Published in final edited form as:

Surgery. 2018 January ; 163(1): 88–96. doi:10.1016/j.surg.2017.10.004.

Cost-Effectiveness of Lobectomy Versus Genetic Testing (Afirma) for Indeterminate Thyroid Nodules: Considering the Costs of Surveillance

Courtney J. Balentine, MD, MPH¹, David J. Vanness, PhD², and David F. Schneider, MD, MS³

¹The Department of Surgery, University of Alabama at Birmingham, Birmingham, AL

²Department of Population Health Sciences, University of Wisconsin, Madison, WI

³Department of Surgery, University of Wisconsin, Madison, WI

Abstract

Background—We evaluated whether diagnostic surgery for indeterminate thyroid nodules would be more cost-effective than genetic testing after including costs of long-term surveillance.

Methods—We used a Markov decision model to estimate cost-effectiveness of thyroid lobectomy versus genetic testing (Afirma®) for evaluation of indeterminate (Bethesda 3–4) thyroid nodules. The base case was a 40 year-old female with a 1 cm indeterminate nodule. Probabilities and estimates of utilities were obtained from the literature. Cost estimates were based on Medicare reimbursements, with a 3% discount rate for costs and quality adjusted life years (QALYs).

Results—Over a 5-year period following diagnosis of indeterminate thyroid nodules, lobectomy was less costly and more effective than Afirma® (Lobectomy: \$6,100; 4.50 QALYs vs. Afirma®: \$9,400; 4.47 QALYs). Only in 253 of 10,000 simulations (2.53%) did Afirma show a net benefit at a cost-effectiveness threshold of \$100,000 per QALY. There was only a 0.3% probability of Afirma® being cost-saving, and a 14.9% probability of improving QALYs.

Conclusions—Our base case result is that diagnostic lobectomy dominates genetic testing as a strategy for ruling out malignancy of indeterminate thyroid nodules. However, these results were highly sensitive to estimates of utilities after lobectomy and living under surveillance after Afirma®.

BACKGROUND

Thyroid nodules are extremely common and pose a dilemma for surgeons and endocrinologists who seek to avoid unnecessary and costly procedures for their patients. Nearly 50% of adults develop thyroid nodules, and increasing use of cross-sectional imaging has led to a substantial increase in the number of patients undergoing fine needle aspiration (FNA) biopsy to evaluate nodules for malignancy^{1,2}. Unfortunately, as many as 30% of

Corresponding Author: Courtney Balentine, 1600 7th Ave South, Lowder Building Suite 500, Birmingham, AL 35233, cbalentine@uabmc.edu, Phone: 205-638-7000, Fax: 205-638-2121.

The manuscript was presented as an oral presentation at the 2016 American Association of Endocrine Surgeons annual meeting in Orlando, Florida.

thyroid nodules will have indeterminate cytology, even though only 5 to 30% of these truly contain a cancer.^{3,4} When faced with an indeterminate thyroid nodule, surgeons have traditionally performed a thyroid lobectomy to obtain the definitive histopathologic diagnosis. Since the vast majority of these procedures reveal benign pathology, many patients are subjected to the risks and costs of surgery without significant benefits.

One approach to avoid unnecessary surgery for indeterminate thyroid nodules is the use of molecular testing to preoperatively identify nodules that are likely to be benign. Advocates of molecular testing for indeterminate nodules have emphasized the cost effectiveness of their approach because a benign result can obviate the need for surgery.⁵ Recent studies and clinical use has focused on one commercially available molecular classifier, Afirma®, that categorizes cytopathology as benign or suspicious based on a proprietary molecular profile. Previous studies have established that Afirma® lowers costs, with the primary benefit being a reduction in surgery for nodules that are ultimately benign.⁶ However, cost modeling for Afirma® was limited to a short follow up period that failed to consider the long-term costs of surveillance for thyroid nodules.⁷ The purpose of this study is to compare the cost-effectiveness of molecular testing (Afirma®) versus diagnostic lobectomy for indeterminate thyroid nodules.

METHODS

Costs & Utilities

Cost estimates were based on average Medicare reimbursement and are shown in Table 1. A 3% discount rate was used for costs and quality adjusted life years (QALYs). Estimates of utilities were drawn from previously published values (Table 1).

Outcomes

Our primary outcome was cost per quality adjusted life years (QALYs), and we assumed a cost-effectiveness threshold of \$100,000/QALY. We also tracked secondary outcomes of unnecessary surgical procedures, cases requiring long-term hormone replacement, and permanent surgical complications. We developed a Markov decision-analytic model to compare thyroid lobectomy with Afirma testing (Figure 1).

Base Case

The base case was a 40-year-old female with a 1 cm indeterminate thyroid nodule (atypia of undetermined significance/follicular lesion of undetermined significance/follicular neoplasm/Hürthle Cell neoplasm). The model assumed that no features obviously concerning for malignancy were present on initial ultrasound and physical examination (cervical lymphadenopathy, vocal cord paresis, tracheal or esophageal invasion). We also assumed that final surgical pathology was always accurate. Parallel arms were constructed to model detected and undetected thyroid cancers (Figure 1). The 2009 American Thyroid Association (ATA) guidelines directed treatment decisions such that cancers greater than 1 cm detected after diagnostic lobectomy received a completion thyroidectomy (1). When the Afirma returned a “suspicious” result, we assumed 66.7% of patients underwent total thyroidectomy and 33.6% underwent an initial lobectomy. For the scenario in which patients

had a repeat indeterminate biopsy result, we assumed that 82.4% would have surgery rather than continued observation (2).

The model was run in 6 month cycles for a 5-year time horizon. Probabilities including transition probabilities and complication rates were obtained from the medical literature and shown in Table 1. Estimates of health state preferences, or utilities also came from the literature, and the most widely used preferences are surrogate preferences, elicited from a sample of healthcare providers familiar with thyroid disease (3). The natural history of thyroid cancer, benign thyroid nodular disease as well as thyroid cancer mortality data were obtained from the literature and the Surveillance, Epidemiology, and End Results (SEER) Program estimates of median survival for a 1cm papillary thyroid cancer for a patient starting at the age of our base case.⁸

Sensitivity Analysis

We used Monte Carlo probabilistic sensitivity analysis with 10,000 replications to derive 95% uncertainty intervals for primary and secondary outcomes and incremental cost-effectiveness ratios. Marginal probabilistic sensitivity analysis was used to assess variation in net benefit (assuming willingness to pay \$100,000 per QALY gained) because of uncertainty in each individual parameter value, holding all other parameters fixed at their base case value. Threshold sensitivity analysis was conducted for parameters found to affect the overall conclusion of net benefit. Scenario sensitivity analyses for cost-effectiveness with respect to the primary outcome (QALYs) were conducted doubling the time horizon to 10 years and assuming that surveillance costs were half their base-case value. Additionally, we conducted a scenario sensitivity analysis with a 10-year time horizon to assess changes over a longer follow up period, and a two-way sensitivity analysis that varied the differences in utility estimates and the cost of surveillance. The decision model was programmed in the R language (version 3.3.2) using base functions and the expm package (version 0.999-0) for matrix exponentiation. Recursive partitioning was conducted using the rpart package (version 4.1-10).

RESULTS

Over a 5-year time horizon, the expected incremental cost of Afirma testing compared with lobectomy was \$3,300 (95% UI: \$1,100 to \$5,400), while expected incremental QALYs were -0.03 (95% UI: -0.07 to 0.03). The difference in life years between the two strategies was minimal (4.6480 for surgery vs. 4.6476 for Afirma). Having higher expected cost and lower expected QALYs, Afirma is said to be dominated by lobectomy; ICER undefined (95% UI: dominated to \$807,000 per QALY gained), with 253 out of 10,000 Monte Carlo simulations indicating positive net benefit of Afirma compared to lobectomy, assuming willingness to pay of \$100,000 per QALY gained. In other words, in only 2.53% of simulations did Afirma show a net benefit at a cost-effectiveness threshold of \$100,000 per QALY.

Afirma reduced the probability of unnecessary surgery from 76.2% (95% UI: 72.8% to 79.4%) to 36.9% (95% UI: 30.9% to 43.1%), with cost per unnecessary surgery averted of \$8,400 (95% UI: \$2,700 to \$14,500). Afirma increased the probability of requiring long

term hormone replacement from 34.7% (95% UI: 31.0% to 38.3%) to 49.2% (95% UI: 44.2% to 54.4%). Afirma left the probability of permanent complications unchanged at 2.0% (95% UI surgery: 1.8% to 2.2%; Afirma : 1.7% to 2.4%).

Sensitivity Analysis

Our primary analysis was based on treatment plans that follow the 2009 ATA guidelines, but new guidelines were issued in 2015⁹. The new guidelines suggested that lobectomy constitutes adequate treatment for most thyroid cancers in the size range of 1–4cm, while previous guidelines recommended completion lobectomy for cancers $\geq$ 1cm. To determine whether following the new guidelines would affect our conclusion, we re-ran the above analysis with the assumption that lobectomy would be sufficient treatment for a 1cm thyroid nodule that was found to be cancer (lobectomy would be both diagnostic and completely therapeutic for the surgical strategy, and cancer diagnosed by Afirma would be treated with lobectomy only). Our sensitivity analysis using the 2015 guidelines found that an Afirma-based strategy is slightly more effective than surgery (+0.0017 QALY's) but at considerably higher cost (+\$3,130) for an incremental cost-effectiveness ratio (ICER) of \$1,843,000. This is substantially higher than the established threshold for cost-effectiveness of \$100,000 per QALY.

Results of the marginal probabilistic sensitivity analysis for the ten parameters whose uncertainty has the greatest effect on net benefit (assuming willingness to pay \$100,000 per QALY) are shown in Figure 2. The three most influential parameters are the differences in utility between the post-lobectomy state and surveillance under Afirma, the cost of lobectomy, and the cost of Afirma. The model is sensitive to the post-lobectomy state utility, with net benefit crossing the zero threshold within the 95% uncertainty interval for that parameter. Threshold analysis determined that, holding all other parameters at their base case value, if the utility of the post-lobectomy health state (without long-term hormone replacement) is less than 0.954, then Afirma is expected to be cost-effective at the \$100,000 per QALY threshold.

Increasing the time horizon to 10 years had no qualitative effect on the cost-effectiveness of Afirma compared to lobectomy. The expected incremental cost of Afirma increased to \$5,400 (95% UI: \$2,700 to \$8,500) while the incremental effectiveness of Afirma fell to -0.05 QALYs (95% UI: -0.14 to 0.048). Afirma remained dominated by lobectomy (95% UI: dominated to \$635,000 per QALY gained), with 218 out of 10,000 Monte Carlo simulations indicating positive net benefit of Afirma compared to lobectomy, assuming willingness to pay of \$100,000 per QALY gained.

Since guidelines for surveillance of nodules classified as benign remain ambiguous, we also varied the intensity of surveillance. Reducing the expected cost of surveillance by half to \$154.50, reflecting a more relaxed surveillance schedule, also had no qualitative effect on the cost-effectiveness of Afirma compared to lobectomy. The expected incremental cost of Afirma fell to \$2,300 (95% UI: \$400 to \$4,100) while the incremental effectiveness of Afirma was unchanged. Afirma remained dominated by lobectomy (95% UI: dominated to \$659,000 per QALY gained), with 529 out of 10,000 Monte Carlo simulations indicating

positive net benefit of Afirma compared to lobectomy, assuming willingness to pay of \$100,000 per QALY gained.

Finally, we ran a two-way sensitivity analysis in which the difference in utility between post-lobectomy state and surveillance under Afirma was varied along with costs of surveillance (Figure 3). The incremental net benefit at willingness to pay of \$100,000 per QALY was again most heavily influenced by varying the estimate of the utility difference. By contrast, the effect of varying surveillance costs across different utility values had a relatively small effect on the incremental net benefit (Figure 3).

DISCUSSION

Our primary finding is that diagnostic lobectomy is both more effective and less costly than genetic testing for evaluation of indeterminate thyroid nodules after accounting for the costs of long-term follow up. The primary appeal of genetic testing (Afirma®) is the ability to avoid under- or over-treatment of indeterminate nodules. If a nodule can be correctly classified as benign, then the patient avoids surgery that carries risks with no clear benefit. However, relying on a genetic test to guide therapy of an indeterminate nodule requires ongoing surveillance to ensure that a malignancy is not being missed. Although the intensity of surveillance is likely to vary among endocrinologists and primary care physicians, the patient will undergo serial ultrasound examinations to monitor for nodule growth or development of other malignant characteristics. If the nodule enlarges significantly, current guidelines recommend that the patient should undergo another FNA evaluation and possibly even a repeat of the genetic test.⁴ Our model demonstrates that over a long period of time, the costs of follow-up and subsequent evaluation eventually exceed the costs of surgery. Although the short-term costs of surgery can be high, there are relatively few long-term costs because definitively diagnosing a nodule as benign obviates the need for further ultrasound follow-up.

Another important factor when comparing the cost effectiveness of surgery versus genetic testing is the value of utility estimates for various health states associated with each strategy. For lobectomy, decreases in quality of life are primarily related to postoperative complications. Many of these utility estimates are derived from a small sample of healthcare providers or surveys of patients, leaving considerable room for improvement.¹⁰ Additionally, we were unable to find any studies that directly assessed the utilities associated with surveillance after genetic testing of indeterminate thyroid nodules. Instead, we were forced to rely on values based on utilities for patients under surveillance for malignancy or benign nodules. The uncertainty related to valuation of health states after surgery or genetic testing was reflected in the conclusions from our two-way sensitivity analysis and marginal probabilistic sensitivity analysis. The modeling of cost-effectiveness was highly sensitive to the value of utilities, suggesting that these estimates need to be more precisely measured in order to compare the two treatment strategies. While it is easy to understand how quality of life decreases as a result of postoperative complications, it is less clear how the uncertainty of long-term surveillance might affect utilities. One could easily imagine that waiting on yearly ultrasounds to evaluate for changes in indeterminate nodules could become stressful and diminish quality of life, but no utility or quality of life data exists for this health state.

Alternatively, patients might place great trust in the results of genetic testing so that serial surveillance has little effect on their lives as long as they believe in the initial test results. Even in this situation, however, it is possible to have a significant negative effect on quality of life if tumors are missed and allowed to progress over time. While thyroid cancer is generally less aggressive than other tumors, discovering that a “benign” lesion is actually a cancer can certainly be an unpleasant surprise for patients even if the disease has not progressed. This issue needs to be further clarified before surgeons and endocrinologists firmly commit to either treatment strategy.

Our work differs from previous studies that suggested genetic testing was more cost-effective than surgery for indeterminate thyroid nodules.⁷ The most likely reason that we reached a different conclusion was inclusion of the associated surveillance for indeterminate nodules classified as benign. We modeled several different strategies for ultrasound surveillance of a nodule that was classified as benign by genetic testing, and we considered the possibility that primary care physicians or endocrinologists could order a second genetic test if the nodule changed and FNA biopsy was again indeterminate. We were also able to incorporate more accurate information on the likelihood that a benign nodule would change substantially enough to trigger further evaluation.¹¹ Equally important, our base case included a 5-year time horizon, but we also evaluated a longer (10-year followup) period than the 5-year span examined by Li, et al, though a 10-year analysis did not change the underlying conclusion.⁷ Because we anticipate that patients with thyroid nodules will have excellent long-term survival, it is important to consider the lifetime costs associated with surgery versus genetic testing. As more time passes and patients undergo repeat ultrasounds and possibly FNA biopsies, the costs of a genetic testing strategy continue to accumulate. Our approach of examining long-term outcomes is ultimately more representative than a short followup period after the initial FNA. The most recent American Thyroid Association (ATA) guidelines recommend surveillance for at least two years, but recommendations for the frequency of surveillance beyond that point remain ambiguous.⁴ Finally, we considered the possibility that surgeons who trusted genetic tests might offer patients total thyroidectomy rather than lobectomy for a 1cm lesion classified as suspicious. We assumed that true believers in genetic testing might regard a test result of “cancer” as equivalent to an FNA showing cancer. Since total thyroidectomy for ≥ 1 cm cancer is a fairly common choice for operation, many surgeons might opt for up front total thyroidectomy in this situation. This assumption for the model likely explains how an Afirma based strategy could potentially result in more permanent complications and a need for thyroid hormone replacement. Notably, when we varied the rate of total thyroidectomy after a positive Afirma test, surgery remained a more cost-effective strategy.

There are several important limitations of our work that should be acknowledged. First, Markov modeling represents an idealized and highly simplified scenario that may not accurately represent costs and outcomes for any particular patient. However, in the absence of a randomized trial, economic modeling is the only method for comparing strategies for thyroid nodules. Like any economic model, certain assumptions were made, and we acknowledge that we lack modern data for modeling the growth of “missed” cancers for the rare scenario of false negative Afirma results. Additionally, we relied on literature value for utilities, and there are methodologic issues with how these were generated, as we mentioned

above. More robust efforts to estimate the value of health states associated with surgery and genetic testing may yield different results. We feel the proxy utility for a post-lobectomy state of 0.99 may be too high as it does not reflect true complication rates and decision-regret for patients who require thyroid hormone supplementation. The surveillance health state is also poorly understood; surveillance probably imparts some degree of psychological distress on patients due to repeated cycles of uncertainty and anxiety. Additionally, each surveillance episode represents an economic burden on the patient to take time off of work, obtain an ultrasound and/or FNA, and visit the clinic. Because of our own doubts regarding these utility values, our sensitivity analysis included testing the models over a wide range of values for utility, and the vast majority of the simulations indicated that lobectomy was the more cost-effective strategy. We also chose to base our analysis on the 2009 ATA guidelines rather than the 2015 guidelines, since there is a significant lag time from issuing guidelines to actual practice changes¹². However, it should be noted that our sensitivity analysis under conditions defined by the 2015 guidelines still did not indicate that the use of genetic testing was a cost-effective strategy.

Conclusions

After including costs of long-term surveillance, our analysis predicts that diagnostic lobectomy dominates genetic testing as a strategy for ruling out malignancy of indeterminate thyroid nodules. However, the results are highly sensitive to estimates of utilities, highlighting the need for more rigorous assessment of patient values and preferences in order to guide decisions about treatment for indeterminate thyroid nodules.

Acknowledgments

Dr. Schneider was supported by NIH UL1TR000427 and NIH KL2TR000428, and Dr. Balentine was funded by an AHRQ K12 Career Development Award (K12 HS023009-3). Dr. Vanness receives consulting fees from Evidera and Novartis for work unrelated to the manuscript.

References

1. Mortensen JD, Woolner LB, Bennett WA. Gross and microscopic findings in clinically normal thyroid glands. *J Clin Endocrinol Metab.* 1955; 15(10):1270–1280. [PubMed: 13263417]
2. Yousem DM, Huang T, Loevner LA, Langlotz CP. Clinical and economic impact of incidental thyroid lesions found with CT and MR. *AJNR Am J Neuroradiol.* 1997; 18(8):1423–1428. [PubMed: 9296181]
3. Yassa L, Cibas ES, Benson CB, et al. Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. *Cancer.* 2007; 111(6):508–516. [PubMed: 17999413]
4. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid.* 2016; 26(1):1–133. [PubMed: 26462967]
5. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med.* 2012; 367(8):705–715. [PubMed: 22731672]
6. Wu JX, Lam R, Levin M, Rao J, Sullivan PS, Yeh MW. Effect of malignancy rates on cost-effectiveness of routine gene expression classifier testing for indeterminate thyroid nodules. *Surgery.* 159(1):118–129.
7. Li H, Robinson KA, Anton B, Saldanha IJ, Ladenson PW. Cost-Effectiveness of a Novel Molecular Test for Cytologically Indeterminate Thyroid Nodules. *The Journal of Clinical Endocrinology & Metabolism.* 2011; 96(11):E1719–E1726. [PubMed: 21865367]

8. Surveillance, Epidemiology, and End Results (SEER) Program. SEER. 18. Nov. 2015 Regs Research Data.
9. Haugen BRM, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2015
10. Esnaola NF, Cantor SB, Sherman SI, Lee JE, Evans DB. Optimal treatment strategy in patients with papillary thyroid cancer: A decision analysis. *Surgery*. 130(6):921–930.
11. Durante C, Costante G, Lucisano G, et al. The natural history of benign thyroid nodules. *JAMA*. 2015; 313(9):926–935. [PubMed: 25734734]
12. Damschroder LJ. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*. 2009; 4:50. [PubMed: 19664226]
13. Schneider DF, Cherney Stafford LM, Brys N, et al. Gauging the Extent of Thyroidectomy for Indeterminate Thyroid Nodules: An Oncologic Perspective. *Endocr Pract*. 2017; 23(4):442–450. [PubMed: 28095042]
14. Gharib H, Goellner JR, Johnson DA. Fine-needle aspiration cytology of the thyroid. A 12-year experience with 11,000 biopsies. *Clin Lab Med*. 1993; 13(3):699–709. [PubMed: 8222583]
15. Gharib H, Papini E, Valcavi R, et al. American Association of Clinical Endocrinologists and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Endocr Pract*. 2006; 12(1):63–102.
16. Pu RT, Yang J, Wasserman PG, Bhuiya T, Griffith KA, Michael CW. Does Hurthle cell lesion/neoplasm predict malignancy more than follicular lesion/neoplasm on thyroid fine-needle aspiration? *Diagn Cytopathol*. 2006; 34(5):330–334. [PubMed: 16604553]
17. Sullivan PS, Hirschowitz SL, Fung PC, Apple SK. The impact of atypia/follicular lesion of undetermined significance and repeat fine-needle aspiration: 5 years before and after implementation of the Bethesda System. *Cancer Cytopathol*. 2014; 122(12):866–872. [PubMed: 25078420]
18. Durante C, Costante G, Lucisano G, et al. The natural history of benign thyroid nodules. *JAMA*. 2015; 313(9):926–935. [PubMed: 25734734]
19. Singh Ospina N, Maraka S, Espinosa de Ycaza AE, et al. Prognosis of patients with benign thyroid nodules: a population-based study. *Endocrine*. 2016; 54(1):148–155. [PubMed: 27142412]
20. Ajmal S, Rapoport S, Ramirez Battle H, Mazzaglia PJ. The natural history of the benign thyroid nodule: what is the appropriate follow-up strategy? *J Am Coll Surg*. 2015; 220(6):987–992. [PubMed: 25667137]
21. Dralle H, Sekulla C, Haerting J, et al. Risk factors of paralysis and functional outcome after recurrent laryngeal nerve monitoring in thyroid surgery. *Surgery*. 2004; 136(6):1310–1322. [PubMed: 15657592]
22. Promberger R, Ott J, Kober F, Karik M, Freissmuth M, Hermann M. Normal parathyroid hormone levels do not exclude permanent hypoparathyroidism after thyroidectomy. *Thyroid*. 2011; 21(2): 145–150. [PubMed: 21190447]
23. Merchavy S, Marom T, Forest VI, et al. Comparison of the incidence of postoperative hypocalcemia following total thyroidectomy vs completion thyroidectomy. *Otolaryngol Head Neck Surg*. 2015; 152(1):53–56. [PubMed: 25358344]
24. Dionigi G, Van Slycke S, Rausei S, Boni L, Dionigi R. Parathyroid function after open thyroidectomy: A prospective randomized study for ligasure precise versus harmonic FOCUS. *Head Neck*. 2013; 35(4):562–567. [PubMed: 22715020]
25. Alvarado R, Sywak MS, Delbridge L, Sidhu SB. Central lymph node dissection as a secondary procedure for papillary thyroid cancer: Is there added morbidity? *Surgery*. 2009; 145(5):514–518. [PubMed: 19375610]
26. Roh JL, Yoon YH, Park CI. Recurrent laryngeal nerve paralysis in patients with papillary thyroid carcinomas: evaluation and management of resulting vocal dysfunction. *Am J Surg*. 2009; 197(4): 459–465. [PubMed: 18848696]

27. Ambe PC, Bromling S, Knoefel WT, Rehders A. Prolonged duration of surgery is not a risk factor for postoperative complications in patients undergoing total thyroidectomy: a single center experience in 305 patients. *Patient Saf Surg.* 2014; 8(1):45. [PubMed: 25610494]
28. Burn JJ, Taylor SF. Natural history of thyroid carcinoma. A study of 152 treated patients. *Br Med J.* 1962; 2(5314):1218–1223. [PubMed: 14017111]
29. Noureldine SI, Olson MT, Agrawal N, Prescott JD, Zeiger MA, Tufano RP. Effect of Gene Expression Classifier Molecular Testing on the Surgical Decision-Making Process for Patients With Thyroid Nodules. *JAMA Otolaryngol Head Neck Surg.* 2015; 141(12):1082–1088. [PubMed: 26606459]
30. Chan WF, Lang BH, Lo CY. The role of intraoperative neuromonitoring of recurrent laryngeal nerve during thyroidectomy: a comparative study on 1000 nerves at risk. *Surgery.* 2006; 140(6): 866–872. discussion 872–863. [PubMed: 17188132]
31. Lorente-Poch L, Sancho JJ, Munoz-Nova JL, Sanchez-Velazquez P, Sitges-Serra A. Defining the syndromes of parathyroid failure after total thyroidectomy. *Gland Surg.* 2015; 4(1):82–90. [PubMed: 25713783]
32. Stoll SJ, Pitt SC, Liu J, Schaefer S, Sippel RS, Chen H. THYROID HORMONE REPLACEMENT AFTER THYROID LOBECTOMY. *Surgery.* 2009; 146(4):554–560. [PubMed: 19789012]
33. Vaiman M, Nagibin A, Hagag P, Kessler A, Gavriel H. Hypothyroidism following partial thyroidectomy. *Otolaryngol Head Neck Surg.* 2008; 138(1):98–100. [PubMed: 18165001]
34. Bauer PS, Murray S, Clark N, Pontes DS, Sippel RS, Chen H. Unilateral thyroidectomy for the treatment of benign multinodular goiter. *J Surg Res.* 2013; 184(1):514–518. [PubMed: 23688788]
35. Li H, Robinson KA, Anton B, Saldanha IJ, Ladenson PW. Cost-effectiveness of a novel molecular test for cytologically indeterminate thyroid nodules. *J Clin Endocrinol Metab.* 2011; 96(11):E1719–1726. [PubMed: 21865367]
36. Esnaola NF, Cantor SB, Sherman SI, Lee JE, Evans DB. Optimal treatment strategy in patients with papillary thyroid cancer: a decision analysis. *Surgery.* 2001; 130(6):921–930. [PubMed: 11742318]

Figure 1.

Markov decision-analytic model comparing diagnostic lobectomy versus genetic testing with Afirma for evaluation of indeterminate thyroid nodule. Post-operative states consist of the presence or absence of complications. Surveillance represents ongoing observation of an initially benign nodule monitored for growth or malignant characteristics that trigger repeat evaluation. Both the Undetected Disease and the Surveillance nodes include the potential for future thyroidectomy, and the Undetected Disease node includes potential for disease progression and death.

Figure 2.

Marginal probabilistic sensitivity analysis of net-benefit as each parameter is sampled 1,000 times from its probability distribution while other parameters are held constant. Boxes represent the interquartile range of net benefit; vertical bars within boxes represent the median; whiskers represent the 95% uncertainty interval and circles represent outliers. The top ten most influential parameters are presented: u_{AL} is the utility after lobectomy (including possibility of hormone replacement); u_{Surv} is the utility of surveillance; u_{AT} is the utility after thyroidectomy; c_{TL} is the cost of lobectomy; c_{Vera} is the cost of the Afirma test; c_{Surv} is the cost of surveillance; c_{TT} is the cost of thyroidectomy; p_{specA} is the specificity of the Afirma test; u_{Vera} is the utility of Afirma surveillance; p_{STT} is the probability of thyroidectomy after suspicious Afirma result.

Figure 3.

Recursive partitioning analysis of 100,000 Monte Carlo replications of the Markov decision model. Boxes represent subsets of replications where parameter values meet all splitting criteria listed above the box. Within each box, the top number represents the proportion of replications in which Afirma is cost-effective at a threshold of \$100,000 per QALY gained (which can be interpreted as the probability of cost-effectiveness; the bottom number represents the proportion of replications meeting the splitting criteria. u_{dif} is equal to the utility of Afirma minus the utility of post-lobectomy (without hormone replacement) health states; u_{DL} is the utility of detected local cancer health state; c_{TL} is the cost of lobectomy.

Table 1

Estimates of probabilities, utilities, costs, and genetic testing characteristics

	Base Case Value (range)	Distribution	95% UI	Source
Operating Characteristics				Alexander et al.
Sensitivity of Afirma for follicular neoplasm/Hurthle cell neoplasm	0.90	Beta	(0.80, 0.99)	Alexander et al. ⁵
Specificity of Afirma for follicular neoplasm/Hurthle cell neoplasm	0.517	Beta	(0.444, 0.588)	Alexander et al. ⁵
Sensitivity of Afirma for AUS/FLUS	0.90	Beta	(0.74, 0.98)	Alexander et al. ⁵
Specificity of Afirma for AUS/FLUS	0.53	Beta	(0.43, 0.63)	Alexander et al. ⁵
Probabilities				Alexander et al. Durante et al. Esnaola et al. Yassa et al.
Probability of cancer in follicular neoplasm/Hurthle cell neoplasm	0.25 (.14 – .33)	Beta	(0.10, 0.30)	Alexander et al. ⁵ Schneider et al. ¹³ Gharib et al. ^{14,15} Pu et al. ¹⁶ Sullivan et al. ¹⁷ Haugen et al. ⁴
Probability of cancer in AUS/FLUS	0.24 (0.06 – .36)	Beta	(0.15, 0.30)	Alexander et al. ⁵ Schneider et al. ¹³ Sullivan et al. ¹⁷ Haugen et al. ⁹
Benign on Repeat FNA after growth	0.989 (.982-.997)	Beta	(0.976, 0.997)	Durante et al. ¹⁸ Alexander et al. ⁵ Ospina et al. ¹⁹
Cancer on Repeat FNA after growth	0.003 (0.0002 – 0.018)	Beta	(0, 0.013)	Durante et al. ¹⁸ Alexander et al. ⁵ Ospina et al. ¹⁹ Haugen et al. ⁹
Death - 6 month hazard Stage 4 papillary thyroid cancer	0.018	Beta	(0.016, 0.021)	SEER 18, Seerstat 8.3.2 (10.5% mortality at 3 years)
Detect local cancer per cycle of Routine surveillance	0.100	Beta	(0.05, 0.166)	Expert opinion (authors expert opinion) Durante et al. ⁵ Ajmal et al. ²⁰
Growth after initial benign result	0.017 (0.013 – 0.52)	Beta	(0.014, 0.019)	Durante et al. ¹⁸ Alexander et al. ⁵ Ospina et al. ¹⁹ Ajmal et al. ^{20,21}
Permanent complication from lobectomy	0.014 (0.0029 – 0.032)	Beta	(0.012, 0.016)	Dralle et al. ²¹ Promberger et al. ²²
Permanent complication from thyroidectomy	0.039 (0.0020 – 0.046)	Beta	(0.033, 0.045)	Dralle et al. ²¹ Promberger et al. ²² Merchavy et al. ²³ Dionigi et al. ²⁴ Alvarado et al. ²⁵ Roh et al. ²⁶ Ambe et al. ²⁷
Progression local cancer per cycle	0.005	Beta	(0, 0.024)	Extrapolation from SEER 18, Seerstat 8.3.2 (incidence of distant mets at detection)

	Base Case Value (range)	Distribution	95% UI	Source
				Durante et al. ¹⁸ Burn et al. ²⁸
Local cancer in 1cm indeterminate nodule	0.238 (0.06–0.36)	Beta	(0.206, 0.272)	Alexander et al. ⁵ Haugen et al. ⁹ Schneider et al. ¹³ Gharib et al. ¹⁴ Pu et al. ¹⁶ Sullivan et al. ¹⁷
Surgery growth and subsequent indeterminate FNA	0.824	Beta	(0.776, 0.869)	Noureldine et al. ²⁹
Thyroidectomy After Suspicious Afirma test	0.667	Beta	(60.5, 0.728)	Noureldine et al. ²⁹
Temporary complication from lobectomy	0.074 (0.054–0.095)	Beta	(0.054, 0.095)	Chan et al. ³⁰ Lorente-Poch et al. ³¹ Promberger et al. ²²
Temporary complication from thyroidectomy	0.218 (0.19–0.28)	Beta	(0.19, 0.247)	Chan et al. ³⁰ Lorente-Poch et al. ³¹ Promberger et al. ²²
Hormone replacement required after Lobectomy	0.143 (0.11–0.19)	Beta	(0.115, 0.173)	Stoll et al. ³² Vaiman et al. ³³ Bauer et al. ³⁴
Health State Utilities				Li et al. ³⁵ Esnaola et al. ³⁶
U(Post-lobectomy, no hormone replacement required)	0.990	Beta	(0.95, 1)	Esnaola et al. ³⁶
U(Post-lobectomy, hormone replacement required)/ U(Thyroidectomy)	0.950	Beta	(0.877, 0.991)	Li et al. ³⁵
U(Distant cancer)	0.700	Beta	(0.569, 0.815)	Esnaola et al. + author's extrapolation
U(Detected local cancer)	0.950	Beta	(0.877, 0.991)	Esnaola et al. + author's extrapolation
U(Permanent complications of lobectomy)	0.700	Beta	(0.575, 0.814)	Li et al. ³⁵
U(Permanent complications of thyroidectomy)	0.650	Beta	(0.523, 0.767)	Li et al. ³⁵
U(6 months before death from distant cancer)	0.350	Beta	(0.24, 0.467)	
U(Routine Surveillance)	0.980	Beta	(0.928, 0.999)	Li et al. ³⁵
U(Temporary complications of lobectomy or thyroidectomy)	0.950	Beta	(0.86, 0.987)	Li et al. ³⁵
U(Surveillance with Afirma)	0.980	Beta	(0.928, 0.999)	Li et al. ³⁵
Temporary decrease in utility due to lobectomy/thyroidectomy	0.013	Beta	(0, 0.057)	0 on Day of Surgery ; Health Utilities Index: 11222223 for 3 days; 11122122 for 4 days; 11122111 for 7 days
Costs (2016 US Dollars)				Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
One-time:				
Lobectomy	3538	Gamma	(1832, 5804)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD

	Base Case Value (range)	Distribution	95% UI	Source
Thyroidectomy	4102	Gamma	(2081, 6726)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Completion thyroidectomy	3639	Gamma	(1896, 6059)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
FNA	325	Gamma	(167, 538)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Neck dissection	7731	Gamma	(3948, 12449)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Initial treatment of complications from lobectomy	5515	Gamma	(2813, 9107)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Initial treatment of permanent complications from thyroidectomy	3280	Gamma	(1690, 5326)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Initial treatment of temporary complications from thyroidectomy	400	Gamma	(209, 663)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Terminal cycle distant cancer	3668	Gamma	(1893, 5975)	https://costprojections.cancer.gov/ all other sites, adjusted with Medical CPI to 2016 USD
Radiation for distant cancer	869	Gamma	(447, 1408)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Recurring (per cycle):				Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Routine Surveillance	309	Gamma	(159, 510)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Replacement hormone (per cycle)	180	Gamma	(93, 297)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Ongoing treatment of permanent complications of lobectomy	959	Gamma	(492, 1576)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD
Ongoing treatment of permanent complications of thyroidectomy	558	Gamma	(286, 919)	Center for Medicare & Medicaid Services (https://costprojections.cancer.gov/), adjusted with Medical CPI to 2016 USD