RESEARCH

Health-related quality of life following FDG-PET/CT for cytological indeterminate thyroid nodules

Elizabeth J de Koster[1], Olga Husson[2,3,4], Eveline W C M van Dam[5], G Sophie Mijnhout[6], Romana T Netea-Maier[7], Wim J G Oyen[1,8,9], Marieke Snel[10], Lioe-Fee de Geus-Oei[1,11,12], Dennis Vriens[11] for the EfFECTS trial study group

[1]Department of Radiology and Nuclear Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands
[2]Department of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands
[3]Division of Medical Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands
[4]Division of Clinical Studies, Institute of Cancer Research, London, UK
[5]Department of Internal Medicine, Division of Endocrinology, location VU University Medical Center, Amsterdam University Medical Centers, Amsterdam, The Netherlands
[6]Department of Internal Medicine, Isala Hospital, Zwolle, The Netherlands
[7]Department of Internal Medicine, Division of Endocrinology, Radboud University Medical Centre, Nijmegen, The Netherlands
[8]Department of Radiology and Nuclear Medicine, Rijnstate Hospital, Arnhem, The Netherlands
[9]Department of Biomedical Sciences and Humanitas Clinical and Research Centre, Department of Nuclear Medicine, Humanitas University, Milan, Italy
[10]Department of Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands
[11]Department of Radiology, Section of Nuclear Medicine, Leiden University Medical Center, Leiden, The Netherlands
[12]Biomedical Photonic Imaging Group, University of Twente, Enschede, The Netherlands

Correspondence should be addressed to O Husson: olga.husson@icr.ac.uk

Abstract

Objective: This study assessed the health-related quality of life (HRQoL) in patients undergoing 2-[18F]fluoro-2-deoxy-D-glucose (FDG)-PET/CT for an indeterminate (Bethesda III/IV) thyroid nodule. FDG-PET/CT accurately rules out malignancy and prevents 40% of futile diagnostic surgeries in these nodules.
Design: Secondary analyses of HRQoL data from a randomised controlled multicentre trial (NCT02208544) in 126 patients from 15 hospitals in the Netherlands were done.
Methods: Longitudinal HRQoL assessment was performed using the EuroQol 5-dimension 5-level (EQ-5D-5L), the RAND 36-item Health Survey v2.0 (RAND-36), and the Thyroid Patient-Reported Outcome (ThyPRO) questionnaire on baseline, 3, 6, and 12 months, relative to the date of the FDG-PET/CT scan.
Results: Patients who were randomised to active surveillance following an FDG-negative nodule instead of diagnostic surgery reported stable HRQoL scores throughout the year. Univariate analysis indicated better HRQoL for patients undergoing surveillance than surgical patients with benign histopathology on multiple physical and psychosocial domains. Univariate within-group analysis suggested both temporary and continued HRQoL deteriorations in patients with benign histopathology over time. Multivariate within-group analysis demonstrated no significant longitudinal HRQoL changes in patients undergoing active surveillance. In contrast, in patients with benign histopathology, worse HRQoL was observed with regard to ThyPRO cognitive impairment ($P = 0.01$) and cosmetic complaints ($P = 0.02$), whereas goitre symptoms ($P < 0.001$) and anxiety ($P = 0.04$) improved over time. In patients with malignant histopathology, anxiety also decreased ($P = 0.05$).
Conclusions: The reassurance of a negative FDG-PET/CT resulted in sustained HRQoL throughout the first year of active surveillance. Diagnostic surgery for a nodule with benign

Key Words

- health-related quality of life
- thyroid nodule
- indeterminate cytology
- Bethesda
- FDG-PET/CT
- diagnostic thyroid surgery

https://ec.bioscientifica.com
https://doi.org/10.1530/EC-22-0014

histopathology resulted in more cognitive impairment and physical problems including cosmetic complaints, but improved goitre symptoms and anxiety. Anxiety was also reduced in patients with malignant histopathology.

Endocrine Connections (2022) **11**, **e220014**

Introduction

Palpable thyroid nodules have a prevalence of approximately 5%. Due to the increased use of imaging techniques for indications unrelated to the thyroid, however, occult thyroid nodules will be detected in up to approximately 65% of the general population (1, 2). Ultrasound and fine-needle aspiration cytology (FNAC) are the first steps in the diagnostic workup (3). Approximately 25% of cytology results are indeterminate, including atypia of undetermined significance or follicular lesion of undetermined significance (Bethesda III) and (suspicious for a) follicular neoplasm or Hürthle cell neoplasm (Bethesda IV) (1, 4). Following an established indeterminate cytological diagnosis (including repeat FNAC in Bethesda III nodules) and consideration of clinical features, ultrasound characteristics, and the patient's treatment preferences, diagnostic thyroid surgery is often recommended (3).

To improve the workup of indeterminate thyroid nodules and potentially avoid diagnostic surgery for approximately 75% benign nodules, a range of additional diagnostics is currently available, including various ultrasound classification systems, molecular testing, and 2-[^{18}F]fluoro-2-deoxy-D-glucose (FDG)-PET/CT (4, 5, 6). Our recent randomised controlled multicentre trial confirmed that a negative FDG-PET/CT scan accurately rules out malignancy with a 94% sensitivity, avoiding 40% of futile diagnostic surgeries for benign nodules (6, 7). Moreover, FDG-PET/CT proved cost-effective in a Dutch setting, with lower societal lifelong costs while maintaining the long-term health-related quality of life (HRQoL) in patients undergoing FDG-PET/CT-driven management as compared to routine diagnostic surgery or molecular testing (8, 9).

HRQoL has been studied extensively in both benign and malignant thyroid diseases and is often impaired as compared to the general population (10, 11, 12). Thyroid surgery negatively affects short-term HRQoL, recovering as the time since surgery passes (10, 13, 14, 15, 16, 17). The consequences of surgery, including long-term postoperative thyroid hormone substitution therapy, potential voice changes, scar cosmesis, and surgical complications such as permanent hypoparathyroidism and vocal cord paralysis, also impair HRQoL (10, 18, 19, 20, 21). In thyroid carcinoma, long-term HRQoL may be worse than in other types of cancer with worse prognosis, partly due to periodic thyroid hormone withdrawal during the follow-up (10, 11, 16). In papillary thyroid microcarcinoma, patients managed by active surveillance had fewer HRQoL problems than patients who underwent hemithyroidectomy (22).

There are few studies, however, on HRQoL in patients managed with or without surgery and additional diagnostics for indeterminate thyroid nodules. Two recent studies showed that molecular testing in Bethesda III/IV nodules resulted in sustained HRQoL for patients managed without surgery following a benign molecular test result, with HRQoL levels similar to patients with a benign cytology result. Patients with a suspicious as compared to a benign molecular test result initially showed poorer HRQoL, which improved after diagnostic surgery (23, 24). Our cost-utility study reported that the general 1-year and lifelong HRQoL, assessed using the EuroQol 5-dimension 5-level (EQ-5D-5L) questionnaire, was similar in patients undergoing FDG-PET/CT-driven management or diagnostic surgery and that cost-effectiveness was most sensitive to changes in HRQoL following hemithyroidectomy for a histopathologically benign nodule (8). Therefore, we hypothesised that temporary and/or domain-specific HRQoL differences may exist between patients undergoing active surveillance following a negative FDG-PET/CT scan and patients undergoing diagnostic surgery, specifically for a benign nodule.

In the current study, we performed a comprehensive evaluation of the HRQoL in patients undergoing FDG-PET/CT for a Bethesda III/IV thyroid nodule.

Material and methods

Study population

All patients who participated in the previously reported prospective, randomised controlled multicentre trial on the efficacy of FDG-PET/CT in cytologically indeterminate (Bethesda III/IV) thyroid nodules (EfFECTS trial, registered as NCT02208544) were assessed for eligibility. This trial was performed in eight academic and seven community hospitals in the Netherlands between July 2015 and

https://ec.bioscientifica.com
https://doi.org/10.1530/EC-22-0014

December 2019 and included adult patients with a Bethesda III or IV thyroid nodule, without increased risk of thyroid malignancy based on their presentation or medical history, who were literate in Dutch or English. Based on cytology, clinical characteristics, and ultrasound features, diagnostic thyroid surgery was scheduled for all patients, in accordance with current international guidelines (3). Patients who could not be randomised due to a strong preference for either surgery (e.g. for goitre symptoms or cosmetic complaints) or active surveillance were not included in the trial.

The current study comprised secondary analyses of HRQoL data acquired during this trial. Trial participants were only eligible for inclusion in the *current* study if they had completed all trial procedures *per protocol*; patients with nonadherence to the trial protocol were excluded. The EfFECTS trial, including the current study, was approved by the Medical Research Ethics Committee on Research Involving Human Subjects region Arnhem-Nijmegen, Nijmegen, the Netherlands. Written informed consent was obtained from each of the participants prior to any study activity.

Study design and procedures

The full methodology of the EfFECTS trial was reported prior, including comprehensive descriptions regarding patient eligibility, selection, randomisation, blinding, FDG-PET/CT procedures, and sample size calculation (6). In summary, all patients underwent FDG-PET/CT and were randomly allocated to an FDG-PET/CT-driven group or diagnostic surgery group. Dependent on their allocation and the FDG-PET/CT result, patients were advised to proceed with the scheduled diagnostic thyroid surgery (i.e. patients with an FDG-positive nodule in the FDG-PET/CT-driven group and all patients allocated to the diagnostic surgery group) or were advised to refrain from surgery and undergo active surveillance (i.e. patients allocated to the FDG-PET/CT-driven group with an FDG-negative nodule). Treatment allocation and the result of the FDG-PET/CT were undisclosed in case diagnostic thyroid surgery was advised. Patients who were advised active surveillance were counselled that their negative FDG-PET/CT result had a remaining risk of malignancy of less than 5%, based on performance characteristics of FDG-PET/CT that were established in the literature (7) and confirmed by the trial (6). Active surveillance consisted of at least one ultrasound procedure and a follow-up visit with the patient's local endocrinologist 1 year after the FDG-PET/CT; additional intermediate follow-up visits were allowed at the discretion of the local physicians. In both study groups, the patient and his/her local physician were free to deviate from the study treatment advice at any time. Postoperative management adhered to the Dutch national guidelines (25). The current study adhered to the *local* histopathological diagnosis as a reference standard, as this diagnosis determined the individual patients' postoperative course of treatment and likely best reflects the patient's illness perception and perceived HRQoL. For the primary trial endpoints, a central histopathology review was performed, causing minor discrepancies between the current and previous papers (6). Patients with a non-invasive follicular thyroid neoplasm with papillary-like nuclear features or follicular tumour of uncertain malignant potential were categorised as benign, as patients were counselled and postoperatively treated as such.

For the current study, patients were categorised into three groups: patients who underwent diagnostic surgery and had (i) benign or (ii) malignant histopathology, and (iii) patients who had an FDG-negative nodule and underwent active surveillance.

HRQoL assessment

HRQoL was assessed through self-administered questionnaires at baseline and 3, 6, and 12 months, counted from the date of the FDG-PET/CT scan. Upon study inclusion, patients indicated whether they preferred to receive electronic questionnaires with email invitations or printed questionnaires that were sent to their home address with a stamped return envelope included. At each time point, in case of an incomplete or missing HRQoL assessment, patients were reminded up to four times by email or phone call, starting approximately 2 weeks and ending 2 months after the mailing date.

Three questionnaires were used: the Thyroid Patient-Reported Outcome (ThyPRO), the EQ-5D-5L, and the RAND 36-item Health Survey v2.0 (RAND-36) questionnaire (26, 27, 28, 29). The 98-item ThyPRO is a validated, thyroid-specific HRQoL questionnaire. It uses a five-point scale to address 14 domains over a 4-week recall period: 4 thyroid disease-related symptom domains (symptoms of neck-related goitre, hyper- and hypothyroidism, and eye symptoms), 5 domains of physical and mental functioning and well-being, and 5 domains examining the impact of the thyroid disorder on essential HRQoL aspects. Within each domain, the five-point-scale item scores are averaged and linearly transformed to a 0–100 scale, with *higher scores indicating worse health states* (29). A composite score is additionally calculated, summarizing 22 items (30). We used the predefined minimally important change (MIC)

values to assess observed HRQoL differences within and between groups (31). The MIC represents 'the smallest difference in a score from a HRQoL questionnaire that is perceived by patients as important, either beneficial or harmful', and that is therefore likely meaningful to patients and physicians (32). Descriptions of the EQ-5D-5L and RAND-36 questionnaires are presented in the Supplementary Data (see section on supplementary materials given at the end of this article).

Statistical analysis

Baseline characteristics were compared between groups using the Pearson chi-square test for categorical data and the one-way ANOVA, Kruskal–Wallis test, or Mann–Whitney *U* test for (log-)normally or non-normally distributed continuous data, where appropriate. We used multiple imputation to account for possible selectively missing data in the questionnaires, using age at baseline, sex, allocation (i.e. study arm in EfFECTS trial), EQ-5D-5L utilities, and time-dependent variables for thyroid surgery, and the benign or malignant local histopathological diagnosis as predictor variables. Univariate comparison of HRQoL scores between groups was performed using independent samples *t*-test with unequal variances; within-group differences over time were assessed using paired *t*-tests. Next, we constructed a mixed linear model for each of the EQ-5D-5L, RAND-36, and ThyPRO scales to analyse repeated measures and evaluate longitudinal changes in HRQoL within and between groups. First, we constructed a model to assess the changes over time within each of the three groups separately. The final model was a random intercept model that used longitudinal data from all four HRQoL assessments and adjusted for fixed effects of the time of the questionnaire, and for a random component of within-subject correlations between the repeated measurements. Secondly, we constructed a random intercept model to assess between-group differences over time, which additionally included fixed effects for the group and the interaction between time and group. For all analyses, a P value <0.05 was considered statistically significant. All statistical analyses were performed using SPSS Statistics (version 27, IBM Corp.).

Results

Of the 132 patients who participated in the EfFECTS trial, 126 were included in the current study (Fig. 1). Patients with thyroid malignancy were less frequently female as compared to benign histopathology and patients under surveillance ($P<0.001$). Other baseline characteristics were evenly distributed across the three groups (Table 1). One hundred fifteen (91.3%) patients fully completed the baseline assessment, and 108 (85.7%), 104 (82.5%), and 106 (84.1%) patients completed the 3-month, 6-month, and 12-month assessments, respectively. The baseline HRQoL assessment was completed 3 days (median, IQR 0–17) after the FDG-PET/CT was performed. Patients under surveillance completed the 12-month assessment 38 days (median, IQR 10–69) after their protocolled 1-year follow-up ultrasound.

ThyPRO

Longitudinal, *within*-group comparison (Fig. 2, Supplementary Table 3) indicated that patients in all three groups experienced an improvement of the negative influence of their thyroid disorder on their QoL over time. Besides that, patients who underwent active surveillance for an FDG-negative nodule generally reported stable HRQoL scores throughout the year; no statistically significant or clinically relevant longitudinal changes were observed in this group.

In all patients undergoing diagnostic surgery, goitre symptoms improved from 3 to 12 months. In patients with benign histopathology, goitre symptoms deteriorated between baseline and 3 months but had improved at 12 months as compared to baseline.

Furthermore, patients with benign histopathology reported more tiredness, social impairment, daily life impairment, and sex life impairment at 3 and/or 6 months as compared to baseline. At 12 months, these differences were not statistically significant relative to baseline. Daily life impairment significantly improved from 3 to 12 months. In contrast, more cognitive impairment and cosmetic complaints were observed at 3, 6, and 12 months relative to baseline. The ThyPRO composite QoL score showed worse HRQoL at 3 months as compared to baseline. Finally, an unexpected increase in hyperthyroid symptoms was observed at 12 months as compared to baseline in patients with benign histopathology. Subgroup analysis showed that this difference could not be explained by the 10 of 74 (14%) patients who developed hypothyroidism and started levothyroxine substitution following their diagnostic surgery: these patients reported a similar increase in hyperthyroid symptoms as compared to the 64 (86%) patients without postoperative hypothyroidism (mean difference between baseline and 12 months +3.9 and +4.7, respectively, $P=0.86$).

https://ec.bioscientifica.com
https://doi.org/10.1530/EC-22-0014

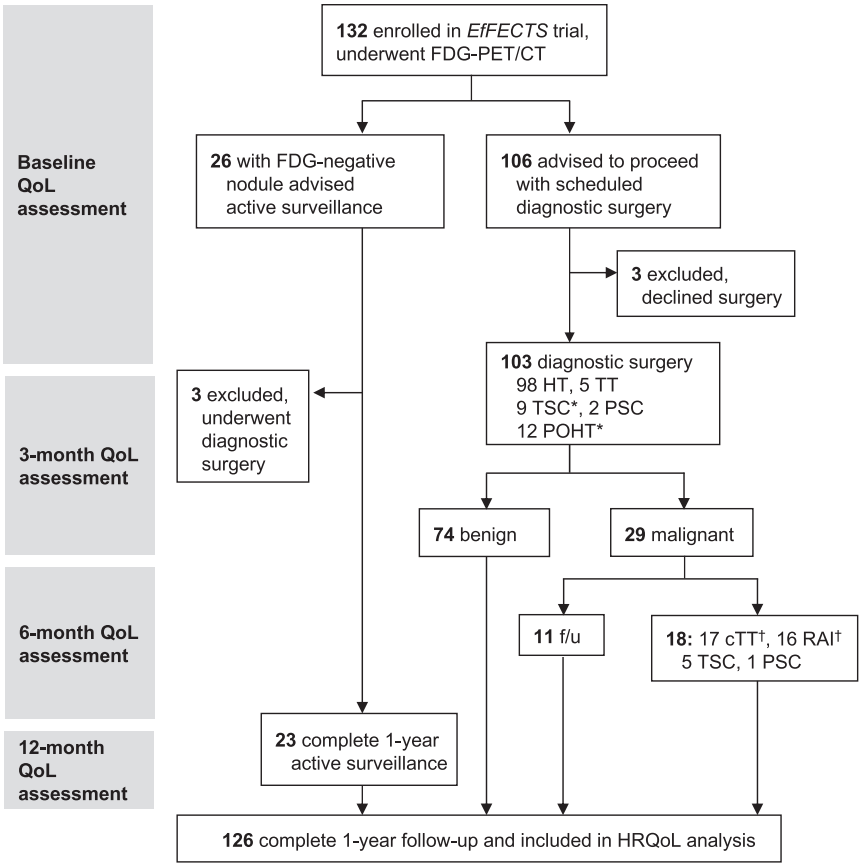

Figure 1
Study flowchart of the first year, visualizing the study procedures, observed treatment, treatment outcomes, and excluded patients. *One patient had a TSC and POHT. †One patient underwent RAI after initial TT for malignancy; two patients underwent cTT for malignancy but no RAI. cTT, completing total thyroidectomy; F/u, follow-up; HT, hemithyroidectomy, including isthmus resection ($n = 3$) and hemithyroidectomy plus nodulectomy ($n = 2$); POHT, postoperative levothyroxine-dependent hypothyroidism after partial thyroidectomy procedure; PSC, permanent surgical complication, including recurrent nerve paralysis ($n = 3$); RAI, adjuvant radioiodine therapy; TT, total thyroidectomy; TSC, transient surgical complication, including haematoma with re-exploration surgery, wound infection, seroma, and transient hypoparathyroidism.

Patients with malignant histopathology experienced more cognitive, social, and daily life impairment at 6 months as compared to baseline. They reported less anxiety at 12 months as compared to baseline.

Univariate *between*-group comparison (Table 2) suggested better HRQoL in patients under active surveillance than patients with benign histopathology, with statistically significant and/or clinically important differences on multiple domains and measurement points. Patients under active surveillance reported less goitre symptoms, tiredness, anxiety, emotional susceptibility, social impairment, cosmetic complaints, and negative influence on QoL at 3 and/or 6 months. At 3, 6, and 12 months, less cognitive impairment, depressivity, and daily life impairment were reported. The ThyPRO composite QoL score also showed statistically significant and/or clinically important differences at 3, 6, and 12 months between these groups.

Between histopathologically benign and malignant groups, only a difference in anxiety at 12 months was observed. No other statistically significant and/or clinically relevant differences were observed between these two groups.

EQ-5D-5L and RAND-36

Results concerning the univariate and multivariate analysis of the EQ-5D-5L and RAND-36 are presented in the Supplementary Data.

Multivariate assessment of changes in HRQoL over time

Multivariate *within*-group analysis showed that hyperthyroid symptoms (B = 1.6, $P = 0.01$), cognitive impairment (B = 2.1, $P = 0.01$), and cosmetic complaints (B = 1.9, $P = 0.02$) significantly worsened over time for surgical patients with benign histopathology (Supplementary Table 4). Goitre symptoms (B = –2.7, $P < 0.001$) and anxiety (B = –1.8, $P = 0.04$) improved between baseline and 12 months. Patients with malignancy reported a statistically significant improvement in anxiety (B = –2.5, $P = 0.05$). There were no significant longitudinal changes in patients undergoing active surveillance. Lastly, on multivariate analysis no statistically significant differences were observed in the HRQoL patterns over time between the three groups.

https://ec.bioscientifica.com
https://doi.org/10.1530/EC-22-0014

Table 1 Patient baseline demographics, clinical characteristics, and HRQoL timing.

	Diagnostic surgery				Active surveillance		*P*
	Malignant		Benign				
	n = 29	%	*n* = 74	%	*n* = 23	%	
Female	16	55%	65	88%	21	91%	<0.001[a]
Age (years) (mean ± S.D.)	56.3 ± 14.5		53.5 ± 12.4		54.7 ± 16.0		0.61[b]
General medical history	26	90%	63	85%	19	83%	0.75[a]
Cardiovascular disease (incl. stroke)	11	38%	21	28%	6	26%	0.57[a]
Malignancy	3	10%	7	9%	2	9%	0.98[a]
Haematological disease (incl. malignancy)	3	10%	6	8%	1	4%	0.73[a]
Neurological disease (excl. stroke)	6	21%	15	20%	6	26%	0.83[a]
Otolaryngology disorders	5	17%	14	19%	4	17%	0.97[a]
Lung disease	4	14%	11	15%	3	13%	0.97[a]
Gastro-intestinal disease	8	28%	21	28%	5	22%	0.82[a]
Urological or gynaecological disease	9	31%	27	36%	7	30%	0.80[a]
Endocrine disease (excl. thyroid)	6	21%	9	12%	4	17%	0.52[a]
Musculoskeletal disorder	14	48%	26	35%	7	30%	0.35[a]
Psychiatric disorder	2	7%	4	5%	1	4%	0.92[a]
Thyroid function							
TSH, mU/L (median, IQR)	1.61 (1.09–2.35)		1.54 (0.92–2.40)		1.42 (0.48–2.20)		0.46[c]
fT4, pmol/L (median, IQR)	14.45 (13.18–16.20)		14.30 (13.10–15.95)		15.05 (13.43–17.18)		0.62[c]
Thyroid US							
Solitary nodule	22	76%	53	72%	13	57%	0.28[a]
Multinodular disease	7	24%	21	28%	10	43%	
US nodule size, mm (median, IQR)	37 (27–50)		35 (25–40)		36 (16–47)		0.47[c]
TNM stage							
T1a	1	3%					
T1b	5	17%					
T2	13	45%					
T3	10	34%					
N0/x	27	93%					
N1a	2	7%					
N1b	0	0%					
M0	29	100%					
Highest educational level completed (*n* = 118)							
Primary school	0	0%	4	6%	2	9%	0.12[a]
Pre-vocational secondary education	7	10%	7	25%	6	27%	
High school	1	4%	5	7%	2	9%	
Vocational education	7	25%	12	18%	4	18%	
Higher professional education	5	18%	30	44%	6	35%	
University	8	29%	10	15%	2	9%	
Employment status (*n* = 115)							
Fulltime job	7	26%	14	21%	3	14%	0.15[a]
Parttime job	9	33%	37	55%	8	38%	
Unemployed	11	41%	16	24%	10	48%	
Average work hours (hours/week) (median, IQR)	18 (0–36)		24 (6–32)		8 (0–30)		0.21[c]
Time from baseline HRQoL assessment in months (median, IQR)							
to FDG-PET/CT scan (*n* = 115)	−0.1 (−0.3–0)		−0.2 (−0.7–0)		0 (−0.6–0)		0.28[c]
to diagnostic surgery (*n* = 94)	1.9 (1.3–3.6)		1.8 (1.2–3.5)				0.52[d]
to cTT (*n* = 17)	5.6 (2.7–6.8)						
to RAI (*n* = 16)	6.3 (3.6–8.4)						
to follow-up US (*n* = 21)					10.6 (9.3–11.4)		
to 3-month HRQoL assessment (*n* = 105)[e]	3.1 (3.0–3.4)	*n* = 23	3.0 (2.8–3.3)	*n* = 62	3.0 (2.9–3.6)	*n* = 20	0.30[c]
to 6-month HRQoL assessment (*n* = 103)[e]	5.9 (5.9–6.0)	*n* = 23	6.0 (5.8–6.4)	*n* = 60	6.0 (5.8–6.6)	*n* = 20	0.81[c]
to 12-month HRQoL assessment (*n* = 104)[e]	11.9 (11.8–12.1)	*n* = 23	11.9 (11.8–12.2)	*n* = 60	12.0 (11.8–12.3)	*n* = 21	0.95[c]

[a]Pearson's chi-square test. [b]one-way ANOVA. [c]Kruskal–Wallis test. [d]Mann–Whitney *U* test. [e]Three, one, and two patients completed the 3-, 6-, and 12-month HRQoL assessment, respectively, but not the baseline HRQoL assessment.

cTT, completing total thyroidectomy; Excl., excluding; fT4, free thyroxine; reference range approximately 10–25 pmol/L (sex and age dependent); HRQoL, health-related quality of life; Incl., including; IQR, interquartile range; RAI, radioiodine therapy; TSH, thyroid-stimulating hormone; reference range 0.4–4.0 mU/L; US, ultrasound.

https://ec.bioscientifica.com
https://doi.org/10.1530/EC-22-0014

Discussion

To the best of our knowledge, our prospective multicentre study was the first to assess HRQoL in patients undergoing an FDG-PET/CT-driven diagnostic workup and subsequent surgical management or active surveillance for a Bethesda III/IV thyroid nodule. These HRQoL evaluations are essential to improve the understanding of the impact that diagnostics and successive management consequences have on patients and to include HRQoL considerations in shared-decision-making processes.

We observed sustained HRQoL during the first year of active surveillance in patients with a cytological indeterminate thyroid nodule and negative FDG-PET/CT scan. These results suggest that the negative FDG-PET/CT reassured the patients of the low malignancy risk and that no meaningful concerns arose during active surveillance. Patients undergoing diagnostic surgery, however, experienced worsened HRQoL on various physical and psychosocial domains throughout the 1-year follow-up, especially during the first 3–6 months. This was primarily observed in patients with benign histopathology, even though visual assessment of Fig. 2 and Supplementary Figs 1, 2 suggested quite similar patterns for patients with malignancy without reaching statistical significance, possibly as this group was smaller ($n = 29$ vs $n = 74$). As hypothesised, the decrease in HRQoL in patients with benign histopathology appeared temporary on most scales, gradually recovering over time and no longer showing a statistically significant difference at 12 months relative to baseline. On some scales, however, HRQoL was still worse than baseline after 1 year, including cognitive impairment and cosmetic complaints scales, and the EQ-5D-5L and RAND-36 physical component score. Goitre symptoms and anxiety had improved beyond baseline levels at the 12-month assessment.

The observed deteriorations in HRQoL during the 3-month assessment are likely attributable to the diagnostic surgery and the first weeks postoperative, considering the 4-week recall periods of the RAND-36 and ThyPRO (26, 27, 29). As presented in Table 1, diagnostic surgery was performed after a median of 1.8 months in patients with benign histopathology and 1.9 months in patients with malignancy. Our findings are in line with previous studies that demonstrated a temporary decrease in HRQoL following thyroid surgery and subsequent improvement up to 12 months postoperatively, including an improvement in goitre symptoms and anxiety (10, 13, 14, 15, 16, 17, 23).

Visual assessment of Fig. 2 and Supplementary Figs 1, 2 suggests that patients with malignant histopathology perceived a persistent decrease in HRQoL at 6 months, which was statistically significantly lower as compared to baseline for the ThyPRO cognitive impairment and impaired daily life scales, EQ-5D-5L, and RAND-36 physical role functioning. These within-group HRQoL effects may be related to the completion of thyroidectomy followed by radioactive iodine (RAI) therapy that approximately half of the patients with malignancy underwent after a median of 5.6 and 6.3 months relative to baseline, respectively. Subgroup analysis was not appropriate due to the limited subgroup sizes. In previous studies, RAI was related to decreased HRQoL, mostly due to physical problems. Worse HRQoL is observed with higher doses of RAI and following thyroid hormone withdrawal as compared to recombinant human thyrotropin admission (10, 33, 34). In the current study, 12 of 16 (22%) patients received RAI after thyroid hormone withdrawal for at least 4 weeks, the remaining after recombinant TSH administration.

HRQoL was recently assessed in patients undergoing molecular testing for a Bethesda III or IV nodule using the ThyPRO (39-item version) (23, 24). Similar to the results of the current study, patients under ultrasound surveillance following a benign molecular test result showed unchanged HRQoL from baseline to median of 15 months of follow-up (23). HRQoL scores in patients with a benign molecular test result were similar to patients with a benign cytology result, suggesting that a benign molecular test result provided similar reassurance of a low risk of malignancy (24). Patients who underwent diagnostic surgery for a suspicious molecular test result showed deteriorated goitre, depression, and anxiety scores at the first assessment after molecular testing, all of which had improved 8 months after surgery. Short-term postoperative HRQoL measurements were not available (23).

It was previously hypothesised that patients undergoing active surveillance may experience some degree of long-term, cyclic psychological distress concentrated around their periodic follow-up, caused by the suspense and uncertainty of repeated ultrasound and/or FNAC procedures and the outcome thereof (35). No indications of such an effect were observed during the 1-year follow-up of the current study, which included one protocolled follow-up visit.

We assessed the MIC to estimate which HRQoL differences could be perceived as important by the patient, but only for the ThyPRO. Following a systematic literature search, we identified no suitable country- and/or disease-specific MIC values for the EQ-5D-5L and

https://ec.bioscientifica.com
https://doi.org/10.1530/EC-22-0014

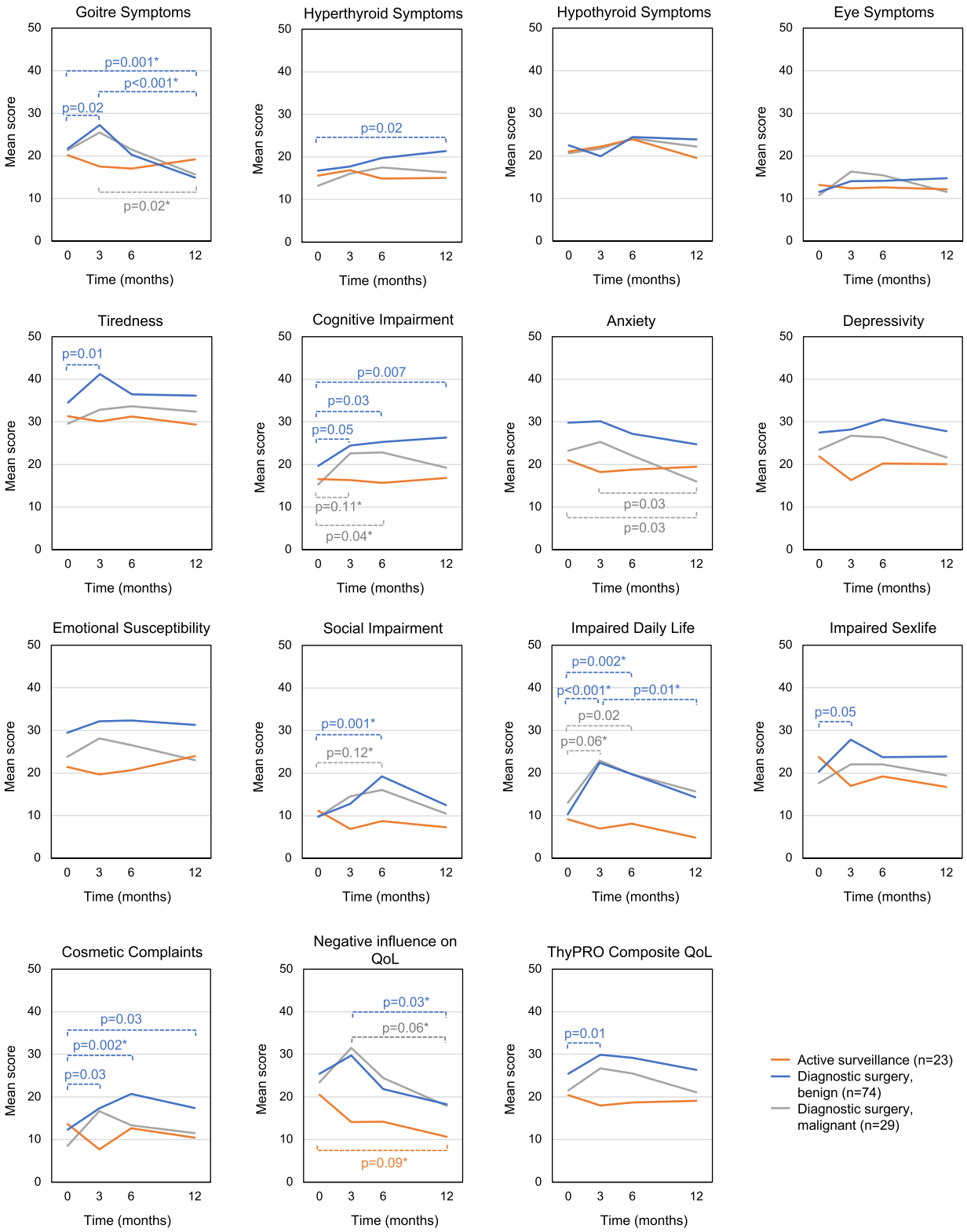

Figure 2

Within-group changes in mean ThyPRO scores over time. For the ThyPRO, scores range from 0 (best score) to 100 (worst score). The *P* value indicates the statistically significant within-group differences between two assessments over time, using a paired samples *t*-test. Full data, including mean differences and other *P* values, are presented in Supplementary Table 3. *Indicates clinically relevant difference using the MIC values established by Nordqvist *et al.* (31). MIC, minimally important change; ThyPRO, Thyroid Patient-Reported Outcome.

https://ec.bioscientifica.com
https://doi.org/10.1530/EC-22-0014

Table 2 Univariate between-group differences in ThyPRO scale scores.

	Patient score, between-group difference											
	Baseline			3 months			6 months			12 months		
ThyPRO	Mean	S.D.	*P*[a]	Mean	S.D.	*P*[a]	Mean	S.D.	*P*[a]	Mean	S.D.	*P*[a]
Goitre symptoms												
Malignant	21.4	18.5	0.93	25.5	21.5	0.71	21.5	21.5	0.80	15.7	17.7	0.84
Benign	21.7	16.2	ref.	27.3	21.2	ref.	20.3	20	ref.	14.9	17.4	ref.
Active surveillance	20.2	14.1	0.67	17.5	14.9	0.01[b]	17.0	12.0	0.34	19.2	14.9	0.25
Hyperthyroid symptoms												
Malignant	13.3	16.3	0.33	16.1	19.0	0.67	17.5	19.8	0.61	16.4	16.9	0.21
Benign	16.8	18.0	ref.	17.8	18.0	ref.	19.8	21.2	ref.	21.4	21.1	ref.
Active surveillance	15.6	14.8	0.75	16.9	15.4	0.81	14.9	15.2	0.22	15.1	13.9	0.10
Hypothyroid symptoms												
Malignant	20.6	23.4	0.72	21.7	25.6	0.74	24.1	25.4	0.96	22.2	23	0.75
Benign	22.5	25.0	ref.	19.9	23.3	ref.	24.4	26.2	ref.	23.9	26.1	ref.
Active surveillance	21.0	27.2	0.82	22.2	24.0	0.69	23.9	22.3	0.92	19.5	22.5	0.43
Eye symptoms												
Malignant	10.8	18.5	0.84	16.3	21.3	0.62	15.5	19.7	0.75	11.5	14.8	0.34
Benign	11.5	17.4	ref.	14.0	18.9	ref.	14.1	16.5	ref.	14.8	17.3	ref.
Active surveillance	13.2	17.6	0.69	12.4	18.3	0.71	12.7	15.6	0.69	12.2	17.8	0.54
Tiredness												
Malignant	29.6	24.1	0.35	32.9	23.9	0.12	33.7	23.6	0.59	32.4	23.4	0.47
Benign	34.5	23.1	ref.	41.2	25.0	ref.	36.5	25.5	ref.	36.1	23.3	ref.
Active surveillance	31.3	20.4	0.52	30.1	19.1	0.02	31.3	20.2	0.31	29.4	17.5	0.14
Cognitive impairment												
Malignant	15.3	20.8	0.34	22.6	23.8	0.72	22.8	22.9	0.62	19.2	18.7	0.10
Benign	19.7	20.9	ref.	24.5	21.8	ref.	25.3	21.8	ref.	26.3	23.1	ref.
Active surveillance	16.6	20.0	0.52	16.4	20.0	0.09[b]	15.7	21.5	0.06[b]	16.8	20.0	0.05[b]
Anxiety												
Malignant	23.2	21.1	0.15	25.3	23.8	0.36	22.1	22.7	0.31	16.0	18.5	0.05
Benign	29.8	20.1	ref.	30.1	24.3	ref.	27.2	25.1	ref.	24.7	23.4	ref.
Active surveillance	21.0	19.8	0.06	18.2	21.1	0.02[b]	18.8	18.3	0.08	19.5	22.4	0.33
Depressivity												
Malignant	23.5	24.0	0.44	26.7	24.5	0.78	26.4	24.9	0.44	21.7	19.0	0.17
Benign	27.5	22.8	ref.	28.2	23.1	ref.	30.6	26.2	ref.	27.8	23.1	ref.
Active surveillance	21.9	19.7	0.25	16.3	15.8	0.005[b]	20.2	17.6	0.03[b]	20.1	17.8	0.09[b]
Emotional susceptibility												
Malignant	23.8	23.2	0.25	28.1	22.7	0.42	26.6	23.5	0.27	23.0	20.5	0.08
Benign	29.5	19.5	ref.	32.2	22.7	ref.	32.3	24.4	ref.	31.3	23.3	ref.
Active surveillance	21.4	18.9	0.08	19.7	18.1	0.007[b]	20.7	14.9	0.005[b]	24.0	18.4	0.12
Social impairment												
Malignant	9.8	16.5	0.99	14.6	19.7	0.68	16.1	24.1	0.55	10.6	18.9	0.65
Benign	9.8	16.8	ref.	12.8	18.7	ref.	19.3	26.2	ref.	12.5	22.8	ref.
Active surveillance	11.2	14.8	0.70	6.9	15.2	0.12	8.8	17.5	0.03[b]	7.3	13.0	0.16
Impaired daily life												
Malignant	13.0	21.4	0.54	22.9	28.2	0.93	19.7	22.8	1	15.7	23.8	0.77
Benign	10.3	15.9	ref.	22.4	26.2	ref.	19.7	26.2	ref.	14.3	21.6	red
Active surveillance	9.1	16.6	0.76	7.0	12.6	<0.001[b]	8.1	15.2	0.007[b]	4.9	11.2	0.006[b]
Impaired sex life												
Malignant	17.6	28.4	0.67	22.0	31.4	0.40	22	33.1	0.81	19.4	32.6	0.53
Benign	20.3	29.9	ref.	27.8	31.7	ref.	23.8	31.9	ref.	23.8	31.7	ref.
Active surveillance	23.8	34.1	0.66	17	28.7	0.12	19.2	29.8	0.53	16.7	31.6	0.35
Cosmetic complaints												
Malignant	8.5	15.5	0.30	16.7	21.9	0.89	13.4	22.7	0.15	11.5	20.4	0.20
Benign	12.3	18.5	ref.	17.3	20.3	ref.	20.7	24.5	ref.	17.4	22.8	ref.
Active surveillance	13.7	15.7	0.73	7.7	11.6	0.004[b]	12.7	18.9	0.10	10.5	13.8	0.08
Negative influence on QoL												
Malignant	23.5	33.7	0.78	31.5	35.3	0.82	24.5	29.0	0.68	17.9	30.7	0.96
Benign	25.5	32.7	ref.	29.7	40.7	ref.	21.9	29.1	ref.	18.3	30.8	ref.
Active surveillance	20.5	28.2	0.49	14.1	27.6	0.03[b]	14.2	22.9	0.20	10.7	18.3	0.15
ThyPRO composite QoL												
Malignant	21.5	19.7	0.33	26.7	20.9	0.48	25.5	19.7	0.40	21.1	17.1	0.17
Benign	25.5	15.7	ref.	29.9	19.4	ref.	29.2	20.5	ref.	26.4	18.8	ref.
Active surveillance	20.4	16.0	0.19	18.0	15.6	0.002[b]	18.7	13.7	0.005[b]	19.1	14.2	0.05

For the ThyPRO, scores range from 0 (best score) to 100 (worst score).
[a]*P* value indicates the between-group difference for that assessment, using an independent samples *t*-test and using the benign histopathology group as reference category (ref.); [b]Indicates clinically relevant difference using the MIC values established by Nordqvist *et al.* (31).

https://ec.bioscientifica.com
https://doi.org/10.1530/EC-22-0014

RAND-36 questionnaires for our study population. As cultural differences play an important role in the perception of HRQoL and the EQ-5D-5L uses country-specific algorithms to calculate utilities, we did not use the available generic EQ-5D-5L MIC values for other countries (36, 37, 38). For the ThyPRO questionnaire, MIC values were recently established in 435 Danish patients with benign thyroid disease, including 135 (31%) nontoxic goitre patients (31). The ThyPRO does not apply country-specific algorithms to calculate its scale scores (29). Finally, because the ThyPRO was considered the most relevant for the current study and because of the limitations of the small group sizes, we chose not to apply Norman's rule of thumb to assess the MIC for the EQ-5D-5L and RAND-36 (39).

Our previous intention-to-treat analysis of the EQ-5D-5L results demonstrated non-inferiority of an FDG-PET/CT-driven workup with regard to general HRQoL based on QALYs (6, 8). For the current comprehensive HRQoL study, a per-protocol analysis with exclusion of crossovers was considered most appropriate, despite its known drawbacks with regard to bias (40). As we aimed to compare the perceived HRQoL of active surveillance after a negative FDG-PET/CT (i.e. reassurance by a negative test result that the risk of malignancy was low, followed by the effects of active surveillance in absence of a definitive histopathological diagnosis) to the HRQoL of diagnostic surgery (i.e. a definitive diagnosis at the cost of effects caused by recovery from surgery and potential surgical complications), patients who crossed over between strategies fitted neither group in light of the counselling and treatment they received.

A main limitation of the current study is the limited sample size of the malignancy and active surveillance groups. Designed as a secondary, explorative analysis of HRQoL data acquired during a randomised controlled trial, the study was not *a priori* powered to distinguish HRQoL effects (6). *Post hoc* sensitivity analysis, however, showed that in the smallest cohort (i.e. active surveillance group, $n=23$) we have 80% power to detect an effect size of approximately 8.9 points on the ThyPRO scales (i.e. 0.539 standard deviations of the observed mean difference between two dependent groups), a number smaller than considered *clinically relevant*, predefined by MIC data (31). Yet, the limited group sizes may have caused underreporting of the *statistical significance* of any smaller effects, especially as compared to similar-sized effects in the larger benign histopathology group. Also, it limited the multivariate analysis by lacking the statistical power to construct a mixed linear model with random intercept and slope, much less to assess other covariates using fixed or random effects, resulting in a simple random intercept model (41). Despite evident univariate within- and between-group differences, no significant HRQoL changes were observed between the groups over time on multivariate analysis.

In conclusion, the current study underpins the importance of an accurate diagnostic workup to sustain HRQoL by preventing futile diagnostic surgeries for indeterminate thyroid nodules. Following an FDG-PET/CT-driven workup in the current study, the reassurance of a negative FDG-PET/CT resulted in sustained HRQoL throughout the first year of active surveillance. Diagnostic surgery for a benign Bethesda III/IV nodule resulted in decreased HRQoL after 1 year with regard to cognitive impairment and physical problems, including cosmetic complaints, but improved goitre symptoms and anxiety. Anxiety was also reduced in patients undergoing surgery for a malignant nodule. Further prospective studies in larger populations are needed to address long-term HRQoL changes between groups.

Supplementary materials

This is linked to the online version of the paper at https://doi.org/10.1530/EC-22-0014.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

The EfFECTS trial, including the current study, was supported by a project grant from the Dutch Cancer Society (KUN 2014–6514).

Acknowledgements

The authors like to thank all the patients who participated in the EfFECTS trial, all members of the *EfFECTS* trial consortium, and all others who were involved in any of the study procedures.

References

1 Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ & Cooper DS. The diagnosis and management of thyroid nodules: a review. *JAMA* 2018 **319** 914–924. (https://doi.org/10.1001/jama.2018.0898)

2 Bernet VJ & Chindris AM. Update on the evaluation of thyroid nodules. *Journal of Nuclear Medicine* 2021 **62** 13S–19S. (https://doi.org/10.2967/jnumed.120.246025)

3 Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, *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** 1–133. (https://doi.org/10.1089/thy.2015.0020)

4 Cibas ES & Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* 2017 **27** 1341–1346. (https://doi.org/10.1089/thy.2017.0500)

5 de Koster EJ, de Geus-Oei LF, Dekkers OM, van Engen-van Grunsven I, Hamming J, Corssmit EPM, Morreau H, Schepers A, Smit J, Oyen WJG, *et al.* Diagnostic utility of molecular and imaging biomarkers in cytological indeterminate thyroid nodules. *Endocrine Reviews* 2018 **39** 154–191. (https://doi.org/10.1210/er.2017-00133)

6 de Koster EJ, de Geus-Oei LF, Brouwers AH, van Dam EWCM, Dijkhorst-Oei LT, van Engen-van Grunsven ACH, van den Hout WB, Klooker TK, Netea-Maier RT, Snel M, *et al.* [18F]FDG-PET/CT to prevent futile surgery in indeterminate thyroid nodules: a blinded, randomised controlled multicentre trial. *European Journal of Nuclear Medicine and Molecular Imaging* 2022 **49** 1970–1984. (https://doi.org/10.1007/s00259-021-05627-2)

7 Vriens D, de Wilt JH, van der Wilt GJ, Netea-Maier RT, Oyen WJ & de Geus-Oei LF. The role of [18F]-2-fluoro-2-deoxy-d-glucose-positron emission tomography in thyroid nodules with indeterminate fine-needle aspiration biopsy: systematic review and meta-analysis of the literature. *Cancer* 2011 **117** 4582–4594. (https://doi.org/10.1002/cncr.26085)

8 de Koster EJ, Vriens D, van Aken MO, Dijkhorst-Oei LT, Oyen WJG, Peeters RP, Schepers A, de Geus-Oei LF, van den Hout WB & EfFECTS trial study group. FDG-PET/CT in indeterminate thyroid nodules: cost-utility analysis alongside a randomised controlled trial. *European Journal of Nuclear Medicine and Molecular Imaging* 2022. (https://doi.org/10.1007/s00259-022-05794-w)

9 Vriens D, Adang EM, Netea-Maier RT, Smit JW, de Wilt JH, Oyen WJ & de Geus-Oei LF. Cost-effectiveness of FDG-PET/CT for cytologically indeterminate thyroid nodules: a decision analytic approach. *Journal of Clinical Endocrinology and Metabolism* 2014 **99** 3263–3274. (https://doi.org/10.1210/jc.2013-3483)

10 Husson O, Haak HR, Oranje WA, Mols F, Reemst PH & van de Poll-Franse LV. Health-related quality of life among thyroid cancer survivors: a systematic review. *Clinical Endocrinology* 2011 **75** 544–554. (https://doi.org/10.1111/j.1365-2265.2011.04114.x)

11 Applewhite MK, James BC, Kaplan SP, Angelos P, Kaplan EL, Grogan RH & Aschebrook-Kilfoy B. Quality of life in thyroid cancer is similar to that of other cancers with worse survival. *World Journal of Surgery* 2016 **40** 551–561. (https://doi.org/10.1007/s00268-015-3300-5)

12 Bianchi GP, Zaccheroni V, Solaroli E, Vescini F, Cerutti R, Zoli M & Marchesini G. Health-related quality of life in patients with thyroid disorders. *Quality of Life Research* 2004 **13** 45–54. (https://doi.org/10.1023/B:QURE.0000015315.35184.66)

13 Mirallie E, Borel F, Tresallet C, Hamy A, Mathonnet M, Lifante JC, Brunaud L, Menegaux F, Hardouin JB, Blanchard C, *et al.* Impact of total thyroidectomy on quality of life at 6 months: the prospective ThyrQoL multicentre trial. *European Journal of Endocrinology* 2020 **182** 195–205. (https://doi.org/10.1530/EJE-19-0587)

14 Shah MD, Witterick IJ, Eski SJ, Pinto R & Freeman JL. Quality of life in patients undergoing thyroid surgery. *Journal of Otolaryngology* 2006 **35** 209–215. (https://doi.org/10.2310/7070.2006.0011)

15 Dogan S, Sahbaz NA, Aksakal N, Tutal F, Torun BC, Yildirim NK, Ozkan M, Ozcinar B & Erbil Y. Quality of life after thyroid surgery. *Journal of Endocrinological Investigation* 2017 **40** 1085–1090. (https://doi.org/10.1007/s40618-017-0635-9)

16 Nickel B, Tan T, Cvejic E, Baade P, McLeod DSA, Pandeya N, Youl P, McCaffery K & Jordan S. Health-related quality of life after diagnosis and treatment of differentiated thyroid cancer and association with type of surgical treatment. *JAMA Otolaryngology–Head and Neck Surgery* 2019 **145** 231–238. (https://doi.org/10.1001/jamaoto.2018.3870)

17 Sorensen JR, Watt T, Cramon P, Dossing H, Hegedus L, Bonnema SJ & Godballe C. Quality of life after thyroidectomy in patients with nontoxic nodular goiter: a prospective cohort study. *Head and Neck* 2017 **39** 2232–2240. (https://doi.org/10.1002/hed.24886)

18 Buttner M, Locati LD, Pinto M, Araujo C, Tomaszewska IM, Kiyota N, Vidhubala E, Brannan C, Hammerlid E, Husson O, *et al.* Quality of life in patients with hypoparathyroidism after treatment for thyroid cancer. *Journal of Clinical Endocrinology and Metabolism* 2020 **105** e4652–e4660. (https://doi.org/10.1210/clinem/dgaa597)

19 Kurumety SK, Helenowski IB, Goswami S, Peipert BJ, Yount SE & Sturgeon C. Post-thyroidectomy neck appearance and impact on quality of life in thyroid cancer survivors. *Surgery* 2019 **165** 1217–1221. (https://doi.org/10.1016/j.surg.2019.03.006)

20 Kletzien H, Macdonald CL, Orne J, Francis DO, Leverson G, Wendt E, Sippel RS & Connor NP. Comparison between patient-perceived voice changes and quantitative voice measures in the first postoperative year after thyroidectomy: a secondary analysis of a randomized clinical trial. *JAMA Otolaryngology– Head & Neck Surgery* 2018 **144** 995–1003. (https://doi.org/10.1001/jamaoto.2018.0309)

21 Goswami S, Peipert BJ, Mongelli MN, Kurumety SK, Helenowski IB, Yount SE & Sturgeon C. Clinical factors associated with worse quality-of-life scores in United States thyroid cancer survivors. *Surgery* 2019 **166** 69–74. (https://doi.org/10.1016/j.surg.2019.01.034)

22 Jeon MJ, Lee YM, Sung TY, Han M, Shin YW, Kim WG, Kim TY, Chung KW, Shong YK & Kim WB. Quality of life in patients with papillary thyroid microcarcinoma managed by active surveillance or lobectomy: a cross-sectional study. *Thyroid* 2019 **29** 956–962. (https://doi.org/10.1089/thy.2018.0711)

23 Schumm MA, Nguyen DT, Kim J, Tseng CH, Chow AY, Shen N & Livhits MJ. Longitudinal assessment of quality of life following molecular testing for indeterminate thyroid nodules. *Annals of Surgical Oncology* 2021 **28** 8872–8881. (https://doi.org/10.1245/s10434-021-10375-6)

24 Wong CW, Schumm MA, Zhu CY, Tseng CH, Arasu A, Han J & Livhits MJ. Quality of life following molecular marker testing for indeterminate thyroid nodules. *Endocrine Practice* 2020 **26** 960–966. (https://doi.org/10.4158/EP-2020-0101)

25 Federatie Medish Specialisten. *Evidence Based Nation-wide Guideline Thyroid Carcinoma version 2.0* . Rotterdam, the Netherlands: Integraal Kankercentrum Nederland, 2014. (available at: https://richtlijnendatabase.nl/richtlijn/schildkliercarcinoom/algemeen.html)

26 Ware Jr JE. SF-36 health survey update. *Spine* 2000 **25** 3130–3139. (https://doi.org/10.1097/00007632-200012150-00008)

27 Brazier J, Roberts J & Deverill M. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics* 2002 **21** 271–292. (https://doi.org/10.1016/s0167-6296(01)00130-8)

28 Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G & Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research* 2011 **20** 1727–1736. (https://doi.org/10.1007/s11136-011-9903-x)

29 Watt T, Bjorner JB, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedus L & Feldt-Rasmussen U. Establishing construct validity for the thyroid-specific patient reported outcome measure (ThyPRO): an initial examination. *Quality of Life Research* 2009 **18** 483–496. (https://doi.org/10.1007/s11136-009-9460-8)

30 Watt T, Bjorner JB, Groenvold M, Cramon P, Winther KH, Hegedus L, Bonnema SJ, Rasmussen ÅK, Ware JE, Jr & Feldt-Rasmussen U. Development of a short version of the thyroid-related patient-reported outcome ThyPRO. *Thyroid* 2015 **25** 1069–1079. (https://doi.org/10.1089/thy.2015.0209)

31 Nordqvist SF, Boesen VB, Rasmussen ÅK, Feldt-Rasmussen U, Hegedus L, Bonnema SJ, Cramon PK, Watt T, Groenvold M & Bjorner JB. Determining minimal important change for the thyroid-related quality of life questionnaire ThyPRO. *Endocrine Connections* 2021 **10** 316–324. (https://doi.org/10.1530/EC-21-0026)

32 Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR & Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings* 2002 **77** 371–383. (https://doi.org/10.4065/77.4.371)

33 Ahn J, Jeon MJ, Song E, Kim TY, Kim WB, Shong YK & Kim WG. Quality of life in patients with papillary thyroid microcarcinoma

according to treatment: total thyroidectomy with or without radioactive iodine ablation. *Endocrinology and Metabolism* 2020 **35** 115–121. (https://doi.org/10.3803/EnM.2020.35.1.115)
34 Dingle IF, Mishoe AE, Nguyen SA, Overton LJ & Gillespie MB. Salivary morbidity and quality of life following radioactive iodine for well-differentiated thyroid cancer. *Otolaryngology–Head and Neck Surgery* 2013 **148** 746–752. (https://doi.org/10.1177/0194599813479777)
35 Balentine CJ, Vanness DJ & Schneider DF. Cost-effectiveness of lobectomy versus genetic testing (Afirma(R)) for indeterminate thyroid nodules: considering the costs of surveillance. *Surgery* 2018 **163** 88–96. (https://doi.org/10.1016/j.surg.2017.10.004)
36 McClure NS, Sayah FA, Xie F, Luo N & Johnson JA. Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores. *Value in Health* 2017 **20** 644–650. (https://doi.org/10.1016/j.jval.2016.11.015)
37 Henry EB, Barry LE, Hobbins AP, McClure NS & O'Neill C. Estimation of an instrument-defined minimally important difference in EQ-5D-5L index scores based on scoring algorithms derived using the EQ-VT version 2 valuation protocols. *Value in Health* 2020 **23** 936–944. (https://doi.org/10.1016/j.jval.2020.03.003)
38 Nolte S, Liegl G, Petersen MA, Aaronson NK, Costantini A, Fayers PM, Groenvold M, Holzner B, Johnson CD, Kemmler G, *et al.* General population normative data for the EORTC QLQ-C30 health-related quality of life questionnaire based on 15,386 persons across 13 European countries, Canada and the Unites States. *European Journal of Cancer* 2019 **107** 153–163. (https://doi.org/10.1016/j.ejca.2018.11.024)
39 Norman GR, Sloan JA & Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care* 2003 **41** 582–592. (https://doi.org/10.1097/01.MLR.0000062554.74615.4C)
40 Smith VA, Coffman CJ & Hudgens MG. Interpreting the results of intention-to-treat, per-protocol, and as-treated analyses of clinical trials. *JAMA* 2021 **326** 433–434. (https://doi.org/10.1001/jama.2021.2825)
41 Bell BA, Morgan GB, Schoeneberger JA, Kromrey JD & Ferron JM. How low can you go?: an investigation of the influence of sample size and model complexity on point and interval estimates in two-level linear models. *Methodology* 2014 **10** 1–11. (https://doi.org/10.1027/1614-2241/a000062)

Received in final form 2 June 2022
Accepted 22 June 2022
Accepted Manuscript published online 22 June 2022