VALUE IN HEALTH 15 (2012) 1005–1013

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Cost-Effectiveness of Using a Molecular Diagnostic Test to Improve Preoperative Diagnosis of Thyroid Cancer

Mehdi Najafzadeh, PhD[1], Carlo A. Marra, PharmD, PhD[2,3], Larry D. Lynd, PhD[2,3], Sam M. Wiseman, MD, FRCSC[4,]*

[1]*Harvard/MGH Center on Genomics, Vulnerable Populations, and Health Disparities, Mongan Institute for Health Policy, Massachusetts General Hospital, Boston, MA, USA;* [2]*Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada;* [3]*Centre for Health Evaluation and Outcome Sciences, St. Paul's Hospital, Vancouver, BC, Canada;* [4]*Department of Surgery & St Paul's Hospital, University of British Columbia, Vancouver, BC, Canada*

ABSTRACT

Objective: Fine-needle aspiration biopsy (FNAB) is a safe and inexpensive diagnostic procedure for evaluating thyroid nodules.Up to 25% of the results from an FNAB, however, may not be diagnostic or may be indeterminate, leading to a subsequent diagnostic thyroid surgery. A new molecularly based diagnostic test could potentially reduce indeterminate cytological results and, with high accuracy, provide a definitive diagnosis for cancer in thyroid nodules. The aim of the study was to estimate the cost-effectiveness of utilizing a molecular diagnostic (DX) test as an adjunct to FNAB, compared with NoDX, to improve the preoperative diagnosis of thyroid nodules. **Methods:** We constructed a patient-level simulation model to estimate the clinical and economic outcomes of using a DX test compared with current practice (NoDX) for the diagnosis of thyroid nodules. By using a cost-effectiveness framework, we measured incremental clinical benefits in terms of quality-adjusted life-years and incremental costs over a 10-year time horizon. **Results:** Assuming 95% sensitivity and specificity of the Dx test when used as an adjunct to FNAB, the utilization of the DX test resulted in a gain of 0.046 quality-adjusted life-years (95% confidence interval 0.019–0.078) and a saving of $1087 (95% confidence interval $691–$1533) in direct costs per patient. If the cost of the Dx test is less than $1087 per test, we expect to save quality-adjusted life-years and reduce costs when it is utilized. Sensitivity of the DX test, compared with specificity, had a larger influence on the overall outcomes.

Keywords: cost-effectiveness, gene expression, molecular diagnostic test, thyroid cancer, fine needle aspiration biopsy, The Bethesta System for Reporting Thyroid Cytopathology (TBSRTC).

Introduction

Thyroid nodules are common, affecting from 4% to 7% in the general population [1]. Fortunately, only a small proportion of these nodules (approximately 5%) are cancerous [1,2]. Fine-needle aspiration biopsy (FNAB) is currently the standard of practice for evaluating thyroid nodules due to its low cost, low complication rate, and availability [3]. The introduction of FNAB in the early 1980s has resulted in significant economic savings and improved patient outcomes by reducing the number of diagnostic thyroid operations required and improving the detection rate of thyroid cancer [4].

A major drawback of FNAB of thyroid nodules is the large number of indeterminate or suspicious cancer diagnoses in that often the lesions cannot definitively be classified as benign or malignant following FNAB. Clinical decision making following an indeterminate cytological result is challenging and may lead to either overtreatment or undertreatment of thyroid nodules. This diagnostic uncertainty is a consequence of not only the subjective nature of thyroid cytology but also the overlapping cytomorphologic characteristics of benign and malignant thyroid lesions. The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) is the result of a recent multidisciplinary effort to develop a uniform system for reporting thyroid FNAB cytological results and their clinical significance [5]. The Bethesda thyroid FNAB cytology reporting system is particularly helpful for providing a clear estimate of cancer risk and recommendations for the management of indeterminate cases (i.e., atypia of undetermined significance, follicular neoplasm, and suspicious cases). Even after following the currently recommended diagnostic algorithm, about half of the thyroid nodules operated on for an indeterminate diagnosis are eventually found to be benign [5,6]. Therefore, the development of a new diagnostic test that can serve as an adjunct to FNAB and reduce the number of unnecessary diagnostic thyroid operations would have significant clinical and economic value.

In the current study, we aimed to 1) evaluate the overall performance of FNAB in combination with the TBSRTC (current practice) in directing thyroid nodule surgical management and 2) estimate the cost-effectiveness of using a new molecular diagnostic (DX) test as an adjunct to FNAB compared with best current practice (FNAB in combination with the TBSRTC guidelines, hereafter referred to as NoDX). We performed the analyses in scenarios for only individuals with an indeterminate FNAB cytological diagnosis, as well as for all individuals who present with thyroid nodules (i.e., including nondiagnostic, benign, and malignant cases in the target population).

* *Address correspondence to:* Sam M. Wiseman, Department of Surgery, University of British Columbia, St. Paul's Hospital, C303-1081 Burrard Street, Vancouver, BC V6Z 1Y63, Canada.

E-mail: smwiseman@providencehealth.bc.ca.

1098-3015 – see front matter Copyright © 2012, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2012.06.017

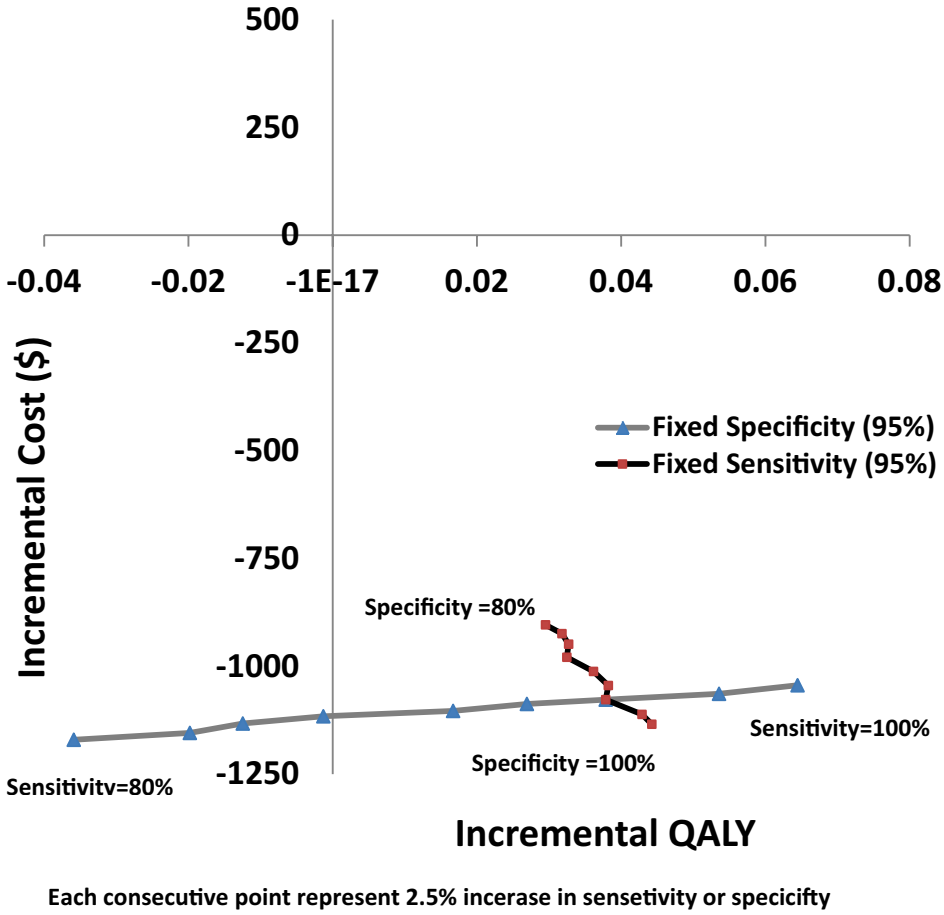

Fig. 1 – Incremental cost and effectiveness for different sensitivity and specificity of DX test. DX, diagnostic; QALY, quality-adjusted life-year.

Methods

Model design

We developed a patient-level discrete event simulation model by using Arena, Version 13.0 (Rockwell Software, Inc., Milwaukee, WI), to calculate the incremental cost-effectiveness of using the DX test in conjunction with FNAB relative to current practice (NoDX) in two simulated cohorts of 10,000 patients with an initial indeterminate FNAB cytological diagnosis. In a patient-level model, one hypothetical patient is created and assigned to the NoDX arm and an identical clone is created and assigned to the DX arm (see Figure 1 available in the Appendix in Supplemental Material found at http://dx.doi.org/10.1016/j.jval.2012.06.017). The model was run over a 10-year time horizon, and results were determined from the societal perspective. Depending on the subsequent diagnostic results in each arm, patients underwent either a total thyroidectomy or hemithyroidectomy, or alternatively were followed clinically, in accordance with the TBSRTC. In particular, we simulated the number of malignant cases across diagnostic categories, and then by comparing the final cytology and histology, each patient was classified as a true-positive, false-positive, true-negative, or false-negative cancer case. The incidence of major morbidity following surgery (i.e., permanent hypoparathyroidism and/or recurrent laryngeal nerve injury [RLNI]) and cancer recurrence was included in the model as was cancer-related mortality.

We needed to explicitly simulate the results of sequential test outcomes for each patient to calculate the overall number of false positives and false negatives in our simulation. We also modeled the time as a continuous variable (rather than using fixed cycles). As such, we used discrete event simulation, which offers the capacity to develop continuous time, patient-level simulations with great flexibility for doing first-order and second-order sensitivity analyses [7,8].

Data sources and assumptions

In the simulation, following the TBSRTC classification of a nodule, there were six possible cytologic outcomes subsequent to an FNAB [5]: 1) nondiagnostic or unsatisfactory, 2) benign, 3) atypia or follicular lesion of undetermined significance, 4) follicular neoplasm, 5) suspicious for malignancy, or 6) malignant. We defined an indeterminate cytological result as an FNAB diagnosis being atypia or follicular lesion of undetermined significance, follicular neoplasm, or suspicious for malignancy. Furthermore, on the basis of review of the literature, we utilized the proportion of patients in each category, and the probabilities of malignancy for each category, to inform the progression through the model (Table 1) [5]. The possible pathways in the model were designed on the basis of proposed management recommendations for each cytological diagnostic category [5].

Unlike the NoDx test arm, the diagnostic results in the DX test arm were confined to two possible diagnoses: benign or malignant. The overall prevalence of malignancy was assumed to be equal in the two arms of the model. Therefore, for a given sensi-

Table 1 – Model assumptions about cytology results.

Parameter	Point estimate (%)	Range	Distribution (α, β)	Reference
Proportion of cytology results, first FNAB*				Cibas and Ali [5]
Nondiagnostic (ND)	11	(2.0, 20.0)	Beta (5, 42)	
Benign	65	(60.0, 70.0)	Beta (236, 127)	
Atypia of undetermined significance (AUS)	4.5	(3.0, 6.0)	Beta (34, 729)	
Follicular neoplasm (FN)	7.3	(6.6, 8.0)	Beta (371, 4708)	
Suspicious for malignancy	7.2	(6.5, 7.9)	Beta (371, 4783)	
Malignant	5	(3.0, 7.0)	Beta (24, 450)	
Probability of malignancy conditional on cytology results				Cibas and Ali [5]
ND	2.5	(1, 4)	Beta (11, 422)	
Benign	1.5	(0, 3)	Beta (4, 258)	
AUS	10	(5, 15)	Beta (14, 129)	
FN	22.5	(15, 30)	Beta (28, 95)	
Suspicious for malignancy	67.5	(60, 75)	Beta (105, 50)	
Malignant	98.5	(97, 99)	Beta (581, 9)	
Proportion of second cytology results if AUS				Cibas and Ali [5]
Benign	61.5	(55.4, 67.7)	Beta (153, 96)	
AUS	20.0	(18.0, 22.0)	Beta (320, 1279)	
FN	6.9	(6.2, 7.6)	Beta (372, 5014)	
Suspicious for malignancy	6.8	(6.1, 7.5)	Beta (373, 5094)	
Malignant	4.7	(4.3, 5.2)	Beta (381, 7668)	
Proportion of second cytology results if ND				Cibas and Ali [5]
ND	36.0	(32.4, 39.6)	Beta (256, 454)	
Benign	49.2	(44.3, 54.2)	Beta (203, 209)	
FN	5.5	(5.0, 6.1)	Beta (378, 6456)	
Suspicious for malignancy	5.5	(4.9, 6.0)	Beta (378, 6556)	
Malignant	3.8	(3.4, 4.2)	Beta (385, 9777)	
Proportion of malignancy types				Davies and Welch [9]
Papillary	79.7	(76.8, 82.6)	Beta (607, 155)	
Follicular	12.1	(9.7, 14.5)	Beta (92, 669)	
Medullary	2.4	(1.3, 3.5)	Beta (18, 743)	
Hurthle cell	4.9	(3.3, 6.5)	Beta (37, 724)	
Anaplastic	0.9	(0.2, 1.6)	Beta (7, 754)	

* Point estimates represent the middle of the reported ranges in Cibas et al. Parameters for beta distributions are calculated on the basis of the reported ranges.

tivity and specificity of the DX test, possible cytological and histological outcomes were simulated in the DX test arm.

Data from National Cancer Institute's Surveillance, Epidemiology, and End Results were utilized to characterize the major types of thyroid malignancy as well as the proportions of different types of cancer conditional on having a malignant nodule (Table 1) [9]. Because there are large differences in the mean age of onset for the different types of thyroid cancer (e.g., individuals diagnosed with anaplastic carcinoma are, on average, more than 20 years older than individuals diagnosed with papillary carcinoma), the expected life years lost vary widely across the different types of thyroid cancer (Table 2). These differences are captured in the simulation as are the differences in the mortality rates associated with different thyroid cancers [10].

Permanent hypoparathyroidism and permanent RLNI were considered to be the two major complications of thyroid surgery. Despite being uncommon events [14], the morbidity associated with these complications is generally chronic and thus impacts clinical outcomes and costs over a lifetime. In the model, we assumed that the failure to detect a malignancy delays necessary surgeries by 28 months on average [13,15] and, therefore, increases the risk of cancer recurrence (relative risk = 2.28) and cancer-related mortality (relative risk = 2.11) [13].

Quality of life

We included 10 health states in the model: perfect health, surgery, pre–radioactive iodine ablation, post–radioactive iodine ablation weeks 0 to 4, post–radioactive iodine ablation weeks 4 to 8, unilateral RLNI, bilateral RLNI, hypoparathyroidism, recurrence of differentiated carcinoma, and death (Table 3). A heath state utility value (six-dimensional health state short form [derived from short form 36 health survey]) was assigned to each health state to facilitate the calculation of quality-adjusted life-years (QALYs) as the final model outcome [20]. For each individual in the model, QALYs were calculated by multiplying the utility values of each health state with the time spent in that health state throughout the model [16].

Mernagh et al. [20] used health-related quality-of-life data (using short form 36 health survey questionnaire) that had been collected during an international multicenter randomized clinical trial with patients who received ablation treatment for their thyroid cancer [21]. We used the six-dimensional health state short form (derived from short form 36 health survey) utilities reported in Mernagh et al. that were derived from short form 36 health survey data in the hypothyroid arm of this clinical trial (N = 30).

Table 2 – Model assumptions about rates of mortality and side effects.

Parameter	Point estimate	Range	Distribution	Reference
Conditional probability of mortality (%)				Shaha [10]
Papillary	7	(6.3, 7.7)	Beta (372, 4941)	
Follicular	15	(13.5, 16.5)	Beta (340, 1926)	
Medullary	24	(21.6, 26.4)	Beta (304, 962)	
Hurthle cell	25	(22.5, 27.5)	Beta (300, 899)	
Anaplastic	86	(77.4, 94.6)	Beta (55, 9)	
Average age of patients (y)*				Hundahl [11]
Papillary	47	(37.6, 56.4)	Gamma (0.47, 100)	
Follicular	50	(40.0, 60.0)	Gamma (0.50,100)	
Medullary	48	(38.4, 57.6)	Gamma (0.48, 100)	
Hurthle cell	49	(39.2, 58.8)	Gamma (0.49, 100)	
Anaplastic	70	(56.0, 84.0)	Gamma (0.70, 100)	
Time between diagnosis and death in cancer-related cases				
High-risk malignancy (mo)	7	(4.0, 6.0)	Gamma (0.07, 100)	Shaha [10]
High-risk malignancy (y)	5	(5.6, 8.4)	Gamma (0.05, 100)	Shaha [10]
Probability of cancer recurrence (%)†	15.5	(8.3, 40.0)	Beta (3, 17)	Kebebew [12]
Effect of false-negative FNAB cytology on outcomes				Sipos and Mazzaferri [13]
Average delay in diagnosis (mo)	28	(25.2, 30.8)	Normal (28, 1.4)	
Relative risk of mortality	2.17	(2.0, 2.4)	Normal (2.17, 0.11)	
Relative risk of long-term recurrence	2.28	(2.1, 2.5)	Normal (2.28, 0.11)	
Probability of major side effects following thyroidectomy (%)				Rosato et al. [14]
Persistent hypoparathyroidism	1.7	(1.5, 1.9)	Beta (393, 22735)	
Permanent laryngeal recurrent nerve (LRN) injury	1	(0.9, 1.1)	Beta (396, 39203)	

* Estimated on the basis of the age distribution of the malignant cases in Hundahl et al. [11].

† Risk of recurrent cancer in 10-year follow-up.

Table 3 – Utility weights and unit costs.

Parameter	Point estimate	Range	Distribution for PSA*	Reference
Unit costs (US $)				
HT, direct costs	$2390	($1912, $2868)	Gamma (23.9, 100)	Soria-Aledo et al. [16]
HT, indirect costs†	$288	($230, $346)	Gamma (2.9, 100)	Soria-Aledo et al. [16]
TT, direct costs	$3058	($2446, $3670)	Gamma (30.6, 100)	Soria-Aledo et al. [16]
TT, indirect costs	$418	($334, $502)	Gamma (4.2, 100)	Soria-Aledo et al. [16]
Ablation, direct costs	$3740	($2992, $4488)	Gamma (37.4, 100)	Wang et al. [17]
Ablation, indirect costs	$784	($627, $941)	Gamma (7.8, 100)	Wang et al. [17]
Ultrasound-guided FNAB	$412	($330, $494)	Gamma (4.1, 100)	Khalid [18]
Follow-up, 5 y	$1111	($889, $1333)	Gamma (11.1, 100)	Shrime [19]
Utilities				
Surgery	0.637	(0.57, 0.70)	Beta (145, 82)	Mernagh et al. [20]
Preablation	0.548	(0.49, 0.60)	Beta (180, 149)	Mernagh et al. [20]
Postablation 0–4 wk	0.637	(0.57, 0.70)	Beta (145, 82)	Mernagh et al. [20]
Postablation 4–8 wk	0.819	(0.74, 0.90)	Beta (72, 16)	Mernagh et al. [20]
Unilateral RLNI	0.627	(0.56, 0.69)	Beta (149, 88)	Kebebew [12]
Bilateral RLNI	0.205	(0.18, 0.23)	Beta (318, 1232)	Kebebew [12]
Hypoparathyroidism	0.778	(0.70, 0.86)	Beta (88, 25)	Kebebew [12]
Cancer recurrence	0.54	(0.49, 0.59)	Beta (183, 153)	Kebebew [12]
Well	1	–	–	Definition
Duration of health states (d)				
Length of hospitalization for TT	4.2	(3.4, 5.0)	Gamma (0.04, 100)	Soria-Aledo et al. [16]
Length of hospitalization for HT	2.9	(2.3, 3.5)	Gamma (0.03, 100)	Soria-Aledo et al. [16]
Time between surgery and ablation	7	(5.6, 8.4)	Gamma (0.07, 100)	Mernagh et al. [20]
Postablation recovery period I	28	(22.4, 33.6)	Gamma (0.28, 100)	Mernagh et al. [20]
Postablation recovery period II	28	(22.4, 33.6)	Gamma (0.28, 100)	Mernagh et al. [20]

FNAB, fine-needle aspiration biopsy; HT, hemithyroidectomy; PSA, probabilistic sensitivity analysis; RLNI, recurrent laryngeal nerve injury; TT, total thyroidectomy.

* Parameters for gamma and beta distributions were calculated by assuming an SD equivalent to 10% of point estimates.

† Indirect costs for TT and HT were calculated on the basis of lengths of hospitalizations and assuming an average income of $99 per day. The average income has been based on men and women annual income and assuming that 12% of the patients were men.

Costs

The model included the cost of the DX test, total thyroidectomy, hemithyroidectomy, clinical follow-up (i.e., yearly physician visits for 5 years), radioactive iodine ablation procedure, and FNAB (Table 3). The cost of the DX test was assumed to include costs of all related procedures including pathology, physician time, and specimen transport and processing.

Direct costs for total thyroidectomy or hemithyroidectomy in our study were based on the results from Soria-Aledo et al. [16]. Indirect costs for total thyroidectomy or hemithyroidectomy were calculated on the basis of average lengths of hospitalizations subsequent to those surgeries and assuming an average income of $99 per day. We calculated the average annual income of men and women on the basis of US census data and assuming that 12% of the patients were men [16]. Our estimation of average income per day is also consistent with the findings from other studies including Wang et al. [17]. Both direct and indirect costs of ablation were based on estimations that have been reported in Wang et al. [17].

Model outcomes

Simulated outcomes for each patient were recorded for the 10-year time horizon of the model (e.g., initial and final cytology, occurrence of cancer, type of cancer, hemithyroidectomy, total thyroidectomy, completion thyroidectomy, complications, and mortality). The health outcomes, in terms of QALYs and costs for each patient, were also accumulated throughout the model stages over the 10-year time horizon. In the base-case analysis, only the direct costs were included and both QALYs and costs were discounted at 3% annually. Sensitivity analyses, however, were conducted by including indirect costs, and by varying the annual discount rate from 3% to 0%. Finally, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the NoDx strategy (FNAB in combination with the TBSRTC) were also estimated and compared with the results of other studies.

We considered the societal perspective in this analysis, and the results were reported in US $, year 2011 values. All the unit costs were inflated to 2011 values in the model by using the annual changes in the consumer price index.

Univariate sensitivity analysis

In the base-case scenario, we generated 10,000 patients for each strategy, with each patient undergoing 1,000 random walks in order to minimize first-level uncertainty, resulting in 10,000,000 simulations per arm [22]. Univariate sensitivity analyses were then conducted to evaluate the effect of varying the value of model parameters over a viable range on the overall outcomes of the model. For a given level of DX test sensitivity, we reran the model by increasing the test specificity from 80% to 100% by 2.5% increments and compared the resulting incremental costs and QALYs. We repeated the same analysis by holding specificity constant and increasing sensitivity 2.5% at each step. We also explored the benefits of expanding the target population of the DX test to include all individuals who present with palpable thyroid nodules.

Probabilistic sensitivity analysis

Values of the parameters in the base-case model are point estimates drawn from various studies and are intrinsically uncertain. To capture the effect of parameter uncertainty on the outcomes, we conducted a probabilistic sensitivity analysis by assigning probability distributions to all model parameters [23].

Results

Results of base-case analysis

In the base-case analysis, among 10,000 simulated patients with an indeterminate diagnosis in the NoDX test arm (current practice), 4,407 benign cases underwent unnecessary diagnostic surgery (false positives) while 116 cancers were not diagnosed (false negatives). In comparison, assuming a DX test sensitivity and specificity of 95%, the number of unnecessary diagnostic thyroid operations was reduced to 323 and 175 cancer cases were not diagnosed (Table 4). Not surprisingly, the substantial reduction in the number of operations was also associated with a reduction in

Table 4 – Base-case results of DX use in indeterminate cases only.

	NoDX (n = 10,000)	DX (n = 10,000)
Final cytology (number of cases)		
Nondiagnostic or unsatisfactory (ND)	0	0
Benign	1479	6182
Atypia of undetermined significance (AUS)	454	0
Follicular neoplasm (FN)	3971	0
Suspicious for malignancy	3987	0
Malignant	109	3818
Final histology (number of cases)		
True negative	1898	6007
True positive	3579	3495
False positive	4407	323
False negative	116	175
Performance measures (probabilities)		
Sensitivity	0.97	0.95
Specificity	0.30	0.95
Positive predictive value (PPV)	0.45	0.92
Negative predictive value (NPV)	0.94	0.97
Malignant cases (number of cases)		
Papillary carcinoma	2924	2936
Follicular carcinoma	503	414
Medullary carcinoma	172	195
Hurthle cell carcinoma	75	92
Anaplastic carcinoma	21	33
Major adverse events (number of cases)		
Persistent hypoparathyroidism	129	68
Permanent laryngeal recurrent nerve injury	89	29
Cancer-related mortality	391	393
Outcomes		
QALY loss	0.306	0.266
Cost (excluding DX test cost, US $)	4638	3558
Incremental QALY		0.040
Incremental cost (US $)*		−1080

DX, diagnostic; QALY, quality-adjusted life-year.
* Negative cost indicates saving.

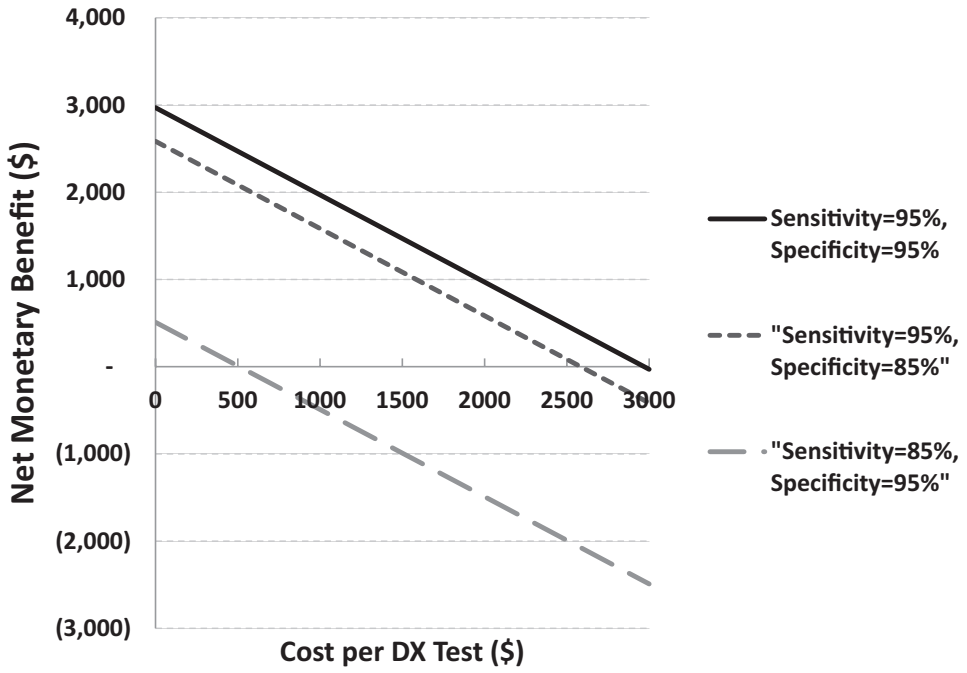

Fig. 2 – Net monetary benefit of DX test for different levels of cost per DX test. Incremental net monetary benefits were calculated on the basis of λΔQALY – ΔC, where Δ QALY and ΔC are the incremental QALYs and incremental costs of DX relative to NoDX, respectively, and λ is society's willingness to pay for an additional QALY gained, which was assumed to be $50,000. DX, diagnostic; QALY, quality-adjusted life-year.

surgical-related complications (Table 4). Specifically, in the NoDx test arm, 129 patients experienced persistent hypoparathyroidism compared with only 68 in the DX test arm. Overall, using the DX test resulted in a gain of 0.046 (95% credible interval 0.019–0.078) QALYs per patient over 10 years (or, 46 QALYs per 1000 patients treated) with a cost saving of $1087 (95% credible interval $691–$1533) in direct costs per patient. Thus, if the cost of the DX test is less than $1087, the DX test strategy would result in QALYs gained and lower costs relative to using the NoDX test strategy.

As expected, varying the sensitivity and specificity of the DX test impacted the results (Fig. 1). Given a specificity of 95%, if the sensitivity of Dx is 87.5% or less, the incremental QALYs associated with using the DX test strategy are negative, suggesting that the DX test strategy is inferior to the NoDX test strategy in terms of clinical outcomes. As shown in Figure 1, the incremental QALYs gained vary widely under different assumptions about the sensitivity of the DX test strategy, while changes in specificity had a smaller effect on the incremental QALYs. Inclusion of indirect costs that are generally related to productivity loss during surgical procedures and inpatient hospital stays increased the estimated savings attributable to the DX test strategy to $1251 per patient in the base-case analysis. Implementation of the DX test strategy was associated with a one-time cost per patient that is part of the overall direct cost in the DX test arm. Figure 2 shows the net monetary benefit of using different costs for the DX test strategy under the assumptions of the base-case scenario and using an arbitrary value of $50,000 per QALY gained as the threshold willingness to pay (λ).

When we simulated 10,000 patients with an indeterminate FNAB diagnosis in the NoDX testing strategy arm (current practice), 3,695 patients had cancer and 3,579 of those were assigned to surgery (either hemithyroidectomy or total thyroidectomy). A total of 4407 patients with benign thyroid nodules also, however, were assigned to receive a diagnostic operation. Thus, we concluded that for indeterminate cases, sensitivity, specificity, PPV, and NPV of current practice (FNAB combined with the TBSRTC) are 97%, 30%, 45%, and 94%, respectively. When we considered all thyroid nodules, the sensitivity, specificity, PPV, and NPV were 94%, 84%, 49%, and 99%, respectively.

Sensitivity analyses

The results of the univariate sensitivity analyses suggest that the cost of hemithyroidectomy, the discount rate, and the risk of permanent hypoparathyroidism have significant effects on the incremental net monetary benefit (Fig. 3). A higher probability of malignancy among individuals with a cytological diagnosis of suspicious for malignancy, follicular neoplasm, and atypia or follicular lesion of undetermined significance was associated with a higher incremental net monetary benefit for the DX testing strategy. The results of univariate sensitivity analyses suggested that neither of these changes in the input parameters could alter the overall outcome that the DX test results in QALYs gained at lower cost over a broad range of assumptions about input parameters.

We repeated the simulation to include all individuals who present with thyroid nodules. By analyzing the results for all individuals, we evaluated the scenario where the DX test can be consistently used in conjunction with all FNABs. In addition to potentially higher sensitivity and specificity, it is expected that the new DX test requires a very small amount of cellular material relative to FNAB cytology, and therefore a significant proportion of nondiagnostic cases may be reclassified into other diagnostic categories by using the DX test. Under this scenario, sensitivity and specificity of the NoDX test strategy was estimated to be 94% and 84%, respectively. Assuming that the DX test could maintain a sensitivity and specificity of 95% in this scenario, it resulted in 0.025 QALYs gained (i.e., 25 per 1000 patients assessed) at a cost of $291 less than that in the NoDx strategy (see Table 1 available in the Appendix in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2012.06.017).

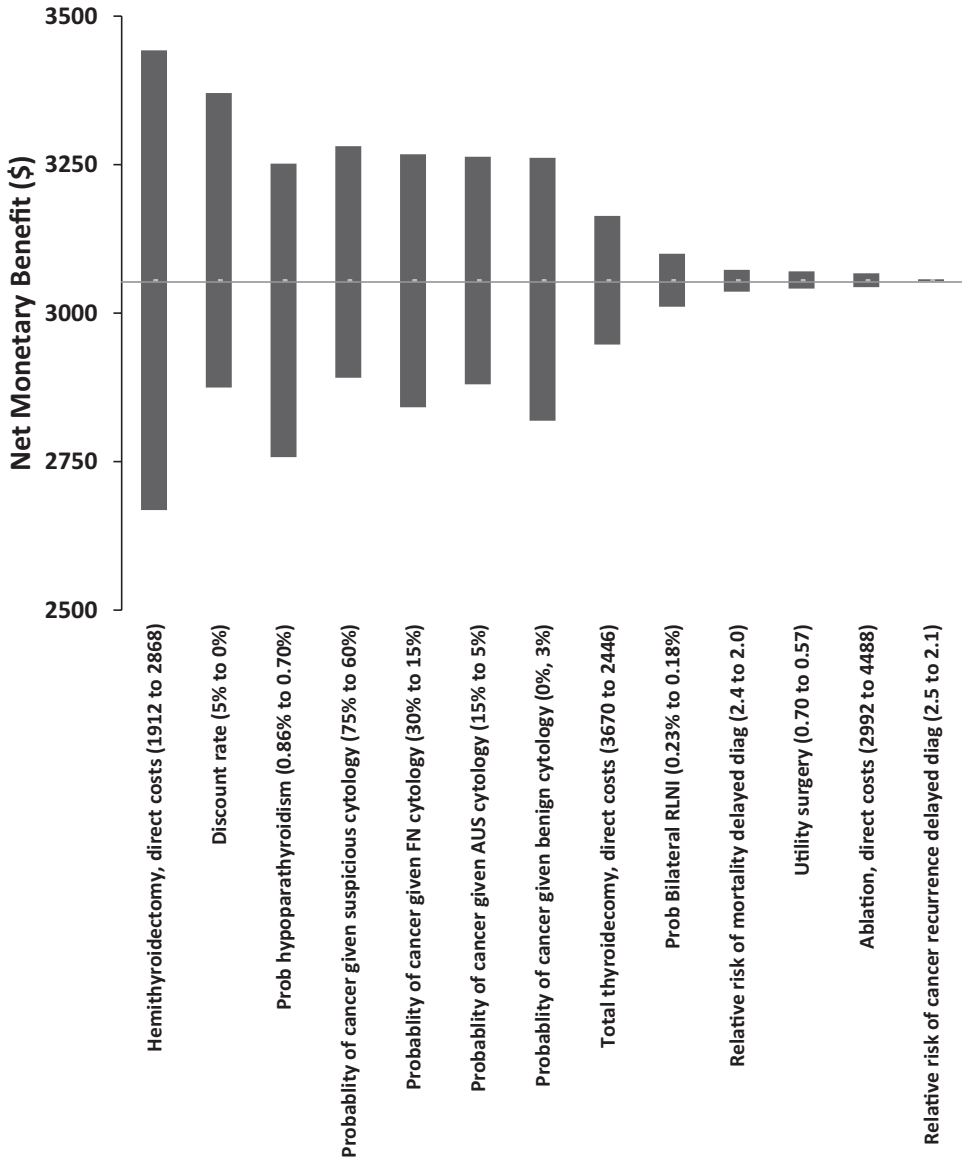

Fig. 3 – Results of univariate sensitivity analysis. AUS, atypia or follicular lesion of undetermined significance; diag, diagnosis; FN, follicular neoplasm; prob, probability; RLNI, recurrent laryngeal nerve injury.

The results of the probabilistic sensitivity analysis also suggest a high probability that the DX test is the dominant strategy (see Figure 2 available in the Appendix in Supplemental Material found at http://dx.doi.org/10.1016/j.jval.2012.06.017). All simulated points fall in the southeast quadrant (the DX testing strategy being more effective but less costly than the NoDx test strategy) if the cost of the Dx test was less than $500. Therefore, assuming 95% sensitivity and specificity, a DX test strategy that costs less than $500 is the dominant strategy with 100% certainty.

Discussion

The current study is the first to evaluate the incremental cost-effectiveness of utilizing a molecular diagnostic testing strategy relative to the currently recommended diagnostic strategy for the diagnosis of thyroid cancer, and it is the first study to evaluate the overall performance of FNAB in combination with the TBSRTC. Given a 95% sensitivity and specificity, the utilization of the DX test strategy to improve the preoperative diagnosis of indeterminate thyroid nodules can substantially reduce the number of diagnostic operations and result in considerable QALY gains and cost savings. The potential cost savings are inversely related to the cost of the DX test strategy but remain beneficial if the cost of the Dx test is less than $1087. Provided that the sensitivities of the DX test and NoDX test strategies are comparable, the DX test strategy remains cost saving if the specificity of the Dx test is more than 80%. In contrast, a small decrement in the sensitivity of the DX test results in a larger negative effect on incremental QALYs gained, because FNAB in conjunction with the TBSRTC already provides high sensitivity at the expense of very low specificity for indeterminate cases and the negative clinical consequences of missing a malignant case are much greater than the expected consequences associated with an unnecessary diagnostic operation for a benign case. Overall, the value of the DX test strategy is largely due to

its ability to prevent unnecessary diagnostic surgical procedures following a false-positive cancer diagnosis that currently comprise approximately half of all thyroid operations [5].

In this analysis, we compared the overall performance of the DX test strategy with that of the NoDX test strategy in terms of the number of undetected cancer cases and the number of patients with benign nodules who were assigned to undergo a thyroid operation. We defined positive cancer diagnosis as a patient who has been referred for consideration of surgery on the basis of the overall evaluation of the diagnostic strategy, as we believe that this definition is more clinically relevant. Therefore, the estimated sensitivity, specificity, PPV, and NPV reflect the overall ability of the strategy to identify cancer cases that require surgery with a high degree of accuracy. For example, the reported sensitivity for the NoDX test strategy reflects characteristics of FNAB in combination with TBSRTC rather than FNAB alone.

Previous studies have evaluated the performance of FNAB alone for the diagnosis of thyroid nodules. Gharib and Goellner [24] suggested that depending on the definition of cytological outcomes, the sensitivity of FNAB can range from 65% to 98% and the specificity can range from 72% to 100%. In a retrospective analysis of 37,895 cases reported by Ravetto et al. [25], the sensitivity and specificity of FNAB was estimated to be 91.8% and 75.5%, respectively. In a review by Tee et al. [26], the observed sensitivity and specificity of FNAB that has been reported also has a broad range, suggesting that the estimated characteristics of FNAB depend on the criteria and the method of measurement. Similar findings have also been reported by several other studies [6,27–30]. We believe that using different assumptions for assigning indeterminate cases into possible diagnostic outcomes has contributed to having wide variation across those studies. In contrast, we have used referral (no referral) to surgery as the criteria for defining positive (negative) diagnosis. Furthermore, in this study, we have estimated characteristics of FNAB in combination with TBSRTC rather than FNAB alone.

We have included the effect of cancer-related mortality on QALYs, but the costs that can occur in final stages of cancer (e.g., palliative care) have not been accounted for in the model. The number of individuals who develop end-stage thyroid cancer, however, was virtually identical in both arms of the model, and therefore would have limited impact on incremental costs or outcomes. The impact of surgery on QALYs was calculated only for the inpatient periods after surgery. In reality, discomfort and limitation in daily activities usually lasts for a period of time after discharge from the hospital, and therefore the impact of surgery on QALYs is likely underestimated in this model. Thus, given a higher frequency of surgeries in the NoDX test strategy, this assumption slightly favored the NoDX test strategy.

Overall, the DX test strategy appears to be the dominant diagnostic strategy, resulting in QALYs gained and reduced costs compared with current practice, if it is shown to have a high sensitivity and specificity with a reasonable cost per test. Regardless of whether the Dx test strategy is based on gene expression, protein expression, transcription analysis, immunocytochemistry, or some other type of molecular study, its ultimate benefit is primarily due to its ability to reduce the number of diagnostic operations performed for benign pathology in individuals who present with thyroid nodules by more accurately diagnosing thyroid cancer without the need for surgery. For clinical relevance and applicability, future studies evaluating molecular diagnostic tests for thyroid cancer must take economic considerations into account.

Source of financial support: This study was supported by the National Research Council Industrial Assistance Program (to Dr. Lesley Esford). Dr. Mehdi Najafzadeh was supported by a Canadian Institute for Health Research Canada (CIHR) Graduate Scholarship during the course of this study.

Acknowledgments

We are thankful for the encouragement that we received from Dr. Elai Davicioni from Genome DX Biosciences, Inc., during this project. We also thank Nicole Tsao and Adam Raymakers for their contribution in reviewing the literature.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2012.06.017 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Mazzaferri EL. Management of a solitary thyroid nodule. N Engl J Med 1993;328:553–9.
[2] Borget I, Vielh P, Leboulleux S, et al. Assessment of the cost of fine-needle aspiration cytology as a diagnostic tool in patients with thyroid nodules. Am J Clin Pathol 2008;129:763–71.
[3] Layfield LJ, Cibas ES, Gharib H, Mandel SJ. Thyroid aspiration cytology: current status. CA Cancer J Clin 2009;59:99–110.
[4] Hamberger B, Gharib H, Melton LJ III, Goellner JR, Zinsmeister AR. Fine-needle aspiration biopsy of thyroid nodules: impact on thyroid practice and cost of care. Am J Med 1982;73:381–4.
[5] Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. Am J Clin Pathol 2009;132:658–65.
[6] Yassa L, Cibas ES, Benson CB, et al. Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. Cancer 2007;111:508–16.
[7] Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ 2006;15:1295–310.
[8] Caro JJ. Pharmacoeconomic analyses using discrete event simulation. Pharmacoeconomics 2005;23:323–32.
[9] Davies L, Welch HG. Thyroid cancer survival in the United States: observational data from 1973 to 2005. Arch Otolaryngol Head Neck Surg 2010;136:440–4.
[10] Shaha AR. Thyroid cancer: extent of thyroidectomy. Cancer Control 2000;7:240–5.
[11] Hundahl SA, Cady B, Cunningham MP, Mazzaferri E, McKee RF, Rosai J, et al. 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 2000;89:202–17.
[12] Kebebew E, Duh QY, Clark OH. 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 2000;24:1295–302.
[13] Sipos JA, Mazzaferri EL. Thyroid cancer epidemiology and prognostic variables. Clin Oncol (R Coll Radiol) 2010;22:395–404.
[14] Rosato L, Avenia N, Bernante P, et al. Complications of thyroid surgery: analysis of a multicentric study on 14,934 patients operated on in Italy over 5 years. World J Surg 2004;28:271–6.
[15] Yeh MW, Demircan O, Ituarte P, Clark OH. False-negative fine-needle aspiration cytology results delay treatment and adversely affect outcome in patients with thyroid carcinoma. Thyroid 2004;14:207–15.
[16] Soria-Aledo V, Flores-Pastor B, Candel-Arenas MF, et al. Evaluation and monitoring of the clinical pathway for thyroidectomy. Am Surg 2008;74:29–36.
[17] Wang TS, Cheung K, Mehta P, et al. 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 2010;95:1672–80.
[18] Khalid AN, Quraishi SA, Hollenbeak CS, Stack BC Jr. Fine-needle aspiration biopsy versus ultrasound-guided fine-needle aspiration biopsy: cost-effectiveness as a frontline diagnostic modality for solitary thyroid nodules. Head Neck 2008;30:1035–9.

[19] Shrime MG, Goldstein DP, Seaberg RM, Sawka AM, Rotstein L, Freeman KL , et al. Cost-effective management of low-risk papillary thyroid carcinoma. Arch Otolaryngol Head Neck Surg 2007;133:1245–53.
[20] Mernagh P, Suebwongpat A, Silverberg J, Weston A. Cost-effectiveness of using recombinant human thyroid-stimulating hormone before radioiodine ablation for thyroid cancer: the Canadian perspective. Value Health 2010;13:180–7.
[21] Pacini F, Ladenson PW, Schlumberger M, et al. Radioiodine ablation of thyroid remnants after preparation with recombinant human thyrotropin in differentiated thyroid carcinoma: results of an international, randomized, controlled study. J Clin Endocrinol Metab 2006;91:926–32.
[22] Halpern EF, Weinstein MC, Hunink MG, Gazelle GS. Representing both first- and second-order uncertainties by Monte Carlo simulation for groups of patients. Med Decis Making 2000;20: 314–22.
[23] O'Hagan A, McCabe C, Akehurst R, et al. Incorporation of uncertainty in health economic modelling studies. Pharmacoeconomics 2005;23: 529–36.
[24] Gharib H, Goellner JR. Fine-needle aspiration biopsy of the thyroid: an appraisal. Ann Intern Med 1993;118:282–9.
[25] Ravetto C, Colombo L, Dottorini ME. Usefulness of fine-needle aspiration in the diagnosis of thyroid carcinoma: a retrospective study in 37,895 patients. Cancer 2000;90:357–63.
[26] Tee YY, Lowe AJ, Brand CA, Judson RT. Fine-needle aspiration may miss a third of all malignancy in palpable thyroid nodules: a comprehensive literature review. Ann Surg 2007;246:714–20.
[27] Amrikachi M, Ramzy I, Rubenfeld S, Wheeler TM. Accuracy of fine-needle aspiration of thyroid. Arch Pathol Lab Med 2001;125:484–8.
[28] Cap J, Ryska A, Rehorkova P, et al. Sensitivity and specificity of the fine needle aspiration biopsy of the thyroid: clinical point of view. Clin Endocrinol (Oxf) 1999;51:509–15.
[29] Morgan JL, Serpell JW, Cheng MS. Fine-needle aspiration cytology of thyroid nodules: how useful is it? ANZ J Surg 2003;73:480–3.
[30] Yang J, Schnadig V, Logrono R, Wasserman PG. Fine-needle aspiration of thyroid nodules: a study of 4703 patients with histologic and clinical correlations. Cancer 2007;111:306–15.