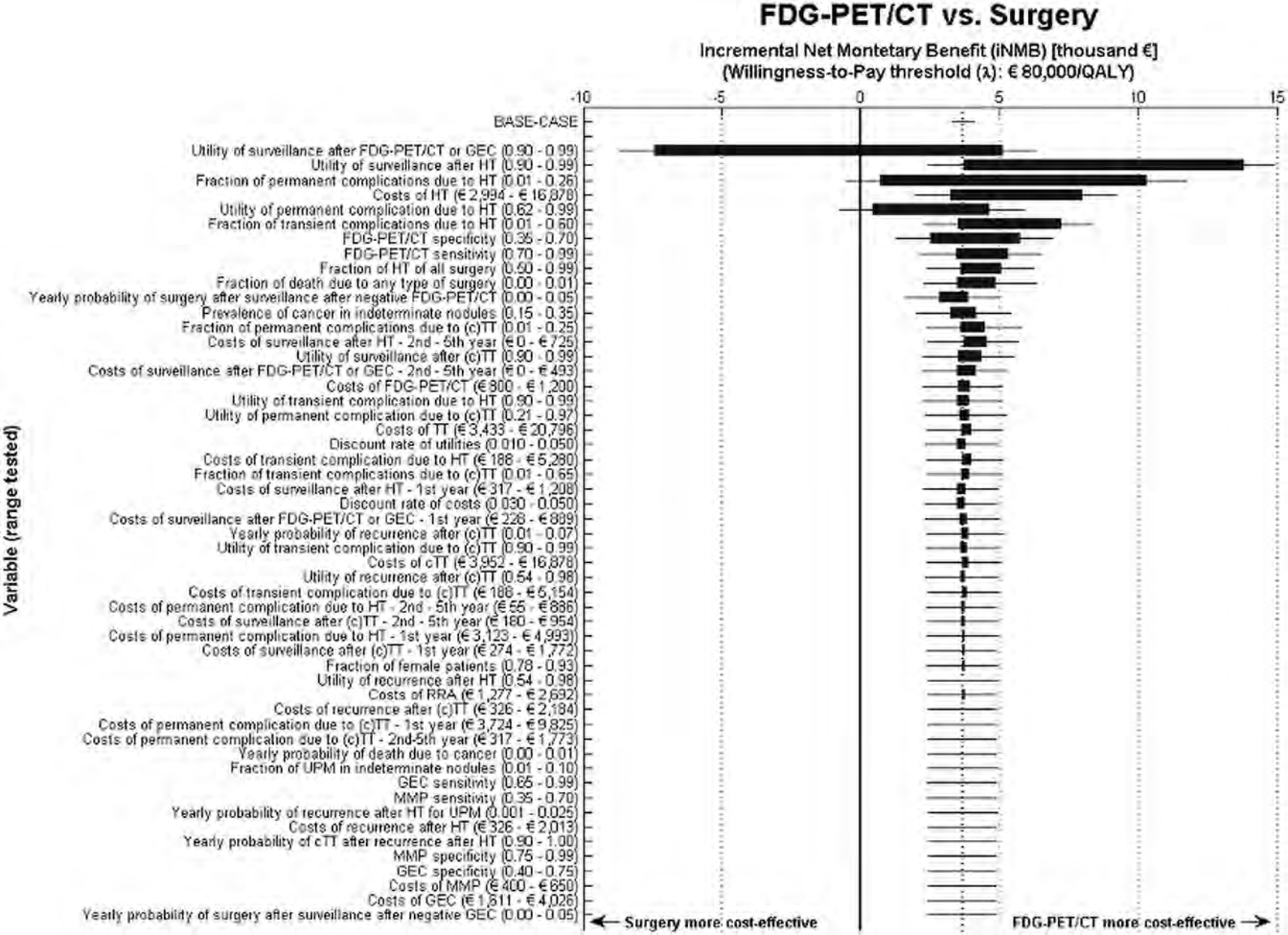
FDG-PET/CT vs. GEC

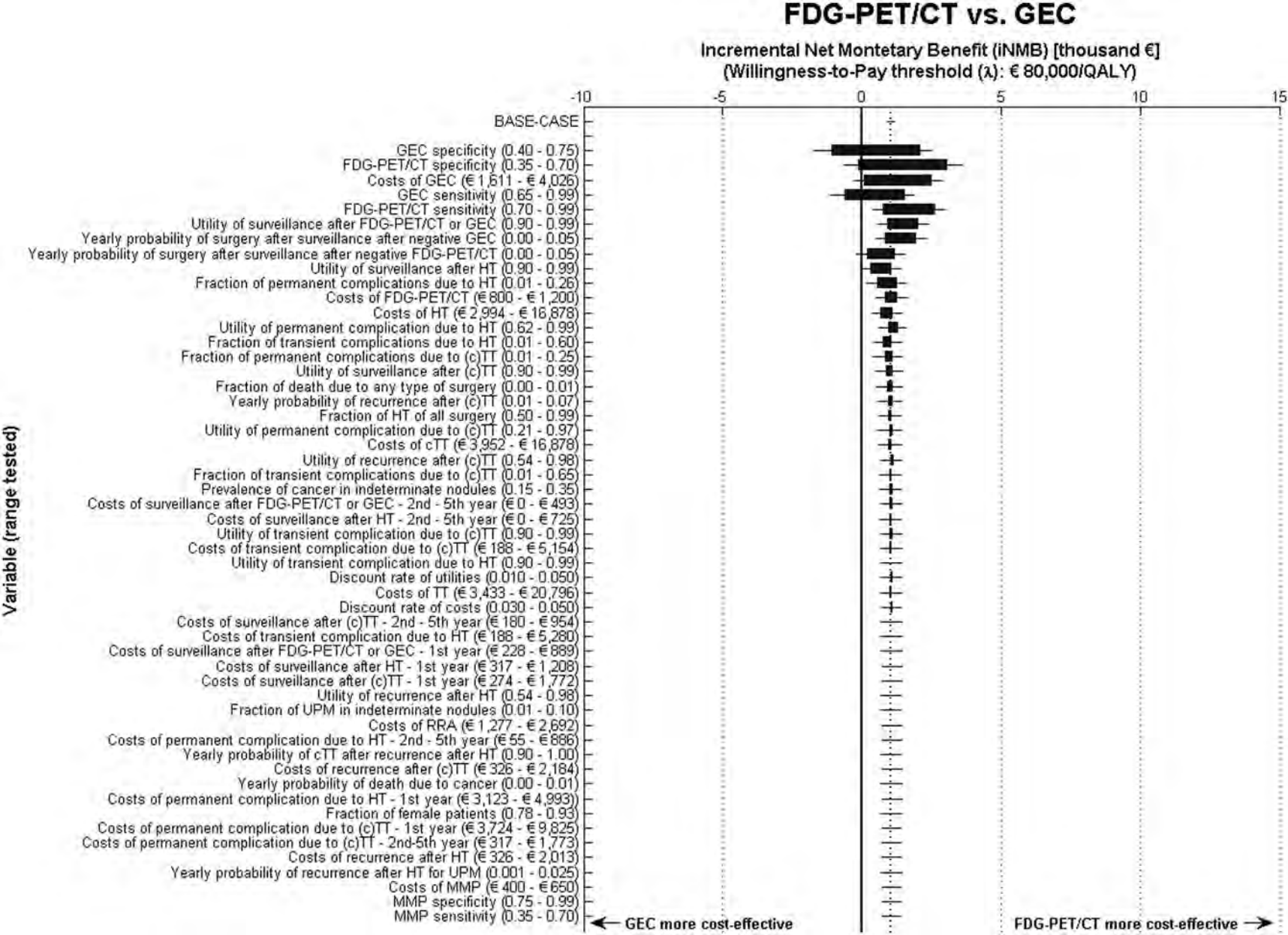


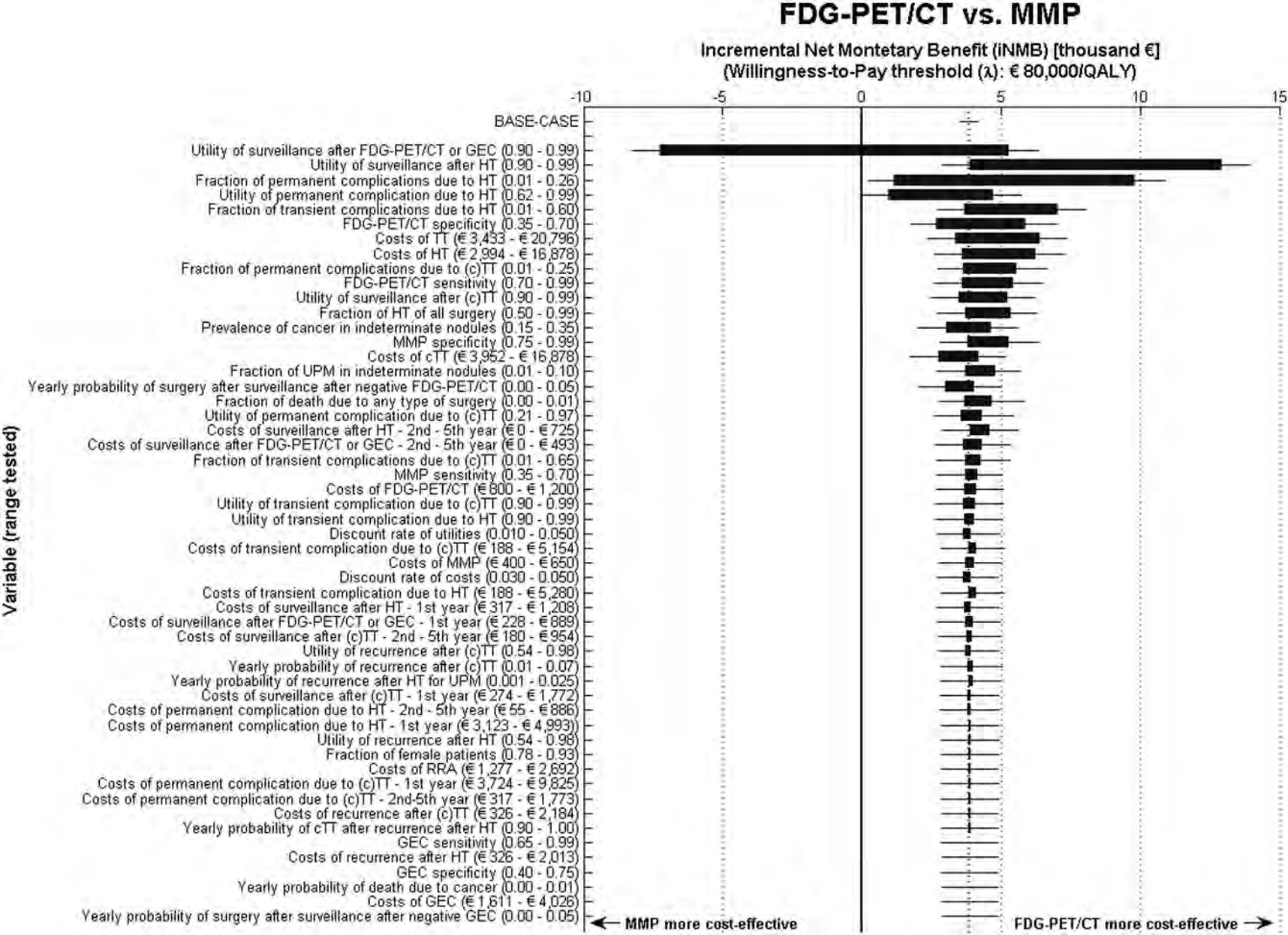
FDG-PET/CT vs. MMP




Probability of positive iNMB as function of Willingness-to-Pay threshold (λ)
FDG-PET/CT vs. Surgery
FDG-PET/CT vs. GEC
FDG-PET/CT vs. MMP
P(iNMB>€0)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
-100
0
100
200
300
400
500
Willingness-to-Pay threshold (λ) [thousand €/QALY]


FDG-PET/CT vs. Surgery
Incremental Net Montetary Benefit (iNMB) [thousand €]
(Willingness-to-Pay threshold (λ): € 80,000/QALY)
-10
-5
0
5
10
15
Variable (range tested)
BASE-CASE
Utility of surveillance after FDG-PET/CT or GEC (0.90 - 0.99)
Utility of surveillance after HT (0.90 - 0.99)
Fraction of permanent complications due to HT (0.01 - 0.26)
Costs of HT (€ 2,994 - € 16,878)
Utility of permanent complication due to HT (0.62 - 0.99)
Fraction of transient complications due to HT (0.01 - 0.60)
FDG-PET/CT specificity (0.35 - 0.70)
FDG-PET/CT sensitivity (0.70 - 0.99)
Fraction of HT of all surgery (0.50 - 0.99)
Fraction of death due to any type of surgery (0.00 - 0.01)
Yearly probability of surgery after surveillance after negative FDG-PET/CT (0.00 - 0.05)
Prevalence of cancer in indeterminate nodules (0.15 - 0.35)
Fraction of permanent complications due to (c)TT (0.01 - 0.25)
Costs of surveillance after HT - 2nd - 5th year (€ 0 - € 725)
Utility of surveillance after (c)TT (0.90 - 0.99)
Costs of surveillance after FDG-PET/CT or GEC - 2nd - 5th year (€ 0 - € 493)
Costs of FDG-PET/CT (€ 800 - € 1,200)
Utility of transient complication due to HT (0.90 - 0.99)
Utility of permanent complication due to (c)TT (0.21 - 0.97)
Costs of TT (€ 3,433 - € 20,796)
Discount rate of utilities (0.010 - 0.050)
Costs of transient complication due to HT (€ 188 - € 5,280)
Fraction of transient complications due to (c)TT (0.01 - 0.65)
Costs of surveillance after HT - 1st year (€ 317 - € 1,208)
Discount rate of costs (0.030 - 0.050)
Costs of surveillance after FDG-PET/CT or GEC - 1st year (€ 226 - € 889)
Yearly probability of recurrence after (c)TT (0.01 - 0.07)
Utility of transient complication due to (c)TT (0.90 - 0.99)
Costs of cTT (€ 3,952 - € 16,878)
Utility of recurrence after (c)TT (0.54 - 0.98)
Costs of transient complication due to (c)TT (€ 188 - € 5,154)
Costs of permanent complication due to HT - 2nd - 5th year (€ 55 - € 836)
Costs of surveillance after (c)TT - 2nd - 5th year (€ 180 - € 954)
Costs of permanent complication due to HT - 1st year (€ 3,123 - € 4,993)
Costs of surveillance after (c)TT - 1st year (€ 274 - € 1,772)
Fraction of female patients (0.78 - 0.93)
Utility of recurrence after HT (0.54 - 0.98)
Costs of RRA (€ 1,277 - € 2,692)
Costs of recurrence after (c)TT (€ 326 - € 2,184)
Costs of permanent complication due to (c)TT - 1st year (€ 3,724 - € 9,825)
Costs of permanent complication due to (c)TT - 2nd-5th year (€ 317 - € 1,773)
Yearly probability of death due to cancer (0.00 - 0.01)
Fraction of UPM in indeterminate nodules (0.01 - 0.10)
GEC sensitivity (0.65 - 0.99)
MMP sensitivity (0.35 - 0.70)
Yearly probability of recurrence after HT for UPM (0.001 - 0.025)
Costs of recurrence after HT (€ 326 - € 2,013)
Yearly probability of cTT after recurrence after HT (0.90 - 1.00)
MMP specificity (0.75 - 0.99)
GEC specificity (0.40 - 0.75)
Costs of MMP (€ 400 - € 650)
Costs of GEC (€ 1,611 - € 4,026)
Yearly probability of surgery after surveillance after negative GEC (0.00 - 0.05)
← Surgery more cost-effective
FDG-PET/CT more cost-effective →


FDG-PET/CT vs. GEC
Incremental Net Monetary Benefit (iNMB) [thousand €]
(Willingness-to-Pay threshold (λ): € 80,000/QALY)
-10
-5
0
5
10
15
Variable (range tested)
BASE-CASE
GEC specificity (0.40 - 0.75)
FDG-PET/CT specificity (0.35 - 0.70)
Costs of GEC (€ 1,611 - € 4,026)
GEC sensitivity (0.65 - 0.99)
FDG-PET/CT sensitivity (0.70 - 0.99)
Utility of surveillance after FDG-PET/CT or GEC (0.90 - 0.99)
Yearly probability of surgery after surveillance after negative GEC (0.00 - 0.05)
Yearly probability of surgery after surveillance after negative FDG-PET/CT (0.00 - 0.05)
Utility of surveillance after HT (0.90 - 0.99)
Fraction of permanent complications due to HT (0.01 - 0.26)
Costs of FDG-PET/CT (€ 800 - € 1,200)
Costs of HT (€ 2,994 - € 16,878)
Utility of permanent complication due to HT (0.62 - 0.99)
Fraction of transient complications due to HT (0.01 - 0.60)
Fraction of permanent complications due to (c)TT (0.01 - 0.25)
Utility of surveillance after (c)TT (0.90 - 0.99)
Fraction of death due to any type of surgery (0.00 - 0.01)
Yearly probability of recurrence after (c)TT (0.01 - 0.07)
Fraction of HT of all surgery (0.50 - 0.99)
Utility of permanent complication due to (c)TT (0.21 - 0.97)
Costs of cTT (€ 3,952 - € 16,878)
Utility of recurrence after (c)TT (0.54 - 0.98)
Fraction of transient complications due to (c)TT (0.01 - 0.65)
Prevalence of cancer in indeterminate nodules (0.15 - 0.35)
Costs of surveillance after FDG-PET/CT or GEC - 2nd - 5th year (€ 0 - € 493)
Costs of surveillance after HT - 2nd - 5th year (€ 0 - € 725)
Utility of transient complication due to (c)TT (0.90 - 0.99)
Costs of transient complication due to (c)TT (€ 188 - € 5,154)
Utility of transient complication due to HT (0.90 - 0.99)
Discount rate of utilities (0.010 - 0.050)
Costs of TT (€ 3,433 - € 20,796)
Discount rate of costs (0.030 - 0.050)
Costs of surveillance after (c)TT - 2nd - 5th year (€ 180 - € 954)
Costs of transient complication due to HT (€ 188 - € 5,280)
Costs of surveillance after FDG-PET/CT or GEC - 1st year (€ 228 - € 889)
Costs of surveillance after HT - 1st year (€ 317 - € 1,208)
Costs of surveillance after (c)TT - 1st year (€ 274 - € 1,772)
Utility of recurrence after HT (0.54 - 0.98)
Fraction of UPM in indeterminate nodules (0.01 - 0.10)
Costs of RRA (€ 1,277 - € 2,692)
Costs of permanent complication due to HT - 2nd - 5th year (€ 55 - € 886)
Yearly probability of cTT after recurrence after HT (0.90 - 1.00)
Costs of recurrence after (c)TT (€ 326 - € 2,184)
Yearly probability of death due to cancer (0.00 - 0.01)
Costs of permanent complication due to HT - 1st year (€ 3,123 - € 4,993)
Fraction of female patients (0.78 - 0.93)
Costs of permanent complication due to (c)TT - 1st year (€ 3,724 - € 9,825)
Costs of permanent complication due to (c)TT - 2nd-5th year (€ 317 - € 1,773)
Costs of recurrence after HT (€ 326 - € 2,013)
Yearly probability of recurrence after HT for UPM (0.001 - 0.025)
Costs of MMP (€ 400 - € 650)
MMP specificity (0.75 - 0.99)
MMP sensitivity (0.35 - 0.70)
← GEC more cost-effective
FDG-PET/CT more cost-effective →


FDG-PET/CT vs. MMP
Incremental Net Monetary Benefit (iNMB) [thousand €]
(Willingness-to-Pay threshold (λ): € 80,000/QALY)
-10
-5
0
5
10
15
Variable (range tested)
BASE-CASE
Utility of surveillance after FDG-PET/CT or GEC (0.90 - 0.99)
Utility of surveillance after HT (0.90 - 0.99)
Fraction of permanent complications due to HT (0.01 - 0.26)
Utility of permanent complication due to HT (0.62 - 0.99)
Fraction of transient complications due to HT (0.01 - 0.60)
FDG-PET/CT specificity (0.35 - 0.70)
Costs of TT (€ 3,433 - € 20,796)
Costs of HT (€ 2,994 - € 16,878)
Fraction of permanent complications due to (c)TT (0.01 - 0.25)
FDG-PET/CT sensitivity (0.70 - 0.99)
Utility of surveillance after (c)TT (0.90 - 0.99)
Fraction of HT of all surgery (0.50 - 0.99)
Prevalence of cancer in indeterminate nodules (0.15 - 0.35)
MMP specificity (0.75 - 0.99)
Costs of cTT (€ 3,952 - € 16,878)
Fraction of UPM in indeterminate nodules (0.01 - 0.10)
Yearly probability of surgery after surveillance after negative FDG-PET/CT (0.00 - 0.05)
Fraction of death due to any type of surgery (0.00 - 0.01)
Utility of permanent complication due to (c)TT (0.21 - 0.97)
Costs of surveillance after HT - 2nd - 5th year (€ 0 - € 725)
Costs of surveillance after FDG-PET/CT or GEC - 2nd - 5th year (€ 0 - € 493)
Fraction of transient complications due to (c)TT (0.01 - 0.65)
MMP sensitivity (0.35 - 0.70)
Costs of FDG-PET/CT (€ 800 - € 1,200)
Utility of transient complication due to (c)TT (0.90 - 0.99)
Utility of transient complication due to HT (0.90 - 0.99)
Discount rate of utilities (0.010 - 0.050)
Costs of transient complication due to (c)TT (€ 188 - € 5,154)
Costs of MMP (€ 400 - € 650)
Discount rate of costs (0.030 - 0.050)
Costs of transient complication due to HT (€ 188 - € 5,280)
Costs of surveillance after HT - 1st year (€ 317 - € 1,208)
Costs of surveillance after FDG-PET/CT or GEC - 1st year (€ 228 - € 889)
Costs of surveillance after (c)TT - 2nd - 5th year (€ 180 - € 954)
Utility of recurrence after (c)TT (0.54 - 0.98)
Yearly probability of recurrence after (c)TT (0.01 - 0.07)
Yearly probability of recurrence after HT for UPM (0.001 - 0.025)
Costs of surveillance after (c)TT - 1st year (€ 274 - € 1,772)
Costs of permanent complication due to HT - 2nd - 5th year (€ 55 - € 886)
Costs of permanent complication due to HT - 1st year (€ 3,123 - € 4,993)
Utility of recurrence after HT (0.54 - 0.98)
Fraction of female patients (0.78 - 0.93)
Costs of RRA (€ 1,277 - € 2,692)
Costs of permanent complication due to (c)TT - 1st year (€ 3,724 - € 9,825)
Costs of permanent complication due to (c)TT - 2nd-5th year (€ 317 - € 1,773)
Costs of recurrence after (c)TT (€ 326 - € 2,184)
Yearly probability of cTT after recurrence after HT (0.90 - 1.00)
GEC sensitivity (0.65 - 0.99)
Costs of recurrence after HT (€ 326 - € 2,013)
GEC specificity (0.40 - 0.75)
Yearly probability of death due to cancer (0.00 - 0.01)
Costs of GEC (€ 1,611 - € 4,026)
Yearly probability of surgery after surveillance after negative GEC (0.00 - 0.05)
← MMP more cost-effective
FDG-PET/CT more cost-effective →