Clinical Endocrinology (2016) **85**, 624–631 doi: 10.1111/cen.13096

ORIGINAL ARTICLE

Utility and cost-effectiveness of molecular testing in thyroid nodules with indeterminate cytology

Emmanuel Labourier

BMDx, Austin, TX, USA

Summary

Context Molecular testing on biopsies from thyroid nodules with indeterminate cytology can improve patient management by preventing unnecessary surgeries on benign nodules.

Objective The aim of the study was to determine the health outcome benefits and cost-effectiveness of molecular testing in nodules with AUS/FLUS or FN/SFN cytology.

Design The initial diagnosis and treatment of a hypothetical cohort of adult U.S. patients with solitary thyroid nodules ≥1 cm was simulated by decision analytic modelling using Medicare cost estimates for three management strategies, standard of care without molecular testing (StC), gene expression classifier (GEC) and mutation and miRNA testing (MMT).

Results Gene expression classifier decreased the rate of unnecessary surgeries by 32% relative to StC, yielding incremental costs of $1008 per patient or $5070 per unnecessary surgery avoided. MMT decreased the surgery rate by 67%, yielding incremental savings of −$1384 per patient or −$3170 per unnecessary surgery avoided. Results remained robust in deterministic sensitivity analyses; MMT was dominant for every variable tested. Independent of cancer prevalence, MMT yielded 52% fewer unnecessary surgeries relative to GEC, 70% fewer two-stage thyroidectomies and correctly identified 70% more benign nodules. Test specificity had to be >68% for molecular testing to be cost-effective and decrease by >50% the rate of unnecessary surgeries performed on benign nodules.

Conclusions Molecular testing with high benign diagnostic yield can generate both positive health outcomes (less surgeries) and positive economic outputs (cost savings). These results are consistent with previously reported cost-utility data and provide valuable insights for informed decision-making by patients, physicians and payers.

(Received 2 March 2016; returned for revision 29 March 2016; finally revised 27 April 2016; accepted 5 May 2016)

Correspondence: Emmanuel Labourier, Austin, Texas, USA.
Tel.: 1 512 923 8518; E-mail: manulabourier@gmail.com

Introduction

Ultrasound imaging followed by cytopathology evaluation of fine-needle aspiration (FNA) biopsies is currently the gold standard for the diagnostic management of adult patients with thyroid nodules.[1] The procedure is extremely efficient with a residual risk of cancer <5% after benign diagnosis and >98% after malignant diagnosis.[2,3] This performance facilitates optimal preoperative patient management by avoiding unnecessary diagnostic lobectomy (DxL) on benign nodules and determining the extent of initial thyroidectomy on malignant nodules. However, cytology has a relatively low diagnostic yield; only a fraction of benign and malignant nodules are identified, leaving up to 35% of aspirates without a definitive diagnosis.[2,3] Several molecular testing strategies have been developed to complement cytology and improve the risk-based stratification of indeterminate nodules. The diagnostic utility and objective of these tests remain the same as for cytology, to decrease the number of unnecessary DxL on benign nodules and avoid completion thyroidectomies (CT) on malignant nodules by identifying the right candidates for initial total thyroidectomy (TT).[4] These clinical needs are particularly relevant for two indeterminate categories, atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS, Bethesda category III) and follicular neoplasm/suspicious for a follicular neoplasm (FN/SFN, Bethesda category IV), where highly variable residual risks of cancer have to be balanced against the risk of surgical complications and adverse events.[1,4]

Several mathematical modelling studies have been conducted to inform healthcare decision makers on the clinical and economic impact of this novel diagnostic modality. Yip *et al.*[5] showed that routine testing for somatic gene mutations associated with thyroid cancer was cost-effective. Savings were achieved mainly by reducing the number of two-stage thyroidectomies on malignant nodules. Li *et al.*[6] also reported very favourable results for the Afirma gene expression classifier (GEC; Veracyte Inc., South San Francisco, CA, USA), a test designed to identify benign nodules and to reduce the number of unnecessary DxL. The study was based on a very a favourable estimate of test specificity (75%) obtained with a classifier distinct from the commercial test. To date, the actual cost-effectiveness of the GEC for the initial diagnosis of indeterminate nodules remains unknown.[4] In an independent study, Lee *et al.*[7] showed that the

standard of care without molecular testing (StC) was less costly than GEC for a subset of patients with two consecutive AUS/FLUS cytology results. The model also predicted that a strategy combining routine GEC followed by selective mutation testing would be more cost-effective than GEC alone. This conclusion was based on the assumption that the performance of mutation testing in GEC-positive cases is the same as in the overall test population, which remains to be demonstrated. To date, no comprehensive clinical study has been performed to assess the performance of the GEC combined with mutation testing.

More recently, the Afirma BRAF test was developed to complement GEC testing.[8] This reflex test consists of a surrogate *BRAF* pV600E gene expression signature designed to identify a subset of malignant cases among nodules with 'suspicious' GEC results. In parallel, the ThyraMIR Thyroid miRNA Classifier (Interpace Diagnostics Group Inc., Parsippany, NY, USA) was developed and optimized to complement testing for somatic gene mutations.[9,10] A straightforward effectiveness analysis further suggested that this strategy could decrease the rate of unnecessary surgeries performed on benign nodules by approximately 70% relative to GEC alone.[9] The present health economics study was therefore undertaken to ascertain for the first time the cost-effectiveness of combination testing for the initial diagnosis of thyroid nodules with AUS/FLUS or FN/SFN cytology in the U.S. Two molecular strategies, GEC combined with Afirma BRAF (collectively referred to as GEC) and mutation testing combined with miRNA testing (MMT), were compared against the current StC. The goals were to predict the impact of molecular testing on healthcare cost and health outcomes and to better understand how underlying clinical parameters and test performance characteristics may contribute to utility and cost-effectiveness.

Materials and methods

Decision model

A decision model was constructed following current best practices[11] to estimate the annual cost of initial diagnosis and treatment of an hypothetical cohort of adult patients with solitary thyroid nodules ≥1 cm in size and indeterminate cytology of AUS/FLUS or FN/SFN. The model was built from a U.S. payer perspective (Medicare) and accounted for all diagnostic, follow-up and surgical costs for up to 1 year after initial diagnosis, including sequelae and adverse events associated with surgery. The cohort was based on a U.S. population of 320 million with a thyroid nodule incidence of 0·1% and a rate of AUS/FLUS and FN/SFN cytology of 20% with 25% thyroid cancer prevalence.[2,3,12–14] Only patients with nodules that needed additional diagnostic evaluation after the initial FNA were modelled. Low-risk patients and patients with high clinical suspicion of malignancy based on individual clinical presentations were assumed to proceed directly to follow-up or surgery.

The model compared three mutually exclusive strategies (Fig. 1) using a set of probability parameters obtained from systematic review of the literature (Table S1). In the StC arm, patients were managed without molecular testing according to the 2015 American Thyroid Association (ATA) guidelines.[1] A fraction of patients with AUS/FLUS cytology received a repeat ultrasound-guided FNA and proceeded to DxL, TT or clinical follow-up according to cytological results while all other patients received a DxL.[15–19] In the molecular arms, all patients were tested with GEC or MMT and patients with non-diagnostic results (insufficient sample amount or test failure) were assumed to proceed to DxL.[4,8–10,20,21] In the GEC arm, all patients with GEC-positive (suspicious) results also received the Afirma BRAF test at no additional cost. All Afirma BRAF-positive patients proceeded to TT and all other patients with suspicious results proceeded to DxL. In the MMT arm, patients were tested for mutations in the *BRAF*, *HRAS, KRAS* and *NRAS* genes and for the PAX8-PPARG, RET-PTC1 and RET-PTC3 fusion transcripts. All patients with negative mutation results received the ThyraMIR Thyroid miRNA Classifier at additional cost and patients with positive results (mutation or miRNA) proceeded to TT. For both molecular arms, patients with negative (benign) results were sent to clinical follow-up. For all strategies, histopathology review after DxL was assumed 100% accurate and all patients with a malignant diagnosis received a CT. Adverse events associated with lobectomy or thyroidectomy were captured in the model and included surgical complications, laryngeal nerve injury (temporary and permanent), hypoparathyroidism and hypothyroidism.[22–29]

Cost-effectiveness analysis

Cost estimates in U.S. dollar were obtained from the Centers for Medicare & Medicaid Services, the Agency of Healthcare Research and Quality and from Lubitz *et al.* (Table S2).[22,30,31] The consequences of the distinct strategies evaluated in the model were compared using the number of unnecessary surgeries avoided as quantifiable health outcome benefit (unit of effectiveness). Base case costs and effectiveness measures were combined into a single summary measure (incremental cost-effectiveness ratio or ICER) defined as the difference in total cost between two mutually exclusive strategies divided by the difference in the number of unnecessary surgeries performed relative to StC ($/unnecessary surgery avoided). Results for which incremental costs were negative (cost saving) and health consequences positive (less unnecessary surgeries performed) were qualitatively referred to as 'dominant'.

Sensitivity analysis

Univariate sensitivity analyses were performed by testing a reasonable range of estimates for every parameter and cost in the model (Tables S1 and S2). Minimum and maximum values were based on ranges obtained from the literature, the Centers for Medicare & Medicaid Services and the Agency of Healthcare Research and Quality. The effect of every variable was tested on ICER, total cost per patient, rates of surgery, two-stage thyroidectomy, unnecessary surgery, or unnecessary adverse events, number of molecular test needed to avoid one surgery and

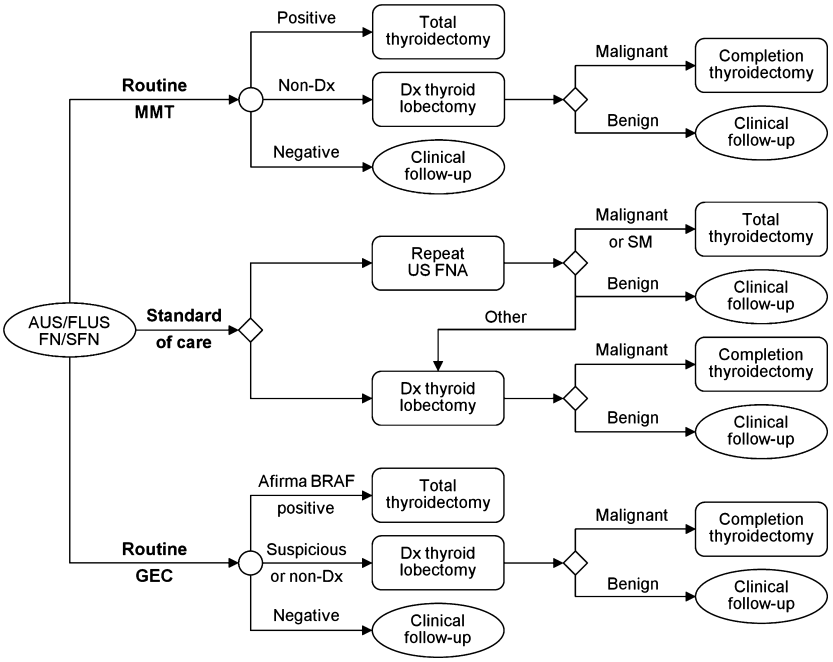

Fig. 1 Decision analytics model for the management of patients with AUS/FLUS or FN/SFN thyroid nodules using three mutually exclusive diagnostic strategies. Ovals represent health states, circles represent molecular tests, boxes represent procedures and diamonds represents decision nodes. MMT, mutation and miRNA testing; Dx, diagnostic; US FNA, ultrasound-guided fine-needle aspiration; SM, suspicious for malignancy; GEC, gene expression classifier.

number of molecular test needed to correctly identify one benign nodule. These metrics are independent of the number of patients included in the model and are insensitive to variables that affect cohort size such as thyroid nodule incidence or rate of nodules with AUS/FLUS or FN/SFN cytology.

Results

In the base case analysis, the cost of initial diagnosis and treatment of an AUS/FLUS or FN/SFN nodule was $11 149 per patient in the absence of molecular testing (StC, Table 1). Approximately half of the cost ($5784 or 52%) resulted from unnecessary surgical procedures and associated adverse events. Use of the GEC decreased the rate of surgery from 88% to 66% of the population tested, the rate of unnecessary surgery from 63% to 43% and the rate of unnecessary adverse events from 7·6% to 5·1% (Fig. 2, left panel). With the added cost of molecular testing, the total cost per patient increased by $1008 to $12 157, corresponding to an incremental cost per unnecessary surgery avoided of $5070 (ICER, Table 1). Use of the MMT further decreased the rate of unnecessary surgery to 20% of the population tested and decreased the rate of two-stage thyroidectomy (DxL followed by CT) from 23% to 6·7%. The total cost per patient decreased to $9765, mainly because of a major reduction in surgery costs relative to StC ($4904 per patient or 51%). Use of the MMT yielded cost savings of −$1384 per patient and −$3170 per unnecessary surgery avoided relative to StC. The MMT was also dominant relative to GEC with an incremental cost saving of −$5652 per unnecessary surgery avoided. Independent of cost, the GEC prevented one surgery for every 4·6 test performed and correctly identified one benign nodule for every 3·1 test performed (Fig. 2, right panel). The MMT achieved the same health outcome benefits for every 1·6 and 1·8 test performed, respectively.

Table 1. Summary of cost distribution and cost-effectiveness results

	StC	GEC	MMT
Cost [$/patient evaluated] (% of total)			
Total cost	$11 149	$12 157	$9765
Repeat FNA	$116 (1·0%)	–	–
Molecular testing	–	$3200 (26%)	$3980 (41%)
Clinical follow-up	$338 (3·0%)	$348 (2·9%)	$317 (3·2%)
Surgeries	$9570 (86%)	$7705 (63%)	$4666 (48%)
Adverse events	$1126 (10%)	$904 (7·4%)	$802 (8·2%)
Unnecessary cost*	$5784 (52%)	$3941 (32%)	$2169 (22%)
ICER [$/unnecessary surgery avoided]			
Molecular *vs* StC	–	$5070	−$3271
MMT *vs* GEC	–	–	−$5652

StC, standard of care; GEC, gene expression classifier; MMT, mutation and miRNA test; ICER, incremental cost-effectiveness ratio.
*Includes cost from unnecessary surgeries and adverse events.

In univariate sensitivity analyses, the cost of molecular testing and the cost of surgical procedures were major contributors to variation in the base case ICER results (Fig. 3). The GEC strategy was mainly sensitive to GEC cost, test specificity, DxL cost and two probability parameters associated with patient management (top panel). Across all variables, use of the GEC decreased the rate of unnecessary surgery by 24–39% relative to StC with an incremental cost of $987 to $13 120 per unnecessary surgery avoided. The MMT strategy was mainly sensitive to the costs of DxL and ThyraMIR (middle panel) and decreased the rate of surgery by 58–77% relative to StC with an incremental cost of −$9731 to +$132 per unnecessary surgery avoided. The MMT was dominant for all ranges of variable tested with an incremental cost saving of −$10 205 to −$2249 relative to GEC (bottom panel).

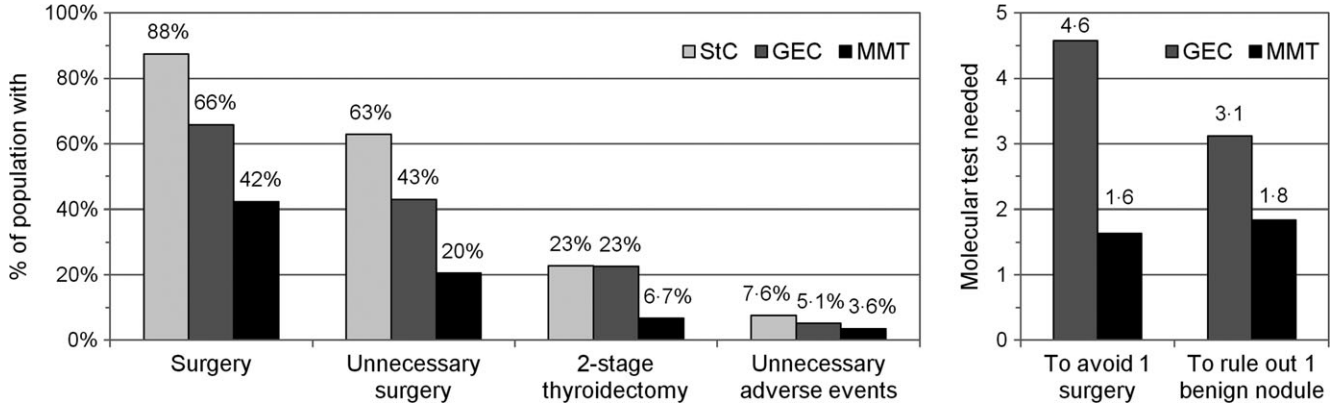

Fig. 2 Summary of health outcome results. The left graph shows the rates of patients with any surgery, with unnecessary surgery on benign nodules, with two-stage thyroidectomy on malignant nodules (diagnostic lobectomy followed by completion thyroidectomy) or with adverse events resulting from unnecessary surgical procedures. The right graph shows the number of molecular test that have to be performed to avoid one surgery or to correctly identify one benign nodule relative to the standard of care without molecular testing. StC, standard ofcare; GEC, gene expression classifier; MMT, mutation and miRNA testing.

Cost-utility in the model was also affected by the prevalence of thyroid cancer in the population tested (Fig. 3). Further analyses showed that for both molecular strategies, the total cost per patient and the rates of surgery and two-stage thyroidectomy decreased with cancer prevalence (Fig. 4). Conversely, the rate of unnecessary surgery performed on benign nodules and the associated costs increased. At any cancer prevalence, the MMT correctly ruled out 1·70-fold (70%) more benign nodules relative to GEC and yielded 2·10-fold (52%) fewer unnecessary surgeries. These metrics were directly proportional to the specificity of molecular testing (Fig. 5). Independent of other test characteristics and for both molecular strategies, incremental improvement in test specificity from 50% to 100% decreased the rate of unnecessary surgery from 43% to 11% of the population tested. The optimal test performance corresponded to an 83% decrease in unnecessary surgeries relative to StC with one benign nodule correctly identified for every 1·6 test performed.

Discussion

The objective of the study was to compare three mutually exclusive clinical management strategies for patients with AUS/FLUS or FN/SFN thyroid nodules. With an estimated population of 57 600 new cases per year, the base case model predicted that the total cost for current practice without molecular testing would be $642 million. With molecular testing, the total annual cost would increase to $700 million for GEC (+9·0%) and decrease to $562 million for MMT (−12%). The relative difference between the two molecular strategies ($138 million or $2392 per patient) would mainly stem from variations in the number of unnecessary primary surgeries performed on benign nodules, from 36 180 surgeries for StC to 24 732 for GEC and 11 804 for MMT. These results highlight specific model parameters that are critical for effective patient management and provide valuable information for the routine implementation of molecular testing in clinical practice.

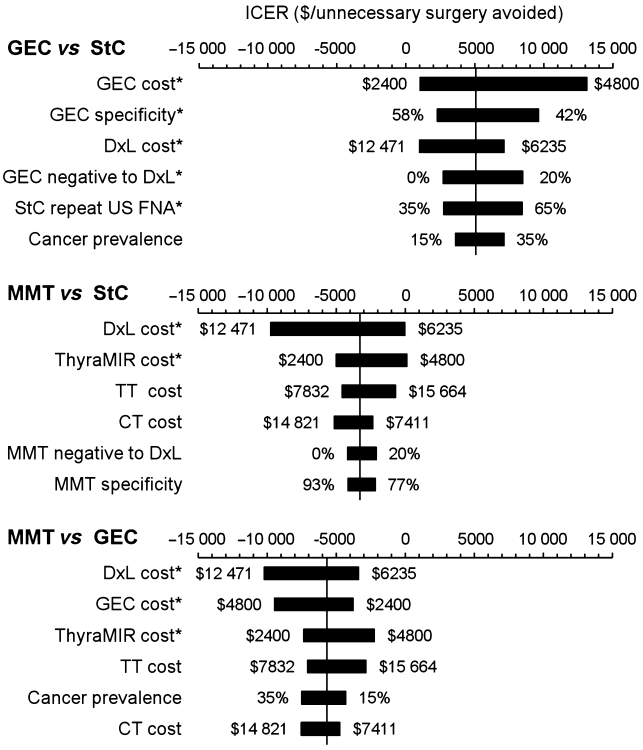

Fig. 3 Summary of univariate sensitivity analyses. The tornado plots show how variation of individual model parameters or costs affected the ICER results (horizontal bars). The range of variable tested is included on each side of the histogram. Negative ICER values indicate cost saving and variables yielding delta ICER >$5000 are labelled with an asterisk (*). The vertical lines represent the base case ICER results. ICER, incremental cost-effectiveness ratio; GEC, gene expression classifier; StC, standard of care; DxL, diagnostic lobectomy; US FNA, ultrasound-guided fine-needle aspiration; MMT, mutation and miRNA testing; TT, total thyroidectomy; CT, completion thyroidectomy.

In cost-utility studies, health effects are usually measured using quantifiable outcomes that are paramount to the program's purpose, such as number of lives saved, life-years gained

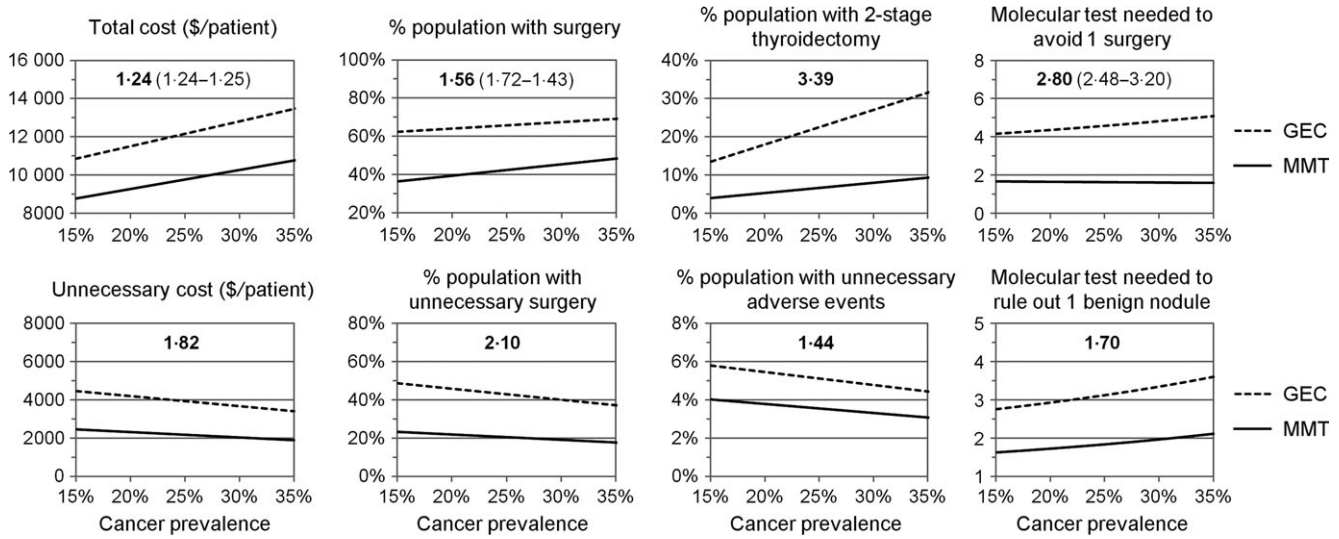

Fig. 4 Impact of thyroid cancer prevalence on cost and health outcome metrics. Each graph shows the results of sensitivity analyses for the indicated metrics when the prevalence of thyroid cancer in the cohort of nodules with AUS/FLUS or FN/SFN cytology varied from 15% to 35%. Numbers in bold indicate the fold difference between the two molecular testing strategies in the base case analysis (25% prevalence). Numbers between parentheses indicate the range obtained if the fold difference was not constant across the range of prevalence tested. GEC, gene expression classifier; MMT, mutation and miRNA testing.

or quality-adjusted life-years. For thyroid cancer, the 5-year survival rate after initial diagnosis is high, 99·9% for the majority of patients with localized disease, and the risk of surgical complications that can negatively affect quality of life is relatively low.[23–27,32] Two independent studies have previously demonstrated minimal differences in quality-adjusted life-years across multiple diagnostic modalities for indeterminate thyroid nodules.[6,7] In this present study, the number of surgery avoided was used as unit of effectiveness. It represents a simple and directly quantifiable measurement of one of the main objectives for diagnostic evaluation of thyroid nodules, the minimization of unnecessary surgeries. With an additional annual cost of $58 million relative to StC and 11 448 fewer unnecessary surgeries (−32%), the GEC strategy yielded an incremental cost of $5070 per unnecessary surgery avoided in the base case analysis. With annual savings of $80 million relative to StC and 24 376 fewer unnecessary surgeries (−67%), the MMT strategy yielded an incremental saving of −$3170 per unnecessary surgery avoided.

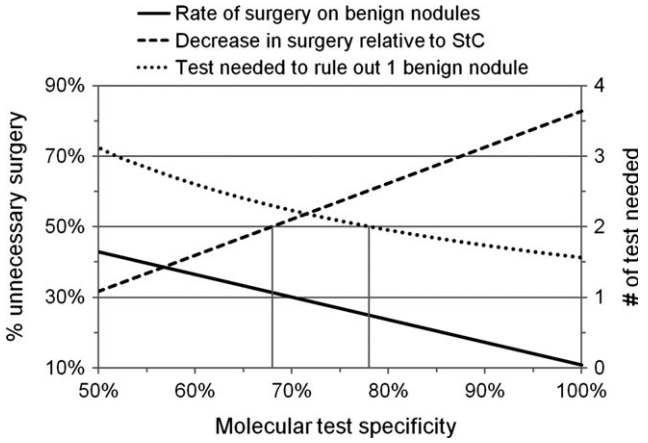

Fig. 5 Impact of test specificity on the effectiveness of molecular testing. The graph shows the results of sensitivity analyses for the indicated metrics when the specificity of either molecular testing strategy varied from 50% to 100%. The rate of unnecessary surgery on benign nodules, the decrease in surgery relative to StC and the number of test needed to rule out one benign nodule were directly proportional to (1-specificity), specificity and (1/specificity), respectively. The vertical grey lines indicate thresholds to obtain a decrease of at least 50% in the rate of unnecessary surgery relative to StC (68% specificity) or to correctly identify one benign nodule for every two tests performed or fewer (78% specificity). StC, standard of care.

Robustness analyses showed that cost-effectiveness in the model was highly sensitive to surgery costs, with lower costs favouring strategies that yielded more surgeries. For example, if the cost of DxL, CT and TT could be significantly reduced by 50%, the total cost of StC would decrease from $642 million to $366 million per year. Molecular strategies would still prevent the same number of unnecessary surgeries but would become more costly than the StC ($478 million for GEC, $428 million for MMT), with MMT remaining dominant over GEC (saving of $2062 per unnecessary surgery avoided). The cost of molecular testing also had a major impact on ICER results. At $2200 per test or less, the GEC would become dominant over StC (lower cost and 32% fewer surgeries, ICER <0) and at $810 the GEC would become cost neutral with MMT (same cost per patient). The other test variable that significantly affected cost-utility was the specificity of the GEC. To be cost-effective relative to the StC, test specificity would have to increase from 50% to 67%. The MMT would remain dominant with savings of $3287 per unnecessary surgery avoided relative to GEC.

The cost-effectiveness of gene expression testing in indeterminate thyroid nodules has previously been investigated. In an

industry-sponsored study, Li *et al.*[6] reported that if the GEC was universally employed in the U.S., 74% fewer surgeries would be performed on patients with benign thyroid nodules. This outstanding performance is actually consistent with the present study results as Li *et al.* used different model parameters that favoured molecular testing, in particular test specificity. When imputing those parameters in the present model, the GEC strategy would also yield 74% fewer unnecessary surgeries relative to StC, suggesting that the two decision models are very similar. When using a more accurate estimate of specificity (50% *vs* 75% in Li *et al.*)[6,20], the GEC would yield only 47% fewer surgeries. In addition, the model from Li *et al.* did not factor in potential test failures and assumed that up to 30% of patients with suspicious GEC results would not receive a surgery. The present model included a conservative estimate of test failure (5%) and followed current ATA patient management recommendations, that is, all patients with suspicious GEC results should be directed to surgery and a fraction of patients with benign results may receive a DxL.[4] With these more practical base case parameters, the GEC yielded 32% fewer unnecessary surgeries relative to StC with an incremental cost of $1008 per patient. The Afirma BRAF reflex test had minimal effect on the model (−$127 per patient) because it only changed the surgery costs from a two-stage thyroidectomy to an initial TT for the small fraction of patients with both GEC-suspicious and BRAF-positive results.

In an independent study comparing distinct molecular testing modalities, Lee *et al.* also previously demonstrated that the GEC was the most costly diagnostic strategy.[7] The reported incremental cost was relatively modest ($201 per patient) but is consistent with differences in study design and cohort. The present decision model simulated a cohort of patients with AUS/FLUS or FN/SFN cytology at initial diagnosis. Per current ATA guidelines[1], patients with a benign repeat FNA did not proceed to surgery, which minimized the rate of unnecessary surgery in the StC arm and optimized its cost-utility. In contrast, the microsimulation model of Lee *et al.* was built to assess the cost-effectiveness of molecular testing after two consecutive AUS/FLUS FNA cytology results. Therefore, all patients in the StC cohort were assumed to receive a DxL, which maximized the rate of unnecessary surgery in the StC arm and favoured the cost-utility of molecular testing. This effect is particularly relevant at low cancer prevalence where a 'surgery only' strategy for the StC would yield a very high rate of unnecessary surgery on benign nodules [equal to (1-prevalence)].

Thyroid cancer prevalence is a critical model parameter because variations across institutions and pathology practices can affect the clinical performance of molecular testing.[2–4] Both studies by Li *et al.* and Lee *et al.* reported that the GEC cost-effectiveness improved at lower cancer prevalence.[6,7] This is consistent with the present work showing that the total costs and the number of patients receiving a surgery decreased with prevalence for both molecular strategies. Sensitivity analyses further showed that prevalence positively or negatively affected every cost, health outcome or effectiveness metrics. Although the GEC cost-effectiveness improved at low prevalence, the rate of unnecessary surgery and the associated costs increased, and were approximately 2-fold higher than for the MMT. Remarkably, the relative differences between the two molecular strategies remained constant across the range of prevalence tested for most metrics.

Broad uncertainty analyses demonstrated that the single test characteristic driving major cost-utility differences across models was test specificity (50% for GEC, 85% for MMT; $P < 0{\cdot}001$ for Fisher exact test). For either molecular strategy, the rate of unnecessary surgery performed on benign nodules was directly proportional to test specificity, with higher specificity resulting in improved health outcome benefits. This observation is in agreement with previous modelling work that demonstrated the mathematical relationship between incremental test performance and test specificity, independent of thyroid cancer prevalence.[9] In this present study, similar equations could be derived to arithmetically verify the accuracy of the model and predict health effects according to test specificity (see Fig. 5 and legend). From these analyses, one can further estimate that to decrease by ≥50% the rate of unnecessary surgery relative to StC, the specificity of any molecular test has to be ≥68%. Similar minimum test performance requirements may be defined for other health outcome or effectiveness metrics, for example, specificity has to be ≥78% for a molecular test to correctly identify one benign nodule every two tests performed or fewer.

As for any model-based evaluation, there remain some limitations to this present study. The probability and cost inputs in the decision model are dependent on the availability of relevant data and on the accuracy and own limitations of the source studies. No parameter was excluded from analysis based on low level or lack of available information. Instead, a conservative approach was adopted by testing a broad range of values reflecting the uncertainty in parameter estimation.[11] With rapidly evolving technologies and clinical management strategies, specific parameter and cost estimates may have to be updated or changed to keep the model current and relevant. Several assumptions in the model also limited the generalization of the results. First, the model was built to assess the cost of diagnosis and treatment of indeterminate thyroid nodules for up to 1 year after initial cytopathology evaluation. It captured only the direct medical costs and did not include the cost of thyroid cancer surveillance or recurrence. Clinical follow-up for nodules with benign repeat cytology or with negative molecular results was restricted to two visits with ultrasound imaging over a 1-year period. Long-term conservative management of nodules with low risk of thyroid cancer may entail additional costs such as repeat FNA cytology and/or repeat molecular testing. Secondly, the simulated cohort consisted exclusively of adult patients with solitary thyroid nodules ≥1 cm in size. The performance and cost-effectiveness of molecular testing in the pediatric population or for multinodular disease remain to be determined. Lastly, the analysis was performed from a U.S. healthcare system perspective using Medicare cost estimates. Results may not be applicable to other countries and to all clinical practices within the U.S. because of potential local variations in patient management algorithms and/or medical costs.

Conclusions

The present health economics model confirmed that molecular evaluation of thyroid nodules with indeterminate cytology generates positive health outcomes by reducing the rate of unnecessary surgery on benign nodules. Depending on its diagnostic yield, molecular testing can lead to either an increase or a decrease in total medical costs in the U.S. The model further predicts that test specificity is a critical parameter for any molecular test to be clinically and economically effective. This requirement is particularly important at low cancer prevalence where most patients have benign nodules and where lower specificity proportionally drives up the rate of unnecessary surgeries and associated adverse events and costs. These results are in agreement with previous studies and lay a framework for the definition of optimal test performance requirements and the development of improved recommendations to inform clinical decision-making.

Acknowledgements

The author thanks Deanna Hertz, Senior Consultant health market access at GfK, for her analysis and input on the probability parameters and costs associated with thyroid surgery and surgical adverse events.

Conflicts of interest

EL was an employee of Asuragen Inc. from March 2006 to March 2014 and a consultant for PDI Inc. from September 2014 to October 2015. The author does not have any direct or indirect employment for and does not hold any equity interest in Interpace Diagnostics Group Inc. or Veracyte Inc.

References

1 Haugen, B.R., Alexander, E.K., Bible, K.C. *et al.* (2016) 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*, **26**, 1–133.

2 Bongiovanni, M., Spitale, A., Faquin, W.C. *et al.* (2012) The Bethesda system for reporting thyroid cytopathology: a meta-analysis. *Acta Cytologica*, **56**, 333–339.

3 Wang, C.C., Friedman, L., Kennedy, G.C. *et al.* (2011) A large multicenter correlation study of thyroid nodule cytopathology and histopathology. *Thyroid*, **21**, 243–251.

4 Ferris, R.L., Baloch, Z., Bernet, V. *et al.* (2015) American thyroid association statement on surgical application of molecular profiling for thyroid nodules: current impact on perioperative decision making. *Thyroid*, **25**, 760–768.

5 Yip, L., Farris, C., Kabaker, A.S. *et al.* (2012) Cost impact of molecular testing for indeterminate thyroid nodule fine-needle aspiration biopsies. *The Journal of Clinical Endocrinology and Metabolism*, **97**, 1905–1912.

6 Li, H., Robinson, K.A., Anton, B. *et al.* (2011) Cost-effectiveness of a novel molecular test for cytologically indeterminate thyroid nodules. *The Journal of Clinical Endocrinology and Metabolism*, **96**, E1719–E1726.

7 Lee, L., How, J., Tabah, R.J. *et al.* (2014) Cost-effectiveness of molecular testing for thyroid nodules with atypia of undetermined significance cytology. *The Journal of Clinical Endocrinology and Metabolism*, **99**, 2674–2682.

8 Diggans, J., Kim, S.Y., Hu, Z. *et al.* (2015) Machine learning from concept to clinic: reliable detection of BRAF V600E DNA mutations in thyroid nodules using high-dimensional RNA expression data. *Pacific Symposium on Biocomputing*, **20**, 371–382.

9 Labourier, E., Shifrin, A., Busseniers, A.E. *et al.* (2015) Molecular testing for miRNA, mRNA, and DNA on fine-needle aspiration improves the preoperative diagnosis of thyroid nodules with indeterminate cytology. *The Journal of Clinical Endocrinology and Metabolism*, **100**, 2743–2750.

10 Wylie, D., Beaudenon, S., Haynes, B.C. *et al.* (2016) Molecular classification of thyroid lesions by combined testing for miRNA gene expression and somatic gene alterations. *The Journal of Pathology: Clinical Research*, **100**, 2743–2750.

11 Caro, J.J., Briggs, A.H., Siebert, U. *et al.* (2012) Modeling good research practices–overview: a report of the ISPOR-SMDM modeling good research practices task force-1. *Medical Decision Making*, **32**, 667–677.

12 Gharib, H. & Papini, E. (2007) Thyroid nodules: clinical importance, assessment, and treatment. *Endocrinology and Metabolism Clinics of North America*, **36**, 707–735.

13 Popoveniuc, G. & Jonklaas, J. (2012) Thyroid nodules. *The Medical Clinics of North America*, **96**, 329–349.

14 Yang, J., Schnadig, V., Logrono, R. *et al.* (2007) Fine-needle aspiration of thyroid nodules: a study of 4703 patients with histologic and clinical correlations. *Cancer*, **111**, 306–315.

15 Yassa, L., Cibas, E.S., Benson, C.B. *et al.* (2007) Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. *Cancer*, **111**, 508–516.

16 VanderLaan, P.A., Marqusee, E. & Krane, J.F. (2011) Clinical outcome for atypia of undetermined significance in thyroid fine-needle aspirations: should repeated fna be the preferred initial approach? *American Journal of Clinical Pathology*, **135**, 770–775.

17 Nayar, R. & Ivanovic, M. (2009) The indeterminate thyroid fine-needle aspiration: experience from an academic center using terminology similar to that proposed in the 2007 National Cancer Institute Thyroid Fine Needle Aspiration State of the Science Conference. *Cancer*, **117**, 195–202.

18 Nagarkatti, S.S., Faquin, W.C., Lubitz, C.C. *et al.* (2013) Management of thyroid nodules with atypical cytology on fine-needle aspiration biopsy. *Annals of Surgical Oncology*, **20**, 60–65.

19 Ustun, H., Astarci, H.M., Altunkaya, C. *et al.* (2012) Fine-needle aspiration of follicular patterned lesions of the thyroid: diagnosis, management, and follow-up according to thyroid Bethesda system. *Acta Cytologica*, **56**, 361–369.

20 Alexander, E.K., Kennedy, G.C., Baloch, Z.W. *et al.* (2012) Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *The New England Journal of Medicine*, **367**, 705–715.

21 Kloos, R.T., Reynolds, J.D., Walsh, P.S. *et al.* (2013) Does addition of BRAF V600E mutation testing modify sensitivity or specificity of the Afirma Gene Expression Classifier in cytologically indeterminate thyroid nodules? *The Journal of Clinical Endocrinology and Metabolism*, **98**, E761–E768.

22 Lubitz, C.C., Kong, C.Y., McMahon, P.M. *et al.* (2014) Annual financial impact of well-differentiated thyroid cancer care in the United States. *Cancer*, **120**, 1345–1352.
23 Qin, C.D., Saha, S., Meacham, R. *et al.* (2014) Surgical risk after unilateral lobectomy versus total thyroidectomy: a review of 47,434 patients. *Surgery: Current Research*, **4**, 208.
24 Shrime, M.G., Goldstein, D.P., Seaberg, R.M. *et al.* (2007) Cost-effective management of low-risk papillary thyroid carcinoma. *Archives of Otolaryngology-Head & Neck Surgery*, **133**, 1245–1253.
25 Landerholm, K., Wasner, A.M. & Jarhult, J. (2014) Incidence and risk factors for injuries to the recurrent laryngeal nerve during neck surgery in the moderate-volume setting. *Langenbeck's Archives of Surgery*, **399**, 509–515.
26 Giordano, D., Valcavi, R., Thompson, G.B. *et al.* (2012) Complications of central neck dissection in patients with papillary thyroid carcinoma: results of a study on 1087 patients and review of the literature. *Thyroid*, **22**, 911–917.
27 Shan, C.X., Zhang, W., Jiang, D.Z. *et al.* (2012) Routine central neck dissection in differentiated thyroid carcinoma: a systematic review and meta-analysis. *The Laryngoscope*, **122**, 797–804.
28 Vaiman, M., Nagibin, A., Hagag, P. *et al.* (2008) Hypothyroidism following partial thyroidectomy. *Otolaryngology-Head and Neck Surgery*, **138**, 98–100.
29 Stoll, S.J., Pitt, S.C., Liu, J. *et al.* (2009) Thyroid hormone replacement after thyroid lobectomy. *Surgery*, **146**, 554–558.
30 Centers for Medicare & Medicaid Services. Medicare reimbursement schedule. Available from: www.cmc.gov (accessed 4 Aug 2015).
31 Agency for Healtcare and Research Quality. Overview of the nationwide inpatient sample from the healthcare cost and utilization project. Available from: www.hcup-us.ahrq.gov. (accessed 4 Aug 2015).
32 National Cancer Institute. Surveillance, Epidemiology and End Results program. Available from: www.seer.cancer.gov. (accessed 4 Aug 2015).

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site.