OPTIMIZATION OF DIAGNOSIS AND FOLLOW-UP OF DIFFERENTIATED THYROID CANCER

JAKOB W. KIST

Optimization of diagnosis and follow-up of differentiated thyroid cancer
PhD Thesis, Utrecht University, The Netherlands

Grant
Part of this thesis was supported by a grant from the Dutch Cancer Society (KWF Kankerbestrijding)

Sponsors
Publication of this thesis was financially supported by Bayer, BV Cyclotron, Chipsoft, Genzyme, Goodlife, Ipsen, Novartis and Olympus.

ISBN
978-90-393-6627-1

Printed
Gildeprint, Enschede

Graphic design
Sanne van de Goor

Optimization of diagnosis and follow-up of differentiated thyroid cancer

Optimalisatie van diagnostiek en follow-up van gedifferentieerd schildkliercarcinoom

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op donderdag 13 oktober 2016 des middags te 2.30 uur

door

Jakob Willem Kist
geboren op 9 mei 1984
te Amsterdam

Promotoren: Prof. dr. I.H.M. Borel Rinkes
Prof. dr. O.S. Hoekstra

Copromotoren: Dr. B. de Keizer
Dr. W.V. Vogel

Contents

Part 3: General discussion and summary

GENERAL INTRODUCTION AND OUTLINE OF THE THESIS

CHAPTER 1

Jakob W. Kist

The thyroid gland

The thyroid gland is located anteriorly of the trachea in the neck. Its primary function is to produce the thyroid hormones thyroxine (T_4) and triiodothyronine (T_3). Both hormones play a crucial role in numerous bodily functions, such as basal metabolic rate, growth, protein synthesis and body temperature regulation. The formation of thyroid hormones is dependent on sufficient dietary iodine intake. Iodine is absorbed to the blood as iodide by the intestines. Iodide is subsequently accumulated by the thyroid cells via the plasma membrane glycoprotein sodium/iodide symporter (NIS). This transport is stimulated by the thyroid stimulating hormone (TSH) produced by the pituitary gland.[1]

Thyroid nodules

The prevalence of thyroid nodules in the general population is very high. Detected by palpation in healthy subjects, it ranges between 4.7/1000 to 51/1000 patients.[2] Smaller, non-palpable thyroid nodules are even more prevalent. On ultrasound (US), the prevalence of thyroid nodules ranges from 33% to 67%.[3-5] These results are concordant with autopsy studies showing prevalence rates of up to 65%.[2]

Overall around 5% of thyroid nodules are malignant, necessitating further evaluation of a thyroid nodule.[6,7] Nodules are firstly imaged by US, which was originally used to identify and count nodules and to measure their size. Later specific US characteristics were identified as markers of thyroid cancer.[8,9] Unfortunately, none of these US characteristics has, by itself, a clinically acceptable diagnostic accuracy to differentiate between a benign and a malignant thyroid nodule. In the last decade more advanced US techniques have been developed.

Of those, real-time elastography (RTE), analyzing the elasticity of the nodule, showed promising results in the first published studies.[10-16] This technique is based on the assumption that malignant nodules are stiffer than benign nodules. This assumption is based on the fact that most malignant nodules have abnormally firm stroma, due to the presence of collagen and myofibroblasts.[17] External force is applied by an US probe and the degree of distortion of the nodule is measured and visualized on the US image. Its definite role in the diagnostic work-up of thyroid nodules has yet to be decided. We therefore conducted a systematic review and meta-analysis, presented in **chapter 2.**

Because of the limited ability of US to discriminate between benign and malignant nodules, fine-needle aspiration (FNA) of the nodule is often indicated. The aspirate is processed via a direct smear and subsequently sent for further analysis by the cytopathologist. The Bethesda System for Reporting Thyroid Cytopathology has become the golden standard to classify thyroid cytology.[18] This six category classification system, of which each category carries an explicit malignancy risk, aimed to standardize FNA reports and

Category	Description category	Risk of malignancy (%)	Usual management
I	Non diagnostic or unsatisfactory	1-4	Repeat FNA
II	Benign	0-3	Clinical follow-up
III	Atypia of undetermined significance or follicular lesion of undetermined significance	~5-15	Repeat FNA / Diagnostic lobectomy
IV	Follicular neoplasm or suspicious for a follicular neoplasm	15-30	Repeat FNA / Diagnostic lobectomy
V	Suspicious for malignancy	60-75	Diagnostic lobectomy
VI	Malignant	97-99	Total thyroidectomy

Table 1: Bethesda classification. Derived from Cibas et al.[18]

thereby guide clinical decision making (Table 1).[18] A FNA sample has to contain at least six groups of epithelial cells, each composed of at least 10 cells, to be classified into one of the diagnostic categories, category II to IV. Non-diagnostic FNA samples are classified as Bethesda category I and are caused by an insufficient number of cells, contamination with blood or other processing issues. Although several guidelines state that no more than 15-20% of FNA samples should be nondiagnostic[19,20], studies have reported rates of over 30%.[21,22]

Rapid onsite adequacy assessment (ROSAA), whereby the FNA sample is assessed for adequacy by a cytopathologist on site, thus providing immediate feedback, has been suggested in literature as a way to improve adequacy rates. It has been successfully introduced in the diagnostic work-up of other tumor types and several studies reported good results with thyroid FNAs.[23-26] Based on these studies ROSAA was implemented a few years ago in the UMC Utrecht. Besides ROSAA, other factors may be of influence on FNA adequacy, like aspiration technique, experience of the operator, type and size of the needle.[22,26,27] In **chapter 4** we present the results of our study investigating what factors are associated with accuracy rates in our center.

Follow-up is advised for benign nodules (Bethesda category II) without compressive symptoms or cosmetic concerns.[18] Because of its very high malignancy rate (around 98%[18,28]) Bethesda category 6 nodules are treated with total thyroidectomy (Table 1).[29,30] Bethesda category III, IV and V FNA results however, pose a dilemma for the clinician. The combination of these three Bethesda categories is also referred to in literature as indeterminate FNA results. It is important to note that in some studies indeterminate nodules

only include category 3 and 4 FNA results, while others include category 5. For indeterminate nodules diagnostic thyroid lobectomy is frequently performed to obtain a final diagnosis. However, final histology turns out benign in the majority of these patients, while they have been exposed to a surgical intervention with associated morbidity and costs. Several potential solutions have been proposed to reduce the number of futile diagnostic lobectomies for indeterminate nodules. Amongst these is the use of the aforementioned US elastography.[31,32] Its accuracy and use for this indication are addressed in **chapter 3**.

Another approach to obtain a conclusive diagnosis of indeterminate nodules is molecular testing of the FNA sample. The 2007 guideline of the American Thyroid Association (ATA) stated that "the use of molecular markers (e.g., BRAF, RAS, RET/PTC, PAX8-PPARγ, or galectin-3) may be considered for patients with indeterminate cytology on FNA to help guide management", but no further recommendations on its application are provided.[29] Of these markers the B-type Raf Kinase Val600Glu mutation, abbreviated as BRAF(V600E), is of particular interest as a mutation in this gene is pathognomonic of thyroid cancer, and mutation assessment can be done on a FNA sample.[33,34] BRAF(V600E) mutations found in indeterminate FNA samples consequently change the clinical approach from repeat FNA or diagnostic lobectomy to therapeutic thyroidectomy, reducing time and costs. The potential relevance of BRAF(V600E) mutation analysis of indeterminate FNA samples in the Netherlands is explored in **chapter 5.**

Epidemiology and treatment of differentiated thyroid cancer

Differentiated thyroid cancer (DTC) is the most prevalent type of thyroid cancer and its incidence is rising. In the Netherlands the incidence of DTC rose from 1.8 per 100.000 in 1990 to 2.7 per 100.000 in 2010.[35] In the United States the incidence of papillary thyroid cancer, the most common subtype of DTC, increased even more from 3.4 per 100.00 in 1975 to 12.5 per 100.000 in 2009.[36] Despite the increase in incidence, mortality rates remained stable over the years, indicating that overdiagnosis is a major factor causing this rise.[37,38]
For DTC staging the TNM (tumor, node, metastasis) classification system is generally recommended.[30] Based on the T-, N- and M-classification and age at time of diagnosis, patients are categorized in different stages (Table 2). The primary treatment for all stages of DTC, except for tumors smaller than 1 cm without known nodal and/or distant metastasis (T1aN0M0-x), consists of a total thyroidectomy followed by ablation therapy with radioactive iodine-131 (^{131}I)(RAI).

^{131}I is a radioisotope with a physical half-life of 8.02 days that returns to its stable state (xenon-131) by emitting both beta particles and gamma rays (γ) consisting of high energy photons. The beta particles inflict damage to the DNA molecule of the thyroid cells, which leads to cell dysfunction and eventually cell death.[39] As the uptake of iodine is highly specific by thyroid

Primary tumor (T-stage)	
TX	Primary tumor cannot be assessed.
To	No evidence of primary tumor.
T1a	Tumor ≤1 cm, limited to the thyroid.
T1b	Tumor >1 cm but ≤2 cm in greatest dimension, limited to the thyroid.
T2	Tumor >2 cm but ≤4 cm in greatest dimension, limited to the thyroid.
T3	Tumor >4 cm in greatest dimension limited to the thyroid or any tumor with minimal extrathyroid extension (e.g., extension to sternothyroid muscle or perithyroid soft tissues).
T4a	Moderately advanced disease.
	Tumor of any size extending beyond the thyroid capsule to invade subcutaneous soft tissues, larynx, trachea, esophagus, or recurrent laryngeal nerve.
T4b	Very advanced disease.
	Tumor invades prevertebral fascia or encases carotid artery or mediastinal vessels.
Regional lymph nodes (N-stage)	
NX	Regional lymph nodes cannot be assessed.
No	No regional lymph node metastasis.
N1	Regional lymph node metastasis.
N1a	Metastases to Level VI (pretracheal, paratracheal, and prelaryngeal/Delphian lymph nodes).
N1b	Metastases to unilateral, bilateral, or contralateral cervical (Levels I, II, III, IV, or V) or retropharyngeal or superior mediastinal lymph nodes (Level VII).
Distant metastasis (M-stage)	
Mo	No distant metastasis.
M1	Distant metastasis.

Stage	T	N	M
Younger than 45 years			
I	Any T	Any N	Mo
II	Any T	Any N	M1
45 years or older			
I	T1	No	Mo
II	T2	No	Mo
III	T3	No	Mo
	T1-3	N1a	Mo
IVa	T4a	No-1	Mo
	T1-3	N1b	Mo
IVb	T4b	Any N	Mo
IVc	Any T	Any N	M1

Table 2: TNM AJCC[70]

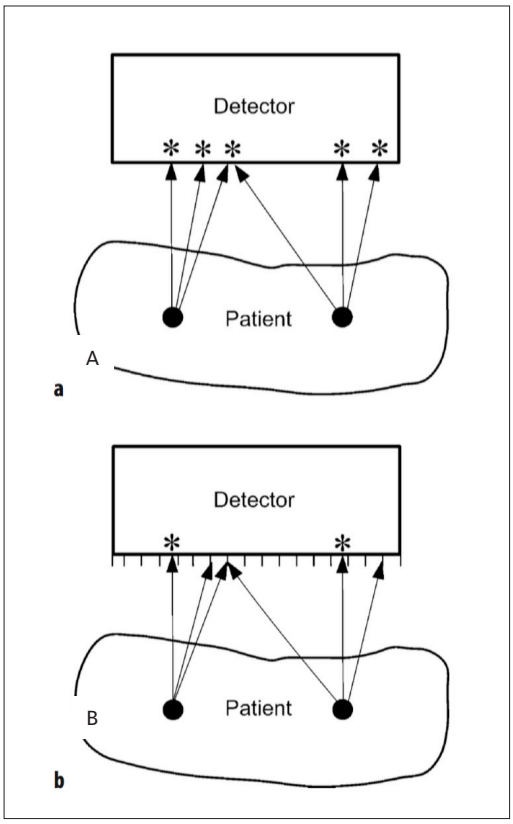
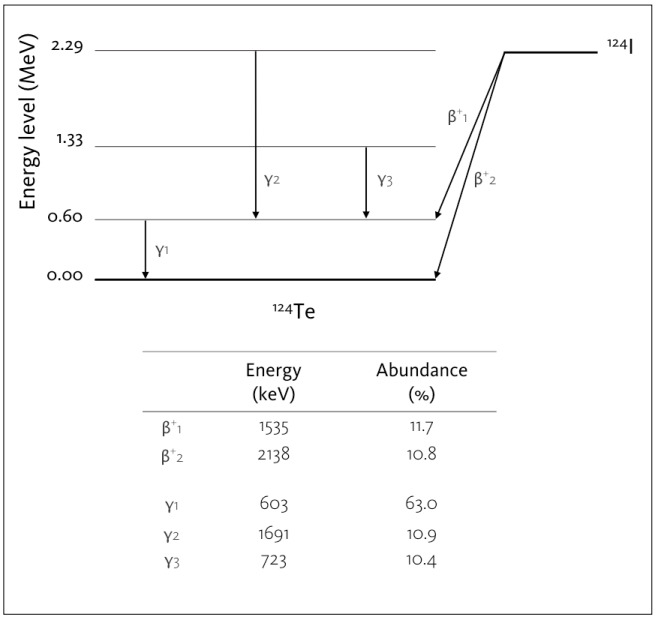


Figure 1: a. In the absence of collimation there is no relationship between the position at which a gamma ray hits the detector and that from which it left the patient. **b.** The parallel-hole collimator forms an image by excluding all gamma rays except those travelling parallel to the hole's axis. Reprinted from Practical Nuclear Medicine by Sharp et al.[44]

cells via NIS, limited doses are delivered to non-thyroid organs. The RAI ablation therapy aims to eradicate thyroid (cancer) cells that remain after total thyroidectomy, and subsequently facilitate follow-up with the biomarker thyroglobulin (Tg): after total thyroidectomy and RAI ablation no thyroid cells should remain, and so Tg-levels should be immeasurable. In addition, RAI ablation enables post-ablation therapy imaging by the emitted photons. This whole body scintigraphy (WBS) can be combined with single photon emission computed tomography / computed tomography (SPECT/CT) and plays an important role in detecting lymph node or distant metastasis.[40-43] The photons are emitted from the patient randomly in all directions and for imaging it is required to locate their origin. This is achieved by a collimator, a lead block with small holes, allowing only the photons emitted in line with the holes to pass through (Fig. 1)[44]. The main disadvantage of this approach is its dramatically reduced sensitivity: approximately only 1 in 100.000 photons reaches the detector.[45] The value of ^{131}I WBS has been further increased by the introduction of 3D imaging with SPECT with anatomical correlation by CT. The detectors of the gamma camera circle around the patient obtaining information of photon emission in every direction. SPECT images can be fused with the anatomical information of the CT scan, enabling a more precise staging after RAI.



	Energy (keV)	Abundance (%)
β^{+}_1	1535	11.7
β^{+}_2	2138	10.8
γ_1	603	63.0
γ_2	1691	10.9
γ_3	723	10.4

Figure 2: Simplified decay scheme of ^{124}I. Multiple prompt high-energy gamma photons are emitted simultaneously with the positrons of which some are accepted within the energy discrimination window of the PET scanner (e.g. 400-650 keV). Image adapted from Lubberink et al.[71]

Suspected recurrence of differentiated thyroid cancer

Despite the excellent survival of DTC patients, a substantial number of DTC patients faces recurrence after their primary treatment during follow-up. Up to 25% faces either locoregional recurrence in lymph nodes or distant metastasis.[46,47] All patients treated for DTC therefore undergo regular Tg measurements and US of the neck to identify possible recurrences. If lymph node metastases are diagnosed on US of the neck surgical resection is indicated.[29,30,48,49] Any increase in Tg-level raises the suspicion of recurrence. In case of an increased Tg-level, but without evidence of locoregional metastasis in lymph nodes on US, additional imaging is warranted.

Diagnostic WBS with a low administered dose of iodine-123 (^{123}I) or ^{131}I was used to identify metastases, however, as their diagnostic accuracy is low, it is no longer advocated.[50-55] Therefore, in these cases guidelines advocate 'blind' treatment, also referred to in literature as 'empiric' treatment, with high dose ^{131}I therapy.[52,56-59] This therapy is effective in a substantial part of the patients. However, in up to half of the patients the post-therapy ^{131}I WBS shows no uptake. [60-63] This means that the therapy can be considered futile, while the patient suffers from side effects. The negative post-therapy ^{131}I WBS is caused by a lack of sufficient iodine uptake by the cancer cells to show up on the scan. Dedifferentiation is the main process reducing iodine avidity of thyroid cancer cells. Genetic alterations in DTC may lead to dedifferentiation. One of the alterations occurring in the process of dedifferentiation is the reduced NIS expression on the membrane, resulting in diminished iodine

avidity. This is of clinical relevance as clinically beneficial effects of ^{131}I therapy depend on sufficient uptake of ^{131}I. Another change in dedifferentiating thyroid cancer cells is its increased (glucose) metabolism. Two alternative diagnostic modalities use these thyroid cancer cell characteristics: iodine-124 (^{124}I) and ^{18}F-fluorodeoxyglucose (^{18}F-FDG). Both tracers are used with Positron Emission Tomography (PET) / Computed Tomography (CT) as they are both positron emitting radioisotopes.

^{124}I has a complex decay scheme with a positron abundance of only 23%. Besides positrons, ^{124}I emits a significant amount of high-energy gamma rays that lead to extra non- or partial annihilation coincidence detections (Fig. 2). Despite these disadvantageous physical properties of ^{124}I as a PET-tracer, ^{124}I PET/CT offers potential advantages above diagnostic ^{131}I and ^{123}I SPECT/CT. Both spatial resolution and sensitivity of PET/CT are higher in comparison to SPECT/CT.[64] Furthermore, its long 4.2-day half-life allows for pre-therapeutic dosimetry.[65,66] ^{18}F-FDG is a tracer that can visualize enhanced glucose metabolism, which typically occurs in dedifferentiated DTC. Uptake of ^{18}F-FDG in metastases is associated with a poorer prognosis.[67] The combination of both modalities is, therefore, of great potential for the work-up of patients with suspected DTC metastases.[68] In **chapter 6** a study protocol aimed at investigating the combined use of ^{18}F-FDG and ^{124}I PET/CT as predictors of the post-therapy ^{131}I WBS outcome is presented.

As DTC is a relatively rare disease and patients with suspected recurrent DTC are even rarer, studies on ^{124}I PET/CT, such as the study described in **chapter 6,** often require a multicenter design. To compare ^{124}I PET scans and especially to compare quantitative ^{124}I PET scans, amongst different centers standardization of scanning and acquisition is required. Previously, such a procedure led to a standard of scanning for ^{18}F-FDG.[69] A similar approach is warranted for ^{124}I. A procedure aimed at achieving this is described in **chapter 7.** The results of the clinical study with the use of calibrated ^{124}I PET/CT as predictors of the post-therapy ^{131}I WBS outcome are presented in **chapter 8.** Finally, in **chapter 9,** a quantitative comparison of the detection limits of ^{124}I PET/CT and ^{131}I SPECT/CT is presented, as the results of **chapter 8** raised questions on the detection limits of ^{124}I PET/CT.

Aims of the thesis

Taken together, two major aims of this thesis can be formulated: firstly, to optimize the diagnosis of the numerous thyroid nodules that are (incidentally) diagnosed **(PART I)**; and secondly, to improve the approach of previously treated DTC patients with biochemical suspicion of recurrence **(PART II)**.

References

1. Gardner D, Shoback D. Greenspan's Basic and Clinical Endocrinology, Ninth Edition. McGraw Hill Professional; 2011.

2. Dean DS, Gharib H. Epidemiology of thyroid nodules. Best Pract Res Clin Endocrinol Metab 2008;22(6):901–11.

3. Reiners C, Wegscheider K, Schicha H, et al. Prevalence of thyroid disorders in the working population of Germany: ultrasonography screening in 96,278 unselected employees. Thyroid 2004;14(11):926–32.

4. Ezzat S, Sarti DA, Cain DR, Braunstein GD. Thyroid incidentalomas. Prevalence by palpation and ultrasonography. Arch Intern Med 1994;154(16):1838–40.

5. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. Eur J Clin Invest 2009;39(8):699–706.

6. Brito JP, Yarur AJ, Prokop LJ, McIver B, Murad MH, Montori VM. Prevalence of Thyroid Cancer in Multinodular Goiter Versus Single Nodule: A Systematic Review and Meta-Analysis. Thyroid 2013;23(4):449–55.

7. Hegedüs L. Clinical practice. The thyroid nodule. N Engl J Med 2004;351(17):1764–71.

8. Sipos JA. Advances in ultrasound for the diagnosis and management of thyroid cancer. Thyroid 2009;19(12):1363–72.

9. Paschke R, Hegedüs L, Alexander E, Valcavi R, Papini E, Gharib H. Thyroid nodule guidelines: agreement, disagreement and need for future research. Nature Reviews Endocrinology 2011;7(6):354–61.

10. Hong Y, Liu X, Li Z, Zhang X, Chen M, Luo Z. Real-time ultrasound elastography in the differential diagnosis of benign and malignant thyroid nodules. J Ultrasound Med 2009;28(7):861–7.

11. Friedrich-Rust M, Sperber A, Holzer K, et al. Real-time Elastography and Contrast-Enhanced Ultrasound for the Assessment of Thyroid Nodules. Exp Clin Endocrinol Diabetes 2009;118(09):602–9.

12. Rubaltelli L, Corradin S, Dorigo A, et al. Differential Diagnosis of Benign and Malignant Thyroid Nodules at Elastosonography. Ultraschall in Med 2008;30(02):175–9.

13. Rago T, Santini F, Scutari M, Pinchera A, Vitti P. Elastography: new developments in ultrasound for predicting malignancy in thyroid nodules. J Clin Endocrinol Metab 2007;92(8):2917–22.

14. Tranquart F, Bleuzen A, Pierre-Renoult P, Chabrolle C, Sam Giao M, Lecomte P. [Elastosonography of thyroid lesions]. J Radiol 2008;89(1 Pt 1):35–9.

15. Bojunga J, Herrmann E, Meyer G, Weber S, Zeuzem S, Friedrich-Rust M. Real-Time Elastography for the Differentiation of Benign and Malignant Thyroid Nodules: A Meta-Analysis. Thyroid 2010;20(10):1145–50.

16. Asteria C, Giovanardi A, Pizzocaro A, et al. US-elastography in the differential diagnosis of benign and malignant thyroid nodules. Thyroid 2008;18(5):523–31.

17. Monpeyssen H, Tramalloni J, Poirée S, Hélénon O, Correas JM. Elastography of the thyroid. Diagnostic and Interventional Imaging 2013;94(5):535–44.

18. Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. Thyroid 2009;19(11):1159–65.

19. Gharib H, Papini E, Paschke R, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. Endocr Pract 2010;16(3):468–75.

20. Pitman MB, Abele J, Ali SZ, et al. Techniques for thyroid FNA: a synopsis of the National Cancer Institute Thyroid Fine-Needle Aspiration State of the Science Conference. Diagn Cytopathol 2008;36(6):407–24.

21. Jing X, Michael CW, Pu RT. The clinical and diagnostic impact of using standard criteria of adequacy assessment and diagnostic terminology on thyroid nodule fine needle aspiration. Diagn Cytopathol 2008;36(3):161–6.

22. Zhu W, Michael CW. How important is on-site adequacy assessment for thyroid FNA? An evaluation of 883 cases. Diagn Cytopathol 2007;35(3):183–6.

23. Layfield LJ, Bentz JS, Gopez EV. Immediate on site interpretation of fine needle aspiration smears. Cancer Cytopathology 2001;

24. Nasuti JF, Gupta PK, Baloch ZW. Diagnostic value and cost effectiveness of on site evaluation of fine needle aspiration specimens: Review of 5,688 cases. Diagn Cytopathol 2002;27(1):1–4.

25. Redman R, Zalaznick H, Mazzaferri EL, Massoll NA. The impact of assessing specimen adequacy and number of needle passes for fine-needle aspiration biopsy of thyroid nodules. Thyroid 2006;16(1):55–60.

26. Ghofrani M, Beckman D, Rimm DL. The value of onsite adequacy assessment of thyroid fine needle aspirations is a function of operator experience. Cancer Cytopathology 2006;

27. Cappelli C, Pirola I, Gandossi E, De Martino E, Agosti B, Castellano M. Fine-Needle Aspiration Cytology of Thyroid Nodule: Does the Needle Matter? Southern Medical Journal 2009;102(5):498–501.

28. Cibas ES, Baloch ZW, Fellegara G, et al. A prospective assessment defining the limitations of thyroid nodule pathologic evaluation. Ann Intern Med 2013;159(5):325–32.

29. Cooper DS, Doherty GM, Haugen BR, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid. 2009;19(11):1167–214.

30. Dutch thyroid cancer managment guidlines (1.1) [Internet]. IKNL. 2007 [cited 2012 Dec 3];:1–161. Available from: http://www.oncoline.nl/richtlijn/doc/index.php?type=pda&richtlijn_id=511

31. Rago T, Scutari M, Santini F, et al. Real-Time Elastosonography: Useful Tool for Refining the Presurgical Diagnosis in Thyroid Nodules with Indeterminate or Nondiagnostic Cytology. J Clin Endocrinol Metab 2010;95(12):5274–80.

32. Lippolis PV, Tognini S, Materazzi G, et al. Is Elastography Actually Useful in the Presurgical Selection of Thyroid Nodules with Indeterminate Cytology? J Clin Endocrinol Metab 2011;96(11):E1826–30.

33. Kebebew E, Weng J, Bauer J, et al. The prevalence and prognostic value of BRAF mutation in thyroid cancer. Ann Surg 2007;246(3):466–70.

34. Xing M, Tufano RP, Tufaro AP, et al. Detection of BRAF mutation on fine needle aspiration biopsy specimens: a new diagnostic tool for papillary thyroid cancer. J Clin Endocrinol Metab 2004;89(6):2867–72.

35. Netherlands Cancer Registry [Internet]. [cited 2015 Aug 27];Available from: http://www.cijfersoverkanker.nl

36. Davies L, Welch HG. Current thyroid cancer trends in the United States. JAMA Otolaryngol Head Neck Surg 2014;140(4):317–22.

37. Davies L, Welch HG. Increasing incidence of thyroid cancer in the United States, 1973-2002. JAMA 2006;295(18):2164–7.

38. Ito Y, Nikiforov YE, Schlumberger M, Vigneri R. Increasing incidence of thyroid cancer: controversies explored. Nature Reviews Endocrinology 2013;:1–7.

39. Robbins RJ, Schlumberger MJ. The evolving role of (131)I for the treatment of differentiated thyroid carcinoma. J Nucl Med 2005;46 Suppl 1(1 suppl):28S–37S.

40. Xue Y-L, Qiu Z-L, Song H-J, Luo Q-Y. Value of 131I SPECT/CT for the evaluation of differentiated thyroid cancer: a systematic review of the literature. Eur J Nucl Med Mol Imaging 2012;40(5):768–78.

41. Chen L, Luo Q, Shen Y, et al. Incremental value of 131I SPECT/CT in the management of patients with differentiated thyroid carcinoma. J Nucl Med 2008;49(12):1952–7.

42. Schmidt D, Szikszai A, Linke R, Bautz W, Kuwert T. Impact of 131I SPECT/spiral CT on nodal staging of differentiated thyroid carcinoma at the first radioablation. J Nucl Med 2009;50(1):18–23.

43. Grewal RK, Tuttle RM, Fox J, et al. The effect of posttherapy 131I SPECT/CT on risk classification and management of patients with differentiated thyroid cancer. J Nucl Med 2010;51(9):1361–7.

44. Sharp PF, Gemmell HG, Murray AD. Practical Nuclear Medicine. Springer Science & Business Media; 2006.

45. Khalil MM. Basic Sciences of Nuclear Medicine. Springer Science & Business Media; 2010.

46. Mazzaferri EL, Jhiang SM. Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer. Am J Med 1994;97(5):418–28.

47. Jonklaas J, Sarlis NJ, Litofsky D, et al. Outcomes of patients with differentiated thyroid carcinoma following initial therapy. Thyroid 2006;16(12):1229–42.

48. Pacini F, Schlumberger M, Dralle H, et al. European consensus for the management of patients with differentiated thyroid carcinoma of the follicular epithelium. 2006. p. 787–803.

49. Cooper DS, Tufano RP. Prophylactic central neck dissection in differentiated thyroid cancer: a procedure in search of an indication. Thyroid 2012;22(4):341–3.

50. de Geus-Oei L-F, Oei H-Y, Hennemann G, Krenning EP. Sensitivity of 123I whole-body scan and thyroglobulin in the detection of metastases or recurrent differentiated thyroid cancer. Eur J Nucl Med Mol Imaging 2002;29(6):768–74.

51. Iwano S, Kato K, Nihashi T, Ito S, Tachi Y, Naganawa S. Comparisons of I-123 diagnostic and I-131 post-treatment scans for detecting residual thyroid tissue and metastases of differentiated thyroid cancer. Ann Nucl Med 2009;23(9):777–82.

52. Pacini F, Agate L, Elisei R, et al. Outcome of differentiated thyroid cancer with detectable serum Tg and negative diagnostic (131) I whole body scan: comparison of patients treated with high (131)I activities versus untreated patients. J Clin Endocrinol Metab 2001;86(9):4092–7.

53. Cailleux AF, Baudin E, Travagli JP, Ricard M, Schlumberger M. Is diagnostic iodine-131 scanning useful after total thyroid ablation for differentiated thyroid cancer? J Clin Endocrinol Metab 2000;85(1):175–8.

54. Pacini F, Capezzone M, Elisei R, Ceccarelli C, Taddei D, Pinchera A. Diagnostic 131-iodine whole-body scan may be avoided in thyroid cancer patients who have undetectable stimulated serum Tg levels after initial treatment. J Clin Endocrinol Metab 2002;87(4):1499–501.

55. Torlontano M, Crocetti U, D'Aloiso L, et al. Serum thyroglobulin and 131I whole body scan after recombinant human TSH stimulation in the follow-up of low-risk patients with differentiated thyroid cancer. Eur J Endocrinol 2003;148(1):19–24.

56. Schlumberger M, Mancusi F, Baudin E, Pacini F. 131I therapy for elevated thyroglobulin levels. Thyroid 1997;7(2):273–6.

57. de Keizer B, Koppeschaar HP, Zelissen PM, et al. Efficacy of high therapeutic doses of iodine-131 in patients with differentiated thyroid cancer and detectable serum thyroglobulin. Eur J Nucl Med 2001;28(2):198–202.

58. Pacini F, Lippi F, Formica N, et al. Therapeutic doses of iodine-131 reveal undiagnosed metastases in thyroid cancer patients with detectable serum thyroglobulin levels. J Nucl Med 1987;28(12):1888–91.

59. Pineda JD, Lee T, Ain K, Reynolds JC, Robbins J. Iodine-131 therapy for thyroid cancer patients with elevated thyroglobulin and negative diagnostic scan. J Clin Endocrinol Metab 1995;80(5):1488–92.

60. Ma C, Xie J, Kuang A. Is empiric 131I therapy justified for patients with positive thyroglobulin and negative 131I whole-body scanning results? J Nucl Med 2005;46(7):1164–70.

61. Pettinato C, Spezi E, Nanni C, et al. Pretherapeutic dosimetry in patients affected by metastatic thyroid cancer using 124I PET/CT sequential scans for 131I treatment planning. Clin Nucl Med 2014;39(8):e367–74.

62. van Tol KM, Jager PL, de Vries EGE, et al. Outcome in patients with differentiated thyroid cancer with negative diagnostic whole-body scanning and detectable stimulated thyroglobulin. Eur J Endocrinol 2003;148(6):589–96.

63. Smallridge RC, Diehl N, Bernet V. Practice trends in patients with persistent detectable thyroglobulin and negative diagnostic radioiodine whole body scans: a survey of American Thyroid Association members. Thyroid 2014;24(10):1501–7.

64. Lubberink M, Abdul Fatah S, Brans B, Hoekstra OS, Teule GJJ. The role of (124) I-PET in diagnosis and treatment of thyroid carcinoma. Q J Nucl Med Mol Imaging 2008;52(1):30–6.

65. Jentzen W, Freudenberg LS, Eising EG, Sonnenschein W, Knust J, Bockisch A. Optimized 124I PET Dosimetry Protocol for Radioiodine Therapy of Differentiated Thyroid Cancer. J Nucl Med 2008;49(6):1017–23.

66. Jentzen W, Hobbs RF, Stahl A, Knust J, Sgouros G, Bockisch A. Pre-therapeutic (124) I PET(/CT) dosimetry confirms low average absorbed doses per administered (131)I activity to the salivary glands in radioiodine therapy of differentiated thyroid cancer. Eur J Nucl Med Mol Imaging 2010;37(5):884–95.

67. Wang W, Macapinlac H, Larson SM, et al. [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography localizes residual thyroid cancer in patients with negative diagnostic (131)I whole body scans and elevated serum thyroglobulin levels. J Clin Endocrinol Metab 1999;84(7):2291–302.

68. Freudenberg LS, Antoch G, Frilling A, et al. Combined metabolic and morphologic imaging in thyroid carcinoma patients with elevated serum thyroglobulin and negative cervical ultrasonography: role of 124I-PET/CT and FDG-PET. Eur J Nucl Med Mol Imaging 2008;35(5):950–7.

69. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging 2009;37(1):181–200.

70. Edge S, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A. AJCC Cancer Staging Manual. Springer; 2010.

71. Lubberink M, Herzog H. Quantitative imaging of 124I and 86Y with PET. Eur J Nucl Med Mol Imaging 2011;38 Suppl 1:S10–8.

PART

diagnosis of
differentiated
thyroid cancer

QUALITATIVE ELASTOGRAPHY CAN REPLACE THYROID NODULE FINE-NEEDLE ASPIRATION IN PATIENTS WITH SOFT THYROID NODULES. A SYSTEMATIC REVIEW AND META-ANALYSIS

CHAPTER 2

Jakob W. Kist
Sjoerd Nell
Thomas P.A. Debray
Bart de Keizer
Timotheus J. van Oostenbrugge
Inne H.M. Borel Rinkes
Gerlof D. Valk
Menno R. Vriens

Eur J Radiol 2015;84(4):652–61.

Abstract

Context

Only a minority of thyroid nodules is malignant; nevertheless, many invasive diagnostic procedures are performed to distinguish between benign and malignant nodules. Qualitative ultrasound elastography is a non-invasive technique to evaluate thyroid nodules.

Objective

To investigate the diagnostic value of qualitative elastography in distinguishing benign from malignant thyroid nodules in patients referred for fine-needle aspiration (FNA).

Data sources

A systematic literature search (PubMed, Embase and Cochrane Library) was performed.

Study selection

Included studies reported thyroid nodule elastography color scores and the related cytologic or histologic findings in patients with a thyroid nodule referred for FNA.

Data extraction

Two independent reviewers extracted study data and assessed study quality. Pooled sensitivities and specificities of different populations were calculated using a bivariate Bayesian framework.

Data synthesis

Twenty studies including 3973 thyroid nodules were analyzed. Pooled results of elastography indicate a summary sensitivity of 85% (95% confidence interval [CI], 79%-90%) and specificity of 80% (95% CI, 73%-86%). The respective pooled negative predictive and positive predictive values were 97% (95% CI, 94%-98%%) and 40% (95% CI, 34 %-48% %). The pretest probability of a benign nodule was 82%. Only 3.7 percent of the false-negative nodules was a follicular thyroid carcinoma. A pooled negative predictive value of 99% (95% CI, 97%-100%) was found when only complete soft nodules (Asteria elastography 1) were classified as benign, which included 14% of the studied population.

Conclusions

Elastography has a fair specificity and sensitivity for diagnostic accuracy. Its major strength entails the detection of benignity, especially when only completely soft nodules are qualified as benign. The outcomes of our analysis show that FNA could safely be omitted in patients referred for analysis of their thyroid nodule when elastography shows it to be completely soft (Asteria elastography 1). This could prevent unnecessary invasive diagnostic procedures in a substantial portion of patients.

Introduction

Approximately 10 to 20 million Americans have clinically detectable thyroid nodules; however, only 4% to 7% of the nodules are malignant.1,2 Ultrasound (US) and fine-needle aspiration (FNA) are commonly used to evaluate the nature of thyroid nodules. Their diagnostic value in routine care is mitigated by several limitations.3,4 In particular, several studies indicated that US lacks accurate criteria for assessing whether a nodule is malignant.3,5 FNA, however, shows good sensitivities and specificities but is not representative in 0.7% to 15% of the cases6, requiring a diagnostic thyroid lobectomy frequently to obtain a final diagnosis. With the increasing detection of nodules in general, it would be of great value if an additional non-invasive tool can determine whether a nodule is benign. This would potentially lead to a reduction in FNAs, which benefits both the patient, as an invasive procedure can be omitted, and society because a cost reduction might be achieved.

Qualitative US elastography has recently been proposed as a new technique for evaluating the elasticity of nodules and identifying whether they are malignant or benign.[7] Its introduction was soon paced by several studies and meta-analyses reporting promising results. However, these studies compared studies with a mixed group of populations (i.e., patients referred for FNA and patients referred for surgery), color systems, and inclusion and exclusion criteria.[8-12] The role of elastography in cystic, calcified and follicular lesions is still a matter of debate. In a meta-analysis published in 2010 4 out of 9 (44%) included follicular carcinomas were false negative.[10] This may because the gross anatomy and cellular patterns of follicular carcinoma overlap with those of benign adenoma. On the other hand, cystic and calcified nodules could be diagnosed false positive, since these lesions are stiff.[13,14]

As suggested by others, the main strength of elastography might lay in diagnosing benignity and, by that, in selecting those that do not need further diagnostic evaluation.[15] This systematic review and meta-analysis focus, therefore, on patients referred for diagnostic evaluation of their thyroid nodule because elastography could prevent unnecessary invasive diagnostic procedures in this population.

The objective of this study is to investigate the diagnostic value of qualitative elastography in distinguishing benign from malignant thyroid nodules in patients referred for FNA. Subanalysis on cystic, calcified and follicular lesions were conducted.

Materials and Methods

Two authors (S.N. and T.O.) independently performed study selection, quality assessment and data extraction.

Search strategy

A systematic search was performed of the PubMed, Embase, and Cochrane Library databases. The search query was limited to December 5th, 2014, and included several synonym terms within the determinant (i.e., elastography) and outcome (i.e., malignancy; Supplemental 1).

Study selection

The relevance of all identified studies was assessed by title and abstract. The full-text paper was retrieved when studies evaluated thyroid malignancies, reported the original study (i.e., no reviews or case reports), and were written in English, German, French, or Dutch. Inclusion of studies was based on population characteristics (patients with a thyroid nodule referred for FNA), index test (qualitative elastography), reference test (thyroid nodule cytology or histology), and full-text publication. A cross-reference check was performed to assess the quality of the search.

Assessment of study quality

The risk of bias for all included studies was evaluated using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) questionnaire.[16] This tool comprises four domains: patient selection, index test, reference standard, and flow and timing. Each domain is assessed for the risk of bias. The first three domains are also assessed for concerns of applicability. The QUADAS-2 was tailored to our analysis, as described in the guideline.16 After tailoring, the tool was applied to a random sample from the available studies to evaluate inter-rater agreement. Finally, two authors (S.N. and T.O.) evaluated all included studies (Supplemental 2), and any disagreement was resolved through re-evaluation by a third author (B.K.) who was blinded to the outcome of the previous assessment.

Data extraction

Population characteristics and nodule classification systems were extracted. Because the included articles used several color classification systems and thresholds to distinguish between benign and malignant nodules, all nodules were classified according to the Asteria elastography (ES) classification.[17] This classification is applicable and widely used in thyroid nodule elastography.[18] Asteria ES is a 4-point scale, in which ES 1 is assigned to nodules with elasticity in the entire examined area, ES 2 is assigned to nodules with elasticity in a large portion of the examined area, ES 3 is assigned to nodules with stiffness in a large portion of the examined area, and ES 4 is assigned to hard nodules (Figure 1).[17] The 4-point Asteria classification was used to generate 4 × 2 tables.

A threshold between Asteria ES 2 and ES 3 was pre-specified and used during the statistical analysis to differentiate between potential malignant and benign nodules. As previously proposed by others, a second threshold analysis (between ES 1 and ES 2) was performed to test whether the diagnostic performance of elastography could be improved.[15]

All classification systems used were assigned independently by two authors (S.N. and J.K.). Subsequently, if needed, study data from included studies not using Asteria classification were reclassified into the Asteria classification. If reclassification was not possible due to the specific classification system used or to the lack of a detailed description of the classification system, the study was excluded from the analysis. Some studies reported data from two elstographic subgroups combined, and, therefore, not all reported data could be included in all analyses. Disagreement at any point of the reclassification process was solved by consulting a third author (B.K.) blinded for the results of the primary assessment.

Elastography subgroup analysis

Further subgroup analyses were planned for studies that (I) excluded calcified and cystic lesions, (II) only excluded cystic lesions, or (III) did not exclude calcified or cystic lesions. The aim of these subgroup analyses was to evaluate the influence of frequently used exclusion criteria on the diagnostic performance of elastography. Finally, an analysis was planned on the incidence of false-negative elastography results in case of follicular thyroid carcinoma (FTC). By this analysis the assumed underperformance of elastography of FTC nodules in literature was tested.

Statistical analysis

Sensitivities (true positive rates), false positive rates (equals 1-specificity), positive predictive values (PPV), negative predictive values (NPV), and corresponding 95% confidence intervals (CIs) of the studies were pooled using methods that preserved the 2-dimensional nature of the data and accounted for variability within and between studies.[19-21] In particular, a bivariate meta-analysis was performed using a Bayesian framework with noninformative prior distributions. The statistical models allowed for joint random effects on the sensitivity and specificity, posterior means were used to present summary measures of true positive rate, false positive rate, PPV, and NPV.

For each statistic, we calculated the 95% prediction interval (95% PI) to interpret the effect of statistical heterogeneity. This interval indicates the possible estimate (e.g., sensitivity) in an individual setting. The univariate I^2 statistics (e.g., $I^2_{E(sens)}$) were calculated to quantify the percentage of the variability in summary estimates that is due to heterogeneity rather than sampling error (chance).[22] We also calculated the bivariate I^2 statistic ($I^2_{E(Biv)}$) to present an overall measure of between-study

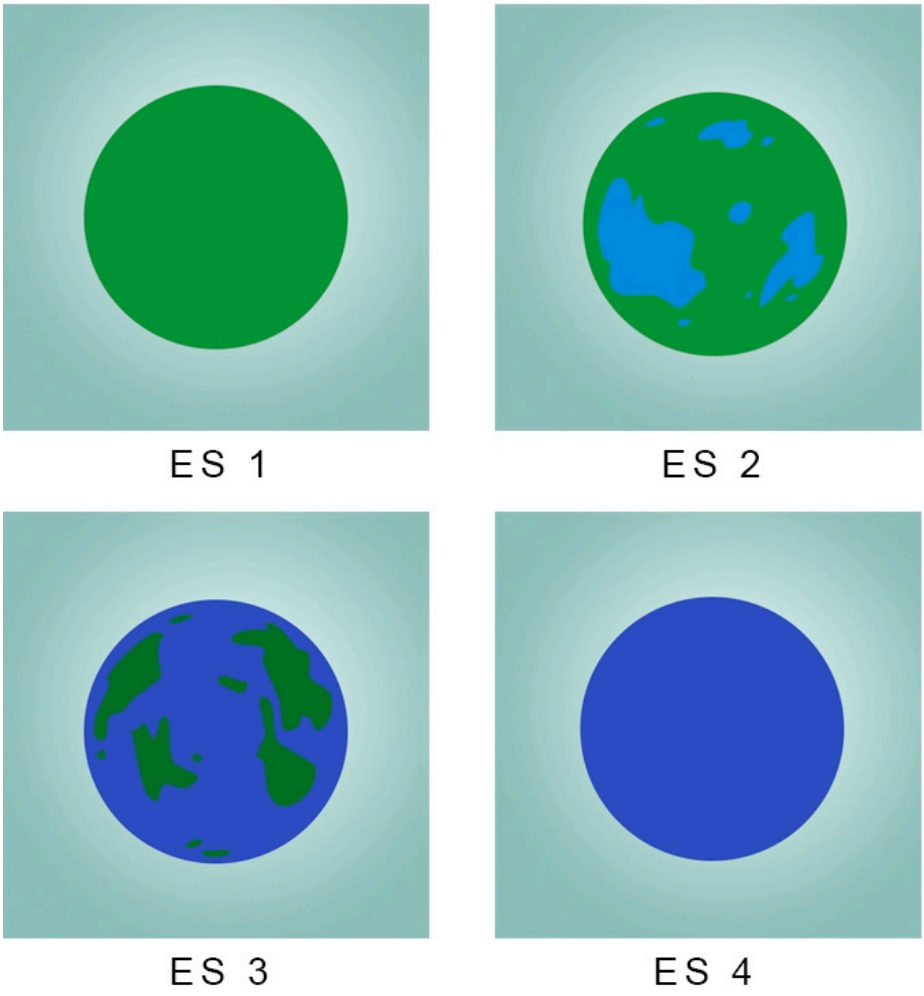


Figure 1: Thyroid nodule elastography (ES) scores according to Asteria classification.[17] Green indicates elastic tissue and blue indicate hard tissue. ES 1: The nodule is displayed homogeneously in green; ES 2: Almost the entire nodule in displayed in light green with some peripheral and/or central blue areas; ES 3: Almost the entire nodule is displayed in hard blue with some green regions; ES 4: The entire nodule is displayed in homogenously hard blue.

heterogeneity. The I^2 statistic may indicate the presence of unimportant (0%-40%), moderate (30%-60%), substantial (50%-90%), and considerable (75%-100%) amounts of between-study heterogeneity.[23] All analyses were performed in R and JAGS software. The corresponding BUGS code is available in Supplemental 3.

Results

Literature search

The search query resulted in 3921 citations. After filtering duplicates, 2461 citations were excluded based on title and abstract (Figure 2). The full text of 116 studies was retrieved for detailed review. From these, 96 studies were excluded: 27 studies evaluated patients referred for surgery before evaluation by elastography, 17 studies did not use qualitative elastography, 14 because quantitative elastography measurements, such as strain indexes, were used, 12 studies only published an abstract, 10 studies because the required data could not be extracted, 8 studies did not concern patients with a thyroid nodule, 3 because the population was restricted (e.g., only patients with a thyroid malignancy or with indeterminate FNA were examined), 2 studies included fewer than 15 patients, 2 studies were a review, and another study did not use the required golden standard.

Quality assessment

The QUADAS-2 tool was applied to the remaining 20 studies. The interobserver agreement rate was 0.59 (95% CI, 0.40-0.79). None of these studies showed a high risk of bias in patient selection, index test, or reference standard according to the QUADAS-2 tool (Figure 3, Supplemental 2). The risk of index test bias was unclear in six studies because they did not specifically indicate whether elastography was interpreted with or without knowledge of the results of the reference standard these studies had a prospective design, where elastography was performed before FNA. In 8 studies (40%), a high risk of flow and timing bias was observed because the interval between the elastography and reference standard was longer than 1 month or not all patients were included in the analysis. All studies had low concerns regarding applicability of patient selection, index test, and reference standard.

Characteristics of the included studies

Of the 3973 thyroid nodules that were evaluated (Table 1), 65 nodules were not suitable for reclassification and were excluded. The number of analyzed nodules in each study ranged from 23 to 912 nodules. The percentage of malignant nodules varied from 5% to 39%. The elastography color scale varied from a 3-point to a 5-point scale. All included studies used cytology or histology as a reference standard.

Pooled results and between-study heterogeneity

A total of 3908 nodules was analyzed (Table 1) of which 2424 (62%) were ES 1 or ES 2. The respective pooled sensitivity and specificity of elastography in differentiating between malignant and benign thyroid nodules were 85% (95% CI, 79%-90%) and 80% (95% CI, 73%-86%). Statistical analyses revealed that $I^2_{E(Biv)}$ was 53%, thereby indicating the presence of a moderate degree of between-study heterogeneity. The 95% PIs ranged from 61% to

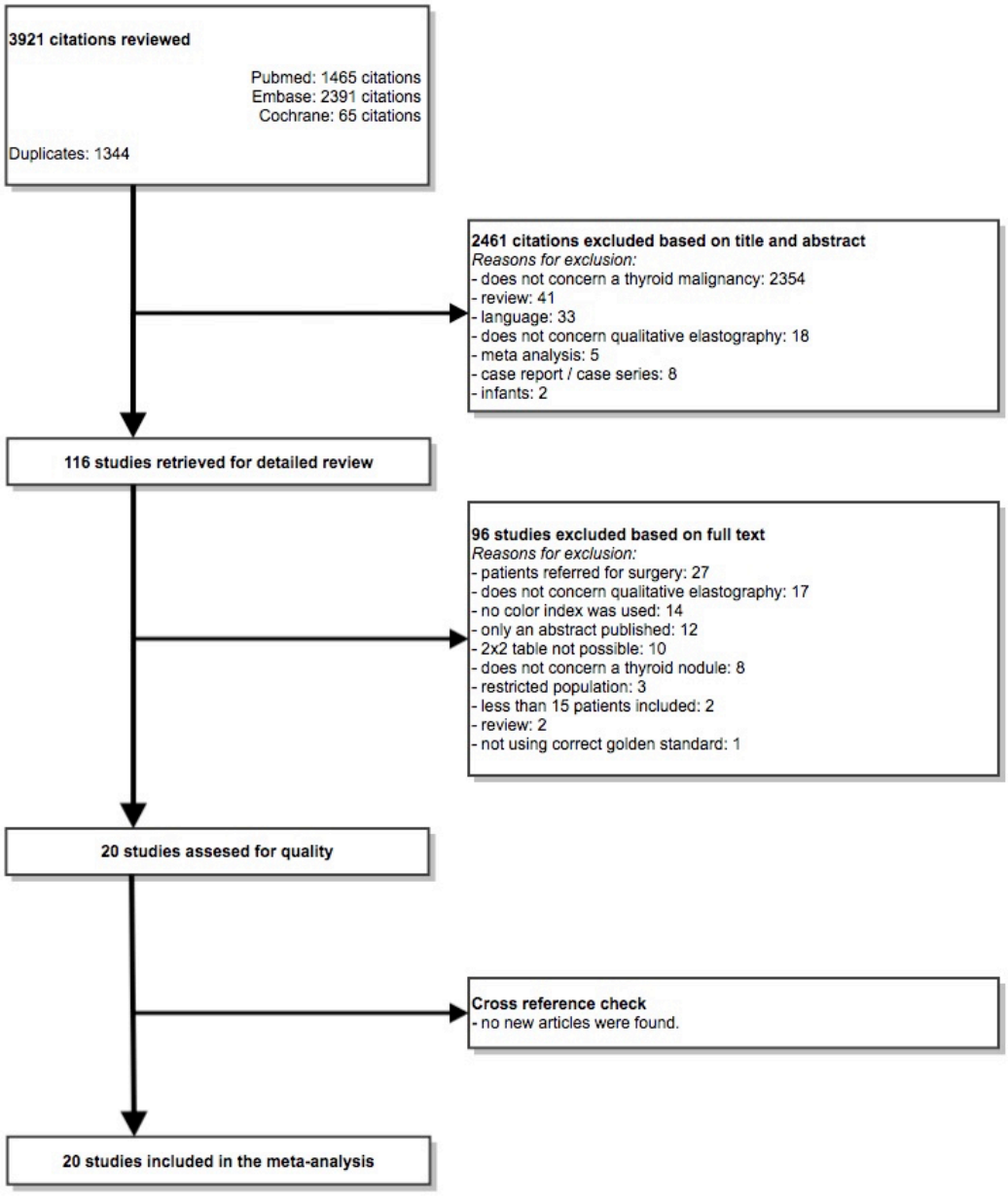


Figure 2: Flowchart of study selection

First author	Year	Country	Nodules analyzed, No.	Malignant nodules, %	Nodules with histology as reference§, %	Elastography technique (probe)	ES scale (used cutoff)	Reference
Guazzaron[52]	2014	Italy	386	13	17	Philips iU22 (5-12 MHz)	4 (2/3)	Cytology & Histology
Tatar[53]	2014	Turkey	150	6	6	Hitachi EUB 8500 (6-13MHz)	5 (3/4)	Cytology & Histology
Azizi[54]	2013	USA	912	9	18	Siemens Acuson S2000 (5.5- 18 MHz)	4 (2/3)	Cytology & Histology
Mehrota[15]	2013	UK	137	7	22	Hitachi EUB 7500HV (7-13 MHz)	3 (1/3)‡	Cytology & Histology
Bojunga[55]	2012	Germany	158	13	36	Hitachi EUB 900 (9MHz)	4 (2/3)	Cytology & Histology
Ciledag[24]	2012	Turkey	74	12	23	Toshiba Apollo (10MHz)	5 (3/4)†	Cytology & Histology
Friedrich-Rust[26]	2012	Germany	60	5	28	Hitachi EUB 900 (9MHz)	4 (2/3)	Cytology & Histology
Mansor[30]	2012	Egypt	45	9	unclear	Hitachi EUB 7500 (10MHz)	4(2/3)	Cytology & Histology
Moon[27]	2012	Korea	703	31	31	Hitachi EUB 7500 (6-14MHz)	4 (2/3)	Cytology & Histology
Trimboli[28]	2012	Italy	498	25	38	Hitachi Logos Hivison (6-14MHz)	4 (2/3)	Cytology & Histology
Unluturk[29]	2012	Turkey	237	24	30	Hitachi EUB 7000 (6-13MHz)	3 (1/2)	Cytology & Histology
Bhatia[13]	2011	China	85	28	36	Siemens AccusoneSie Touch (13.5MHz)	4 (2/3)	Cytology & Histology
Merino[36]	2011	Spain	106	10	17	Siemens Acuson S2000 (13.5MHz)	3 (1/3)‡	Cytology & Histology
Yerli[42]	2011	Turkey	72	14	33	Hitachi EUB 7000 (5-13MHz)	5 (3/4)	Cytology & Histology
Kagoya[58]	2010	Japan	47	23	19	Hitachi Logos EUB 7500 (6-14MHz)	4 (2/3)	Cytology & Histology
Friedrich-Rust[25]	2009	Germany	53	9	34	Hitachi EUB 900 (9MHz)	4 (2/3)	Cytology & Histology
Rubaltelli[37]	2009	Italy	51	22	33	Hitachi Logos Hivison (10MHz)	4 (2/3)	Cytology & Histology
Asteria[17]	2008	Italy	86	20	29	Hitachi EUB 8500 (6-13MHz)	4 (2/3)	Cytology & Histology
Ferrari[34]	2008	Italy	23	39	70	Hitachi EUB 8500 (13MHz)	4 (2/3)	Cytology & Histology
Tranquart[39]	2008	France	108	6	7	Hitachi Elite Doppler (13 MHz)	4 (2/3)	Cytology & Histology

Table 1: Study characteristics.
§, percentage of the analyzed thyroid nodules of which histology was used as reference standard; ES, elastography; UK, United Kingdom; USA, United States of America;
†, an elastography color threshold other than Asteria ES 2/3 was used for the initial statistical analysis in that particular study;
‡, the middlemost color score was not applicable to the Asteria classification.

96% (sensitivity) and from 41% to 96% (specificity). The corresponding I^2 statistics were 27% (95% CI, 14%-43%) and 84% (95% CI, 75%-90%). The respective pooled NPV and PPV were 97% (95%CI, 94%-98%) and 40% (95% CI, 34%-48%), with $I^2_{E(Biv)}$ at 49% (Figure 4). The pretest probability of a benign nodule was 82%.

Subgroup analysis of a different threshold

Fourteen studies reported sufficient information to compare the completely soft (ES 1) with the not completely soft nodules (ES >1). A total of 2012 nodules was analyzed, of which 277 (14%) were ES 1. The pooled sensitivity and specificity of these studies were 99% (95% CI, 96%-100%) and 14% (95 CI, 6%-30%) respectively. The bivariate I^2 statistic indicated a low degree of between-study heterogeneity ($I^2_{E(Biv)}$ = 24%), particularly for the pooled sensitivity ($I^2_{E(Sens)}$ = 5%). The corresponding 95% PIs were 88% to 100% and 0% to 86%. The respective pooled NPV and PPV were 99% (95% CI, 97%-100%; 95% PI, 93%-100%) and 16 % (95% CI, 12%-23%; 95% PI, 5%-45%) with $I^2_{E(Biv)}$ = 14% (Figure 5).

Subgroup analysis of elastography exclusion criteria

After detailed review of included studies we concluded that this subgroup analysis was not possible. Main reason is that the studies excluding cystic and/or calcified nodules all used different criteria: ranging from studies only including completely solid nodules to studies excluding only completely cystic nodules.[15,17,24-29] Similarly, in the four studies excluding calcifications different criteria were applied: eggshell calcifications, macrocalcifications or coarse calcifications.[24,25,28,29] Furthermore, because the populations of the studies that included cystic or calcified nodules consisted only for an unknown minority of these specific nodules, their influence on elastography outcome is not determinable.

Subgroup analysis of FTC nodules

In our cohort of 3908 included nodules, 157 were false negative when using the cut-off between ES 2 and ES 3. Histology was described of 106 of these nodules. Only four false-negative nodules were FTC's (3,7%). The other false-negative nodules consisted of 97 papillary thyroid carcinomas (PTC) (91,5%), three follicular variants of PTC (2,8%), one medullary thyroid carcinoma (0,9%) and one metastasis of another tumor to the thyroid (0,9%). No data was extractable from the included studies to determine the incidence of FTC's in the total population of nodules.

Discussion

Qualitative elastography has been described as a promising new technique in differentiating benign from malignant thyroid nodules.30-42 The aim of this systematic review and meta-analysis was to investigate the diagnostic value of qualitative elastography in distinguishing benign from malignant thyroid nodules in patients referred for FNA and, thereby, potentially prevent unnecessary invasive diagnostic procedures.

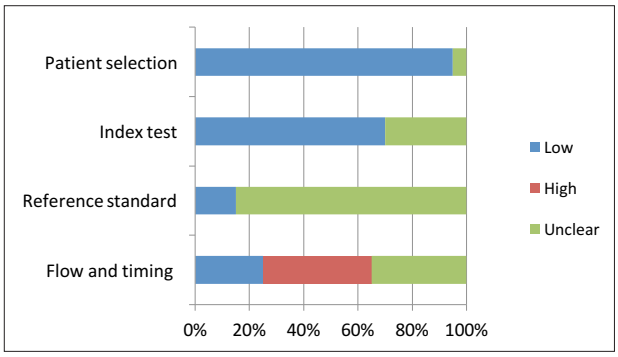
We found 20 studies including 3973 thyroid nodules. By using the standard cutoff (ES 1-2 vs. ES 3-4), we found that thyroid nodule elastography in patients referred for FNA had a pooled sensitivity of 85% and specificity of 90%. The NPV and PPV were 97% and 40%, respectively. An even higher NPV of 99% (95% CI, 97%-100%) was observed when only completely soft nodules were considered as benign (i.e., cutoff between ES 1 and 2). This implies that 99% of patients with a completely soft nodule indeed do have a benign nodule and that further diagnostic procedures can be safely omitted, which means 14% of this study population. Even when accounting for potential between-study heterogeneity, the NPV in a new study is still likely to exceed 93%.

We believe both cutoffs can be used: a cutoff between ES 2 and 3 could lead to a reduction of 62% of the FNAs in this study at the expense of a higher, but still low, number of missed malignancies. Elastography can thus prevent invasive diagnostic FNAs and thyroid lobectomies with associated complications and costs[43], without substantial risk of missing malignant nodules. However, because the PPV is only 40%, a positive elastography result is equivocal and further diagnostic evaluation by FNA, and, if FNA is indifferent or inconclusive diagnostic lobectomy, is unavoidable.

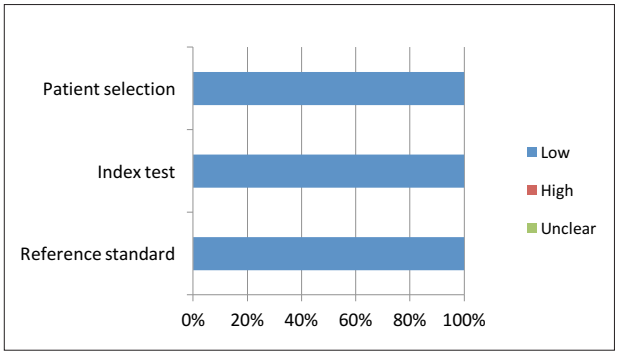
Due to different cystic and/or calcified excluding criteria between studies it was not possible to determine whether exclusion of cystic or calcified nodules improves diagnostic accuracy of elastography. A small study analyzing 85 nodules advocated excluding cystic lesions.[13] This recommendation has been widely used in subsequent studies but has not been studied more thoroughly.[24,27,29] The influence of cystic and calcified lesions has been doubted as well. As this specific question could not be answered by pooling the current available data, more research on this subject is advocated.

Another debate is whether FTCs are evenly well diagnosed with elastography in comparison with other histological subtypes, since 4 out of 9 FTC's (44%) in the meta-analysis of Bojunga et al. were false negative.[10,44] Although we were unable to determine the number of true-positive FTC's, we believe this hypothesis can be refuted. As the overall incidence of FTC's in thyroid carcinomas is around 10% and the false negative elastographies consisted of only 3,7% FTC's we conclude that elastography is diagnosing FTCs as good as any other thyroid malignancy.[45]

Inter- and intraobserver variability is a potential disadvantage of



Proportion of studies with low, high, or unclear risk of bias, %



Proportion of studies with low, high, or unclear concerns regarding the applicability, %

Figure 3: The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool for analysis of included study quality. The risk of bias for the included studies and the level of concern regarding the applicability of included studies are shown.

qualitative elastography, since the observer has to classify a nodule using a qualitative color score. Especially the difference between an ES 2 and ES 3 score could be challenging in some nodules. The difference between elasticity and stiffness in a large portion of the examined area could be small. This hypothesis was confirmed by Park et al. who found no statistical interobserver concordance among 45 thyroid nodules.[46] However, three studies analyzing 106 to 152 nodules, found good inter and intraobserver variabilities, with kappa rates >0.79.[32,36,47] We expect further improvement of the inter- and intraobserver variability using an ES cutoff between ES 1 and ES 2, but this need to be investigated in new prospective studies.

The aim of our meta-analysis was to determine if thyroid nodule elastography could replace fine needle aspirations in specific patients. Previously published meta-analyses focused on the diagnostic value of thyroid nodule elastography in a broad sense. Bojunga, Sun and Veer et al. analyzed pooled data of difference elastography techniques, like qualitative

elastography, strain index elastography and shear wave elastography.[8-10] Ghajarzadeh and Razavi specifically analyzed qualitative elastography but pooled data of patients referred for surgery and patients referred for fine needle aspiration (FNA).[11,12] An increased number of malignant nodules could be expected in patients referred for surgery compared to patients referred for FNA due to selection bias, which makes the pooled data difficult to use in clinical practice. In a subanalysis done by Sun et al. this influence is confirmed, as the sensitivity and specificity of studies only including patients referred for surgery are 90% and 81% respectively, compared to 71 and 76% in patients referred for FNA.[8]

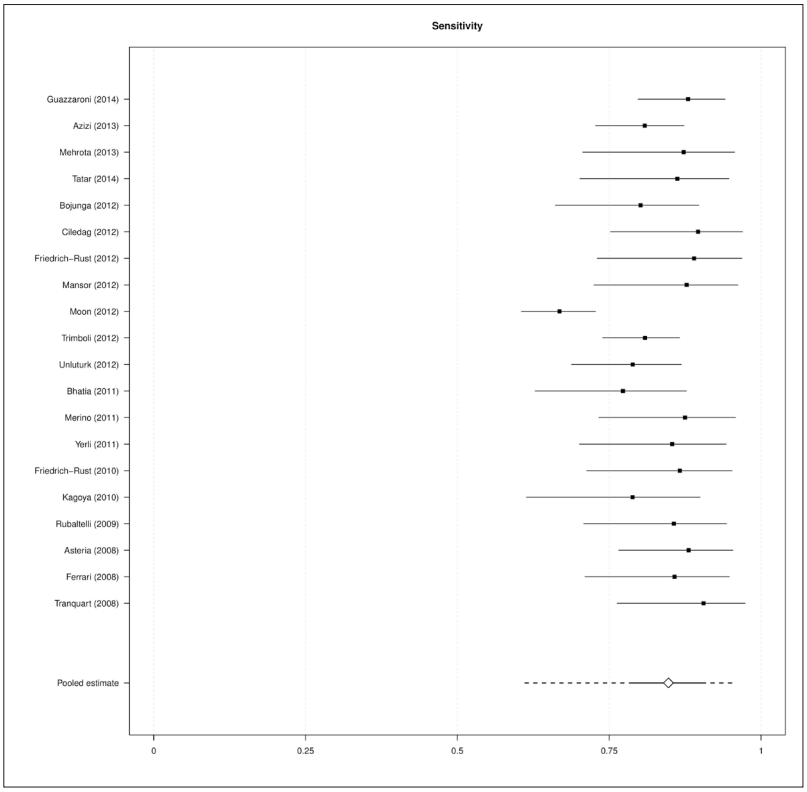
The previously published meta-analyses found a high NPV and negative likelihood ratio's, but these values were a reflection of heterogeneous pooled data and, therefore, difficult to interpret and apply. Furthermore, the published previous meta-analyses had shortcomings. For instance, Veer et al. did not assess the quality of the included studies, did not test statistical heterogeneity and did not describe his mathematical model. The study of Veer et al. is the only meta-analyses displaying his findings in NPV's instead of negative likelihood ratios, but, unfortunately, the brief description of the used methods makes it impossible to reproduce his outcomes.[9] Furthermore, all of the previously published meta-analyses missed several papers. For example, the publication of Ferrari et al. was missed by all except of Veer et al., while Veer et al., however, missed Trimboli's paper of 498 nodules.[8-12,28,34]

Several similar techniques to elastography have been evaluated in the diagnostic work-up of thyroid nodules. Semi-quantitative elastography (SE), strain ratio, compares nodule elasticity to surrounding tissue (e.g., normal thyroid tissue or muscle). A similar method uses carotid artery pulsations as its reference. Shear-wave elastography (SWE) is using acoustic pressure in order to assess elasticity and provides quantitative information on the stiffness of the nodule. This technique is said to be less operator dependent and more consistent.[44] In the meta-analysis of Veer et al. an analysis of SWE has been done showing a sensitivity, specificity, PPV and NPV of 86, 89, 60 and 97 percent, respectively.[9] Compared to our results the sensitivity and specificity improved slightly, the PPV significantly, and the NPV was similar. Razavi et al. included an analysis of different strain ratios.[11] Results showed overall similar results with overlapping CI's. However, the ratio between carotid artery and nodule seemed superior to elastography, although only two studies used this method.[11,48,49] Recently, a preliminary study directly comparing SWE with qualitative elastography showed no difference in diagnostic performance.[50] However, as this study consisted of patients referred for surgery, the authors state that larger studies in non-surgical populations are warranted. Concluding, there are suggestions in literature that SWE and SE might be alternatives with comparable or better diagnostic performance. However, further high-quality studies comparing these techniques with qualitative elastography are warranted and the results of any comparison at this point should be interpreted with care.

The strength of this study, with respect to other meta-analyses on this subject, is the extensive systematic literature search, strict population and diagnostic test selection, and the use of state-of-the-art methods to perform a meta-analysis that appropriately deals with the bivariate nature of diagnostic test accuracy data.[19-21] Despite the harmonization of study populations, inclusion and exclusion criteria, and types of elastography tests, the presence of between-study heterogeneity remains a limitation.
In particular, we observed a substantial difference in the prevalence of patients undergoing surgery in the included studies. This difference may be attributed to a variation in-patient flow (e.g. due to different protocols) between hospitals and countries and thereby affect the predictive performance of elastography in individual settings. Moreover, other factors such as different elastography devices and local experience may also play a role in the origin of between-study heterogeneity. We have addressed this heterogeneity to a large extent by performing subgroup analyses where only completely soft nodules were qualified as benign. Here, we found that the NPV of elastography remained consistently high across different study populations ($I^2_{(NPV)}$ = 3%). Fortunately, we found that the NPV of elastography remained fairly high even when such heterogeneity was taken into account. It is important to emphasize that our results are only applicable to patients that would have been referred for FNA and not all thyroid nodules widely prevalent in the general population. Finally, it is possible that publication bias occurred although this is less likely in studies of diagnostic test accuracy.[51]

Conclusions

This comprehensive meta-analysis demonstrates that qualitative elastography has fair specificity and sensitivity for diagnostic accuracy for patients with a thyroid nodule referred for FNA. Its major strength entails the detection of benign nodules, with a NPV of 99% for nodules that are completely soft (Asteria ES 1) and a NPV of 97% when predominantly soft nodules are also considered benign (Asteria ES 2). The outcome of our analysis shows that FNA could safely be omitted when elastography shows Asteria ES 1. Consequently, 14% of patients with a thyroid nodule referred for FNA could be saved invasive diagnostic procedures. Even more procedures could be saved when Asteria ES1 and ES 2 are considered to be benign. However, this increases the chance of missing malignancies. Further diagnostic evaluation is recommended when elastography is positive, because in that case no reliable conclusions about the nature of the nodule can be drawn. As current literature consists of mainly small single center studies, we believe that validation of our findings is warranted in powered prospective (multicenter) studies and in comparison with semi-quantitative elastography and/or shear-wave elastography.



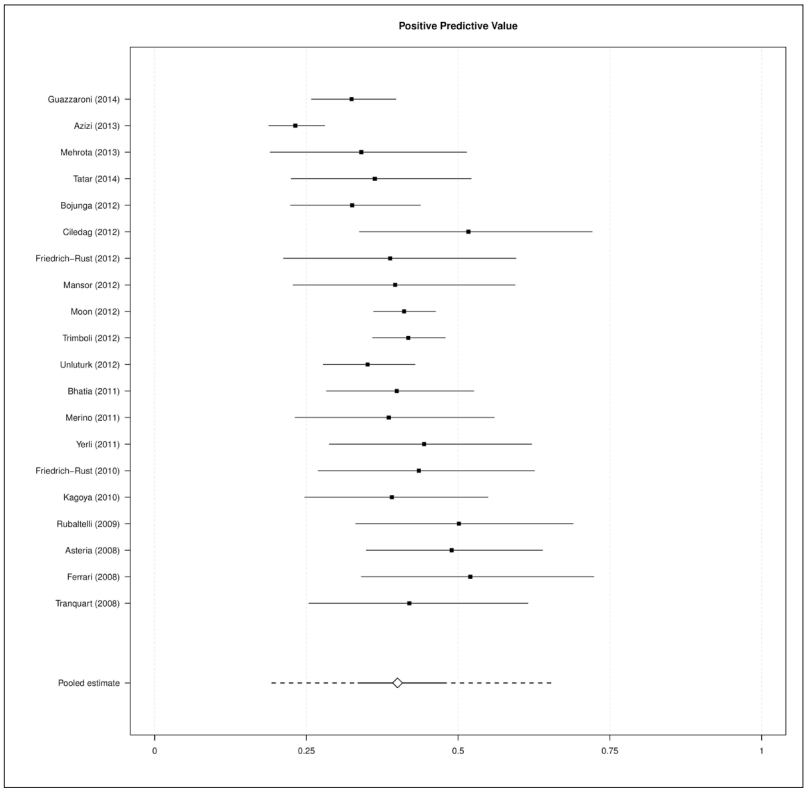
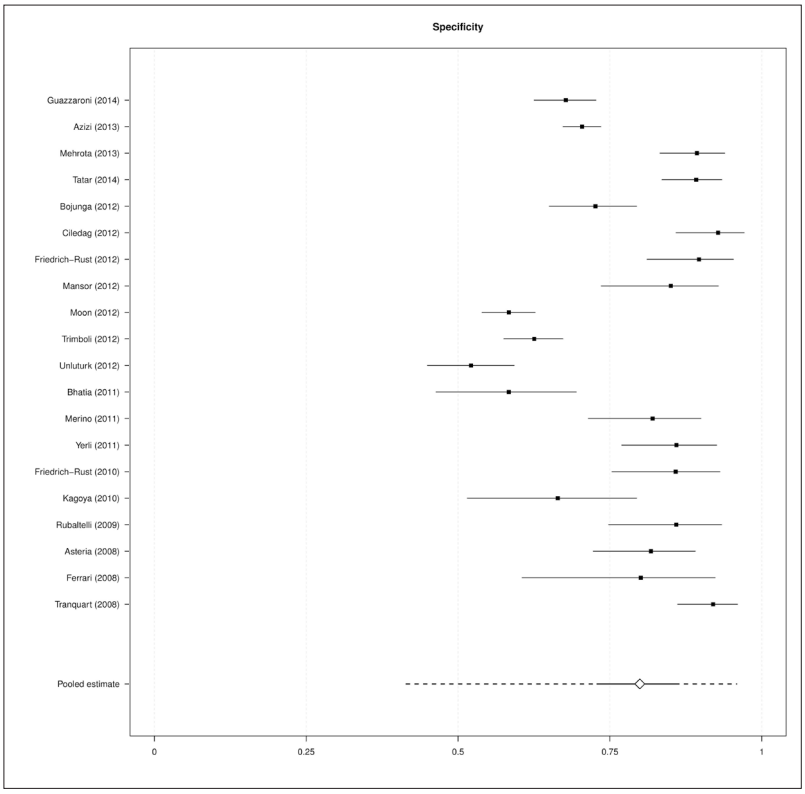
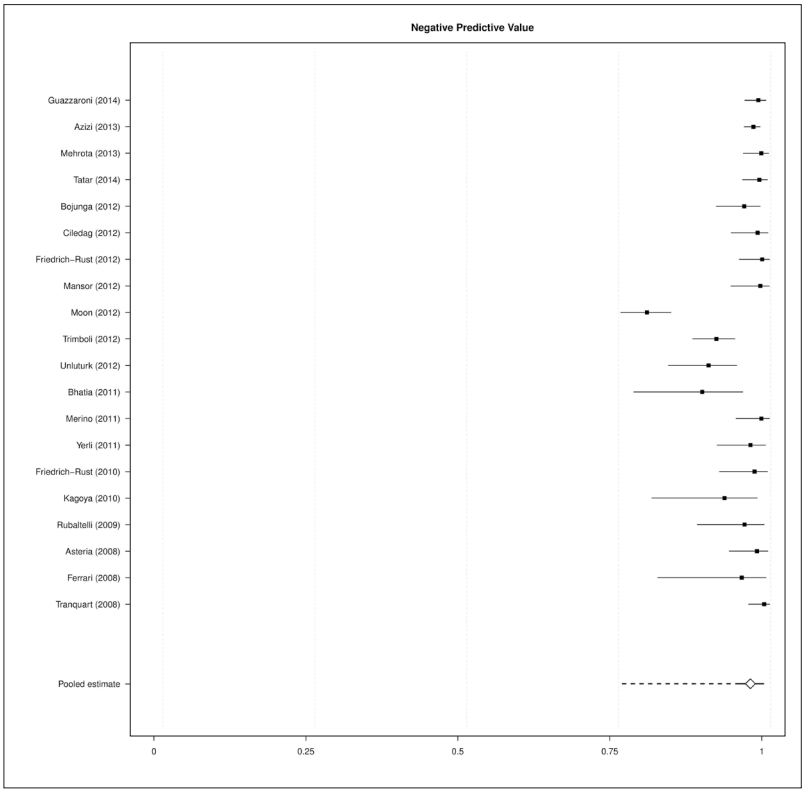
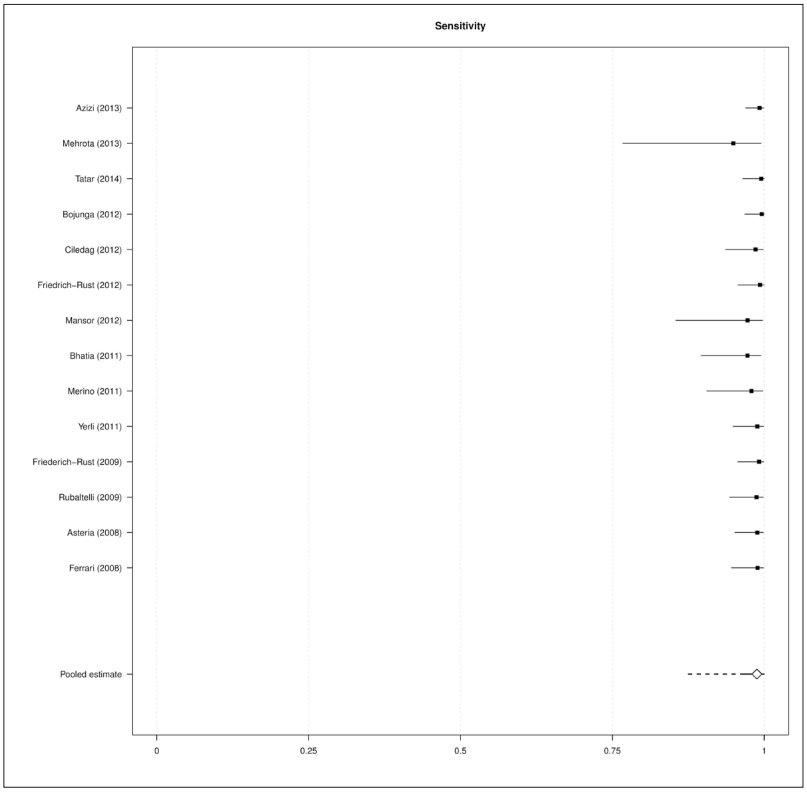


Figure 4: Forest plot and pooled estimate of ES 1,2 versus ES 3,4. Diamond, pooled estimate; unbroken line; 95% confidence interval; dotted line, 95% prediction interval. I^2 statistics: 53% (bivariate), 27% (Sensitivity), 84% (Specificity), 49% (bivariate), 50% (PPV), 50% (NPV)


Specificity
Guazzaroni (2014)
Azizi (2013)
Mehrota (2013)
Tatar (2014)
Bojunga (2012)
Ciledag (2012)
Friedrich−Rust (2012)
Mansor (2012)
Moon (2012)
Trimboli (2012)
Unluturk (2012)
Bhatia (2011)
Merino (2011)
Yerli (2011)
Friedrich−Rust (2010)
Kagoya (2010)
Rubaltelli (2009)
Asteria (2008)
Ferrari (2008)
Tranquart (2008)
Pooled estimate
0
0.25
0.5
0.75
1


Negative Predictive Value
Guazzaroni (2014)
Azizi (2013)
Mehrota (2013)
Tatar (2014)
Bojunga (2012)
Ciledag (2012)
Friedrich−Rust (2012)
Mansor (2012)
Moon (2012)
Trimboli (2012)
Unluturk (2012)
Bhatia (2011)
Merino (2011)
Yerli (2011)
Friedrich−Rust (2010)
Kagoya (2010)
Rubaltelli (2009)
Asteria (2008)
Ferrari (2008)
Tranquart (2008)
Pooled estimate
0
0.25
0.5
0.75
1



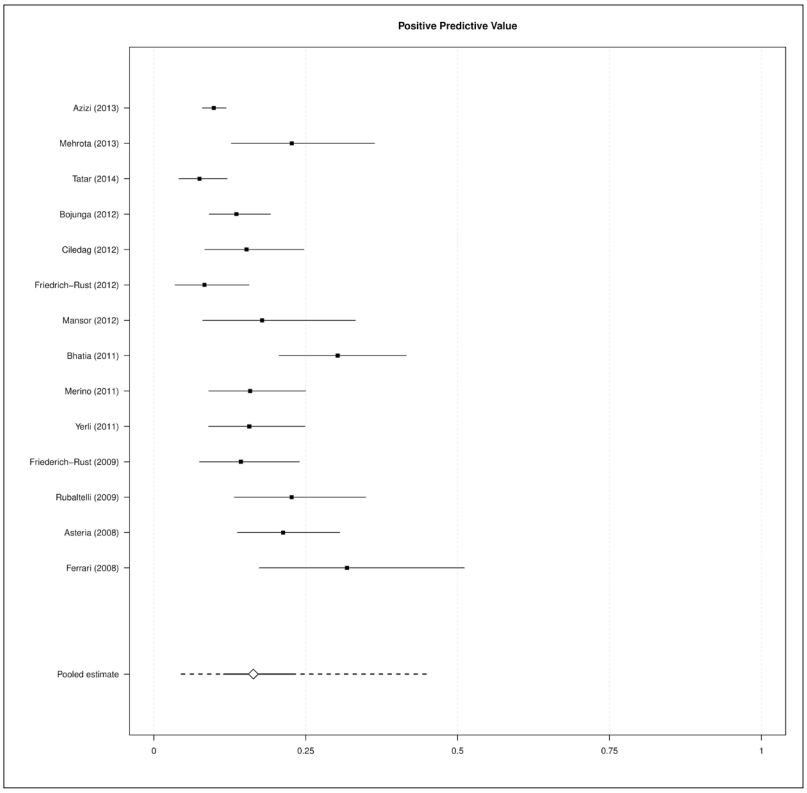
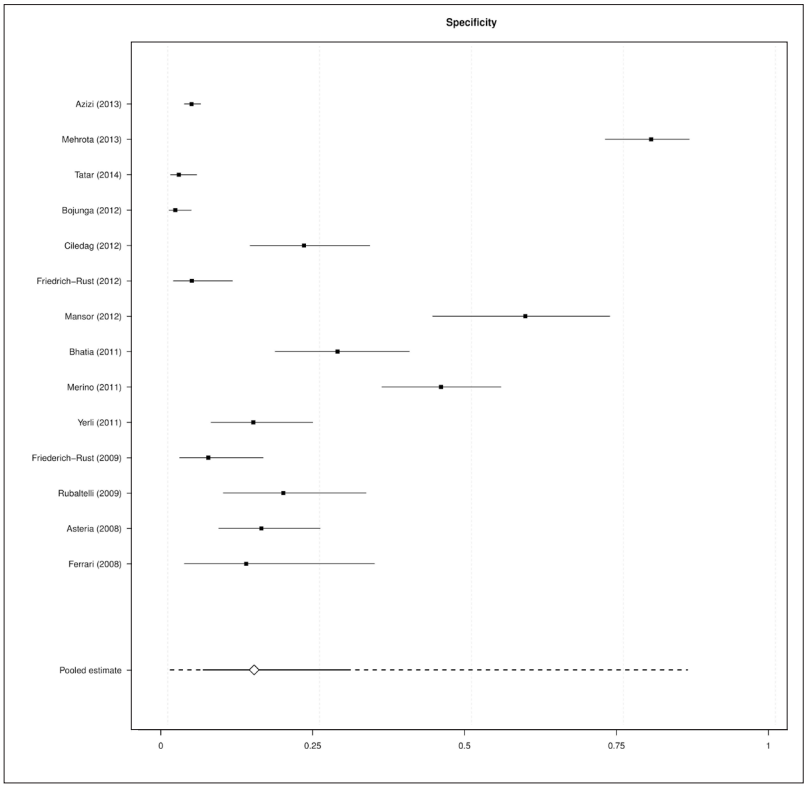
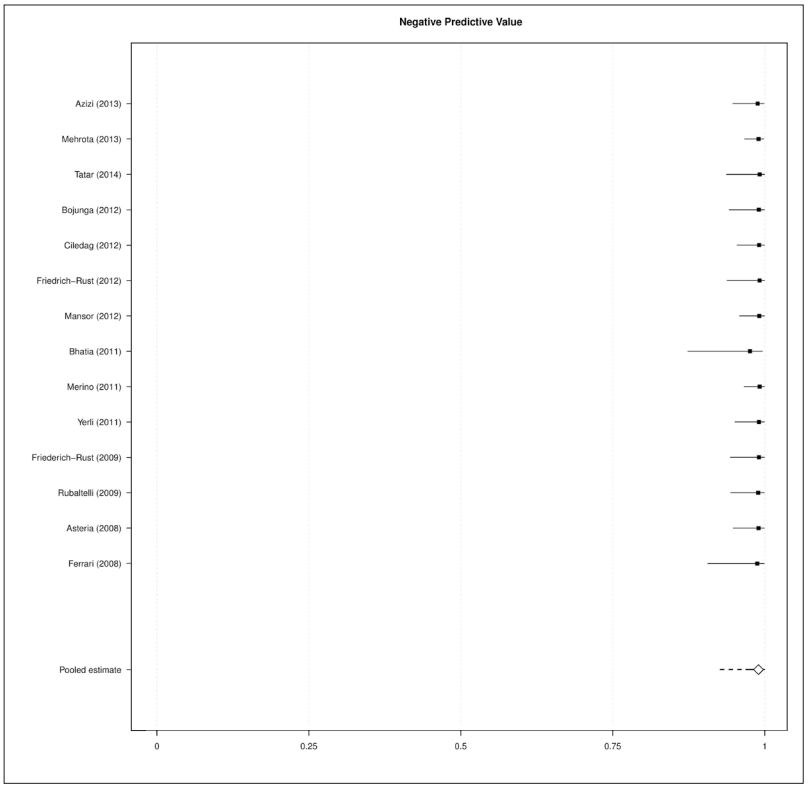


Figure 5: Forest plot and pooled estimate of completely soft nodules (ES1) compared to harder nodules (ES>1). Diamond, pooled estimate; unbroken line; 95% confidence interval; dotted line, 95% prediction interval. I2 statistics: 24% (bivariate), 5% (Sensitivity), 84% (Specificity), 14% (bivariate), 70% (PPV), 2% (NPV)


Specificity
Azizi (2013)
Mehrota (2013)
Tatar (2014)
Bojunga (2012)
Ciledag (2012)
Friedrich−Rust (2012)
Mansor (2012)
Bhatia (2011)
Merino (2011)
Yerli (2011)
Friederich−Rust (2009)
Rubaltelli (2009)
Asteria (2008)
Ferrari (2008)
Pooled estimate
0
0.25
0.5
0.75
1


Negative Predictive Value
Azizi (2013)
Mehrota (2013)
Tatar (2014)
Bojunga (2012)
Ciledag (2012)
Friedrich−Rust (2012)
Mansor (2012)
Bhatia (2011)
Merino (2011)
Yerli (2011)
Friederich−Rust (2009)
Rubaltelli (2009)
Asteria (2008)
Ferrari (2008)
Pooled estimate
0
0.25
0.5
0.75
1

References

1. Cramer H. Fine needle aspiration cytology of the thyroid. Cancer Cytopathology 2000;90(6):325–9.

2. Papini E, Guglielmi R, Bianchini A, et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. J Clin Endocrinol Metab 2002;87(5):1941–6.

3. Iannuccilli JD, Cronan JJ, Monchik JM. Risk for malignancy of thyroid nodules as assessed by sonographic criteria: the need for biopsy. J Ultrasound Med 2004;23(11):1455–64.

4. Gharib H, Papini E, Paschke R, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. Endocr Pract 2010;16(3):468–75.

5. Hoang JK, Lee WK, Lee M, Johnson D, Farrell S. US Features of thyroid malignancy: pearls and pitfalls. Radiographics 2007;27(3):847–60–discussion861–5.

6. Yang GC, Liebeskind D, Messina AV. Ultrasound-guided fine-needle aspiration of the thyroid assessed by Ultrafast Papanicolaou stain: data from 1135 biopsies with a two- to six-year follow-up. Thyroid 2001;11(6):581–9.

7. Lyshchik A, Higashi T, Asato R, et al. Thyroid gland tumor diagnosis at US elastography. Radiology 2005;237(1):202–11.

8. Sun J, Cai J, Wang X. Real-time ultrasound elastography for differentiation of benign and malignant thyroid nodules: a meta-analysis. J Ultrasound Med 2014;33(3):495–502.

9. Veer V, Puttagunta S. The role of elastography in evaluating thyroid nodules: a literature review and meta-analysis. Eur Arch Otorhinolaryngol 2014;

10. Bojunga J, Herrmann E, Meyer G, Weber S, Zeuzem S, Friedrich-Rust M. Real-Time Elastography for the Differentiation of Benign and Malignant Thyroid Nodules: A Meta-Analysis. Thyroid 2010;20(10):1145–50.

11. Razavi SA, Hadduck TA, Sadigh G, Dwamena BA. Comparative effectiveness of elastographic and B-mode ultrasound criteria for diagnostic discrimination of thyroid nodules: a meta-analysis. AJR Am J Roentgenol 2013;200(6):1317–26.

12. Ghajarzadeh M, Sodagari F, Shakiba M. Diagnostic accuracy of sonoelastography in detecting malignant thyroid nodules: a systematic review and meta-analysis. AJR Am J Roentgenol 2014;202(4):W379–89.

13. Bhatia KSS, Rasalkar DP, Lee YP, et al. Cystic change in thyroid nodules: a confounding factor for real-time qualitative thyroid ultrasound elastography. Clin Radiol 2011;66(9):799–807.

14. Rago T, Santini F, Scutari M, Pinchera A, Vitti P. Elastography: new developments in ultrasound for predicting malignancy in thyroid nodules. J Clin Endocrinol Metab 2007;92(8):2917–22.

15. Mehrotra P, McQueen A, Kolla S, Johnson SJ, Richardson DL. Does elastography reduce the need for thyroid FNAs? Clin Endocrinol (Oxf) 2013;78(6):942–9.

16. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529–36.

17. Asteria C, Giovanardi A, Pizzocaro A, et al. US-elastography in the differential diagnosis of benign and malignant thyroid nodules. Thyroid 2008;18(5):523–31.

18. Itoh A, Ueno E, Tohno E, et al. Breast disease: clinical application of US elastography for diagnosis. Radiology 2006;239(2):341–50.

19. Verde PE. Meta-analysis of diagnostic test data: a bivariate Bayesian modeling approach. Stat Med 2010;29(30):3088–102.

20. Leeflang MMG, Deeks JJ, Rutjes AWS, Reitsma JB, Bossuyt PMM. Bivariate meta-analysis of predictive values of diagnostic tests can be an alternative to bivariate meta-analysis of sensitivity and specificity. J Clin Epidemiol 2012;65(10):1088–97.

21. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58(10):982–90.

22. Zhou Y, Dendukuri N. Statistics for quantifying heterogeneity in univariate and bivariate meta-analyses of binary data: the case of meta-analyses of diagnostic accuracy. Stat Med 2014;33(16):2701–17.

23. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557–60.

24. Ciledag N, Arda K, Arıbas BK, Aktas E, Köse SK. The Utility of Ultrasound Elastography and MicroPure Imaging in the Differentiation of Benign and Malignant Thyroid Nodules. AJR Am J Roentgenol 2012;198(3):W244–9.

25. Friedrich-Rust M, Sperber A, Holzer K, et al. Real-time Elastography and Contrast-Enhanced Ultrasound for the Assessment of Thyroid Nodules. Exp Clin Endocrinol Diabetes 2009;118(09):602–9.

26. Friedrich-Rust M, Romenski O, Meyer G, et al. Acoustic Radiation Force Impulse-Imaging for the evaluation of the thyroid gland: a limited patient feasibility study. Ultrasonics 2012;52(1):69–74.

27. Moon HJ, Sung JM, Kim EK, Yoon JH, Youk JH, Kwak JY. Diagnostic Performance of Gray-Scale US and Elastography in Solid Thyroid Nodules. Radiology 2012;262(3):1002–13.

28. Trimboli P, Guglielmi R, Monti S, et al. Ultrasound sensitivity for thyroid malignancy is increased by real-time elastography: a prospective multicenter study. J Clin Endocrinol Metab 2012;97(12):4524–30.

29. Unlütürk U, Erdo an MF, Demir O, Güllü S, Ba kal N. Ultrasound Elastography Is Not Superior to Grayscale Ultrasound in Predicting Malignancy in Thyroid Nodules. Thyroid 2012;

30. Mansor M, Okasha H, Esmat S, Hashem AM, Attia KA, El-din Hussein H. Role of ultrasound elastography in prediction of malignancy in thyroid nodules. Endocr Res 2012;37(2):67–77.

31. Nacamulli D, Nico L, Barollo S, et al. Comparison of the diagnostic accuracy of combined elastosonography and BRAF analysis vs cytology and ultrasonography for thyroid nodule suspected of malignancy. Clin Endocrinol (Oxf) 2012;77(4):608–14.

32. Ragazzoni F, Deandrea M, Mormile A, et al. High diagnostic accuracy and interobserver reliability of real-time elastography in the evaluation of thyroid nodules. Ultrasound Med Biol 2012;38(7):1154–62.

33. Stoian D, Cornianuz M, Dobrescu A, Lazar F. Nodular thyroid cancer. Diagnostic value of real time elastography. Chirurgia (Bucur) 2012;107(1):39–46.

34. Ferrari FS, Megliola A, Scorzelli A, Guarino E, Pacini F. Ultrasound examination using contrast agent and elastosonography in the evaluation of single thyroid nodules: Preliminary results. Journal of Ultrasound 2008;11(2):47–54.

35. Hong Y, Liu X, Li Z, Zhang X, Chen M, Luo Z. Real-time ultrasound elastography in the differential diagnosis of benign and malignant thyroid nodules. J Ultrasound Med 2009;28(7):861–7.

36. Merino S, Arrazola J, Cardenas A, et al. Utility and interobserver agreement of ultrasound elastography in the detection of malignant thyroid nodules in clinical care. AJNR Am J Neuroradiol 2011;32(11):2142–8.

37. Rubaltelli L, Corradin S, Dorigo A, et al. Differential Diagnosis of Benign and Malignant Thyroid Nodules at Elastosonography. Ultraschall in Med 2008;30(02):175–9.

38. Shuzhen C. Comparison analysis between conventional ultrasonography and ultrasound elastography of thyroid nodules. European Journal of Radiology 2012;81(8):1806–11.

39. Tranquart F, Bleuzen A, Pierre-Renoult P, Chabrolle C, Sam Giao M, Lecomte P. [Elastosonography of thyroid lesions]. J Radiol 2008;89(1 Pt 1):35–9.

40. Wang A. The investigation of conventional ultrasound and uselastosonography for thyroid nodules. J Med Imaging Radiat Sci 2012;43(1):S51.

41. Wang Y, Dan H-J, Dan H-Y, Li T, Hu B. Differential diagnosis of small single solid thyroid nodules using real-time ultrasound elastography. J Int Med Res 2010;38(2):466–72.
42. Yerli H, Yilmaz T, Oztop I. Clinical importance of diastolic sonoelastographic scoring in the management of thyroid nodules. AJNR Am J Neuroradiol 2013;34(3):E27–30.
43. 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(5):763–71.
44. Cantisani V, Lodise P, Grazhdani H, et al. Ultrasound elastography in the evaluation of thyroid pathology. Current status. Eur J Radiol 2014;83(3):420–8.
45. Davies L, Welch HG. Increasing incidence of thyroid cancer in the United States, 1973-2002. JAMA 2006;295(18):2164–7.
46. Park SH, Kim SJ, Kim E-K, Kim MJ, Son EJ, Kwak JY. Interobserver Agreement in Assessing the Sonographic and Elastographic Features of Malignant Thyroid Nodules. AJR Am J Roentgenol 2009;193(5):W416–23.
47. Friedrich-Rust M, Meyer G, Dauth N, et al. Interobserver agreement of Thyroid Imaging Reporting and Data System (TIRADS) and strain elastography for the assessment of thyroid nodules. PLoS ONE 2013;8(10):e77927.
48. Dighe M, Kim J, Luo S, Kim Y. Utility of the ultrasound elastographic systolic thyroid stiffness index in reducing fine-needle aspirations. J Ultrasound Med 2010;29(4):565–74.
49. Dighe M, Bae U, Richardson ML, Dubinsky TJ, Minoshima S, Kim Y. Differential diagnosis of thyroid nodules with US elastography using carotid artery pulsation. Radiology 2008;248(2):662–9.
50. Liu B-X, Xie X-Y, Liang J-Y, et al. Shear wave elastography versus real-time elastography on evaluation thyroid nodules: a preliminary study. Eur J Radiol 2014;83(7):1135–43.
51. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58(9):882–93.
52. Guazzaroni M, Spinelli A, Coco I, Del Giudice C, Girardi V, Simonetti G. Value of strain-ratio on thyroid real-time sonoelastography. Radiol Med 2014;119(3):149–55.
53. Tatar IG, Kurt A, Yilmaz KB, Do an M, Hekimoglu B, Hucumenoglu S. The role of elastosonography, gray-scale and colour flow Doppler sonography in prediction of malignancy in thyroid nodules. Radiol Oncol 2014;48(4):348–53.
54. Azizi G, Keller J, Lewis M, Puett D, Rivenbark K, Malchoff C. Performance of Elastography for the Evaluation of Thyroid Nodules: A Prospective Study. Thyroid 2013;23(6):734–40.
55. Bojunga J, Dauth N, Berner C, et al. Acoustic radiation force impulse imaging for differentiation of thyroid nodules. PLoS ONE 2012;7(8):e42735.
56. Kagoya R, Monobe H, Tojima H. Utility of elastography for differential diagnosis of benign and malignant thyroid nodules. Otolaryngol Head Neck Surg 2010;143(2):230–4.

Supplementary data

(((Cancer[Title/Abstract])) OR (cancers[Title/Abstract])) OR (carcinoma[Title/Abstract])) OR (carcinomas[Title/Abstract])) OR (cancerous[Title/Abstract])) OR (neoplasm[Title/Abstract])) OR (neoplasms[Title/Abstract])) OR (neoplasia[Title/Abstract])) OR (neoplastic[Title/Abstract])) OR (oncology[Title/Abstract])) OR (oncologic[Title/Abstract])) OR (oncological[Title/Abstract])) OR (tumor[Title/Abstract])) OR (tumors[Title/Abstract])) OR (tumour[Title/Abstract])) OR (tumours[Title/Abstract])) OR (tumourous[Title/Abstract])) OR (malignant[Title/Abstract])) OR (malignancy[Title/Abstract])) OR (malignancies[Title/Abstract])) OR (microcarcinoma[Title/Abstract])) OR (microcarcinomas[Title/Abstract])) OR (benign neoplasm[Title/Abstract])) OR (benign neoplasms[Title/Abstract])) OR (adenoma[Title/Abstract])) OR (adenomas[Title/Abstract])) OR (follicular[Title/Abstract])) OR (papillary[Title/Abstract])) OR (anaplastic[Title/Abstract])) OR (Hurtle cell[Title/Abstract])) OR (tall cell[Title/Abstract])) OR (pathology[Title/Abstract])) OR (pathological[Title/Abstract])) OR (pathologic[Title/Abstract])) OR (histology[Title/Abstract])) OR (histologic[Title/Abstract])) OR (histological[Title/Abstract])) OR (histopathologic[Title/Abstract])) OR (histopathology[Title/Abstract])) OR (cytology[Title/Abstract])) OR (cytologic[Title/Abstract])) OR (cytological[Title/Abstract])) OR (medullary[Title/Abstract]))) AND (((((((((((((((Elastography[Title/Abstract])) OR (elastographies[Title/Abstract])) OR (elastosonography[Title/Abstract])) OR (elastograph[Title/Abstract])) OR (elastographic[Title/Abstract])) OR (elastosonographic[Title/Abstract])) OR (elasto[Title/Abstract])) OR (elastogram[Title/Abstract])) OR (elastograms[Title/Abstract])) OR (sonoelastographic[Title/Abstract])) OR (real-time ultrasound elastography[Title/Abstract])) OR (tissue elasticity imaging[Title/Abstract])) OR (elasticity imaging technique[Title/Abstract])) OR (elasticity imaging techniques[Title/Abstract])))) AND (“1945/01/01”[Date - Publication] : “2014/12/05”[Date - Publication])

Supplemental 1: Search syntax for elastography and malignancy in the Pubmed database. A comparable syntax with the same terms was used for the Embase and Cochrane library.

QUADAS-2 Tool Signaling Questions[16]

A. Evaluation of bias

1. Patient selection
    a. Was a consecutive sample of patients enrolled?
    b. Was a case-control design avoided?
    c. Was selection bias avoided by including patients with a thyroid nodule without previous diagnostic work-up such as fine-needle aspiration (FNA)?
2. Index test (thyroid nodule elastography)
    a. Was the elastography outcome interpreted without knowledge of the cytologic or histologic outcome?
    b. Was the thyroid nodule elastography colors scale pre-specified?
3. Reference standard (thyroid nodule cytology or histology)
    a. Was thyroid nodule cytology or histology used as a reference standard?
    b. Were the reference standard results interpreted without knowledge of the elastography results?
4. Flow and timing
    a. Was the elastography performed within one month before the cytology or histology?
    b. Did all patients receive a reference standard?
    c. Were all patients included in the analysis?

B. Evaluation of applicability

1. Patient selection
    a. Was the patient group studied a typical clinical practice in an outpatient setting?
    b. Were patients included before they were referred for FNA, and not when they were already referred for surgery?
2. Index test (thyroid nodule elastography)
    a. Was a thyroid nodule elastography device used?
    b. Was a high-frequency linear probe used to perform the elastography?
    c. Was a thyroid nodule elastography color scale used?
    d. Was the thyroid nodule elastography performed by an experienced operator?
3. Reference standard (thyroid nodule cytology or histology)
    a. Was cytology or histology of the thyroid nodule used as reference standard?

Supplemental 2: Quality assessment of the included studies

Supplemental 3: QUADAS-2 tool for quality assessment of the included studies. The risk of bias for the included articles and the level of concern regarding the applicability of included articles are shown. Ref, reference; J = low risk, L = high risk, ? = unclear risk

Quality assessment of the included studies

Study	Year of publication	Risk of bias				Applicability concerns		
		Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Guazzaroni[52]	2014	☺	☺	?	?	☺	☺	☺
Tatar[53]	2014	☺	☺	☺	?	☺	☺	☺
Azizi[54]	2013	☺	?	?	?	☺	☺	☺
Mehrota[15]	2013	☺	?	?	?	☺	☺	☺
Bojunga[55]	2012	☺	☺	☺	?	☺	☺	☺
Ciledag[24]	2012	☺	☺	?	☺	☺	☺	☺
Friedrich-Rust[26]	2012	☺	☺	?	☹	☺	☺	☺
Mansor[30]	2012	☺	?	?	☹	☺	☺	☺
Moon[27]	2012	☺	?	?	☹	☺	☺	☺
Trimboli[28]	2012	?	☺	?	☹	☺	☺	☺
Unluturk[29]	2012	☺	☺	☺	?	☺	☺	☺
Bhatia[13]	2011	☺	☺	?	☹	☺	☺	☺
Merino[36]	2011	☺	☺	?	☹	☺	☺	☺
Yerli[42]	2011	☺	?	?	☹	☺	☺	☺
Kagoya[56]	2010	☺	☺	?	☺	☺	☺	☺
Friedrich-Rust[25]	2009	☺	☺	?	☹	☺	☺	☺
Rubaltelli[37]	2009	☺	☺	?	☺	☺	☺	☺
Asteria[17]	2008	☺	☺	?	☺	☺	☺	☺
Ferrari[34]	2008	☺	?	?	?	☺	☺	☺
Tranquart[39]	2008	☺	☺	?	☺	☺	☺	☺

Supplemental 4: Meta-analysis

JAGS Model

We specified a bivariate meta-analysis model in JAGS to calculate summary senstivities, specificities, PPV and NPV. For each analysis, the model was initiated using 4 independent chains with 10000000 iterations each. Posterior densities were based on 50000 iterations from each chain. Non-informative prior distributions were used for the (logit of the) summary statistics and between-study covariance. In particular, we used a Normal distribution with mean zero and variance 100000 for the logit of the summary TPR, FPR, PPV and (1-NPV). We used an inverse Wishart prior for the between-study covariance with a diagonal identity scale matrix and 3 degrees of freedom.

The model below can be used to calculate summary sensitivities and specificities, or summary PPV and NPV. De data is given by 'y', 'z', 'm' and 'n', where 'y' and 'z' indicate the number of true positives and true negatives respectively. When calculating the pooled sensitivity and specificity, 'm' and 'n' represent the total amount of positives (TP+FN) and total amount of negatives (TN +FP). Conversely, when calculating the pooled PPV and NPV, 'm' and 'n' are only calculated for cohort studies and represent (TP+FP) and, respectively, (TN+FN).

Parameters
W = diag(2) #prior precision matrix

```
model{
 for (i in 1:N){
  y[i] ~ dbin(pi[i], n[i])
  z[i] ~ dbin(pj[i], m[i])
  logit(pi[i]) <- obs[i,1]
  logit(pj[i]) <- obs[i,2]

  # Reciprocal of sample size in ith study
  inv.N1[i] <- 1/n[i]
  inv.N2[i] <- 1/m[i]

  # Random effects
  obs[i,1:2] ~ dmnorm(mu[1:2],Tau[,])
 }

 mu[1] ~ dnorm(0, 0.001)
 mu[2] ~ dnorm(0, 0.001)

#Use scaled inverse Wishart for Tau
 Tau[1:2, 1:2] ~ dwish(W[,], 3)
 Sigma[1:2,1:2] <- inverse(Tau[,])

 #Pooled sens and spec
 sens <- 1/(1+exp(-mu[1]))
 spec <- 1/(1+exp(-mu[2]))
 fpr  <- 1-(1/(1+exp(-mu[2])))
 rho  <- Sigma[1,2]/sqrt(Sigma[1,1]*Sigma[2,2])

 #Prediction intervals for sens and spec
 obs.new ~ dmnorm(mu[1:2],Tau[,])
 sens.new <- 1/(1+exp(-obs.new[1]))
 spec.new <- 1/(1+exp(-obs.new[2]))
 fpr.new  <- 1-(1/(1+exp(-obs.new[2])))

  ########################################
########################################
##########
 #Calculation of I2
  # Zhou, Y. Dendukuri, N (2014) Statistics for
quantifying heterogeneity in univariate
  # and bivariate meta-analyses of binary data:
the case of meta-analyses of diagnostic
 # accuracy. Statistics in Medicine 33:2701-2717.
  ########################################
########################################
##########

 # mean of inverse sample size in each study
 avgN1 <- mean(inv.N1[])
 avgN2 <- mean(inv.N2[])

 #Expected within-study variance
 sigmasqA <- (exp(Sigma[1,1]/2-mu[1])+exp(Sigm
a[1,1]/2+mu[1])+2)*avgN1
  sigmasqB <- (exp(Sigma[2,2]/2-mu[2])+exp(Sig
ma[2,2]/2+mu[2])+2)*avgN2

 # univariate I2
 I2E[1] <- Sigma[1,1]/(Sigma[1,1]+sigmasqA)
 I2E[2] <- Sigma[2,2]/(Sigma[2,2]+sigmasqB)

 # bivariate I2
 I2E.bivariate <- sqrt(exp(logdet(Sigma[,]))/

(sqrt(exp(logdet(Sigma[,])))+sqrt(sigmasqA-

*sigmasqB))

}
```

Supplemental 4: Meta-analysis

THE ROLE OF QUALITATIVE ELASTOGRAPHY IN THYROID NODULE EVALUATION: EXPLORING ITS TARGET POPULATIONS

CHAPTER 3

Jakob W. Kist
Sjoerd Nell
Bart de Keizer
Gerlof D. Valk
Inne H.M. Borel Rinkes
Menno R. Vriens

Endocr 2015;50(2):265–7.

Editorial

Thyroid nodules are frequently found and pose a dilemma to the clinician, as only a few harbor a malignancy, and the majority of nodules are benign. The standard work-up of thyroid nodules consists of ultrasound and fine needle aspiration (FNA), both having their limitations.[1,2] In particular, ultrasound lacks criteria for determining whether a nodule is malignant.[3] FNA shows good sensitivities and specificities, but inconclusive and indeterminate results are frequently found, resulting in the need for a diagnostic thyroid lobectomy to obtain a final diagnosis.[4] In the long lasting search to reduce the number of invasive diagnostic procedures, real-time qualitative elastography has been proposed to fulfill this need. Elastography determines the elasticity of the thyroid nodule. Soft nodules are assumed to be benign, whereas hard nodules are considered to be malignant. Qualitative elastography represents the elasticity of the nodule in a colored image projected over the ultrasound image. Multiple elasticity scoring systems are used, which makes reviewing literature challenging. Most common is the 4-point scale developed by Asteria et al., in which elastography 1 (ES 1) is assigned to nodules with elasticity in the entire nodule, ES 2 is assigned to nodules with elasticity in a large portion of the nodule, ES 3 is assigned to nodules with stiffness in a large portion of the nodule, and ES 4 is assigned to hard nodules. A cut-off between ES 2 and ES 3 is widely accepted to discriminate benign from malignant nodules.[5]

Studies on thyroid nodule elastography have focused on different target populations: 1) patients referred for FNA with the aim to reduce the number of FNAs, and 2) patients with an indeterminate FNA result (i.e. Bethesda classification III or IV), with the aim to reduce the number of futile lobectomies. Recently, we performed a meta-analysis of studies that investigated the first population: patients referred for FNA. The aim of this study was to determine whether elastography could determine the nature of thyroid nodules and thereby identify those that require further analysis by FNA.[6] In this study, analyzing twenty reports including 3908 nodules, two different cut-offs were examined. The first was the standard cut-off between ES 2 and 3. The second used a cut-off between ES 1 and 2, meaning that only the completely soft nodules were considered benign and the rest of the nodules as potentially malignant. Both cut-offs showed that elastography is an excellent tool to diagnose benignity, with a respective negative predictive value (NPV) of 97% and 99%. Based on these outcomes it was concluded that in these cases FNA could be omitted safely. However, considering the modest positive predictive value (PPV) of only 40% of elastography, this implies that any nodule with an elastography score above ES 2 requires further analysis.[6]

In the current issue of Endocrine, Trimboli et al. published an extensive review and meta-analysis on nodules with an indeterminate FNA.[7] Although the majority of these nodules are benign, around one in four harbors a malignancy.[4] Diagnostic lobectomies are often performed to obtain conclusive histology. To reduce the number of futile surgeries, it would be of great

interest to find a tool that can determine which nodules should be operated on. In this meta-analysis, almost 500 thyroid nodules with indeterminate FNA derived from eight studies were analyzed. Three of these studies especially focused on these particular nodules, while the rest consisted of a subgroup of studies investigating the overall performance of elastography. All studies were relatively small with a maximum of 142 nodules included. The pooled estimated diagnostic parameters are disappointing with a sensitivity of 69%, a specificity of 75% and a respective NPV and PPV of 82% and 63%. The heterogeneity between the studies was moderate to high in all diagnostic parameters (I2 65%-95%). Taken together these findings lead to the justified conclusion that, for this population, qualitative elastography is not able to guide the clinician in his decision-making and that a diagnostic lobectomy should be performed to obtain a final diagnosis in all cases.

There is an apparent difference between the outcome of our meta-analysis and the meta-analysis of Trimboli et al.[6,7] The excellent NPV of 97% for elastography in nodules referred for FNA was not reproduced in nodules with an indeterminate FNA. Trimboli et al. found a NPV of 82%.[7]

The main difference between both meta-analyses is obviously the target population: nodules referred for FNA versus nodules with an indeterminate FNA outcome. However, it is unclear why this would influence the NPV. The higher a priori chance of malignancy in the meta-analysis of Trimboli et al. might have lowered the NPV slightly, but this does not explain this significant difference. Other potential causes are the relatively small number of included studies, the nodules analyzed per study and overall number of analyzed nodules. Moreover, the different outcomes could be based on the high heterogeneity of the included studies in the meta-analysis by Trimboli. This might be caused by publication bias, as suggested by the authors, but as well by the classification of nodules as indeterminate, since several studies showed that FNA cytopathology is complicated and requires expertise.[8] The included indeterminate nodules could be influenced by this phenomenon and, therefore, be not as homogeneous as desirable. Based on the currently published data, a meta-analysis of the studies included by Trimboli et al. with an alternative cut-off (between ES 1 and 2), like performed in our meta-analysis, was not possible. In the future studies focusing on this alternative should be performed.

At this point, it is uncertain whether the described modest diagnostic performance of qualitative elastography of thyroid nodules with an indeterminate FNA represents the ground truth or that future studies specifically focusing on the indeterminate thyroid nodules will reveal better results. We propose that such future studies should be done using the novel, more advanced quantitative elastography techniques, such as Shear Wave Elastography and Acoustic Radiation Force Impulse Imaging. These techniques have been shown to have promising accuracy results for malignancy, said to be less operator dependent and more consistent, although they have not been studied extensively in indeterminate nodules specifically.[9] Another topic of future studies could aim to evaluate the combination of ultrasound and

elastography scoring systems to guide the necessity of FNA in the ES positive nodules.[10,11]

In conclusion, qualitative elastography can replace fine needle aspirations in patients with (nearly) complete soft thyroid nodules, but its role in guiding clinicians in case of indeterminate cytology remains to be determined.

References

1. 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(5):714–20.

2. Samulski TD, Shutty C, LiVolsi VA, Montone K, Baloch Z. The reliability of thyroid nodule ultrasound features and size to predict malignancy in fine needle aspiration specimens: Practical utility for the evaluating pathologist. Diagn Cytopathol 2015;

3. Brito JP, Gionfriddo MR, Nofal Al A, et al. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. J Clin Endocrinol Metab 2014;99(4):1253–63.

4. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. Acta Cytologica 2012;56(4):333–9.

5. Asteria C, Giovanardi A, Pizzocaro A, et al. US-elastography in the differential diagnosis of benign and malignant thyroid nodules. Thyroid 2008;18(5):523–31.

6. Nell S, Kist JW, Debray TPA, et al. Qualitative elastography can replace thyroid nodule fine-needle aspiration in patients with soft thyroid nodules. A systematic review and meta-analysis. Eur J Radiol 2015;84(4):652–61.

7. Trimboli P, Treglia G, Sadeghi R, Romanelli F, Giovanella L. Reliability of real-time elastography to diagnose thyroid nodules previously read at FNAC as indeterminate: a meta-analysis. Endocr 2014;

8. Davidov T, Trooskin SZ, Shanker BA, Yip D, Eng O. Routine second-opinion cytopathology review of thyroid fine needle aspiration biopsies reduces diagnostic thyroidectomy. Surgery 2010;

9. Cantisani V, Lodise P, Grazhdani H, et al. Ultrasound elastography in the evaluation of thyroid pathology. Current status. Eur J Radiol 2014;83(3):420–8.

10. Russ G, Royer B, Bigorgne C, Rouxel A, Bienvenu-Perrard M, Leenhardt L. Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. Eur J Endocrinol 2013;168(5):649–55.

11. Trimboli P, Guglielmi R, Monti S, et al. Ultrasound sensitivity for thyroid malignancy is increased by real-time elastography: a prospective multicenter study. J Clin Endocrinol Metab 2012;97(12):4524–30.

THYROID US-FNA: POSITIVE INFLUENCE OF ON-SITE ADEQUACY ASSESSMENT AND NUMBER OF NEEDLE PASSES ON DIAGNOSTIC CYTOLOGY RATE

CHAPTER 4

Elizabeth J. de Koster
Jakob W. Kist
Menno R. Vriens
Inne H. M. Borel Rinkes
Gerlof D. Valk
Gerard Stapper
Bart de Keizer

Acta Cytologica 2016;60(1):39–45.

Abstract

Objective

Nondiagnostic cytology is the most important limitation of thyroid ultrasound-guided fine needle aspiration (US-FNA). The current study aimed to identify factors associated with adequacy rate of thyroid US-FNA.

Study Design

Consecutive thyroid US-FNAs (2006-2013) were retrospectively included. Attending radiologists, radiology fellows and radiology residents performed US-FNA, usually involving 2-3 needle passes. During more recent years, rapid on-site adequacy assessment (ROSAA) was performed to ensure specimen adequacy. Ultrasound characteristics, procedural variations and cytology results were extracted from ultrasound and pathology reports and statistically evaluated.

Results

Diagnostic cytology was obtained in 64.6% of 1.381 included thyroid US-FNAs. Factors associated with nondiagnostic cytology were ROSAA (74.6% diagnostic cytology, OR 0.55, 95% CI 0.42-0.71), a third clinic visit or more for US-FNA of the same thyroid nodule (54.7%, OR 1.56, 95% CI 1.16-2.10) and increased intranodular vascularization (51.8%, OR 1.73, 95% CI 1.17-2.57). With ROSAA an increasing number of needle passes demonstrated improving adequacy rates. The adequacy rate was not operator dependent.

Conclusion

This study demonstrates that ROSAA improves the adequacy rate of thyroid US-FNA. Without ROSAA, we recommend performing at least 3 needle passes. Less diagnostic cytology is obtained from nodules with increased intranodular vascularization or from those undergoing US-FNA for the third time or more.

Introduction

Prevalence of thyroid nodules on ultrasound (US) is reported as high as 30-70%. As only 5% harbor a malignancy, fine needle aspiration (FNA) is essential for their evaluation.[1-4] Adequate cytological diagnoses provided by FNA have resulted in a reduction of futile surgeries for benign thyroid nodules of up to 50%.[1,2,5]

Although safe and cost-effective, the most important limitation of thyroid FNA is its high rate of nondiagnostic cytology with reported rates up to 34%.[1,2,6-11] The established definition for adequate cytology is the presence of at least six follicular cell groups, each containing 10-15 cells, preferably obtained from at least two aspirates of a nodule.[2,12,13] Nondiagnostic cytology leads to repeat FNA procedures, futile diagnostic hemithyroidectomies and associated additional costs. Rapid on-site adequacy assessment (ROSAA) during the FNA procedure provides immediate feedback and allows for repeat aspirates to ensure specimen adequacy. Studies evaluating the added diagnostic value of ROSAA consistently demonstrated that ROSAA decreases the rates of nondiagnostic specimens and false-negative cytology, and the necessary number of needle passes.[4,9,10,14-17]

Other factors that have been associated with nondiagnostic US-FNA results, are cystic content of a thyroid nodule, poor slide preparation with smears too thick, blood obscuring visualization of follicular cells, poor specimen fixation or inadequate staining technique.[1,6,16,18] Modified or alternative techniques, such as using a different type and size of the biopsy needle or additional liquid based cytology, are related to better adequacy rates.[19-22] Also, both experience-based operator dependency as well as inter-observer variance among cytotechnologists and cytopathologists have repeatedly been demonstrated. [23-26]

This study was designed to evaluate the adequacy rate of all thyroid ultrasound-guided FNAs (US-FNAs) at our hospital and to identify factors associated with nondiagnostic cytology. Specifically, we hypothesized that performance of thyroid US-FNA by an experienced radiologist or with additional ROSAA would improve adequacy rates.

Materials and Methods

Patient Selection

With approval by the local Institutional Medical Ethics Committee, all consecutive thyroid US-FNAs in adult patients between 2006 and 2013 at the University Medical Center Utrecht were retrospectively included. Data were extracted from both US and cytopathology reports. Data recorded included the operator of the US-FNA, whether ROSAA was performed, number of aspirates during one procedure, thyroid nodule US characteristics, and number of clinic visits for repeated US-FNA of the same thyroid nodule. Four

groups were used to categorize the 127 operators: 107 radiology residents, 13 radiology fellows, six attending radiologists (with 5-20 years of experience) and a separate category for our single most experienced radiologist (over 20 years of experience). US-FNAs were excluded when it concerned residual tissue after previous thyroid surgery.

US-FNA procedure

US was performed using a Philips iU22 Ultrasound System (Philips Healthcare, Eindhoven, the Netherlands) with a linear 12- to 5-MHz transducer. FNA was performed using either a 21-gauge, 1.5 inch (BD Microlance™ 3, nr. 2) or a 23-gauge, 1.25 inch needle (BD Microlance™ 3, nr. 14; Becton, Dickinson and Company, Breda, the Netherlands). If a cytotechnologist was present for ROSAA, for each needle pass direct smears were made, air dried and stained with the Diff-Quik method. Immediate microscopic review of the slides determined the cytological adequacy based on presence of a minimum of six follicular cell groups of 10-15 cells.[2,12,13] In case of an inadequate specimen and if time and patient comfort allowed, an additional aspirate was obtained and ROSAA repeated. The number of aspirates was registered. Finally, all obtained cytology was submitted to the cytopathology department for definite analysis. If no cytotechnologist was present, direct smears were made, marked and submitted unstained. Two needle passes were standard procedure for US-FNAs without ROSAA.

Cytologic assessment

At the cytologic laboratory May-Grünwald-Giemsa staining was applied to all smears. Any additional residual material, placed in a container of CytoLyt® solution (Cytyc Corporation, Marlborough, MA) during US-FNA, was processed using the Papanicolaou stain. Descriptive pathology reports were drafted for each specimen. Smears that could not be interpreted due to quantitative or qualitative reasons were classified as 'nondiagnostic'. The reports were descriptive and no further distinct diagnostic categories were reported.

Statistical analysis

Data analysis was performed using IBM SPSS Statistics (version 20.0; IBM Corp, Armonk, NY). Contingency tables were constructed to compare the different variables to the cytological outcome (diagnostic or nondiagnostic). Chi-squared (2), Fisher's exact and independent samples t-tests were used for categorical data, categorical data plus small sample sizes, and continuous or numeric data, respectively. A p-value <0.05 was considered statistically significant. Multivariate analysis was performed for all factors significant on cross tabulation. The Wald test was used to reduce the multivariate model until solely factors that were independently and significantly associated with the outcome remained.

	Nondiagnostic	Diagnostic	p value
FNAs performed (n = 1,381)	35.4 (489)	64.6 (892)	
Sex, female	35.7 (424)	64.3 (765)	0.627
Age, years	54.1 ± 14.1	52.8 ± 14.0	0.101
Procedural variations			
Clinic visit for thyroid US-FNA			
1st	33.0 (249)	67.0 (506)	0.001
2nd	33.8 (128)	66.2 (251)	
3rd or more	45.3 (112)	54.7 (135)	
Operator of US-FNA			
Radiology resident	36.1 (338)	63.9 (597)	0.576
Radiology fellow	29.7 (35)	70.3 (83)	
Most experienced radiologist X	35.8 (77)	64.2 (138)	
Other attending radiologists	34.5 (39)	65.5 (74)	
Rapid on-site adequacy assessment			
No	39.7 (384)	60.3 (583)	<0.001
Yes	25.4 (105)	74.6 (309)	
Thyroid nodule characteristics			
Maximum diameter, mm	25.8 ± 13.6	27.1 ± 13.6	0.120
Nodularity			
Solitary	32.5 (102)	67.5 (212)	0.292[a]
Multinodular	36.4 (385)	63.6 (672)	
Unknown	20.0 (2)	80.0 (8)	
Consistency			
Solid	27.6 (47)	72.4 (123)	0.054
Mixed solid-cystic	36.5 (107)	63.5 (186)	
Cystic	29.8 (34)	70.2 (80)	
Unknown	37.4 (301)	62.6 (503)	
Sonographic appearance			
Hypoechogenic	36.9 (58)	63.1 (99)	0.813
Isoechogenic	28.9 (11)	71.1 (27)	
Hyperechogenic	37.8 (14)	62.2 (23)	
Unknown	35.3 (406)	64.7 (743)	
Homogeneity			
Homogenous	30.8 (4)	69.2 (9)	0.816[b]
Inhomogenous	34.2 (119)	65.8 (229)	
Unknown	35.9 (366)	64.1 (654)	
Increased intranodular vascularization			
Yes	48.2 (54)	51.8 (58)	0.006
No	25.5 (12)	74.5 (35)	
Unknown	34.6 (423)	65.4 (799)	
Intranodular calcifications			
Yes	32.8 (44)	67.2 (90)	0.324
No	26.9 (14)	73.1 (38)	
Unknown	36.1 (431)	63.9 (764)	

[a] mean ± SD. [b] Calculated using Fisher's exact test.

Table 1. Results for all thyroid FNAs. Demographics, procedural variations and sonographic characteristics of all thyroid US-FNAs (n = 1381), and comparison of all variables according to diagnostic or nondiagnostic cytology result

Results

In total, 1381 thyroid US-FNAs (mean maximum nodule diameter 26.7 ± 13.6 mm) were included in the study (Table 1). These were performed in 682 patients (86% female, mean age 53.3 years (SD ± 14.1). Diagnostic cytology was obtained in 892 (64.6%). Benign cytology was reported in 793 (57.4%) cases, 54 (3.9%) had cytology of undetermined significance, including atypia and follicular or hurtle cell proliferations, and 45 (3.3%) were suspicious or positive for malignancy.

Cytological adequacy

Cytological adequacy was significantly higher with performance of ROSAA (309/414, 74.6%) than without it (583/967, 60.3%, p<0.001)(Table 1). The first and second clinic visits for US-FNA of the same thyroid nodule yielded similar rates of adequate cytology (67.0% vs. 66.2%, respectively), but third and subsequent visits less often resulted in a diagnostic specimen (54.7%, p<0.001). Increased intranodular vascularization on Doppler US was associated with a significantly higher nondiagnostic rate (48.2%, p=0.006). Other US characteristics, such as nodule size and consistency, showed no significant association. Finally, no operator dependency was demonstrated. The mean number of US-FNAs performed by each operator group was 8.7 (range: 1-31) for radiology residents, 9.0 (range: 1-19) for radiology fellows and 18.8 (range: 1-45) for attending radiologists. Our single most experienced radiologist performed 215 US-FNAs.

Multivariate analysis

Multivariate analysis revealed significant association with nondiagnostic cytology for three or more clinic visits (OR 1.56, 95% CI 1.16-2.10), ROSAA (OR 0.55, 95% CI 0.42-0.71) and strong intranodular vascularization (OR 1.73, 95% CI 1.17-2.57) (Table 2).

Number of needle passes

Additionally, subanalysis was performed of the number of needle passes per procedure. This was only registered for US-FNAs when ROSAA was performed. A mean number of 1.75 needle passes (SD ±0.86) was needed to acquire a diagnostic specimen. With three or less aspirates, 72.0% (298/414) of all US-FNAs were diagnostic, while this was 59.7% (247/414) in case of two or less aspirates (Figure 1).

	OR adj.	(95% CI)	p-value
Clinic visit for thyroid US-FNA			
1st	1	ref.	
2nd	1.01	(0.77 - 1.31)	0.950
3rd or more	1.56	(1.16 - 2.10)	0.003
Rapid On-Site Adequacy Assessment			
No	1	ref.	
Yes	0.55	(0.42 - 0.71)	<0.001
Increased intranodular vascularisation			
No or unknown	1	ref.	
Yes	1.73	(1.17 - 2.57)	0.006

OR adj.: adjusted odds ratio. CI: confidence interval.

Table 2. Multivariate analysis
Adjusted odds ratios for variables associated with nondiagnostic cytology

Discussion

Despite its reliability and cost-effectiveness for distinguishing between benign and malignant thyroid nodules, the nondiagnostic rate of thyroid FNA - varying from 2-34% in literature - remains a major clinical issue.[1,2,6-10,19] The current study identified multiple factors that were significantly associated with adequacy rate: adequacy increased with ROSAA, but decreased with three or more clinic visits and strong intranodular vascularization.

Consistently higher adequacy rates of thyroid FNA with ROSAA have led to recommendations for its routine application.[4,9,10,15,16,27] The absolute 14.3% (39.7% to 25.4%) decrease in nondiagnostic rate that our study demonstrated is larger than the average of 9% that was reported in a meta-analysis by Witt et al[28]. This result supports their finding that the increase in adequacy rate with ROSAA is correlated with the adequacy rate without ROSAA, and that adequacy improves more when the initial nondiagnostic rate is higher.[28]

Before ROSAA was implemented, common practice at our institution was to acquire no more than two aspirates per FNA procedure. This resulted in 60.3% adequate cytology. Subanalysis of our results demonstrated that 72.0% diagnostic specimens were obtained when ROSAA was combined with three or fewer needle passes. This was 59.7% with ROSAA and two needle passes or fewer - similar to the former adequacy rate. Without ROSAA the operating radiologist blindly estimates the number of needle passes needed for adequate cytology. More aspirates then likely result in higher adequacy rates. Studies that performed 4 to 6 passes as standard protocol reported nondiagnostic rates as low as 5.5%.[17,29,30] By providing feedback for immediate repeat biopsies and/or redirection of the needle within the thyroid nodule, ROSAA might not only improve adequacy, but also decrease the necessary number of needle

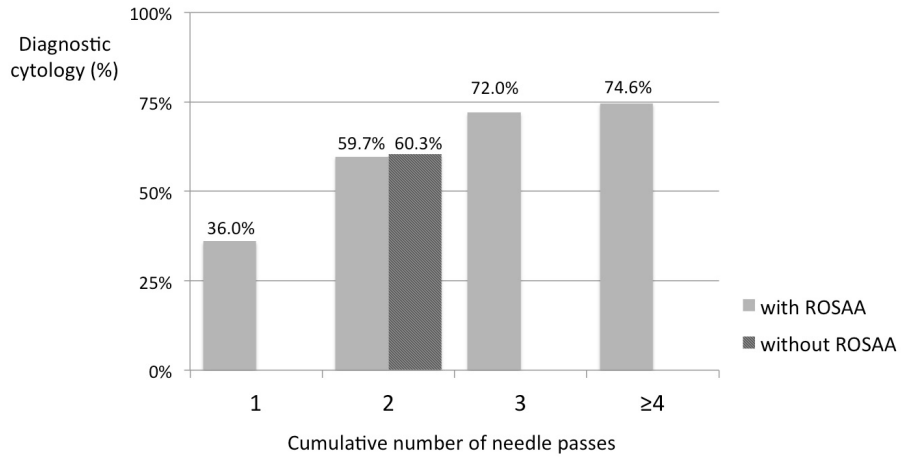


Figure 1. US-FNA adequacy rate (%) compared to cumulative number of needle passes

passes for institutions where protocol prescribes acquiring relatively many aspirates.[17] However, we propose that solely obtaining more aspirates per nodule – without ROSAA – might achieve an equal improvement in adequacy rate. Hereby minimizing additional costs, effort and procedure time, this might be the preferred, more cost-effective option for some institutions instead of implementing ROSAA. We would recommend performing at least 3 to 4 needle passes.

The influence of operator experience on adequacy rate was not demonstrated in the current study, despite vast evidence of operator dependency in literature - for FNA of the thyroid as well as other organs.[9,24,26] Ljung et al. defined an experienced operator as one who performs at least 100 FNAs of various body sites per year.[26] For thyroid FNAs specifically, it has been suggested that an operator has to perform at least 200 procedures before achieving high adequacy rates.[9,24,30] As more than one hundred operators performed the thyroid US-FNAs in our study (most of whom were radiology residents), the individual and general level of experience was limited and may have had a negative effect on the overall diagnostic rate. Even our most specialized radiologist merely performed 215 out of 1381 thyroid US-FNAs over the course of eight years, an average of 27 procedures per year with a success rate similar to the mean (64.2% vs. 64.6%). Centralizing thyroid FNA procedures to specialized centers and educating a small number of dedicated head-and-neck radiologists could substantially improve adequacy rates by reducing the influence of operator inexperience.[25,26]

The current study demonstrated that strong intranodular vascularization on Doppler US was significantly associated with lower adequacy rates. Other

studies confirm what we experience in daily practice: blood contamination is one of the main qualitative reasons for nondiagnostic thyroid cytology and FNA of a hypervascular nodule will often result in bloody aspirates.[4,19] Yet, strong intranodular vascularization is rarely mentioned in literature as a predictor of nondiagnostic cytology.[31] It is, however, described as a possible predictor of malignancy, emphasizing the need for adequate cytology.[7,12,32] In contrast to the results of previous studies, other US characteristics, such as consistency or nodule size, were not significantly related to the adequacy rate of thyroid US-FNAs in our study.[6,33]

Even though the subsequent clinic visits after the second US-FNA were related to a significantly lower adequacy rate, nevertheless, a third or fourth US-FNA still yielded diagnostic cytology for a substantial 54.7% of patients in our population. Cost-benefit analysis, consequences and associated morbidity of surgical intervention, as well as risks associated with delayed treatment for undetected malignancies, cause a clinical dilemma of deciding between diagnostic surgery or an additional FNA.[34] A diagnostic hemithyroidectomy is recommended for nodules with a repeatedly nondiagnostic US-FNA and suspicious US or clinical characteristics, because approximately 10% harbor a malignancy.[5,27] Weighing the low risk of aggressive thyroid cancer against the roughly 50% prevented surgeries upon diagnostic cytology of the third US-FNA, we believe that performing a third US-FNA is beneficial and safe, provided that the time interval between subsequent US-FNAs is appropriately minimalized.

The ultimate goal of this study was to improve the adequacy rate of thyroid US-FNA at our hospital by first addressing those factors that could improve quality with minor adjustments or minimal additional resources. Moreover, our study shows the value of institutional assessment of quality of care. We believe that quality assessment of the existing primary diagnostic tools should always be done before more complex measures are added to the diagnostic routine. The resolution might be as simple as performing a few extra needle passes per procedure, leaving additional diagnostics such as molecular testing for BRAF mutations – promising but costly – for the fewer cases that remain nondiagnostic despite improved basic procedures. Guidelines for improvement – like the ones existing for breast cancer – could support this process.[35] Nevertheless, it starts with a critical assessment of the institution's own performance and acknowledgement of the bottlenecks. The 35% nondiagnostic rate in our study might not be as exceptionally high as it seems compared to previously published rates. Large inter-institutional variations indivertibly exist, but publication bias presumably leaves worse rates unpublished and the general nondiagnostic rate higher than the reported average of 17%.[28,36]

The most important limitation of this study was its retrospective design. Data regarding multiple variables were missing for many US-FNAs, mostly US characteristics due to limited reports. Also amongst these were the number of needle passes for most non-ROSAA US-FNAs, impeding full evaluation of our

hypothesis regarding number of needle passes in this subset.

Secondly, even when ROSAA was performed, US-FNA procedure was often discontinued before adequate material was acquired. Possible reasons for this are time limitations for one procedure, repeatedly bloody specimens, patient stress and discordance between on-site and final adequacy assessments. Further research is needed at our institution to analyze the optimal number of needle passes and its presumed relation to the adequacy rate with and without ROSAA.

Potential confounders were addressed during data collection and analysis. Other than the procedural variations analyzed, no differences in US-FNA or cytology assessment procedures were demonstrated throughout the years. A potential confounding factor is that our most experienced radiologist was specifically requested as operator for a number of thyroid FNAs, mostly repeated procedures. Performing the more difficult FNAs may have deteriorated his adequacy rate. This possible relation could not be objectified due to limited reports.

Conclusions

This study demonstrated that ROSAA improved the adequacy rate of thyroid US-FNA. When ROSAA is not performed, we recommend obtaining three or more aspirates. Diagnostic cytology is less often obtained from nodules with increased intranodular vascularization or during US-FNA that is performed for the third time or more. Adequacy rate of thyroid US-FNA was not operator dependent in this study. Almost equally important, the results of our study underline the importance of institutional quality assessment of basic diagnostic procedures.

References

1. Abraham TM, de las Morenas A, Lee SL, Safer JD: In thyroid fine-needle aspiration, use of bedside-prepared slides significantly increased diagnostic adequacy and specimen cellularity relative to solution-based samples. Thyroid 2011;21:237–42.

2. Gharib H, Goellner JR: Fine-needle aspiration biopsy of the thyroid: an appraisal. Ann Intern Med 1993;118:282–9.

3. Mazzaferri EL: Management of a solitary thyroid nodule. N Engl J Med 1993;328:553–9.

4. Redman R, Zalaznick H, Mazzaferri EL, Massoll NA: The impact of assessing specimen adequacy and number of needle passes for fine-needle aspiration biopsy of thyroid nodules. Thyroid 2006;16:55–60.

5. Cibas ES, Ali SZ: The Bethesda System for Reporting Thyroid Cytopathology. Thyroid 2009;19:1159–65.

6. Alexander EK, Heering JP, Benson CB, Frates MC, Doubilet PM, Cibas ES, et al.: Assessment of nondiagnostic ultrasound-guided fine needle aspirations of thyroid nodules. J Clin Endocrinol Metab 2002;87:4924–7.

7. Carmeci C, Jeffrey RB, McDougall IR, Nowels KW, Weigel RJ: Ultrasound-guided fine-needle aspiration biopsy of thyroid masses. Thyroid 1998;8:283–9.

8. Danese D, Sciacchitano S, Farsetti A, Andreoli M, Pontecorvi A: Diagnostic accuracy of conventional versus sonography-guided fine-needle aspiration biopsy of thyroid nodules. Thyroid 1998;8:15–21.

9. Ghofrani M, Beckman D, Rimm DL: The value of onsite adequacy assessment of thyroid fine-needle aspirations is a function of operator experience. Cancer 2006;108:110–3.

10. Jing X, Michael CW, Pu RT: The clinical and diagnostic impact of using standard criteria of adequacy assessment and diagnostic terminology on thyroid nodule fine needle aspiration. Diagn Cytopathol 2008;36:161–6.

11. Eedes CR, Wang HH: Cost-effectiveness of immediate specimen adequacy assessment of thyroid fine-needle aspirations. Am J Clin Pathol 2004;121:64–9.

12. Cooper DS, Doherty GM, Haugen BR, Hauger BR, Kloos RT, Lee SL, et al.: Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid 2009;19:1167–214.

13. Baloch ZW, Sack MJ, Yu GH, Livolsi VA, Gupta PK: Fine-needle aspiration of thyroid: an institutional experience. Thyroid 1998;8:565–9.

14. Baloch ZW, Tam D, Langer J, Mandel S, LiVolsi VA, Gupta PK: Ultrasound-guided fine-needle aspiration biopsy of the thyroid: role of on-site assessment and multiple cytologic preparations. Diagn Cytopathol 2000;23:425–9.

15. Breeze J, Poller DN, Gibson D, Tilley EA, Cooke L, Soar E, et al.: Rapid on-site assessment of specimens by biomedical scientists improves the quality of head and neck fine needle aspiration cytology. Cytopathology 2014;25:316–21.

16. Zhu W, Michael CW: How important is on-site adequacy assessment for thyroid FNA? An evaluation of 883 cases. Diagn Cytopathol 2007;35:183–6.

17. Cengic I, Tureli D, Ahiskali R, Bugdayci O, Aydin H, Aribal E: Thyroid fine needle aspiration biopsy: do we really need an on-site cytopathologist? Eur J Radiol 2014;83:680–3.

18. Massoll NA, Nizam MS, Mazzaferri EL: Cystic thyroid nodules: Diagnostic and therapeutic dilemmas. Endocrinologist 2002;12:185–198.

19. Cappelli C, Pirola I, Gandossi E, De Martino E, Agosti B, Castellano M: Fine-needle aspiration cytology of thyroid nodule: does the needle matter? South Med J 2009;102:498–501.

20. Rossi ED, Morassi F, Santeusanio G, Zannoni GF, Fadda G: Thyroid fine needle aspiration cytology processed by ThinPrep: an additional slide decreased the number of inadequate results. Cytopathology 2010;21:97–102.

21. Saleh H, Bassily N, Hammoud MJ: Utility of a liquid-based, monolayer preparation in the evaluation of thyroid lesions by fine needle aspiration biopsy: comparison with the conventional smear method. Acta Cytol 2009;53:130–6.

22. Samir AE, Vij A, Seale MK, Desai G, Halpern E, Faquin WC, et al.: Ultrasound-guided percutaneous thyroid nodule core biopsy: clinical utility in patients with prior nondiagnostic fine-needle aspirate. Thyroid 2012;22:461–7.

23. Cibas ES, Baloch ZW, Fellegara G, LiVolsi VA, Raab SS, Rosai J, et al.: A prospective assessment defining the limitations of thyroid nodule pathologic evaluation. Ann Intern Med 2013;159:325–32.

24. De Fiori E, Rampinelli C, Turco F, Bonello L, Bellomi M: Role of operator experience in ultrasound-guided fine-needle aspiration biopsy of the thyroid. Radiol Med 2010;115:612–8.

25. Berner A, Sigstad E, Pradhan M, Grøholt KK, Davidson B: Fine-needle aspiration cytology of the thyroid gland: comparative analysis of experience at three hospitals. Diagn Cytopathol 2006;34:97–100.

26. Ljung BM, Drejet A, Chiampi N, Jeffrey J, Goodson WH, Chew K, et al.: Diagnostic accuracy of fine-needle aspiration biopsy is determined by physician training in sampling technique. Cancer 2001;93:263–8.

27. Haugen BR, Alexander EK, Bible KC, Doherty G, Mandel SJ, Nikiforov YE, et al.: 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2015;thy.2015.0020.

28. Witt BL, Schmidt RL, B.L. W, R.L. S: Rapid onsite evaluation improves the adequacy of fine-needle aspiration for thyroid lesions: A systematic review and meta-Analysis. Thyroid 2013;23:428–35.

29. Karadeniz Cakmak G, Emre AU, Tascilar O, Gultekin FA, Ozdamar SO, Comert M: Diagnostic adequacy of surgeon-performed ultrasound-guided fine needle aspiration biopsy of thyroid nodules. J Surg Oncol 2013;107:206–10.

30. Sidiropoulos N, Dumont LJ, Golding AC, Quinlisk FL, Gonzalez JL, Padmanabhan V: Quality improvement by standardization of procurement and processing of thyroid fine-needle aspirates in the absence of on-site cytological evaluation. Thyroid 2009;19:1049–52.

31. Grani G, Calvanese A, Carbotta G, D'Alessandri M, Nesca A, Bianchini M, et al.: Intrinsic factors affecting adequacy of thyroid nodule fine-needle aspiration cytology. Clin Endocrinol (Oxf) 2013;78:141–4.

32. Morris LF, Ragavendra N, Yeh MW: Evidence-based assessment of the role of ultrasonography in the management of benign thyroid nodules. World J Surg 2008;32:1253–1263.

33. de Meer SGA, Schreinemakers JMJ, Zelissen PMJ, Stapper G, Sie-Go DMDS, Rinkes IHMB, et al.: Fine-needle aspiration of thyroid tumors: identifying factors associated with adequacy rate in a large academic center in the Netherlands. Diagn Cytopathol 2012;40 Suppl 1:E21–6.

34. 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.

35. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L: European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition--summary document. Ann Oncol 2008;19:614–22.

36. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR: Publication bias in clinical research. Lancet 1991;337:867–72.

CORRESPONDENCE: PREOPERATIVE BRAF(V600E) MUTATION SCREENING IS UNLIKELY TO ALTER INITIAL SURGICAL TREATMENT OF PATIENTS WITH INDETERMINATE THYROID NODULES - A PROSPECTIVE CASE SERIES OF 960 PATIENTS

CHAPTER 5

Lutske Lodwijk
Jakob W. Kist
Gerlof D. Valk
Menno R. Vriens
Inne H.M. Borel Rinkes

Cancer 2014;120(7):1083–4.

Correspondence

With great interest we read the article of Kleiman et al. addressing the putative added value of preoperative BRAF(V600E) analysis on the initial surgical strategy.[1] As the basic surgical strategy for indeterminate thyroid nodules differs between many US centers and Europe, we propose that the data of Kleiman et al. might lead to an opposite conclusion regarding the use of preoperative BRAF testing in the European setting. To substantiate this hypothesis, we have used data from the Netherlands Cancer Registry.

As pointed out by Kleiman et al., the standard surgical approach to indeterminate thyroid nodules in the USA is to perform a total thyroidectomy on all patients with Bethesda category V and on all nodules with "worrisome cytologic features", eg. nuclear grooves and pseudoinclusions. As 12 of the 13 BRAF mutants had already undergone total thyroidectomy as the initial procedure based on the cytology, routine preoperative BRAF testing would have altered surgery in only one patient.

In contrast to the situation in many US centers, in the Netherlands, as well as in large parts of Europe, a diagnostic hemithyroidectomy is performed routinely on all patients with indeterminate FNA results. When the final histology reveals malignancy a completion thyroidectomy is done as second stage operation.

In the Netherlands 459 patients were diagnosed with DTC in 2011. In their flowchart (figure 1) Kleiman et al did not differentiate between Bethesda category III/IV and Bethesda category V. Nonetheless, we could extract this from the text and tables. In short, 36 (12%) of the indeterminate results were Bethesda category V, of those 36, 26 (72%) were malignant, and of those 26, 11 (31%) were BRAF positive. When we apply these results to our nationwide data, using the same distribution of relative ratios as found by Kleiman, we would have had 42 patients with Bethesda category V. Of those 42 patients, 30 (72%) would have had a malignancy, 13 of which (31%) would comprise a BRAF(V600E) mutation. Hence, by performing preoperative BRAF(V600E) analysis in the group of patients with Bethesda category V, these 13/42 patients (31%) would have benefited as their initial surgical treatment would have been changed to total thyroidectomy, and a two-stage procedure could have been avoided. This would involve a serious benefit, including a shortened period of uncertainty for the patient, avoidance of a second operation, and a significant reduction of the time between diagnosis and final treatment. Of course, we realize that our analysis is not statistically flawless. For practical reasons, we have relied on the assumption that cytology scores in our country are not significantly different from those in the USA. This assumption is supported by two other studies, describing comparable ratios of indeterminate cytology results and BRAF(V600E) positivity.[2,3]

Subsequently, we performed a rough estimate as to whether the BRAF(V600E) mutation analysis could be cost-effective. If only nodules with Bethesda category V are tested, 42 BRAF(V600E) tests have to be performed

to detect 13 BRAF(V600E) mutations. A calculation based on the costs of hemi- and total thyroidectomy in our institution showed that this would not lead to a significant increase in costs.

Based on the notion that the Dutch approach is comparable to the strategy of large parts of Europe the introduction of preoperative BRAF testing on Bethesda category V nodules could make a difference: a considerable amount of patients might be spared unnecessary two stage surgery without increasing the total costs of the treatment. We therefore propose that the impact of preoperative BRAF(V600E) testing of indeterminate thyroid nodules on initial surgical management is predominantly dependent on the routine initial surgical strategy adhered to.

References

1. Kleiman DA, Sporn MJ, Beninato T, et al. Preoperative BRAF(V600E) mutation screening is unlikely to alter initial surgical treatment of patients with indeterminate thyroid nodules: a prospective case series of 960 patients. Cancer 2013;119(8):1495–502.

2. Moses W, Weng J, Sansano I, et al. Molecular Testing for Somatic Mutations Improves the Accuracy of Thyroid Fine-needle Aspiration Biopsy. World J Surg 2010;34(11):2589–94.

3. Musholt TJ, Fottner C, Weber MM, et al. Detection of papillary thyroid carcinoma by analysis of BRAF and RET/PTC1 mutations in fine-needle aspiration biopsies of thyroid nodules. World J Surg 2010;34(11):2595–603.

PART 2

diagnosis of recurrent differentiated thyroid cancer

RECURRENT DIFFERENTIATED THYROID CANCER: TOWARDS PERSONALIZED TREATMENT BASED ON VALUATION OF TUMOR CHARACTERISTICS WITH PET (THYROPET STUDY): STUDY PROTOCOL OF A MULTICENTER OBSERVATIONAL COHORT STUDY

CHAPTER 6

Jakob W. Kist
Bart de Keizer
Marcel P.M. Stokkel
Otto S. Hoekstra
Wouter V. Vogel
On behalf of the Thyropet study group*

BMC Cancer 2014;14(1):405.

* THYROPET Study group: John M.H. de Klerk, Harm van Tinteren, Jan Paul de Boer, Hans Morreau, Marc C. Huisman, Eef G.W.M. Lentjes, Thera P. Links, Jan W.A. Smit, Jules Lavalaye, Piet L. de Jager, Corneline J. Hoekstra, Martin Gotthardt, Vanessa J.R. Schelfhout, Friso M. van der Zandt, Dyde A. Huysmans, Adrienne H. Brouwers, Wieger I. de Bruin, Ferida Sivro, Judit A. Adam, Ha T.T. Phan, Gerrit W. Slooff, and Nils R.L. Wagenaar

Abstract

Background

After initial treatment of differentiated thyroid carcinoma (DTC) patients are followed with regular thyroglobulin (Tg) measurements to detect a possible recurrence. In case of elevated levels of Tg and negative neck ultrasonography, patients are treated "blindly" with Iodine-131 (^{131}I). However, in up to 50% of patients, the post-therapy whole body scan reveals no ^{131}I-targeting of tumor lesions. Such patients derive no benefit from the blind therapy but are exposed to its toxicity. Alternatively, iodine-124 (^{124}I) Positron Emission Tomography/Computed Tomography (PET/CT) has become available to visualize DTC lesions and without toxicity. In addition to this, ^{18}F-luorodeoxyglucose (^{18}F-FDG) PET/CT detects the recurrent DTC phenotype, which does not accumulate iodine or has lost that capacity. Taken together, the combination of ^{124}I and ^{18}F-FDG PET/CT has potential to stratify patients for treatment with ^{131}I.

Methods/Design

In a multicenter prospective observational cohort study the hypothesis that the combination of ^{124}I and ^{18}F-FDG PET/CT can avoid futile ^{131}I treatments in patients planned for 'blind' therapy with high activity of ^{131}I, is tested.

One hundred patients planned for ^{131}I undergo both ^{124}I and ^{18}F-FDG PET/CT after rhTSH stimulation. Independent of the outcome of the scans, all patients will subsequently receive, after thyroid hormone withdrawal, the planned ^{131}I therapy. The post ^{131}I therapeutic scintigraphy will be compared with the outcome of the ^{124}I and ^{18}F-FDG PET/CT in order to evaluate the diagnostic value of the combined PET modalities.

This study primary aims to reduce the number of futile ^{131}I therapies. Secondary aims are the nationwide introduction of ^{124}I PET/CT by a quality assurance and quality control (QA/QC) program, to correlate imaging outcome with histopathological features, to compare ^{124}I PET/CT after rhTSH and after withdrawal of thyroid hormone, and to compare ^{124}I and ^{131}I dosimetry.

Discussion

This study aims to evaluate the potential value of the combination of ^{124}I and ^{18}F-FDG PET/CT in the prevention of futile ^{131}I therapies in patients with biochemically suspected recurrence of DTC. To our best knowledge no studies addressed this in a prospective cohort of patients. This is of great clinical importance as a futile ^{131}I is a costly treatment associated with morbidity and therefore should be restricted to those who will likely benefit from this treatment.

Background

Differentiated thyroid cancer (DTC) is the most frequent endocrine tumor, with an annual incidence per 100.000 individuals of 1 – 3 in men and 2 – 4 in women.[1] In general, DTC has an excellent prognosis, and only 5 to 10% will die of their disease.[2,3] Prognosis is less favorable when the disease recurs after primary treatment. Local or regional recurrence occurs in 5 – 20% of patients.[4] Distant metastases develop in up to 10%, usually in the lungs and bones.[5] Recurrences are usually detected during the early years of follow-up, but may occur years later.[6] As in many diseases, early detection of recurrence improves outcome and survival, because limited disease load may allow surgical resection and/or effective treatment with radioactive iodine (^{131}I). Follow-up is therefore necessary throughout the patients' life. Therefore, even though DTC incidence is low, many patients are currently under surveillance for a possible recurrence (estimated 500.000 in the United States).[7,8]

The serum marker Thyroglobulin (Tg) plays a pivotal role in the follow-up of differentiated thyroid cancer. Serum Tg should be undetectable in DTC patients following effective thyroid remnant ablation with ^{131}I, so that any detectable level reflects (neoplastic) thyroid tissue.[9] The level of serum Tg is related to the amount of neoplastic thyroid tissue; it has been estimated that 1 g of neoplastic thyroid tissue corresponds with a serum Tg of 1 ng/ml during thyroid hormone replacement therapy, and with 2 – 10 ng/ml following recombinant human thyroid hormone stimulating hormone (rhTSH) stimulation.[10,11] A serum Tg cut-off level ≥ 2 ng/ml following rhTSH is highly sensitive to identify patients in whom persistent tumor may be found with imaging techniques.[12,13]

When recurrent DTC is suspected because of serum Tg above the cut-off level, several imaging tests may be performed to detect the exact sites of recurrence. The sodium/iodide symporter (NIS) mediates iodide uptake in the thyroid gland and thyroid cancer cells.[14] The ability of the thyroid to accumulate Iodide via NIS is the basis for scintigraphic thyroid imaging with radioiodine (using the gamma-emitting ^{123}I) as well as for therapy using the beta-emitter ^{131}I, which targets and destroys iodide-transporting benign and malignant thyroid cells. In thyroid cancer, the primary therapy is total thyroidectomy, which in practice is near-total to spare adjacent nerves and parathyroids. Postoperatively, ^{131}I is used to ablate these postoperative thyroid remnants, and to detect (using post ^{131}I whole body scintigraphy) and treat potential metastases.[15-17] With this approach, highly selective radiation doses can be achieved in tumor tissue, often much higher than with external radiotherapy.

Historically, the follow-up of patients with DTC included scintigraphy after a low activity of ^{131}I if serum Tg was elevated, but this is no longer recommended because of poor sensitivity.[18-23] To date, whole body scintigraphy after 'blind' administration of high 'therapeutic' activity of ^{131}I is performed in these patients, during withdrawal of thyroid hormone replacement to stimulate uptake of iodine in cells of thyroidal origin, both to diagnose and re-

stage the potential recurrence and to initiate its treatment.[18,24-28] This strategy can be effective, but an estimated 38% - 50% of patients will have a negative post-therapeutic ^{131}I whole body scan and/or no objective therapy effect.[29,30] Such patients will have received a total body irradiation of 450 millisievert (mSv) and may have suffered from side effects such as nausea, sialoadenitis, loss of taste, or reduced spermatogenesis. Furthermore, their risk of secondary malignancies has increased.[31,32] All induced by a treatment from which they derived no benefit. Also, the prolonged thyroid hormone withdrawal and subsequent hypothyroidism necessary for ^{131}I therapy have major impact on quality of life, with a majority of patients suffering from significant changes in physical, psychological, and social well-being.[33-37] The high frequency of high activities ^{131}I from which patients do not derive any benefit but are exposed to its toxicity and potential adverse oncological effects, has led to a search for new diagnostic tools to improve the selection of patients for such treatment.

Nowadays, ultrasound of the neck is applied to detect a local recurrence or regional lymph node metastases allowing direct biopsy to confirm the diagnosis. Nonetheless, ultrasound is limited to the neck only, and when negative in the presence of detectable Tg, whole body evaluation is required.

Recently, Iodine-124 (^{124}I) has become available as a novel radionuclide for whole body Positron Emission Tomography/Computed Tomography (PET/CT) in the follow-up of DTC[38-41], with a promising diagnostic accuracy and a considerably lower radiation exposure than whole body scintigraphy after therapeutic activity of ^{131}I.[39] Furthermore, recent experience has shown that ^{124}I PET/CT images may be representative for the biodistribution and radiation dosimetry of subsequent therapy with ^{131}I.[42,43] Thus, ^{124}I PET/CT may allow for more accurate restaging of patients in a whole body procedure, perform dosimetry for subsequent ^{131}I therapy and predict the outcome of the treatment. However, some recurrent DTC lesions do not accumulate iodine, which is correlated with tumor dedifferentiation and this implies a poor prognosis.[5] Patients suspected of non-iodine accumulating DTC, so far only evident after futile blind ^{131}I therapy, require restaging before local or systemic therapy may be applied. Metabolic PET imaging with the glucose analogon ^{18}F-fluorodeoxyglucose (^{18}F-FDG), especially during (rh)TSH stimulation, has a high sensitivity to detect recurrent DTC in patients with detectable Tg and negative iodine scintigraphy.[44] It may correlate with a more aggressive tumor behavior and poor prognosis[45], and can help to select patients for other treatment modalities (surgery, external beam radiotherapy or multikinase inhibitors[46-48]). ^{18}F-FDG PET/CT is currently applied only when prior treatment and imaging with therapeutic activity of ^{131}I has proven to be ineffective.[49] The value of ^{18}F-FDG PET/CT before ^{131}I treatment has not been tested.

At the biological level, ^{124}I and ^{18}F-FDG uptake is related to expression of the sodium iodine symporter (NIS)[16], while ^{18}F-FDG uptake is related to hexokinase-I (HKI) and Hypoxia-inducible factor 1-alpha (HIF-1α) activity.[50,51] The evaluation of the relation of ^{124}I and ^{18}F-FDG PET/CT imaging findings and histopathological parameters (such as thyroglobulin, TTF1, Ki-67 and

Cytokeratine-19 staining) and response to ^{131}I treatment will give more insight in the fundamental knowledge about DTC.

The present study aims to test the power of combined for detect and characterize DTC lesions in patients with suspected recurrence. Based on the characteristics of ^{124}I and ^{18}F-FDG PET/CT, it is reasonable to assume that a combined strategy of imaging and histopathological evaluation at the time of suspected recurrence will yield adequate information on the disease stage prior to treatment with ^{131}I, regardless of tumor dedifferentiation, with a potential impact on clinical decision making. The combination of both entities has been suggested in proof of concept studies[52], illustrated in figure 1, but needs proper testing, to increase fundamental knowledge about DTC and further improve treatment.

The multicenter design of this study requires highly standardized procedures for ^{124}I PET/CT. Previously nationwide standardization was done for ^{18}F-FDG PET/CT in the Netherlands, which eventually evolved into the European EARL accreditation system.[53,54] In order to compare the scans between centers calibration and standardization of the ^{124}I PET/CT scans prior to the start of the study will be done a in quality assurance and quality control (QA/QC) program.

In summary, therapy with high activities of ^{131}I for recurrent DTC is effective in many cases, but the current blind approach also leads to

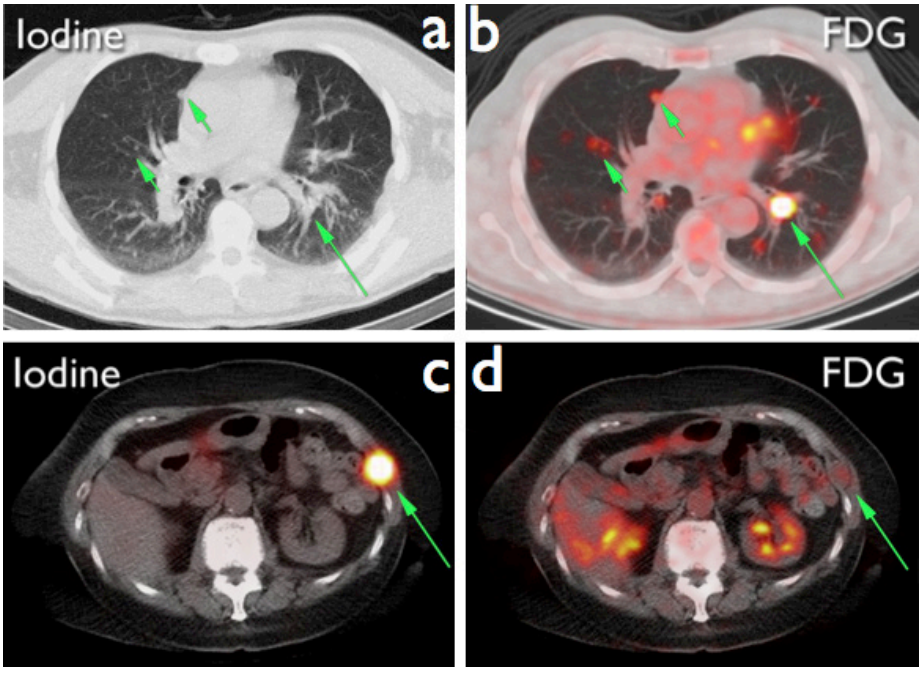


Figure 1: Images from two different patients scanned with both ^{124}I and ^{18}F-FDG PET/CT. The ^{124}I PET/CT of patient 1 (1a) shows multiple 124I negative pulmonary nodules, which are evidently ^{18}F-FDG positive (1b). The thoracic wall lesion of patient 2 is clearly ^{124}I avid (1c) and showing no uptake on the FDG-PET/CT.

overtreatment, delay, and unnecessary decrease in quality of life in a significant number of cases. As we described, a combination of diagnostic tests has a potential to allow earlier and better restaging and selection for treatment. The proposed trial aims to test the value and optimal implementation of these new tests, standalone and in combination, to derive parameters for a new personalized strategy for diagnosis and treatment of patients with (suspected) recurrent DTC.

Methods and design

Study objectives

The primary aim of the study is to evaluate the value of combined imaging with ^{124}I and ^{18}F-FDG PET/CT in the prevention of futile treatment with high therapeutic activity of ^{131}I. Interpretation of both PET-scans will lead to a positive or negative treatment proposal. This will be compared with the actual response on therapy. The definition of a futile treatment will be a negative post blind ^{131}I therapy scintigraphy.

We define four secondary aims. Firstly, our aim is to organize a synchronized introduction and QA/QC of ^{124}I PET/CT in the Netherlands. More specifically, we aimed to create a procedure for the cross-calibration of ^{124}I PET/CT in a multicenter setting, which guarantees reliable and comparable quantification, and is practical to use. The procedure should result in calibration factors per scanner and an indication of a measurement threshold of the scanner, which is defined as the lowest activity that can be reliably quantified. The measurement threshold will be determined per vendor.

Secondly, translational correlation of ^{124}I and ^{18}F-FDG PET/CT with histopathology (where available) and treatment outcome will be done, in an explorative setting. The outcome of the treatment is defined as a positive or negative post-therapy scan. This scan and both ^{124}I and ^{18}F-FDG PET/CT will be correlated with histopathological features. The expression of different markers will be quantified in the samples. In this way we aim to determine which histopathological features of both primary tumor and metastatic lesions can predict outcome of the scans.

Thirdly, the study aims to investigate whether ^{124}I PET/CT has the same diagnostic, dosimetric and prognostic yield during stimulation with rhTSH as with hormone withdrawal combined with low-iodine diet. Because ^{124}I PET/CT will be performed both after stimulation with rhTSH and after withdrawal from levothyroxine it is possible to determine any differences in outcome from the two scan preparation strategies. Both visual assessment as the quantifiable data will be compared. As simultaneous administration of ^{131}I and ^{124}I is required this can only be done in selected.

Fourthly, we aim to compare ^{124}I PET/CT and ^{131}I scintigraphy dosimetry and correlate the results with clinical outcome. As ^{124}I-PET cannot be considered as the golden standard for dosimetry of iodine therapy the

dosimetry based on ^{124}I-PET will be compared with ^{131}I-scintigraphy dosimetry. An additional phantom study will be performed to correlate the results.

Study design

This study is designed as a nationwide multicenter observational cohort study. The study population includes patients with biochemically suspicion (i.e. increase Tg levels) of recurrence of their previously completely removed thyroid carcinoma without evidence of local recurrence, planned for 'blind' therapeutic activity of ^{131}I.

The patients to be included in the study should meet the following inclusion criteria:

1. Patients with a history of differentiated thyroid cancer
2. After complete thyroidectomy and ablation of functional remnants with $_{131}$I.
3. Planned for 'blind' treatment with high activity of ^{131}I based on biochemically suspected recurrence, defined as a Tg-level above 2.0 ng/ml.
4. Ultrasonography of the neck performed < 2 months prior to inclusion.

If one of the following criteria is met patients will be excluded from the study:

1. Age < 18 years
2. Pregnancy
3. Incapacitated subjects
4. Contrast enhanced CT performed < 4 months prior to inclusion
5. ^{131}I therapy performed < 12 months prior to inclusion
6. Indication for other therapy modality (i.e. surgery in case of a positive ultrasonography, radiotherapy, embolization or chemotherapy)

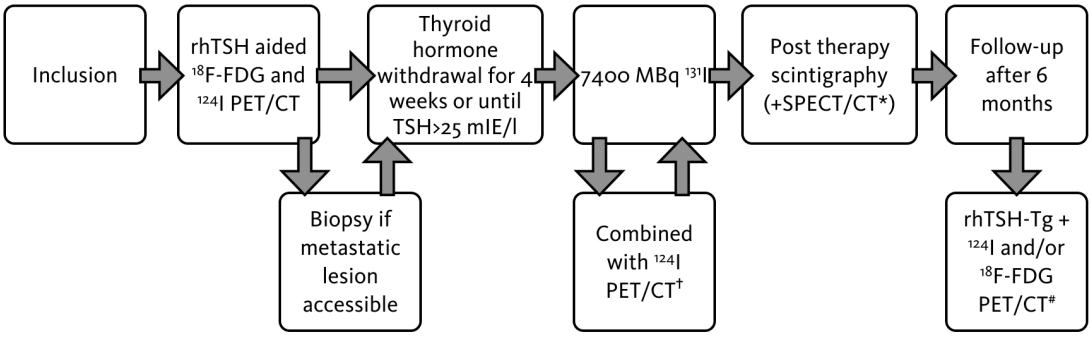


Figure 2: Flow chart THYROPET study.
Only in selected centers; if allowed according to local radiation safety regulations; * if available in center; # ^{124}I and ^{18}F-FDG PET/CT only if pre-therapy scan was positive

Study endpoints

Primary endpoint is the number of futile high-dose ^{131}I treatments that could have been avoided by implementation of pre-therapy imaging based on post-therapy scintigraphy

Four secondary endpoints were defined: (1) Synchronized QA/QC of ^{124}I-PET in the Netherlands,

(2) correlation of ^{124}I-PET/CT and FDG-PET/CT with histopathological parameters, (3) correlation between ^{124}I-PET/CT findings during rhTSH and withdrawal combined with low-iodine diet and (4) correlation between ^{124}I-PET/CT and ^{131}I-scintigraphy dosimetry

Study procedures

The study consists of four phases: pre-therapy, between pre-therapy and therapy, therapy and follow-up phase. For each phase the main study procedures are described below. Figure 2 shows an overview of the most important procedures.

Pre-therapy phase

Patients with biochemically confirmed recurrent DTC, will undergo ^{18}F-FDG and ^{124}I PET/CT imaging after pre-treatment with two injections of rhTSH. ^{18}F-FDG will be administered and ^{18}F-FDG PET/CT will be performed 60 minutes post injection. Subsequently, 74 megabecquerel (MBq) of ^{124}I is administered intravenously. ^{124}I PET/CT scans are then performed 24 and 96 hours after administration of ^{124}I.

Between 'pre-therapy phase' and 'therapy phase'

If either the ^{18}F-FDG PET/CT or the ^{124}I PET/CT shows metastatic lesions and it is possible to acquire a biopsy from the lesion, this will be done to correlate histopathological characteristics with both the result of the scans and the resection specimen of the original tumor. If multiple metastatic lesions are present on either of the scans, a biopsy will be pursued to acquire from every lesion, but only if the ^{124}I or FDG uptake differs between the different lesions. This will be done in easily accessible metastatic lesions without large risks of complications and/or discomfort for the subject.

After the pre-therapy phase, subjects will start thyroid hormone withdrawal 4 weeks prior to ^{131}I therapy. A low-iodine diet (LID) will be prescribed one week before the therapy.

Therapy phase

Subjects will undergo ^{131}I therapy with 7400 MBq of ^{131}I orally. In a subgroup of subjects (in selected centers) additional ^{124}I PET/CT scans will be performed for dosimetric evaluation. Furthermore, the influence of the method of preparation for the scan, either withdrawal of thyroid hormone or rhTSH stimulation, will be evaluated. Seven days after administration of ^{131}I a post-therapy scintigraphy is made, combined with SPECT/CT if available.

Follow-up phase

Six months after therapy both Tg and TSH levels will be determined after rhTSH administration. If the previous ^{18}F-FDG-PET/CT or the ^{124}I PET/CT showed pathological uptake, that specific PET modality will be repeated. If both PET techniques were positive during the pre-therapy phase, both the ^{18}F-FDG-PET/CT and the ^{124}I PET/CT will be repeated.

If another treatment modality, e.g. surgery, external beam radiotherapy or multikinase inhibitors, is indicated after the ^{131}I therapy the data of this additional therapy will be collected as well. If a metastatic lesion is removed surgically the histopathological specimen will be collected for additional staining and reviewing by an expert endocrine pathologist.

Additional protocol information

Histopathology thyroidectomy specimen

From every inclusded subject original resection specimens of the thyroid will be collected and if possible additional staining will be done. All specimens will be reviewed and scored by an expert endocrine pathologist.

Histopathology biopsies

If one or more biopsies are acquired from the subjects between the pre-therapy and therapy phase they will be stored fresh-frozen and analyzed later.

Review panel

The local nuclear physician will assess all scans and, additionally, an expert review panel consisting of experienced nuclear physicians will assess every scan and every lesion individually as either positive or negative. Finally, the expert panel will discuss their disagreements to reach consensus on every scan and of every lesion in each scan.

Sample size calculation

The power calculation is based on the (conservative) assumption that 40% of patients currently undergo a futile treatment. With approximately 50 evaluable patients per year in the Netherlands, we estimate we are able to include a minimum of 100 patients in 3 years. With a sample size of exactly 100 evaluable patients, a two-sided 95.0% confidence interval for a single proportion using the Pearson-Klopper method for constructing the confidence interval (exact binomial CI) will extend 10% from the observed proportion for an expected proportion of 40%.

Recruitment and consent

The patients will be selected for potential participation by the endocrinologist. After consultation with on whether the patient is eligible the local principal investigator of the study the endocrinologist informs the patient. Informed consent is acquired at least a week later by the local principal investigator.

Withdrawal of individual subjects

Subjects of the study can leave the study at any time for any reason without any consequences. The investigator can decide to withdraw a subject from the study for urgent medical reasons. For every subject that decides to withdraw from the study a new subject will be included. In this way the number of subjects included will not be changed. If subjects withdraw from the study they will be offered regular follow-up.

Follow-up of patients

Patients will receive standard follow-up according to the Dutch guidelines after the subject has completed the study.

Premature termination of the study

The study relies on ^{124}I PET/CT being predictive for ^{131}I-treatment outcome. When 3 patients have been encountered with negative ^{124}I PET/CT and positive post-therapy scintigraphy, the main clinical hypothesis can no longer be supported and the study will be stopped.

Statistical analysis

Patient demographic data, tumor characteristics and data derived from the scans will be described in frequency tables. χ^2-tests and trend tests (for ordered scales) will be used to determine whether a significant reduction in futile treatments could have been achieved by applying the ^{124}I and ^{18}F-FDG PET/CT. More in detail: interpretation of both PET-scans will lead to a positive or negative treatment proposal. This will be compared with the actual response on therapy. The definition of a futile treatment will be a negative post 'blind' ^{131}I therapy scintigraphy. Additionally, accuracy measures such as sensitivity, specificity, positive and negative predictive value will be calculated from this data. Multivariate analysis will be performed whenever appropriate using logistic regression.

Discussion

Since ^{124}I has become available for PET scanning, the interest for its use in DTC has been high. More and more studies addressed its potential use in these patients. Furthermore, it is well known that during dedifferentiation of DTC, its tumor cells may become FDG avid and multiple studies have correlated ^{18}F-FDG PET/CT with aggressiveness of DTC and the loss of iodine avidity. To our best knowledge no studies however addressed in a large prospective cohort of patients with recurrent thyroid cancer the additional value of these scan modalities in the prevention of futile ^{131}I therapies. This is of great clinical importance as a futile ^{131}I treatment is costly and not without short- and long-term side effects and should therefore be restricted to those who will likely to benefit from this treatment.

References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin 2010;60(5):277–300.

2. Schvartz C, Bonnetain F, Dabakuyo S, et al. Impact on overall survival of radioactive iodine in low-risk differentiated thyroid cancer patients. J Clin Endocrinol Metab 2012;97(5):1526–35.

3. Welsh L, Powell C, Pratt B, et al. Long-term outcomes following low-dose radioiodide ablation for differentiated thyroid cancer. J Clin Endocrinol Metab 2013;98(5):1819–25.

4. Shaha AR. Recurrent differentiated thyroid cancer. Endocr Pract 2012;18(4):600–3.

5. Schlumberger M, Challeton C, De Vathaire F, et al. Radioactive iodine treatment and external radiotherapy for lung and bone metastases from thyroid carcinoma. J Nucl Med 1996;37(4):598–605.

6. Mazzaferri EL, Kloos RT. Clinical review 128: Current approaches to primary therapy for papillary and follicular thyroid cancer. J Clin Endocrinol Metab 2001;86(4):1447–63.

7. Husson O, Haak HR, Oranje WA, Mols F, Reemst PHM, van de Poll-Franse LV. Health-related quality of life among thyroid cancer survivors: a systematic review. Clin Endocrinol (Oxf) 2011;75(4):544–54.

8. The SEER Cancer Statistics Review 1975-2010 [Internet]. [cited 2014 Jul 15]. Available from: http://seer.cancer.gov/csr/1975_2011/browse_csr.php?sectionSEL=26&pageSEL=sect_26_table.21.html

9. Francis Z, Schlumberger M. Serum thyroglobulin determination in thyroid cancer patients. Best Pract Res Clin Endocrinol Metab 2008;22(6):1039–46.

10. Bachelot A, Cailleux AF, Klain M, et al. Relationship between tumor burden and serum thyroglobulin level in patients with papillary and follicular thyroid carcinoma. Thyroid 2002;12(8):707–11.

11. Demers LM, Spencer CA, editors. Laboratory Medicine Practice Guidelines: Laboratory Support for the Diagnosis and Monitoring of Thyroid Disease. 2002;:1–125. Available from: http://www.nacb.org

12. Haugen BR, Pacini F, Reiners C, et al. A comparison of recombinant human thyrotropin and thyroid hormone withdrawal for the detection of thyroid remnant or cancer. J Clin Endocrinol Metab 1999;84(11):3877–85.

13. Mazzaferri EL, Robbins RJ, Spencer CA, et al. A consensus report of the role of serum thyroglobulin as a monitoring method for low-risk patients with papillary thyroid carcinoma. 2003. p. 1433–41.

14. Castro MR, Bergert ER, Goellner JR, Hay ID, Morris JC. Immunohistochemical analysis of sodium iodide symporter expression in metastatic differentiated thyroid cancer: correlation with radioiodine uptake. J Clin Endocrinol Metab 2001;86(11):5627–32.

15. Dadachova E, Carrasco N. The Na/I symporter (NIS): imaging and therapeutic applications. Semin Nucl Med 2004;34(1):23–31.

16. Filetti S, Bidart JM, Arturi F, Caillou B, Russo D, Schlumberger M. Sodium/iodide symporter: a key transport system in thyroid cancer cell metabolism. Eur J Endocrinol 1999;141(5):443–57.

17. de Meer SG, Vriens MR, Zelissen PM, Borel Rinkes IH, de Keizer B. The role of routine diagnostic radioiodine whole-body scintigraphy in patients with high-risk differentiated thyroid cancer. J Nucl Med 2011;52(1):56–9.

18. Pacini F, Agate L, Elisei R, et al. Outcome of differentiated thyroid cancer with detectable serum Tg and negative diagnostic (131) I whole body scan: comparison of patients treated with high (131)I activities versus untreated patients. J Clin Endocrinol Metab 2001;86(9):4092–7.

19. Cailleux AF, Baudin E, Travagli JP, Ricard M, Schlumberger M. Is diagnostic iodine-131 scanning useful after total thyroid ablation for differentiated thyroid cancer? J Clin Endocrinol Metab 2000;85(1):175–8.

20. Pacini F, Capezzone M, Elisei R, Ceccarelli C, Taddei D, Pinchera A. Diagnostic 131-iodine whole-body scan may be avoided in thyroid cancer patients who have undetectable stimulated serum Tg levels after initial treatment. J Clin Endocrinol Metab 2002;87(4):1499–501.

21. Pacini F, Molinaro E, Castagna MG, et al. Recombinant human thyrotropin-stimulated serum thyroglobulin combined with neck ultrasonography has the highest sensitivity in monitoring differentiated thyroid carcinoma. J Clin Endocrinol Metab 2003;88(8):3668–73.

22. Frasoldati A, Pesenti M, Gallo M, Caroggio A, Salvo D, Valcavi R. Diagnosis of neck recurrences in patients with differentiated thyroid carcinoma. Cancer 2003;97(1):90–6.

23. Torlontano M, Crocetti U, D'Aloiso L, et al. Serum thyroglobulin and 131I whole body scan after recombinant human TSH stimulation in the follow-up of low-risk patients with differentiated thyroid cancer. Eur J Endocrinol 2003;148(1):19–24.

24. Schlumberger M, Mancusi F, Baudin E, Pacini F. 131I therapy for elevated thyroglobulin levels. Thyroid 1997;7(2):273–6.

25. de Keizer B, Koppeschaar HP, Zelissen PM, et al. Efficacy of high therapeutic doses of iodine-131 in patients with differentiated thyroid cancer and detectable serum thyroglobulin. Eur J Nucl Med 2001;28(2):198–202.

26. Pacini F, Lippi F, Formica N, et al. Therapeutic doses of iodine-131 reveal undiagnosed metastases in thyroid cancer patients with detectable serum thyroglobulin levels. J Nucl Med 1987;28(12):1888–91.

27. Pineda JD, Lee T, Ain K, Reynolds JC, Robbins J. Iodine-131 therapy for thyroid cancer patients with elevated thyroglobulin and negative diagnostic scan. J Clin Endocrinol Metab 1995;80(5):1488–92.

28. Pachucki J, Burmeister LA. Evaluation and treatment of persistent thyroglobulinemia in patients with well-differentiated thyroid cancer. Eur J Endocrinol 1997;137(3):254–61.

29. Ma C, Xie J, Kuang A. Is empiric 131I therapy justified for patients with positive thyroglobulin and negative 131I whole-body scanning results? J Nucl Med 2005;46(7):1164–70.

30. van Tol KM, Jager PL, de Vries EGE, et al. Outcome in patients with differentiated thyroid cancer with negative diagnostic whole-body scanning and detectable stimulated thyroglobulin. Eur J Endocrinol 2003;148(6):589–96.

31. Sawka AM, Thabane L, Parlea L, et al. Second primary malignancy risk after radioactive iodine treatment for thyroid cancer: a systematic review and meta-analysis. Thyroid 2009;19(5):451–7.

32. Brown AP, Chen J, Hitchcock YJ, Szabo A, Shrieve DC, Tward JD. The Risk of Second Primary Malignancies up to Three Decades after the Treatment of Differentiated Thyroid Cancer. J Clin Endocrinol Metab 2008;93(2):504–15.

33. Schroeder PR. A Comparison of Short-Term Changes in Health-Related Quality of Life in Thyroid Carcinoma Patients Undergoing Diagnostic Evaluation with Recombinant Human Thyrotropin Compared with Thyroid Hormone Withdrawal. J Clin Endocrinol Metab 2005;91(3):878–84.

34. Luster M, Felbinger R, Dietlein M, Reiners C. Thyroid hormone withdrawal in patients with differentiated thyroid carcinoma: a one hundred thirty-patient pilot survey on consequences of hypothyroidism and a pharmacoeconomic comparison to recombinant thyrotropin administration. Thyroid 2005;15(10):1147–55.

35. Dow KH, Ferrell BR, Anello C. Quality-of-life changes in patients with thyroid cancer after withdrawal of thyroid hormone therapy. Thyroid 1997;7(4):613–9.

36. Bianchi GP, Zaccheroni V, Solaroli E, et al. Health-related quality of life in patients with thyroid disorders. Qual Life Res 2004;13(1):45–54.

37. Botella-Carretero JI, Galán JM, Caballero C, Sancho J, Escobar-Morreale HF. Quality of life and psychometric functionality in patients with differentiated thyroid carcinoma. Endocr Relat Cancer 2003;10(4):601–10.

38. Freudenberg LS, Antoch G, Jentzen W, et al. Value of (124)I-PET/CT in staging of patients with differentiated thyroid cancer. Eur Radiol 2004;14(11):2092–8.

39. Lubberink M, Abdul Fatah S, Brans B, Hoekstra OS, Teule GJJ. The role of (124) I-PET in diagnosis and treatment of thyroid carcinoma. Q J Nucl Med Mol Imaging 2008;52(1):30–6.

40. Lee J, Nah KY, Kim RM, et al. Effectiveness of [124I]-PET/CT and [18F]-FDG-PET/CT for Localizing Recurrence in Patients with Differentiated Thyroid Carcinoma. J Korean Med Sci 2012;27(9):1019.

41. Van Nostrand D, Moreau S, Bandaru VV, et al. (124)I positron emission tomography versus (131)I planar imaging in the identification of residual thyroid tissue and/or metastasis in patients who have well-differentiated thyroid cancer. Thyroid 2010;20(8):879–83.

42. Lubberink M, van Schie A, de Jong HWAM, van Dongen GAMS, Teule GJJ. Acquisition settings for PET of 124I administered simultaneously with therapeutic amounts of 131I. J Nucl Med 2006;47(8):1375–81.

43. Freudenberg LS, Jentzen W, Görges R, et al. 124I-PET dosimetry in advanced differentiated thyroid cancer: therapeutic impact. Nuklearmedizin 2007;46(4):121–8.

44. Rivera M, Ghossein RA, Schoder H, Gomez D, Larson SM, Tuttle RM. Histopathologic characterization of radioactive iodine-refractory fluorodeoxyglucose-positron emission tomography-positive thyroid carcinoma. Cancer 2008;113(1):48–56.

45. Alzahrani AS, Al-Zahrani AS, Abouzied M-EM, et al. The role of F-18-fluorodeoxyglucose positron emission tomography in the postoperative evaluation of differentiated thyroid cancer. Eur J Endocrinol 2008;158(5):683–9.

46. Mirallié E, Guillan T, Bridji B, et al. Therapeutic impact of 18FDG-PET/CT in the management of iodine-negative recurrence of differentiated thyroid carcinoma. Surgery 2007;142(6):952–8–discussion952–8.

47. Frilling A, Tecklenborg K, Görges R, Weber F, Clausen M, Broelsch EC. Preoperative diagnostic value of [(18)F] fluorodeoxyglucose positron emission tomography in patients with radioiodine-negative recurrent well-differentiated thyroid carcinoma. Ann Surg 2001;234(6):804–11.

48. Brose MS, Nutting CM, Sherman SI, et al. Rationale and design of decision: a double-blind, randomized, placebo-controlled phase III trial evaluating the efficacy and safety of sorafenib in patients with locally advanced or metastatic radioactive iodine (RAI)-refractory, differentiated thyroid cancer. BMC Cancer 2011;11:349.

49. Jentzen W, Freudenberg LS, Eising EG, Sonnenschein W, Knust J, Bockisch A. Optimized 124I PET Dosimetry Protocol for Radioiodine Therapy of Differentiated Thyroid Cancer. J Nucl Med 2008;49(6):1017–23.

50. Kim M-H, Ko SH, Bae JS, et al. Non-FDG-avid primary papillary thyroid carcinoma may not differ from FDG-avid papillary thyroid carcinoma. Thyroid 2013;23(11):1452–60.

51. Hooft L, van der Veldt AAM, van Diest PJ, et al. [18F]fluorodeoxyglucose uptake in recurrent thyroid cancer is related to hexokinase i expression in the primary tumor. J Clin Endocrinol Metab 2005;90(1):328–34.

52. Freudenberg LS, Antoch G, Frilling A, et al. Combined metabolic and morphologic imaging in thyroid carcinoma patients with elevated serum thyroglobulin and negative cervical ultrasonography: role of 124I-PET/CT and FDG-PET. Eur J Nucl Med Mol Imaging 2008;35(5):950–7.

53. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging 2009;37(1):181–200.

54. Boellaard R, Oyen WJG, Hoekstra CJ, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. Eur J Nucl Med Mol Imaging 2008;35(12):2320–33.

MULTICENTER CALIBRATION OF PET/CT SCANNERS FOR ^{124}I IN THE NETHERLANDS

CHAPTER 7

Jakob W. Kist
Manfred van der Vlies
Otto S. Hoekstra
Henry N.J.M. Greuter
Bart de Keizer
Marcel Stokkel
Wouter Vogel
Marc C. Huisman
Arthur van Lingen

EJNMMI Res 2016;6(1):39.

Abstract

Background

Studies on imaging of differentiated thyroid cancer (DTC) using ^{124}I often require a multi-center approach, as the prevalence of DTC is low. Calibration of participating scanners is required to obtain comparable quantification. As determination of a well-defined range of recovery coefficients is complicated for various reasons, a simpler approach based on the assumption that the iodine uptake is highly focal with a background that significantly lacks radioactivity might be more efficient. For each scanner a linear conversion between known and observed activity can be derived, allowing quantification that can be traced to a common source for all scanners within one study-protocol. The aim of this paper is to outline a procedure using this approach in order to set up a multicenter calibration of PET/CT scanners for ^{124}I.

Methods

A cylindrical polyethylene phantom contained six 2 ml vials with reference activities of ~2, 10, 20, 100, 400, 2000 kBq, produced by dilution from a known activity. The phantom was scanned twice on PET/CT scanners of participating centers within one week. For each scanner the best proportional and linear fit between measured and known activities were derived and based on statistical analyses of the results of all scanners it was determined which fit should be applied. In addition, a Bland-Altman analyses was done on calibrated activities with respect to reference activities to asses the relative precision of the scanners.

Results

Nine Philips (vendor A) and nine Siemens (vendor B) PET/CT scanners were calibrated in a time period of 3 days before and after the reference time. No significant differences were detected between the two subsequent scans on any scanner. Six fitted intercepts of vendor A were significant different from zero, so the linear model was used. Intercepts ranged from -8 kBq to 26 kBq and slopes ranged from 0.80 to 0.98. Bland-Altman analyses of calibrated and reference activities showed that the relative error of calibrated activities was smaller than of uncalibrated activities.

Conclusion

A simplified multicenter calibration procedure for PET/CT scans that show highly focal uptake and negligible background is feasible and results in more precise quantification. Our procedure can be used in multicenter ^{124}I PET scans focusing on (recurrent) DTC.

Introduction

Iodine-124 (^{124}I) is currently of great interest as a PET(/CT) tracer in patients with (metastasized) differentiated thyroid cancer (DTC) for pre-therapeutic assessment of iodine avidity of lesions and for dosimetric purposes.[1-4] Dosimetry requires reliable quantification, and, as gathering strong clinical evidence in this relatively rare disease requires multicenter studies to ensure sufficient patient enrollment, calibration of scanners is required.[5] The concurrent emergence of the EANM Research Ltd (EARL) accreditation procedure for 2-[18F]fluoro-2-deoxy-D-glucose (^{18}F-FDG) imaging aims to achieve comparable scanner performances across multiple sites through harmonization of the acquisition of PET/CT scans.[6] However, in case of ^{124}I used in DTC patients, a standardization strategy as used for ^{18}F-FDG is not adequate, because recovery of the partial volume effect is difficult to determine.[7] Due to the combination of the low positron abundance (around 25%) and the presence of 602 keV non-annihilation photons, image-derived activity concentrations are inaccurate.[7,8] Documented recovery coefficients depend on object volume and shape, background activity, voxel size and number of effective iterations.[7] Due to the high specificity of iodine for thyroid tissue, the uptake in the background is negligible. This allows to determine the total activity within the lesions by drawing an oversized volume of interest (VOI) around the imaged target, avoiding the unknown influence of the partial volume effect. The lesion uptake in units of activity concentration or standardized uptake value (SUV, %) can be calculated from the total activity and the lesion volume, determined from anatomical imaging, e.g. a CT scan.

By measuring a range of ^{124}I activities in a phantom experiment, a linear relation between the reference and measured activities can be derived per scanner. Hence, measured activities can be converted to calibrated activities for all scanners used in a multicenter study. Knowledge of the inaccuracies of individual scanners allows for benchmarking and thereby determining underperforming scanners. Excluding these scanners will improve the overall accuracy of the quantification of ^{124}I in a multicenter study and thereby its quality. The aim of this paper is to outline a procedure for multicenter calibration of the total activity of ^{124}I in focal, low background areas.

Methods

A cylindrical polyethylene phantom containing six 2 ml cylindrical glass vials (figure 1), representing typical lesion volumes was developed in-house. By weighing and diluting from a known activity of ^{124}I (BV Cyclotron, Amsterdam, the Netherlands) reference activities (A_r) of approximately 2, 10, 20, 100, 400 and 2000 kBq at reference time (T_r) were obtained and put into the vials. The activity-series was based on the Thyropet protocol, the optimized dosimetry protocol by Jentzen et al. and calculations with the iodine kinetic model

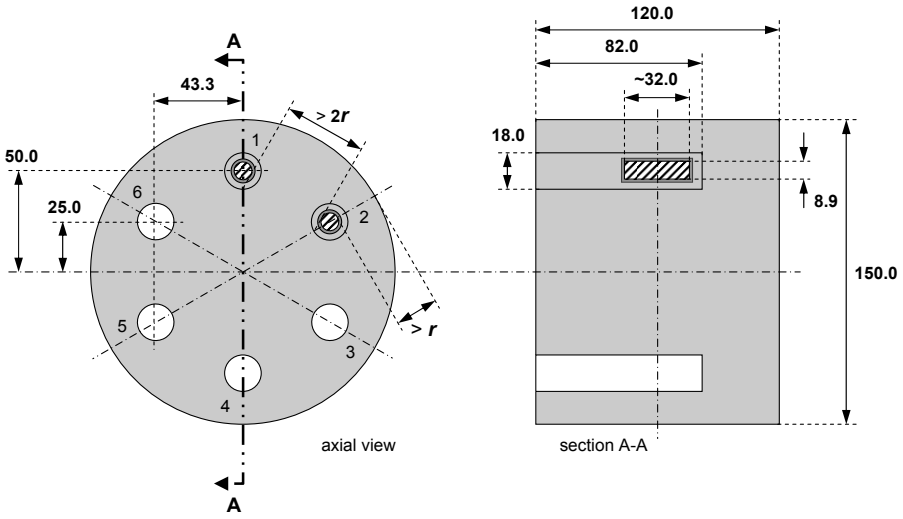


Figure 1: Design of the phantom. All sizes are in millimeters. The height is lower than the smallest axial field-of-view of the PET/CT scanners included for calibration. The phantom consists of solid polyethylene (light gray) and contains six openings (indicated 1 to 6; 3 to 6 colored white). In each opening, a polyethylene socket can be placed tightly, enclosing an amber glass vial (dark gray, wall thickness about 1.5 mm) of 2 mL inner volume, filled with ^{124}I activity (shaded). The position of the vials in the phantom is such that the ^{124}I positrons (maximum range r) do not interfere between the vials and do not reach the outside of the phantom. The vials have a neck and screw cap, so the shape and height are an indication.

described in ICRP publication 53.[3,9-11] No activity in the background was used. A non-radioactive solution of approximately 1 mg/mL iodine$^-$ was used for dilution to prevent $^{124}I^-$ from sticking to the walls. Reference activities at the time of calibration of the scanners (T_c) were obtained by correction for decay of ^{124}I (half-life 100.2 hours)

The phantom was scanned on 18 PET/CT scanners, nine from Philips (Philips Healthcare, Best, the Netherlands) (vendor A) and nine from Siemens (Siemens Medical Solutions, Erlangen, Germany) (vendor B), during a period of three days before and three days after the reference time (Table 1). Before the calibration a dedicated ^{124}I scan protocol was implemented and tested on each scanner. In the clinical setting the duration of a whole body patient scan preferably is limited to 30 minutes. Therefore, the scan time per axial field of view (FOV) was two minutes for PET/CT scanners of vendor A and four minutes for scanners of vendor B. This difference is a consequence of the difference in length of the effective FOV of the two vendors: ~9 cm for vendor A and ~14-18 cm for vendor B. The standard energy window (approximately 350-680 keV) was applied. All common acquisition corrections were applied, i.e. normalization and corrections for scatter and attenuation, decay and dead

Site no.	Vendor	ToF	Slices CT	EARL*	Reconstruction protocol^	Voxel size (mm^3)
1	A	Yes	16	Yes	BLOB-OS-TF	64
2	A	No	16	No	LOR-RAMLA	64
3	A	Yes	16	No	BLOB-OS-TF	64
4	A	Yes	16	Yes	BLOB-OS-TF	64
5	A	Yes	16	Yes	BLOB-OS-TF	64
6	A	Yes	64	No	BLOB-OS-TF	64
7	A	Yes	64	Yes	BLOB-OS-TF	64
8	A	Yes	64	Yes	BLOB-OS-TF	64
9	A	Yes	16	No	BLOB-OS-TF	64
10	B	No	16	Yes	OSEM 2D	14,2
11	B	Yes	40	No	PSF+TOF	49,8
12	B	Yes	64	Yes	PSF+TOF	20,2
13	B	Yes	40	Yes	PSF+TOF	30,4
14	B	Yes	64	Yes	PSF+TOF	11,6
15	B	No	16	No	OSEM 3D	82,9
16	B	No	16	No	OSEM 3D	82,9
17	B	No	40	Yes	OSEM 3D	21,4
18	B	No	40	No	OSEM 3D	49,7

Table 1: Characteristics of included scanners.
*Scanner accredited by EANM Research Ltd (EARL). ^Reconstruction protocol name as named by vendor in DICOM header

time. If the scanner was EARL accredited, EARL reconstruction parameters were used [12]. Else, 2D or 3D OSEM reconstructions with sufficient convergence and a 5 mm full width at half maximum Gaussian reconstruction filter were utilized. If available, the time of flight option on the scanners was applied for acquisitions and in reconstruction.

Two axial FOVs were centered around the vials and scanned subsequently. The start time of the scan was defined as the calibration time (T_c). On each scanner the phantom was scanned twice: after the first scan it was turned 180 degrees around the axial axis to determine the reproducibility of the quantification procedure.

A large VOI of 110 mL (46 mm diameter, 66 mm height) was placed around each 2 mL vial (figure 2a and 2d). No interobserver variation was expected and images were analyzed by one person (JK), using the open source software Osirix™ (version 4.1.2, 64 bit).[13] Measured activities ($A_{m,o}$ and $A_{m,180}$) were obtained for each acquisition (0 and 180 degrees) and vial.

Per scanner the data points $A_{m,o}$ and $A_{m,180}$ were compared with a two-

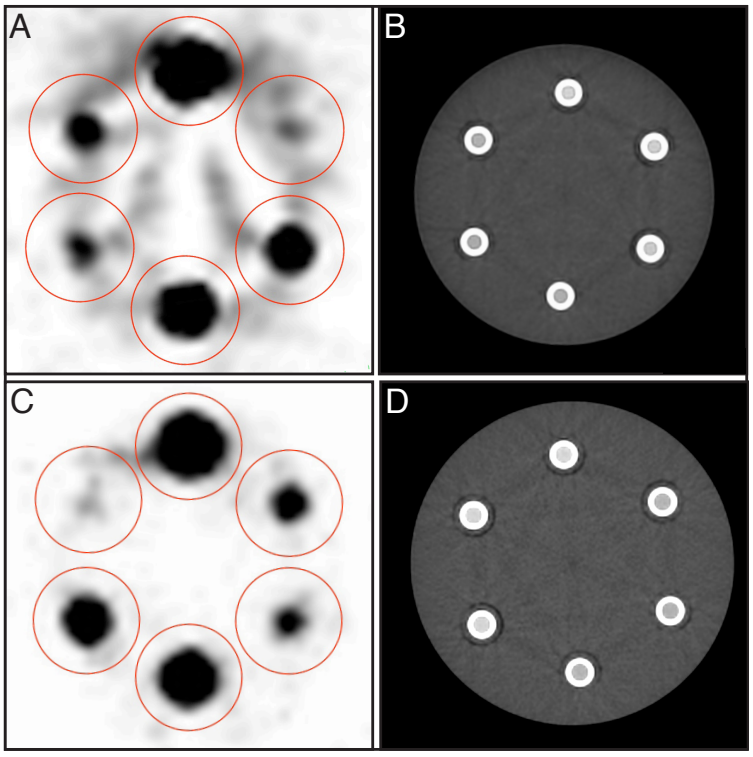


Figure 2: Examples of PET and CT images of axial slices acquired from scanner of vendor A (2A-B; obtained from scanner number 5) and vendor B (2C-D; obtained from scanner number 15) to emphasize the differences between the two scanner vendors. Both scanners were calibrated shortly after each other, so decay of the reference activities was negligible. At the time of calibration, the top vials contained a reference activity A_r of 2.49 MBq. Clockwise, the other vials contained 2.6, 127, 514, 13 and 26 kBq, respectively. Figures 2A and 2C: CTAC images, with circular regions of interest matching with the VOIs (in red). Figures 2B and 2D; corresponding CT images.

sided paired T-test. If no significant difference ($p < 0.05$) was obtained, the measurements were considered reproducible and only $A_{m,o}$ was used for further analyses. Data points from $A_{m,o}$ were fitted by linear regression to the function $A_{m,o} = \alpha_{cal} + \beta_{cal} \cdot A_r$, with and without the additional constraint that the intercept α equals zero, so both a proportional and a linear model were investigated. The standard error of the intercept, $\Delta\alpha$, was calculated with a two-sided Student's t-distribution ($p < 0.05$). If and only if, none of the intercepts differed statistically significant from zero, the model without intercept, i.e. $A_{m,o} = \beta_{cal} \cdot A_r$, was applied, otherwise the model with intercept was used ($A_{m,o} = \alpha_{cal} + \beta_{cal} \cdot A_r$).

The estimated slope and, if applicable, intercept of each scanner were used to calculate the calibrated activity, A_{cal}, for each measured activity in each vial, so $A_{cal} = A_{m,o,cal}/\beta_{cal}$ or $A_{cal} = (A_{m,o} - \alpha_{cal})/\beta_{cal}$.

A Bland-Altman analysis was performed for the calibrated activities to assess the relative precision of the scanners. The relative errors between calibrated activity (A_{cal}) and reference activity (A_r) were calculated for each scanner as an absolute value: $E_{cal} = |(A_r - A_{cal}) / A_r|$. Subsequently, the measured activities were analyzed with Bland-Altman, again expressed as an absolute value: $E_{m,o} = |(A_r - A_{m,o}) / A_r|$, to investigate whether calibration improved the measured activities.

All analyses were carried out in Microsoft Excel for Windows, version 2003 (Microsoft corp., Redmond, WA, USA).

Results

The phantom was scanned on 18 PET/CT scanners, 9 scanners of vendor A and 9 of vendor B, in 16 hospitals in the Netherlands (table 1). Figure 2 shows typical images obtained from scanners of vendor A and B (figure 2A and 2D).

No significant difference between the measured activities $A_{m,o}$ and $A_{m,180}$ were found in any of the vials of any of the scanners tested. The intercepts (α_{cal}) were significantly different from zero in six out of nine scanners of vendor A (range α_{cal}: 3-26 kBq, range $\Delta\alpha_{cal}$: 4-29 kBq). None of the intercepts derived from scanners of vendor B differed significantly from zero (range α_{cal}: -8-11 kBq, range $\Delta\alpha_{cal}$: 1-18 kBq). The significant differences between intercept and zero found in six scanners implied the use of the linear model for calibration for all scanners. Applying this model, the β_{cal} for scanners of vendor A ranged from 0.80 to 0.98, whereas the β_{cal} for scanners of vendor B ranged from 0.85 to 0.97. Two example plots, one scanner of each vendor, of the measured activity ($A_{m,o}$) against the reference activity A_r, including fit to the linear model, are displayed in figure 3.

Figure 4 shows the relative errors of the measured activities before calibration (blue data points) and the calculated activities after calibration (green data points). The calibration procedure reduced the relative error in all decades of reference activities. This is depicted by the green and blue line in figure 4, indicating the average relative error of the measured and calibrated activities, respectively, for each decade of reference activities. The data in figure 4 also indicate in a higher precision the majority of scanners of vendor B. If the relative errors of the measured activities by the scanners are compared to a reasonably chosen threshold of 50%, figure 4 shows that the first scanner of vendor A exceeds 50% at 43 kBq, while the first scanner of vendor B exceeds this percentage at 16 kBq. After calibration these activities are 12 kBq (vendor A) and 2.6 kBq (vendor B).

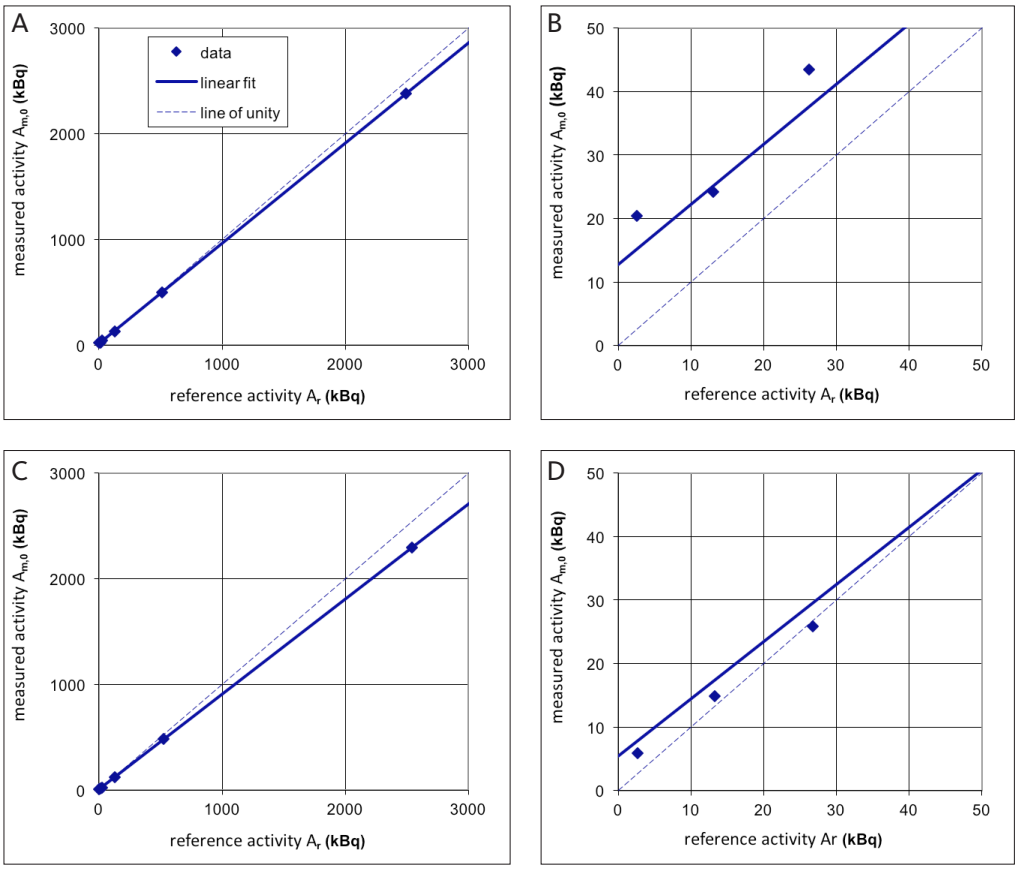


Figure 3: Graphs reported to centers of the involved scanners as example of the differences between the scanner vendors. Linear fit of measured and reference activity data according to $A_{m,o} = \alpha cal + \beta_{cal} \cdot Ar$. Top row: scanner number 5 (vendor A; α_{cal} = 13 kBq and β_{cal} = 0.947), bottom row: scanner number 15 (vendor B; α_{cal} = 5.5 kBq and β_{cal} = 0.901). Note that the graphs on the right side are the same as the corresponding ones on the left side, apart from the scales of the axes. Furthermore, the larger offset α_{cal} of scanner 5 compared to scanner 15 can be foreseen from the larger background signal, shown in figure 2A and 2C respectively.

Discussion

In this study we present a calibration procedure attaining calibrated scanners for multicenter studies using ^{124}I and focusing on DTC, and proved it to be feasible. Eighteen PET/CT scanners across the Netherlands were calibrated within one week with a convenient and easy to handle phantom, using a single set of ^{124}I reference activities and predefined scanner specific acquisition and reconstruction protocols. The estimated parameters of the applied model were reproducible, making the procedure robust. The calibration resulted in a

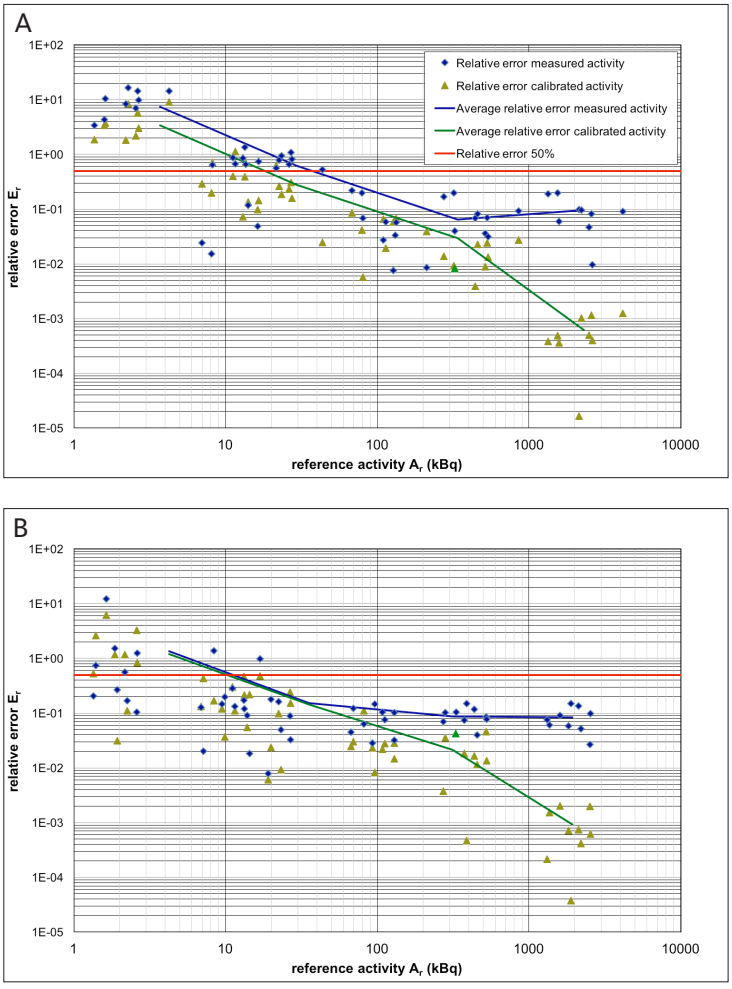


Figure 4: Bland-Altman plots of measured activity $A_{m,o}$ and calibrated activity A_{cal}, versus reference activity A_r for all scanners, shown per vendor (Figure 4A: vendor A; figure 4B: vendor B). For presentation purposes, axes are on a log-log scale. and the relative errors are calculated as an absolute value: $E_r = |(A_r - A_m) / A_r|$ and $E_r = |(A_r - A_{cal}) / A_r|$. To assess improvement of calibration, the average of reference activities A_r and corresponding average relative errors are calculated for each reference activity decade. The blue and green solid lines indicate the average relative errors in measured and calibrated activities, respectively

decrease in relative errors of calibrated activities compared to (non-calibrated) measured activities (figure 4), making quantitative data comparable amongst different centers.

Our relatively simple approach could be used in other clinical multicenter studies focusing on other diseases and tracers, provided the assumptions of focal uptake and negligible background are fulfilled. This might be the case for other ^{124}I tracers and for ^{68}Ga, ^{18}F and ^{81}Zr tracers.

Standardized and reliable quantification of ^{124}I PET imaging is essential in the design of multicenter and dosimetry studies. For example, in a mono-center study by Ho et al. dosimetric analysis of multiple ^{124}I PET/CT scans before and after selumetinib treatment of patients with radioiodine refractory metastases was performed. The outcome of this analysis determined whether a new high dose ^{131}I treatment would be beneficial.[1] The results of this study are

promising, however, in order to expand and validate these, multicenter studies are warranted. This underlines the need for multicenter standardization of scanning and quantification.

The phantom and the procedure were designed with the aim of being reliable, efficient and safe. Therefore, the design of the phantom used no ^{124}I background activity. Although different from previously published calibration procedures for other isotopes[6,14], the absence of background activity seemed justified from a clinical perspective and it simplified the calibration procedure significantly. Iodine uptake is highly specific for thyroid tissue and very limited to only a few non-thyroid tissues, like the salivary glands, gastric mucosa and choroid plexus.[15] These are, however, not in areas of clinical relevance, so uptake in these organs is extraneous for patient image analysis. The main advantage of an empty background was that the partial volume effect could be dealt with straightforward. No partial volume correction due to spill in of ^{124}I background signal into the signal of the vials was necessary. Additionally, correction for spill out from the vials was possible by drawing relatively large fixed VOIs.[16] By using this method variable lesion volumes in the phantom were not a requisite. Some practical issues of the calibration procedure could be handled more straightforward by omitting background activity. For instance, no solution with a well-known ^{124}I background activity had to be produced, making the phantom preparation less complex. Furthermore, legislation for the transportation of the reference activities to all involved centers, as well as for the handling of the activities at the centers, could be met with limited effort due to the lower total activity.

The reference activities used to determine the calibration parameters are traceable to one standard activity. The calibration procedure was designed to support a multicenter study using the same supplier of ^{124}I for the study-related patient scans as for the calibration procedure.[3,11] Since production and activity assessment protocols used by the supplier are standardized, activities used in the calibration procedure and for patient scans in the multicenter study are traceable to a higher standard. In this way difficulties with activity assessment by dose calibrators as described by Beattie et al. are overcome.[17] If in the future, however, multicenter studies using isotopes produced by more than one supplier, reference activities should be related to a higher, preferably international, primary standard.[18] Potentially, organizations like EARL or EATRIS could play a pivotal role in the development of these standards.[12,19]

The process of dilution and weighing used to produce the reference activities had the risk that an inexact reference activity concentration propagated to the next lower concentration. However, this method is most likely more accurate than by direct measurement with, e.g. a dose calibrator, due to the low signal to noise ratio at low activities.[17] Therefore, this process was preferred, while the risk of propagation of inaccurate concentrations was minimized with an indicative, direct measurement of the produced activity concentrations in each process step.

From the assessment of the precision it appears that after calibration

the relative error of the scanners of both vendors exceeds the 50% level at activity levels lower then 2-20 kBq (figure 4). It should therefore be kept in mind, if in the clinical setting the uptake is below these lower limits, reliable quantification becomes inaccurate and should not be used for dosimetric calculations. Additionally, this method can be used to assess individual scanners precision in order to exclude underperforming scanners from multicenter studies, providing that multiple measurements per scanner are used.

Conclusions

A simplified multicenter calibration procedure for ^{124}I PET/CT scans in DTC is feasible and results in smaller relative errors in ^{124}I quantification. In the future quantification will be of growing importance especially in multicenter clinical trials and therefore, standardized calibration procedures need to become applied widely. Our procedure can be used in multicenter ^{124}I PET scans focusing on (recurrent) DTC.

References

1. Ho AL, Grewal RK, Leboeuf R, et al. Selumetinib-enhanced radioiodine uptake in advanced thyroid cancer. N Engl J Med 2013;368(7):623–32.
2. Pettinato C, Spezi E, Nanni C, et al. Pretherapeutic dosimetry in patients affected by metastatic thyroid cancer using 124I PET/CT sequential scans for 131I treatment planning. Clin Nucl Med 2014;39(8):e367–74.
3. Kist JW, de Keizer B, Stokkel MPM, Hoekstra OS, Vogel WV, THYROPET study group. Recurrent differentiated thyroid cancer: towards personalized treatment based on evaluation of tumor characteristics with PET (THYROPET Study): study protocol of a multicenter observational cohort study. BMC Cancer 2014;14(1):405.
4. Jentzen W, Hoppenbrouwers J, van Leeuwen P, et al. Assessment of Lesion Response in the Initial Radioiodine Treatment of Differentiated Thyroid Cancer Using 124I PET Imaging. J Nucl Med 2014;55(11):1759–65.
5. Lassmann M, Reiners C, Luster M. Dosimetry and thyroid cancer: the individual dosage of radioiodine. Endocr Relat Cancer 2010;17(3):R161–72.
6. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42(2):328–54.
7. Jentzen W. Experimental investigation of factors affecting the absolute recovery coefficients in iodine-124 PET lesion imaging. Phys Med Biol 2010;55(8):2365–98.
8. Freudenberg LS, Jentzen W, Görges R, et al. 124I-PET dosimetry in advanced differentiated thyroid cancer: therapeutic impact. Nuklearmedizin 2007;46(4):121–8.
9. Jentzen W, Freudenberg LS, Eising EG, Sonnenschein W, Knust J, Bockisch A. Optimized 124I PET Dosimetry Protocol for Radioiodine Therapy of Differentiated Thyroid Cancer. J Nucl Med 2008;49(6):1017–23.
10. ICRP. Radiation Dose to Patients from Radiopharmaceuticals. Ann ICRP 1988;(18):1–4.
11. Kist JW, de Keizer B, van der Vlies M, et al. 124I PET/CT to predict the outcome of blind 131I treatment in patients with biochemical recurrence of differentiated thyroid cancer; results of a multicenter diagnostic cohort study (THYROPET). J Nucl Med 2015;
12. ResEARch 4 Life, EANM Research Ltd. EARL [Internet]. [cited 2014 Jan 19]; Available from: http://earl.eanm.org/cms/website.php
13. Rosset A, Spadola L, Ratib O. OsiriX: an open-source software for navigating in multidimensional DICOM images. J Digit Imaging 2004;17(3):205–16.
14. Makris NE, Boellaard R, Visser EP, et al. Multicenter harmonization of 89Zr PET/CT performance. J Nucl Med 2014;55(2):264–7.
15. Zimmermann MB. Iodine deficiency. Endocr Rev 2009;30(4):376–408.
16. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. J Nucl Med 2007;48(6):932–45.
17. Beattie BJ, Pentlow KS, O'Donoghue J, Humm JL. A Recommendation for Revised Dose Calibrator Measurement Procedures for 89Zr and 124I. PLoS ONE 2014;9(9):e106868.
18. Pommé S. Methods for primary standardization of activity. Metrologia 2007;44(4):17–26.
19. van Dongen GA, Ussi AE, de Man FH, Migliaccio G. EATRIS, a European initiative to boost translational biomedical research. Am J Nucl Med Mol Imaging 2013;3(2):166–74.

^{124}I PET/CT TO PREDICT THE OUTCOME OF BLIND ^{131}I TREATMENT IN PATIENTS WITH BIOCHEMICAL RECURRENCE OF DIFFERENTIATED THYROID CANCER; RESULTS OF A MULTICENTER DIAGNOSTIC COHORT STUDY (THYROPET)

CHAPTER 8

Jakob W. Kist
Bart de Keizer
Manfred van der Vlies
Adrienne H. Brouwers
Dyde A. Huysmans
Friso M. van der Zant
Rick Hermsen
Marcel P.M. Stokkel
Otto S. Hoekstra
Wouter V. Vogel
On behalf of the Thyropet study group*

J Nucl Med 2016;57(5):701–7.

* THYROPET Study group: John M.H. de Klerk, Harm van Tinteren, Jan Paul de Boer, Hans Morreau, Marc C. Huisman, Eef G.W.M. Lentjes, Thera P. Links, Jan W.A. Smit, Jules Lavalaye, Piet L. de Jager, Corneline J. Hoekstra, Martin Gotthardt, Vanessa J.R. Schelfhout, Wieger I. de Bruin, Ferida Sivro, Judit A. Adam, Ha T.T. Phan, Gerrit W. Slooff, and Nils R.L. Wagenaar

Abstract

Introduction

Patients with suspected recurrence from differentiated thyroid carcinoma (DTC), based on an increased thyroglobulin (Tg) level and negative neck ultrasound (US), pose a clinical dilemma. Since standard imaging has a low yield identifying potential recurrence, 'blind' ^{131}I treatment is often applied. However, a tumor-negative ^{131}I whole body scintigraphy (WBS) prevails in 38-50% of patients. We performed a prospective multicenter observational cohort study to test the hypothesis that ^{124}I PET/CT can identify the patients with a tumor negative post-therapy ^{131}I WBS.

Methods

Our study was designed to include 100 patients with detectable Tg and a negative neck US, who were planned for 'blind' ^{131}I therapy. All patients underwent ^{124}I PET/CT after rhTSH stimulation. Subsequently, after 4-6 weeks of thyroid hormone withdrawal patients were treated with 5.5-7.4 GBq ^{131}I, followed by WBS a week later. The primary endpoint was the number of ^{131}I therapies that could have been omitted using the predicted outcome of the ^{124}I PET/CT, operationalized as the concordance of tumor detection by ^{124}I PET/CT, using post-^{131}I therapy WBS as the reference test. The study would be terminated if three patients had a negative ^{124}I PET/CT and a positive post-therapy ^{131}I.

Results

After inclusion of 17 patients we terminated the study preliminarily, as the stopping rule had been met. Median Tg-level at ^{131}I therapy was 28 µg/L (interquartile range: 129). Eight post-therapy WBS were negative (47%), all of which correctly predicted by negative ^{124}I PET/CT. Nine post-therapy WBS showed iodine avid tumor, of which four also had positive ^{124}I PET/CT findings. Sensitivity, specificity, negative predictive value and positive predictive value of ^{124}I PET/CT were 44% (CI 14-79%), 100% (CI 63-100%), 62% (CI 32-86%) and 100% (CI 40-100%), respectively. Implementation of ^{124}I PET in this setting would have led to 47% (8/17) less futile ^{131}I treatments, but 29% of patients (5/17) would have been denied potentially effective therapy.

Conclusion

In patients with biochemical evidence of recurrent DTC and a tumor negative neck ultrasound, the high false negative rate of rhTSH stimulated ^{124}I PET/CT as implemented in this study precludes its use as a scouting procedure to prevent futile blind ^{131}I therapy.

Introduction

Differentiated thyroid cancer (DTC) incidence is rising and it is the most prevalent endocrine cancer. Although patients have an excellent 10-year survival rate of over 95%[1,2], up to 25% of patients will face recurrent locoregional disease or distant metastasis.[3] Prognosis is less favorable when recurrences and metastases occur.[1] When recurrence is suspected based on serum Thyroglobulin (Tg) levels, without clinical evidence of locoregional metastasis, patients are treated empirically with high dose iodine-131 (^{131}I).[4,5] However, this 'blind' therapy can be considered as futile in the 38-50% of patients with a tumor negative post-therapeutic whole body scintigraphy (WBS).[6-9] On top of that, ^{131}I therapy induces substantial short- and long-term morbidity due to hormone withdrawal associated hypothyroidism, early and late sialoadenitis in up to 30% which can lead to xerostomia, dental cariës and stomatitis.[10,11] Moreover, societal costs are considerable due to productivity loss.[12,13] Therefore, there is a need for a diagnostic modality to predict which patients are likely to benefit from high dose ^{131}I therapy. The yield of low dose ^{131}I- and of iodine-123 (^{123}I) WBS in this setting is low so that they are no longer recommended.[14-17]

Iodine-124 (^{124}I) positron emission tomography/computed tomography (PET/CT) has been investigated in DTC patients for several years.[18-20] Apart from the superior imaging characteristics of PET (compared to SPECT), ^{124}I PET/CT offers lower radiation exposure for the patient in comparison to blind ^{131}I therapy, and it may be a tool for pre-therapeutic dosimetry.[8,18,21] Taken together, ^{124}I PET/CT potentially allows for accurate restaging of DTC, prediction of ^{131}I therapy outcome, and better selection of patients for ^{131}I treatment. However, prospective studies with clear eligibility criteria and standardized procedures, comparing ^{124}I PET/CT outcomes with post-therapy ^{131}I WBS in patients treated 'blind' with ^{131}I are lacking.

Iodine uptake in metastases can be stimulated exogenously, using recombinant human thyroid stimulating hormone (rhTSH)(Thyrogen®, Genzyme Corporation, Cambridge, MA, USA) or endogenously by thyroid hormone withdrawal (THW). As the latter induces morbidity associated with severe hypothyroidism and may stimulate tumor growth, pre-therapeutic scanning, aiming to predict uptake on ^{131}I WBS (and potential therapy efficacy), is preferably done after rhTSH stimulation. However, it is unclear whether and to which extent patient preparation with rhTSH rather than THW affects the diagnostic accuracy of ^{124}I PET/CT.

The clinical research question of this study was to determine whether futile ^{131}I treatment could be prevented by pre-therapeutic imaging with ^{124}I PET/CT, in patients with biochemical suspicion of recurrence without clinical evidence of locoregional metastases. Furthermore, we aimed to investigate the effect of the TSH stimulation method on diagnostic performance of ^{124}I PET/CT in this clinical setting.

Methods

Study design

The THYROPET study was designed as a prospective nationwide multicenter diagnostic cohort study (Clinicaltrials.gov identifier: NCT01641679).[22]. The study was approved by the institutional review board of the Netherlands Cancer Institute and by the local institutional review boards of the 17 participating centers. This study was investigator initiated, and conducted in accordance with the principles of the Declaration of Helsinki and Good Clinical Practice guidelines. All participants provided written informed consent.

Patients

Adult patients with biochemical suspicion of recurrence of previously treated DTC (serum Tg >2 ng/ml), without evidence of locoregional recurrence (negative ultrasound of the neck), who were planned for 'blind' treatment with high dose ^{131}I were eligible for inclusion (Table 1). Main exclusion criteria were recent ^{131}I therapy (<12 months before inclusion) and any indication for another treatment modality (e.g. surgery).

Procedures

Pre-therapeutic phase.

Patients received 0.9 mg rhTSH intramuscularly on two consecutive days, one day after the second rhTSH administration, 74 MBq of ^{124}I was administered intravenously, followed by ^{124}I PET/CT 24 and 96 hours later.

Inclusion criteria	
1.	Patients with a history of DTC
2.	After complete thyroidectomy and ablation of functional remnants with ^{131}I
3.	Planned for 'blind' treatment with high activity of ^{131}I based on biochemically suspected recurrence, defined as a Tg-level above 2.0 ng/ml
4.	Ultrasonography of the neck performed < 2 months prior to inclusion

Exclusion criteria	
1.	Age < 18 years
2.	Pregnancy
3.	Incapacitated subjects
4.	Contrast enhanced CT performed < 4 months prior to inclusion
5.	^{131}I therapy performed < 12 months prior to inclusion
6.	Indication for other therapy modality (i.e. surgery in case of a positive ultrasonography, radiotherapy, embolization or chemotherapy)

Table 1: In- and exclusion criteria of THYROPET study

Therapeutic phase

Patients were requested to keep a one-week low iodine diet prior to ^{131}I therapy. When TSH was ≥25 mIU/L or after at least 4 weeks of THW, patients were admitted for ^{131}I therapy (5.5 or 7.4 GBq). In centers where local radiation safety regulations allowed such, patients underwent repeat ^{124}I PET/CT during ^{131}I treatment: a second dose of ^{124}I was intravenously injected directly after the administration of ^{131}I, and a PET/CT was performed after 24 and 96 hours. One week after ^{131}I all patients underwent post-therapy WBS and/or SPECT/CT. In addition (data not presented here), all patients underwent ^{18}F-FDG PET/CT in the pre-therapeutic phase, and ^{18}F-FDG and ^{124}I PET/CT six months after therapy.[22]

Image acquisition

PET/CT scanners of participating centers were calibrated for ^{124}I to ensure adequate image quality.[23] All ^{124}I PET/CT scans were performed according to the scan protocol associated to this study, which included an administered dose of 74 MBq and optimized settings for ^{124}I. Scan trajectories (PET and WBS) covered perineum-skull vertex, for a scan time of 30 minutes. A standard energy window was applied for all PET acquisitions. EARL reconstruction parameters were used.[24] If available, time of flight acquisition and reconstruction was performed. WBS was performed as planar scintigraphy, combined with SPECT/CT according to standard local procedures.

Image interpretation

An expert panel of three independent experienced nuclear medicine physicians reviewed all PET/CT scans, blinded to the ^{131}I WBS results. The post-therapy ^{131}I WBS were assessed separately by two other experienced nuclear medicine physicians. All scans were scored as malignant, equivocal or non-suspicious. Any disagreement between reviewers was discussed in order to reach consensus. If a scan was scored as equivocal it was considered malignant in the analysis to obtain maximal sensitivity.

Quantitative analysis of ^{124}I PET/CT was performed using manually drawn VOIs, measuring the total activity (kBq) within the lesion, corrected with the calibration factor determined for that scanner at calibration.[23]

Outcomes

The primary endpoint of this study was the accuracy of ^{124}I PET/CT to predict, at a patient level, the post-therapy ^{131}I WBS test result, as an operationalization of the impact of the implementation of ^{124}I PET/CT as a scouting procedure to set the indication for ^{131}I therapy. Secondary end point was a quantitative and visual comparison between ^{124}I PET/CT performed after rhTSH stimulation (rhTSH-^{124}I PET/CT) and after thyroid hormone withdrawal combined with low-iodine diet (THW-^{124}I PET/CT).

Statistical methods

The study was designed to define the accuracy of ^{124}I PET/CT to predict the result of post high dose ^{131}I WBS. The power calculation was based on the (conservative) assumption of 40% futile 'blind' ^{131}I treatments (6,7). With a sample size of 100 evaluable patients, a two-sided 95.0% confidence interval (CI) for a single proportion using the Pearson-Clopper method for constructing the confidence interval (exact binomial CI) would extend 10% from the observed proportion for an expected proportion of 40%.

As wrongfully withholding potentially curative treatment for these patients is unacceptable, we decided to stop the study if in 3 patients the ^{124}I PET/CT scan turned out false-negative (one-sided 95%-CI upper limit 12%). There was continuous monitoring of false negative ^{124}I PET/CTs by the local study coordinators, followed by central adjudication by the central expert panel.

Patient demographic data, tumor characteristics and data derived from the scans are described in frequency tables. Accuracy measures such as sensitivity, specificity, positive predictive value and negative predictive value were calculated using the per-patient result of ^{124}I PET/CT and ^{131}I WBS as index and reference test, respectively.

Results

The study was open for inclusion from December 2012 to May 2014. In May 2014, a third false-negative ^{124}I PET/CT (vs. ^{131}I WBS) was reported to the study coordinators, and when this finding was confirmed at central review, the study was preliminarily terminated according to the predetermined stopping rule. As safety was not compromised, study procedures were completed in all 19 patients included until then. All patients underwent rhTSH-^{124}I PET/CT in the pre-therapeutic phase, and five of them THW-^{124}I PET/CT in the therapeutic phase.

Of these 19 patients, two were excluded from the ^{124}I PET/CT vs. ^{131}I WBS comparison: one because other imaging, performed because of clinical signs and symptoms, had shown numerous distant metastases, among which vertebral metastases threatening the spinal cord. In agreement with an escape rule in the protocol (i.e. allowing the the attending physician to withdraw a subject from the study for urgent medical reasons), it was decided to refrain from ^{131}I therapy and to start immediate palliative radiotherapy on the spinal metastases. Another patient was excluded since the elevated Tg-level at time of inclusion was not reproduced at an additional assessment during the study before ^{131}I therapy.

The baseline characteristics of the 17 evaluable patients are listed in Table 2. Nine patients had previously received more than one ^{131}I treatment, with a mean interval between last ^{131}I treatment and inclusion of 5.6 years; the median serum Tg-level at ^{131}I therapy was 28 µg/L (interquartile range 129).

	n or mean (±SD)	% or range
Age at inclusion (y)	56 (± 16)	23-80
Sex		
Male	9	53
Female	8	47
Histopathology		
Papillary TC	10	59
Follicular TC	3	18
Minimally invasive follicular TC	1	6
Follicular variant of papillary TC	2	12
Unknown*	1	6
Stage TNM†		
T1-2N0M0	5	29
T1-2N1M0	4	24
T1-2N0M1	1	6
T3-4N0M0	4	24
T3-4N1M0	3	18
Time since primary tumor (mo)	88 (± 82)	15-305
Time since last ^{131}I therapy (mo)	67 (± 72)	15-304
Number of previous ^{131}I therapies		
1	8	47
2	4	24
3	3	18
4	2	12
Cumulative dose ^{131}I before inclusion (GBq)	10 (± 7.2)	1.9-29.6
Tg at time of therapy (µg/L)‡	99 (± 151)	2.1-531.3

Table 2: Baseline characteristics of subjects at time of inclusion (n=17)
SD: standard deviation; TC: thyroid carcinoma; mo: months; GBq: Gigabecquerel; Tg: thyroglobulin
*Patient was operated abroad, histopathology primary tumor not known, †TNM: Tumor-node-metastasis stage, 7th edition; ‡ After thyroid hormone withdrawal.

Patient-based analysis showed that 9/17 post-therapy ^{131}I scans were tumor positive (53%, CI: 28-77%). ^{124}I PET/CT scans showed uptake compatible with tumor activity in 4/17 scans (Table 3). In these four patients with a positive ^{124}I PET/CT the post-therapy ^{131}I WBS was concordant (case 2, 3, 7 and 16). In the eight patients with negative post-therapy ^{131}I scans, ^{124}I PET/CT scans were also negative. However, of the 13 ^{124}I PET/CT scans with no pathological uptake, five were false negative (Figs. 1A and 1B). One ^{124}I positive lymph node was surgically removed before the ^{131}I therapy (i.e. protocol violation), this

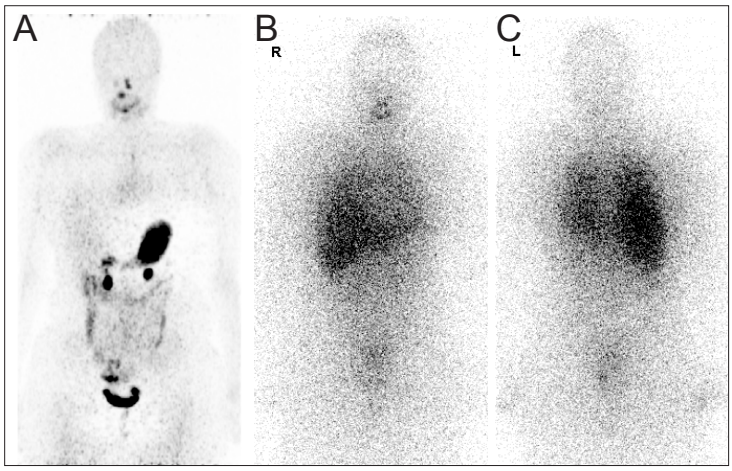


Figure 1A: Patient (case no. 5) with negative ^{124}I PET/CT (A) with disseminated lung metastases on ^{131}I WBS (anterior (B) and posterior (C))

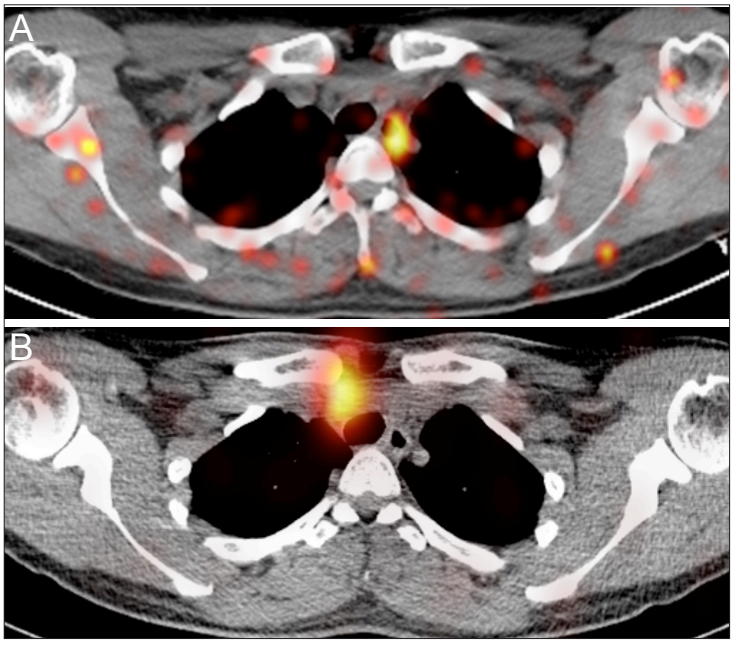


Figure 1B: Patient (case no. 12) without pathological uptake of ^{124}I on PET/CT, only physiological uptake in the esophagus (A) and a positive ^{131}I SPECT/CT, showing uptake in retroclavicular lymph node, which was confirmed as metastasis after surgical removal (B).

lesion was not included in the analysis (case 14). At a patient level, sensitivity, specificity, negative predictive value and positive predictive value of ^{124}I PET/CT (vs. ^{131}I WBS) were 44% (CI 14-79%), 100% (CI 63-100%), 62% (CI 32-86%) and 100% (CI 40-100%), respectively (Table 4).

Lesion-based analysis showed that the post-therapy ^{131}I WBS revealed 14 lesions versus 8 on ^{124}I PET/CT (miliary lung metastases in case 5 were counted as one lesion). The size of three out of seven lesions false negative at ^{124}I PET/CT could not be determined as no anatomical substrate was seen on CT (case

				Pre-therapy (after rhTSH)	Therapy phase (after THW)		
Case	**Histology**	**TNM***	**Tg†**	**^{124}I PET/CT**	**^{124}I PET/CT**	**^{131}I WBS (+SPECT/CT)**	**Remarks and follow-up**
1	PTC	T2N1aM0	5	Negative	ND	N-S: thymus	Tg ↑
2	PTC	T1aN0M1	32	Positive: C4	Positive: C4	Positive: C4	Quantitative analysis ^{124}I uptake: see text Tg =
3	miFTC	T2N0M0	16	Positive: LN retrosternal Equivocal: liver/ thoracic wall	Positive: LN retrosternal	Positive: LN retrosternal Equivocal: liver/ thoracic wall	MRI, bone scan & target US liver/thoracic wall: negative Surgery: i.a. retrosternal LN resected: thymic remnant Tg ↑
4	fvPTC	T2N1bM0	12	Negative	Negative	Negative	Tg decreased after surgery for ^{18}F-FDG avid LNM
5	PTC	T3N1bM0	184	Negative	ND	Disseminated lung metastasis	False-negative ^{124}I PET/CT Tg ↓
6	miFTC	T3N0M0	27	Negative	ND	Negative	Tg ↑
7	FTC	T3N0M0	148	Positive: LN aortopulmonary window Equivocal: nodule lung	ND	Positive: LN aortopulmonary window Equivocal: nodule lung	Tg ↓
8	PTC	T2N0M0	13	Negative	ND	Negative	Tg ↑
9	PTC	T3N0M0	38	Negative	Negative	Positive: nodule lung, no anatomic substrate on CT N-S: thymus	False-negative ^{124}I PET/CT Tg ↓
10	PTC	T2N0M0	2	Negative	ND	Negative	Tg =
11	PTC	T3N1bM0	3	Negative	Negative	Positive: three lung nodules; 7, 5 and 5 mm	False-negative ^{124}I PET/CT Tg ↓
12	PTC	T1bN0M0	107	Negative	ND	Positive: LN neck level 6: 12 mm N-S: thymus	False-negative ^{124}I PET/CT Surgery: node picking ^{131}I avid LN: metastasis Tg ↑
13	PTC	T1bN1bM0	28	Negative	ND	Negative	Tg =
14	FTC	T1mN1bM0	127	Positive: LN supraclavicular‡	ND	Positive: lung nodule, no anatomic substrate on CT	False-negative ^{124}I PET/CT Surgery: node picking supraclavicular: no metastasis‡ Tg ↓
15	NK	T1bN0M0	5	Negative	ND	Negative	Tg =
16	FTC	T4N1bM0	533	Positive: Two LNs neck level 3	ND	Positive: Two LNs neck level 3	Follow-up ^{124}I PET/CT scan: LNs not ^{124}I avid anymore Tg ↓, became undetectable
17	fvPTC	T3N0M0	400	Negative	ND	Negative	Tg ↑

Table 3: overview of included patients. rhTSH: recombinant human thyroid stimulating hormone; THW: thyroid hormone withdrawal; PTC: papillary thyroid cancer; miFTC: minimally invasive follicular thyroid cancer; fvPTC: follicular variant of papillary thyroid cancer; NK: not known; LN: lymph node; N-S: not suspicious; US: neck ultrasound; FNA: fine needle aspiration; ND: not done; Tg: thyroglobulin; *TNM: Tumor-node-metastasis stage, 7th edition; †Tg in µg/L at time of therapy (after THW); Tg ↑/↓: Thyroglublin level increased/decreased during follow-up; Tg= : Thyroglublin level remained stable during follow-up; ‡LN surgically removed before 131I therapy, therefor not included in diagnostics accuracy calculations (i.e. protocol violation)

5, 9 and 16). The other false negative lesions measured 5, 5 and 7 mm (all lung lesions, case 11) and 12 mm in diameter (retroclavicular lymph node, case 12).

The ^{124}I PET/CT scans after 24 and 96 hours were on patient level concordant in all but two patients (case 3 and 14). In those patients the lesions could not be depicted above the background noise after 96 hours. No additional lesions were seen after 96 hours.

Five patients underwent ^{124}I PET/CT after rhTSH stimulation as well as THW during ^{131}I therapy. No additional lesions were seen on THW-^{124}I PET/CT in comparison to the rhTSH-^{124}I PET/CT. Two of these patients (cases 2 and 3) showed pathological ^{124}I uptake on either scan. Case 2 showed enhanced uptake (residual disease) around an orthopedic cage placed in vertebra C4 after resection of a DTC metastasis (Fig. 2). At 24 hours after ^{124}I administration, lesional tracer uptake was 13 kBq following rhTSH stimulation (0,02% of injected dose ^{124}I), vs. 332 kBq after THW (0,56% of injected dose ^{124}I).

The positive lesion in case 3 was located in the mediastinum and was thought to represent a lymph node metastasis a (corresponding with a 3-4 mm node at CT). ^{18}F-FDG PET/CT had identified another suspicious mediastinal lymph node, negative on either iodine scan. At subsequent resection of this focus, the iodine positive lesion was also removed. Histopathology revealed thymic tissue. ^{124}I uptake in this lesion was below 2 kBq, both after THW and rhTSH, precluding reliable quantification.

Discussion

Our results demonstrate that rhTSH stimulated ^{124}I PET/CT as applied in this study is not suited to avoid futile blind ^{131}I therapy, because of its high false-negativity rate (38%; 5/13). If ^{124}I PET/CT results would have been used to guide therapy, potentially beneficial ^{131}I therapy would have been withheld in 5/17 (29%). To our knowledge this is the first prospective study performing a head-to-head comparison of pre-therapeutic ^{124}I PET/CT with post-therapy ^{131}I WBS in

	^{131}I WBS positive	**^{131}I WBS negative**	**Total**
^{124}I PET/CT positive	4	0	4
^{124}I PET/CT negative	5	8	13
Total	9	8	17

Table 4: patient based analysis of outcome of ^{124}I PET/CT and ^{131}I WBS

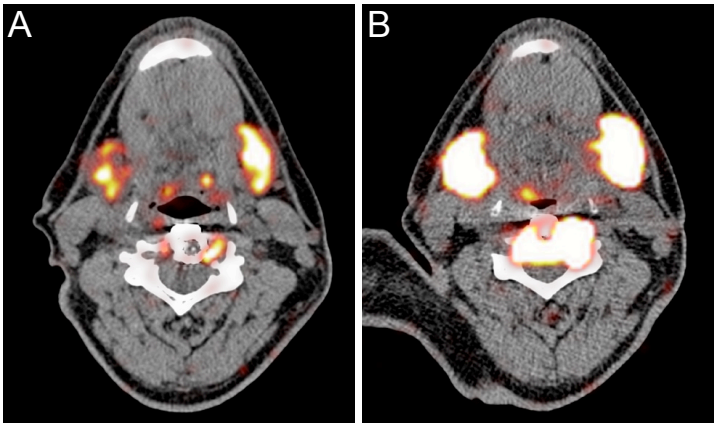


Figure 2: ^{124}I PET/CT with increased uptake of ^{124}I in metastatic lesion in C4 after THW (B) compared to rhTSH (A).

a well-defined cohort of patients planned for 'blind' ^{131}I therapy. The high rate of futile therapies in our study, 47% (8/17), is consistent with published data, corroborating the need for better pre-therapeutic diagnostic modalities.[6,7]

Several publications described comparisons of ^{124}I PET/CT and ^{131}I WBS in thyroid cancer patients, but the results are inconsistent, and difficult to compare with the results of the current study, due to variable clinical settings, study aims and inclusion criteria[8,19,25-29], methodological design [28,30], different ^{131}I dosages [8,29], diagnostic ^{131}I WBS [27,28], and non-standardized and/or variable patient preparation methods [8,25,27,28,30]. Of these studies the recent study by Khorjekar et al. is of interest as it specifically describes a cohort of 12 patients with suspicion of metastatic DTC, selected from different databases, in which a negative ^{124}I PET/CT was followed by a post-therapeutic ^{131}I WBS.[30] In 10 patients the ^{131}I WBS revealed pathological uptake. Another study of interest reported that, similarly to case 5 in our cohort (Fig. 1A), ^{124}I PET can be false negative in case of disseminated lung metastases.[31]

Biological as well as technical issues may have caused ^{124}I PET/CT scans to be false-negative. First, the method of TSH stimulation by rhTSH, instead of by THW in combination with a low iodine diet, might have compromised the sensitivity of ^{124}I PET/CT. This hypothesis seems to be supported by the observation in case 2, in whom after THW and a low-iodine diet the ^{124}I uptake was 23 times higher than after rhTSH stimulation (Fig. 2). Quantification of ^{124}I might be affected by concurrent presence of ^{131}I. However, a recently published phantom study showed that ^{131}I did not impact the accuracy of ^{124}I quantification.[32] One other study has reported that significantly more foci were detected on both ^{124}I PET and ^{131}I WBS after THW stimulation in comparison to rhTSH stimulation.[28] However, in that study inter-patient comparisons were used rather than the head-to-head comparison in our study. Only few cases are published with intra-patient comparisons of imaging after both THW and

rhTSH. Freudenberg et al. described one case in which both an adrenal and a lymph node metastasis were only seen on ^{124}I PET/CT after THW and not after rhTSH stimulation.[33] Another study described four cases with 10 metastatic DTC lesions showing an 9-62% higher iodine uptake after THW compared to rhTSH stimulation.[34] Additionally, a dosimetric analysis of three patients with 22 metastatic lesions revealed an increased absorbed dose after THW in comparison to rhTSH in all but two lesions .[35]. Taken together, our findings combined with previous suggest that rhTSH patient preparation for ^{124}I PET/CT may lead to false negative ^{124}I PET/CT scans, however, future studies with head-to-head comparisons are warranted to confirm this.

Secondly, we can only speculate whether the 74 MBq ^{124}I dosage in our study has contributed to the observed false negativity. Even though this dosage is higher than in most published studies (23-64 MBq)[25,27,30,31,36-38], Ho et al. recently reported the use of 222 MBq ^{124}I in patients with metastatic DTC refractory to radioiodine.[39] Additionally, it is unclear whether improved PET scan technology (scanner design[40] and/or ^{124}I reconstruction protocols) will improve ^{124}I PET performance to a clinically relevant extent in this context. If in vitro data support the notion that detectability significantly improves by such innovations and/or higher ^{124}I dosages, the current study should be repeated.

In our opinion, a scouting procedure using ^{124}I PET/CT still remains the most rational strategy to reduce futile ^{131}I therapies, but optimization is clearly required.

Conclusion

In patients with biochemical recurrence of DTC and a negative ultrasound of the neck, ^{124}I PET/CT after rhTSH stimulation before 'blind' ^{131}I therapy, as applied in this study, does correctly predict tumor positive uptake on post-therapeutic ^{131}I WBS. Due to the high false negative rate of ^{124}I PET/CT, ^{131}I should not be omitted based on a negative ^{124}I PET/CT.

References

1. Davies L, Welch HG. Current thyroid cancer trends in the United States. JAMA Otolaryngol Head Neck Surg 2014;140(4):317–22.

2. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin 2010;60(5):277–300.

3. Shaha AR. Recurrent differentiated thyroid cancer. Endocr Pract 2012;18(4):600–3.

4. Links TP, Huysmans DAKCJM, Smit JWA, et al. [Guideline "Differentiated thyroid carcinoma," including diagnosis of thyroid nodules]. Ned Tijdschr Geneeskd 2007;151(32):1777–82.

5. Cooper DS, Doherty GM, Haugen BR, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid. 2009;19(11):1167–214.

6. Ma C, Xie J, Kuang A. Is empiric 131I therapy justified for patients with positive thyroglobulin and negative 131I whole-body scanning results? J Nucl Med 2005;46(7):1164–70.

7. van Tol KM, Jager PL, de Vries EGE, et al. Outcome in patients with differentiated thyroid cancer with negative diagnostic whole-body scanning and detectable stimulated thyroglobulin. Eur J Endocrinol 2003;148(6):589–96.

8. Pettinato C, Spezi E, Nanni C, et al. Pretherapeutic dosimetry in patients affected by metastatic thyroid cancer using 124I PET/CT sequential scans for 131I treatment planning. Clin Nucl Med 2014;39(8):e367–74.

9. Smallridge RC, Diehl N, Bernet V. Practice trends in patients with persistent detectable thyroglobulin and negative diagnostic radioiodine whole body scans: a survey of American Thyroid Association members. Thyroid 2014;24(10):1501–7.

10. Van Nostrand D. Sialoadenitis secondary to 131I therapy for well differentiated thyroid cancer. Oral Diseases 2011;17(2):154–61.

11. Fard-Esfahani A, Emami-Ardekani A, Fallahi B, et al. Adverse effects of radioactive iodine-131 treatment for differentiated thyroid carcinoma. Nucl Med Commun 2014;35:808–17.

12. Luster M, Felbinger R, Dietlein M, Reiners C. Thyroid hormone withdrawal in patients with differentiated thyroid carcinoma: a one hundred thirty-patient pilot survey on consequences of hypothyroidism and a pharmacoeconomic comparison to recombinant thyrotropin administration. Thyroid 2005;15(10):1147–55.

13. Dow KH, Ferrell BR, Anello C. Quality-of-life changes in patients with thyroid cancer after withdrawal of thyroid hormone therapy. Thyroid 1997;7(4):613–9.

14. Pacini F, Agate L, Elisei R, et al. Outcome of differentiated thyroid cancer with detectable serum Tg and negative diagnostic (131) I whole body scan: comparison of patients treated with high (131)I activities versus untreated patients. J Clin Endocrinol Metab 2001;86(9):4092–7.

15. Cailleux AF, Baudin E, Travagli JP, Ricard M, Schlumberger M. Is diagnostic iodine-131 scanning useful after total thyroid ablation for differentiated thyroid cancer? J Clin Endocrinol Metab 2000;85(1):175–8.

16. Pacini F, Capezzone M, Elisei R, Ceccarelli C, Taddei D, Pinchera A. Diagnostic 131-iodine whole-body scan may be avoided in thyroid cancer patients who have undetectable stimulated serum Tg levels after initial treatment. J Clin Endocrinol Metab 2002;87(4):1499–501.

17. Pacini F, Molinaro E, Castagna MG, et al. Recombinant human thyrotropin-stimulated serum thyroglobulin combined with neck ultrasonography has the highest sensitivity in monitoring differentiated thyroid carcinoma. J Clin Endocrinol Metab 2003;88(8):3668–73.

18. Lubberink M, Abdul Fatah S, Brans B, Hoekstra OS, Teule GJJ. The role of (124) I-PET in diagnosis and treatment of thyroid carcinoma. Q J Nucl Med Mol Imaging 2008;52(1):30–6.

19. Phan HTT, Jager PL, Paans AMJ, et al. The diagnostic value of 124I-PET in patients with differentiated thyroid cancer. Eur J Nucl Med Mol Imaging 2008;35(5):958–65.

20. Freudenberg LS, Jentzen W, Stahl A, Bockisch A, Rosenbaum-Krumme SJ. Clinical applications of 124I-PET/CT in patients with differentiated thyroid cancer. Eur J Nucl Med Mol Imaging 2011;38(S1):48–56.

21. Freudenberg LS, Jentzen W, Görges R, et al. 124I-PET dosimetry in advanced differentiated thyroid cancer: therapeutic impact. Nuklearmedizin 2007;46(4):121–8.

22. Kist JW, de Keizer B, Stokkel MPM, Hoekstra OS, Vogel WV, THYROPET study group. Recurrent differentiated thyroid cancer: towards personalized treatment based on evaluation of tumor characteristics with PET (THYROPET Study): study protocol of a multicenter observational cohort study. BMC Cancer 2014;14(1):405.

23. van der Vlies M, Kist J, Greuter H. Multi-center calibration of PET/CT scanners for iodine-124. Thyroid 2013;

24. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging 2009;37(1):181–200.

25. de Pont C, Halders S, Bucerius J, Mottaghy F, Brans B. ^{12} I PET/CT in the pretherapeutic staging of differentiated thyroid carcinoma: comparison with posttherapy ^{131}I SPECT/CT. Eur J Nucl Med Mol Imaging 2013;40(5):693–700.

26. Freudenberg LS, Antoch G, Jentzen W, et al. Value of (124)I-PET/CT in staging of patients with differentiated thyroid cancer. Eur Radiol 2004;14(11):2092–8.

27. Van Nostrand D, Moreau S, Bandaru VV, et al. (124)I positron emission tomography versus (131)I planar imaging in the identification of residual thyroid tissue and/or metastasis in patients who have well-differentiated thyroid cancer. Thyroid 2010;20(8):879–83.

28. Van Nostrand D, Khorjekar GR, O'Neil J, et al. Recombinant human thyroid-stimulating hormone versus thyroid hormone withdrawal in the identification of metastasis in differentiated thyroid cancer with 131I planar whole-body imaging and 124I PET. J Nucl Med 2012;53(3):359–62.

29. Capoccetti F, Criscuoli B, Rossi G, Ferretti F, Manni C, Brianzoni E. The effectiveness of 124I PET/CT in patients with differentiated thyroid cancer. Q J Nucl Med Mol Imaging 2009;53(5):536–45.

30. Khorjekar GR, Van Nostrand D, Garcia C, et al. Do negative 124I pretherapy positron emission tomography scans in patients with elevated serum thyroglobulin levels predict negative 131I posttherapy scans? Thyroid 2014;24(9):1394–9.

31. Freudenberg LS, Jentzen W, Müller SP, Bockisch A. Disseminated iodine-avid lung metastases in differentiated thyroid cancer: a challenge to 124I PET. Eur J Nucl Med Mol Imaging 2008;35(3):502–8.

32. Braad PEN, Hansen SB, Høilund-Carlsen PF. Impact of high (131)I-activities on quantitative (124)I-PET. Phys Med Biol 2015;60(13):5297–312.

33. Freudenberg LS, Jentzen W, Petrich T, et al. Lesion dose in differentiated thyroid carcinoma metastases after rhTSH or thyroid hormone withdrawal: 124I PET/CT dosimetric comparisons. Eur J Nucl Med Mol Imaging 2010;37(12):2267–76.

34. Pötzi C, Moameni A, Karanikas G, et al. Comparison of iodine uptake in tumour and nontumour tissue under thyroid hormone deprivation and with recombinant human thyrotropin in thyroid cancer patients. Clin Endocrinol (Oxf) 2006;65(4):519–23.

35. Plyku D, Hobbs R, Huang K, et al. 124I-PET/CT based tumor dosimetry for 131I therapy of metastatic differentiated thyroid cancer (DTC) - a comparison of recombinant human thyroid-stimulating hormone vs thyroid hormone withdrawal patient preparation methods. J Nucl Med 2015;56(supplement 3):394–4.

36. Hartung-Knemeyer V, Nagarajah J, Jentzen W, et al. Pre-therapeutic blood dosimetry in patients with differentiated thyroid carcinoma using 124-iodine: predicted blood doses correlate with changes in blood cell counts after radioiodine therapy and depend on modes of TSH stimulation and number of preceding radioiodine therapies. Ann Nucl Med 2012;26(9):723–9.

37. Freudenberg LS, Antoch G, Frilling A, et al. Combined metabolic and morphologic imaging in thyroid carcinoma patients with elevated serum thyroglobulin and negative cervical ultrasonography: role of 124I-PET/CT and FDG-PET. Eur J Nucl Med Mol Imaging 2008;35(5):950–7.

38. Nagarajah J, Jentzen W, Hartung V, et al. Diagnosis and dosimetry in differentiated thyroid carcinoma using 124I PET: comparison of PET/MRI vs PET/CT of the neck. Eur J Nucl Med Mol Imaging 2011;38(10):1862–8.

39. Ho AL, Grewal RK, Leboeuf R, et al. Selumetinib-enhanced radioiodine uptake in advanced thyroid cancer. N Engl J Med 2013;368(7):623–32.

40. Nguyen NC, Vercher-Conejero JL, Sattar A, et al. Image Quality and Diagnostic Performance of a Digital PET Prototype in Patients with Oncologic Diseases: Initial Experience and Comparison with Analog PET. J Nucl Med 2015;56(9):1378–85.

COMPARISON OF ^{124}I PET/CT AND ^{131}I SPECT/CT DETECTABILITY

CHAPTER 9

Casper Beijst
Jakob W. Kist
Matthijs Elschot
Max A. Viergever
Otto S. Hoekstra
Bart de Keizer
Hugo W.A.M. de Jong

J Nucl Med 2016;57(1):103–8.

Abstract

Introduction

Radioiodine therapy with ^{131}I is used for treatment of suspected recurrence of differentiated thyroid carcinoma. Pre-therapy ^{124}I PET/CT with a low activity (~1% of ^{131}I activity) can be performed to determine whether uptake of ^{131}I, and thereby the desired therapeutic effect, may be expected. However, false negative ^{124}I PET/CTs as compared with the post-therapy ^{131}I SPECT/CTs have been reported by several groups. The purpose of this study was to investigate whether the reported discrepancies may be ascribed to a difference in detectability of lesions on ^{124}I PET/CT and ^{131}I SPECT/CT, and hence, to determine whether the administered activity of ^{124}I is sufficient to achieve equal detectability.

Methods

Phantom measurements were performed using the NEMA-2007 image quality phantom. As a measure of detectability the contrast-to-noise ratio (CNR) was calculated. The ^{124}I activity concentration was expressed as a percentage of the ^{131}I activity concentration required to achieve the same CNR. This metric was defined as the detectability equivalence percentage (DEP).

Results

Lower DEPs were obtained for smaller spheres, so that a relatively low ^{124}I activity concentration is sufficient to achieve a similar detectability of lesions with ^{124}I PET/CT as with ^{131}I SPECT/CT. The DEP was 1.5, 1.9, 1.9, 4.4, 9.0 and 16.2% for the spheres with a diameter of 10, 13, 17, 18, 25 and 37 mm respectively, for attenuation and scatter corrected SPECT versus point spread function (PSF) modeled and time-of-flight (TOF) PET. For no-PSF no-TOF PET, the DEP was 3.6, 2.1, 3.5, 7.8, 15.1 and 23.3%.

Conclusion

A relatively low activity of 74 MBq ^{124}I (~1% of ^{131}I activity) is sufficient to achieve similar detectability of lesions on ^{124}I PSF TOF PET/CT and ^{131}I SPECT/CT for small spheres (≤10mm), since the reported DEPs are close to 1%. False negative ^{124}I PET/CTs as compared with the post-therapy ^{131}I SPECT/CTs may be ascribed to differences in detectability for large lesions (>10mm) and for no-PSF no-TOF PET, since DEPs are larger than 1%. Based on DEPs of 3.5% for lesion diameters up to 17mm on no-PSF no-TOF PET, activities as high as 170 MBq of ^{124}I may be warranted to obtain equal detectability.

Introduction

Radioiodine therapy with ^{131}I is used for the primary treatment of differentiated thyroid carcinoma by ablating remnant thyroid tissue and potential residual tumor tissue after thyroid resection. Treatment with 5.5-7.4 GBq of ^{131}I is indicated if a patient is suspected of metastases.[1] The administered activity of ^{131}I is generally empiric and non-patient-specific, since the uptake and hence the therapeutic effectiveness of radioiodine in metastatic lesions is usually not known beforehand. Post-therapy ^{131}I SPECT/CT and/or whole body scintigraphy is routinely performed to assess tumor uptake. Up to 50% of empirically treated patients show no uptake on the post-therapy imaging.[2] To predict whether uptake of radioiodine, and hence a desired therapeutic effect, may be expected, pre-therapy imaging with a low activity is performed.

Several pre-therapy imaging strategies have been suggested to optimize patient-specific treatment activity and to prevent unnecessary ^{131}I therapies. ^{131}I whole-body scintigraphy with a low non-therapeutic activity of ^{131}I (half-life 8 days) can be used for pre-therapy imaging.[3] However, diagnostic low activity scintigraphy with ^{131}I does not adequately predict the results of post-therapeutic high activity ^{131}I scintigraphy.[4] Moreover, septal penetration by the high-energy (364 keV and 637 keV) gamma photons negatively affects the image quality of ^{131}I SPECT and scintigraphy, which makes ^{131}I less suitable for diagnostic purposes. The medium energy (159 keV) gamma emitter ^{123}I has also been suggested for pre-therapy imaging and studies investigating the potential of ^{123}I yielded good results.[5] However, due to the short half-life of ^{123}I (13.2 h) a large fraction of the administered activity has already decayed before the maximum uptake is achieved and imaging is performed.

Alternatively, ^{124}I with a half-life of 4.2 days has been proposed for pre-therapy imaging and assessment of treatment response. Several groups have reported promising results using ^{124}I as a diagnostic agent.[6,7] ^{124}I is a positron emitter and can be imaged with PET, with superior resolution, sensitivity and quantitation as compared with scintigraphy or SPECT used for ^{131}I. Relatively low activities of 20 – 74 MBq are used for diagnostic ^{124}I PET imaging.[8-13]

Unfortunately, clinical experiences have shown that ^{124}I PET/CT does not always predict uptake of ^{131}I reliably, and discrepancies between post-therapy ^{131}I SPECT/CT and pre-therapy ^{124}I PET/CT have been reported by several groups.[8-13] More specifically, in some cases no uptake was found on ^{124}I PET/CT images, whereas uptake was found on ^{131}I SPECT/CT images. An example from a study we performed (unpublished data) is shown in Figure 1.[14] These false negative ^{124}I PET/CTs as compared with ^{131}I SPECT/CTs may be ascribed to a difference in detectability of lesions on ^{124}I PET/CT and ^{131}I SPECT/CT.

The purpose of this study was to investigate whether the reported discrepancies may be ascribed to a difference in detectability of lesions on ^{124}I PET/CT and ^{131}I SPECT/CT, and hence, to determine whether the administered activity of ^{124}I is sufficient to achieve equal detectability. This was done by establishing the activity concentration of ^{124}I expressed as a percentage of the ^{131}I activity concentration at which the contract-to-noise ratio was equal for both modalities.

Materials and methods

Phantom

To compare the detectability of lesions on ^{131}I SPECT/CT and ^{124}I PET/CT images, acquisitions of the IEC NEMA 2007 phantom (PTW, Freiburg, Germany) were performed with varying activity concentrations. The phantom is torso-shaped and has a lid holding refillable thin-walled spheres of 10, 13, 17, 22, 28 and 37 mm in diameter. Separate phantoms were used for ^{124}I and ^{131}I experiments. The phantoms were filled only once and different activity concentrations were obtained by leaving the activity to decay. This approach decreases measurement errors in comparison with refilling the phantom. Two phantoms were used for detectability analysis; one with activity in the background compartment and one without.

Clinically, large variations may occur in lesion to background ratios. Rubello et al. have performed measurements of the lesion-to-background ratio with a gamma probe during radio-guided surgery, and they obtained a mean lesion-to-background ratio of 11.4.[15] Therefore, we used two phantoms with different background concentrations to capture both extremities; no background (1:0) and a high background concentration (10:1). The activity

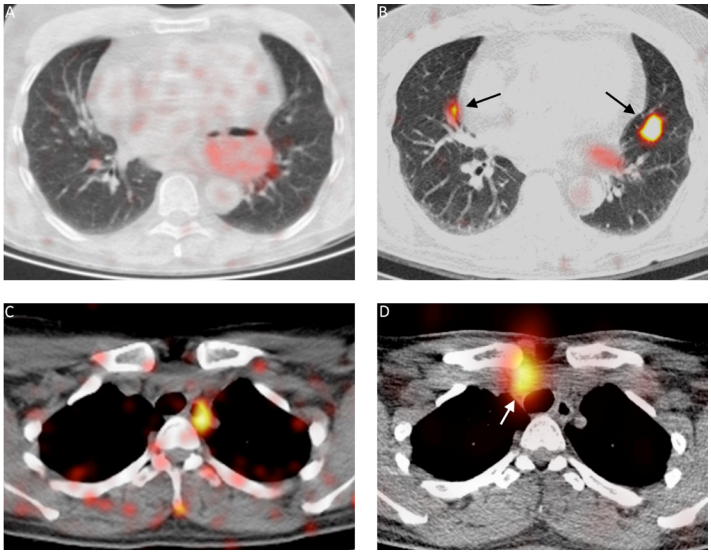


Figure 1. Examples of cases from the THYROPET study[14] where false negative ^{124}I PET/CTs as compared with the 131I SPECT/CT were obtained. ^{124}I PET/CT (A and C) acquired 24 hours after oral administration of 74 MBq 124I and ^{131}I SPECT/CT (B and D) acquired 7 days after oral administration of 7400 MBq ^{131}I, for the first and second patient respectively. The PET/CT scans were acquired using a Philips Gemini in 2 minutes per bed position and a LOR-RAMLA reconstruction without time-of-flight correction (no-PSF no-TOF). The SPECT/CT scans were acquired on a Siemens Symbia T2 in a total scan time of 21 minutes using a Flash3D attenuation corrected reconstruction method (no-SC 6i8s). The arrows indicate the metastatic lesions.

concentration in the spheres ranged from $1.8x10^4$ Bq/ml to 4.6 Bq/ml for ^{124}I and from $9.1x10^5$ Bq/ml to $3.9x10^2$ Bq/ml for ^{131}I. The initial activity concentrations are shown in Supplemental Table 1. The activity concentration in the phantoms was chosen such that it could be used to effectively compare the detectability of lesions in the range around the expected minimum detectable activity (MDA). In total, 45 ^{124}I PET acquisitions were acquired over a period of 50 days. Similarly, 42 ^{131}I SPECT acquisitions were performed over a period of 90 days.

Scanners and acquisition

A Siemens Biograph mCT Time-of-Flight (TOF) PET/CT scanner (Siemens Healthcare, Erlangen, Germany) with TrueV (axial field of view 21.6 cm) was used to acquire PET data. ^{124}I images were acquired using a 435-650 keV energy window in 4 minutes per bed position, in accordance with the clinical protocol.[14] Three bed positions were used to ensure that the sensitivity was uniform along the entire length of the phantom in the axial direction.

A dual-headed Siemens Symbia T16 (Siemens Healthcare, Erlangen, Germany) SPECT/CT system was used to acquire SPECT images. The Siemens BiCoreä high-energy (HE) (Siemens Healthcare, Erlangen, Germany) collimator that was used has 8000 holes, a hole length of 59.7 mm, hole diameter of 4.0 mm and septal thickness of 2.0 mm. For all ^{131}I acquisitions, 128 projections were acquired on a 128 x 128 grid with a 4.8 x 4.8 mm^2 pixel size. In accordance with the clinical protocol, a 25 s acquisition time per view was used, resulting in a total acquisition time of 26 minutes.

Reconstruction

PET images were reconstructed to create 200 x 200 x 316 voxel images with 4.07 x 4.07 x 1.50 mm^3 voxel size. The reconstruction incorporated a ^{124}I prompt gamma correction[16] and a randoms correction by adding the Gaussian-filtered randoms-sinogram to the forward projection during the iterative reconstruction.[17] Two reconstruction methods were used to obtain PET images.

The first method was an ordered subset expectation maximization 3D reconstruction method incorporating time of flight (TOF) information (TrueX) with point spread function (PSF) model based resolution recovery and CT based attenuation and scatter correction, using 4 iterations and 21 subsets in accordance with the clinical protocol. A Gaussian post-reconstruction filter was applied with a full width at half maximum of 5 mm. This method will be referred to as the PSF TOF method.

$$CNR = \frac{C_H - C_B}{\sigma_B}$$

The second method was an ordered subset expectation maximization 3D reconstruction method with attenuation correction, scatter correction, no TOF modeling and no PSF model based resolution recovery. As advised

by the vendor, 4 iterations and 24 subsets were used and a Gaussian post-reconstruction filter was applied with a full width at half maximum of 5 mm. This reconstruction method will be referred to as the no-PSF no-TOF method.

SPECT images were reconstructed to create 128 x 128 x 80 voxel images with 4.8 x 4.8 x 4.8 mm^3 voxel size. The Siemens Flash 3D reconstruction algorithm was used, incorporating attenuation correction and resolution recovery using distance dependent PSFs. Three reconstruction methods were used to obtain SPECT images. The first method used 6 iterations with 8 subsets and triple-energy-window scatter correction (SC 6i8s), the second method used 6 iterations with 8 subsets and did not incorporate scatter correction (no-SC 6i8s) and the third method used 30 iterations with 8 subsets and triple-energy-window scatter correction (SC 30i8s). Reconstructions with 6 iterations and 8 subsets were performed to allow comparison with the clinical trial running at our hospital (14), whereas reconstructions with 30 iterations and 8 subsets were included for increased contrast recovery at the cost of increased image noise.[18,19]

Quantitative analysis

Detectability Equivalence Percentage. As a measure of detectability, we calculated the contrast-to-noise ratio (CNR) for each sphere using where C_H is the mean voxel value in the sphere volume of interest (VOI), C_B is the mean voxel value in the background VOI and is the standard deviation in the background VOI.

Three-dimensional VOI masks were created based on the sphere coordinates determined from the co-registered CT. The position of the spheres in the CT image was determined automatically using a Hough-transform based circle detection method.[20] The background VOI was defined as the entire phantom minus the sphere VOIs. To eliminate the influence of partial volume effects on the background measurement, a 2 cm margin around the spheres and the phantom edges was subtracted from the background VOI by means of binary erosion.

To assess the difference in detectability of spheres on ^{124}I PET/CT and ^{131}I SPECT/CT, the ^{124}I activity concentration was expressed as a percentage of the ^{131}I activity concentration required to achieve the same CNR. This metric was defined as the detectability equivalence percentage (DEP). For example, a DEP of 1% indicates that the same CNR is obtained on the ^{124}I PET/CT image as on the ^{131}I SPECT/CT image if the ^{124}I activity concentration is 1% of the ^{131}I activity concentration. The DEP was determined by calculating the average ratio of the CNR curves (CNR versus activity concentration) for the two isotopes, and multiplied by 100 to obtain a percentage. This was done for each sphere size and sphere-to-background ratio. To reduce the influence of noise on the calculated DEPs, regression analysis of the curves was performed. Cubic spline fits of the activity concentration as a function of the sphere CNR were performed and plotted in the log-log domain, because the range of the curves was several orders of magnitude. The mean ratios were calculated over the

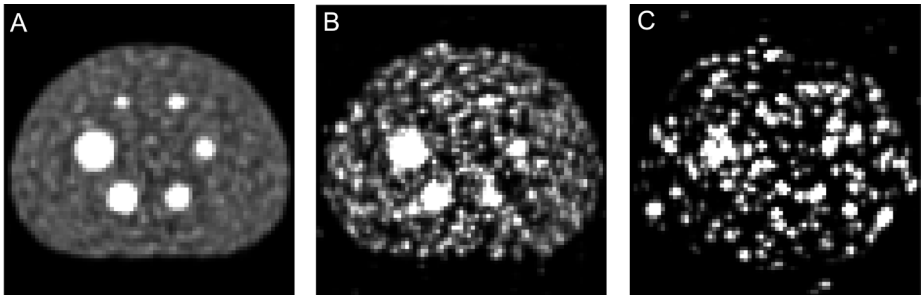


Figure 2: Typical ^{124}I PET/CT images of the phantom with activity in the background compartment reconstructed using the PSF TOF method showing the central slice through the spheres for (A) 2.2x104 Bq/ml, (B) 9.3x102 Bq/ml and (C) 6.5x101 Bq/ml in the spheres. The maximum of the gray scale is five times the mean pixel value of the image.

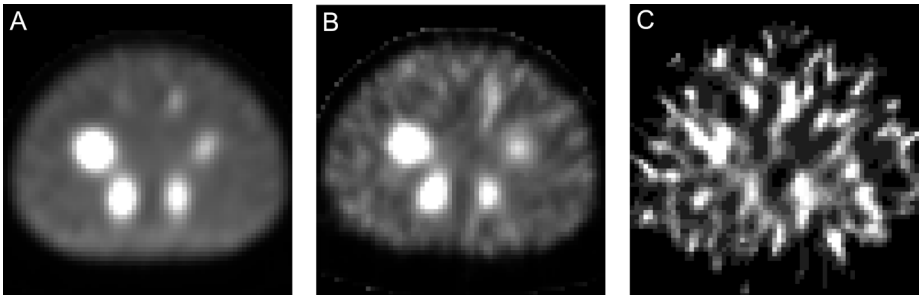


Figure 3: Typical ^{131}I SPECT/CT images of the phantom with activity in the background compartment reconstructed using the SC 6i8s method showing the central slice through the spheres for (A) 9.1x105 Bq/ml, (B) 4.5x104 Bq/ml and (C) 2.2x103 Bq/ml in the spheres. The maximum of the gray scale is five times the mean pixel value of the image.

largest possible interval where CNR data was acquired for both isotopes and where CNR values were greater than 1. For the phantom without activity in the background compartment, the CNR values, where the mean pixel value of the background VOI was smaller than one, were not used for the regression analysis.

Minimum Detectable Activity

False negative ^{124}I PET/CTs as compared with the ^{131}I SPECT/CTs occur when the ^{124}I activity concentration is below minimum detectable activity (MDA) and ^{131}I concentration is above MDA. Therefore, the MDA was calculated by using the Rose criterion.[21] The Rose criterion states that a lesion is no longer visible when the CNR of the lesion falls below a certain threshold value, which can be used

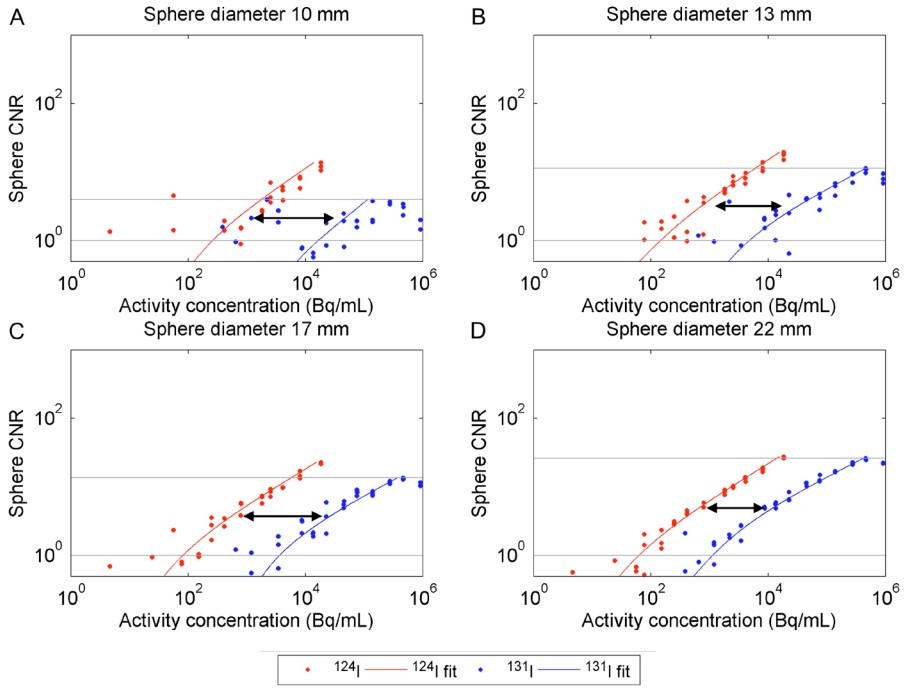


Figure 4: The sphere CNR as a function of activity concentration for the ^{124}I (red) and ^{131}I (blue) phantom with background for spheres with a diameter of (A) 10 mm, (B) 13 mm, (C) 17 mm and (D) 22 mm. The DEP is determined by calculating the ratio of the curves, which is graphically represented as the horizontal shift of the curves (black arrow). The gray lines denote the interval where the average ratio between the curves was determined. PET images were reconstructed using the PSF TOF method and SPECT images using the SC 6i8s method.

to determine the MDA. Because the Rose criterion has been validated for use in 2D only, we calculated the CNR of the (2D) central slice through the spheres. The detectability measure K_{2D} was obtained by correcting the CNR for the lesion size according to where CNR_{2D} is the CNR in 2D, $C_{H,2D}$ is the mean voxel value in the sphere 2D region of interest (ROI), $C_{B,2D}$ is the mean voxel value in the 2D background ROI, $\sigma_{B,2D}$ is the standard deviation in the 2D background ROI and N is the number of pixels in the lesion ROI. A K_{2D} threshold value of 8 was used to determine the MDA.[22,23] The MDA was defined as the activity concentration where the K_{2D} versus activity concentration curve intersected with the threshold value. Regression analysis of the K_{2D} versus activity concentration curves with cubic splines was performed to reduce the influence of noise.

Results

Figure 2 and Figure 3 show typical examples of the ^{124}I PET/CT and ^{131}I SPECT/CT NEMA phantom images for different activity concentrations. The sphere CNR was calculated and plotted as a function of activity concentration, as shown in Figure 4, Supplemental Figure 1 and Supplemental Figure 2 for the phantom with and without background activity, respectively. Defined as the average ratio between ^{124}I and ^{131}I activity concentration to achieve the same CNR, the DEP was calculated for each combination of PET and SPECT reconstruction methods and for each sphere size (Table 1 and Table 2).

In general, lower DEPs were obtained for smaller spheres, so that a relatively low ^{124}I activity concentration is sufficient to achieve a similar detectability of lesions with ^{124}I PET/CT as with ^{131}I SPECT/CT. Furthermore, the DEP depends on the reconstruction method that is used. The DEP is higher for the no-PSF no-TOF PET reconstruction method than for the PSF TOF reconstruction method. Furthermore, the DEP is slightly higher for the SC 6i8s method than for the no-SC 6i8s method, and lower for the SC 30i8s method than for the other SPECT reconstruction methods. The differences between DEPs for the PSF TOF and no-PSF no-TOF reconstruction methods are smaller for the phantom without activity in the background compartment. In general, the DEPs are slightly smaller for the phantom without activity in the background compartment.

Table 3 shows the MDA for the phantom with activity in the background compartment. The MDA of ^{124}I is lower for the PSF TOF reconstruction method than for the no-PSF no-TOF reconstruction method. Similarly, for the 13 and 17 mm spheres, the MDA of ^{131}I is lower for the SC 6i8s reconstruction method than for the no-SC 6i8s reconstruction method. However, for the 10, 22 and 28 mm spheres, the MDA of ^{131}I is higher for the SC 6i8s reconstruction method than for the no-SC 6i8s reconstruction method. The MDAs of the SC 30i8s reconstruction method are larger than for the other SPECT reconstruction methods.

Discussion

A relatively low activity of 74 MBq ^{124}I (~1% of ^{131}I activity) is sufficient to achieve a similar detectability of lesions on ^{124}I PSF TOF PET/CT and ^{131}I SPECT/CT for small spheres (≤10mm), since the reported detectability equivalence percentages (DEPs) are close to 1%. False negative ^{124}I PET/CTs as compared with the post-therapy ^{131}I SPECT/CTs may be ascribed to differences in detectability for large lesions (>10mm) and for no-PSF no-TOF PET, since DEPs are larger than 1%.

The results showed that the DEP was lower for smaller spheres, which indicates that smaller spheres are relatively more easily detected on PET than SPECT images. This can be explained by the fact that the resolution of ^{124}I PET

PET reconstruction method	SPECT reconstruction method	10 mm	13 mm	17 mm	22 mm	28 mm	37 mm
PSF TOF	SC 6i8s	1.5	1.9	1.9	4.4	9.0	16.2
no-PSF no-TOF	SC 6i8s	3.6	2.1	3.5	7.8	15.1	23.3
PSF TOF	no-SC 6i8s	1.4	2.1	1.9	4.1	6.8	11.6
no-PSF no-TOF	no-SC 6i8s	3.4	2.1	3.6	7.4	11.9	18.1
PSF TOF	SC 30i8s	0.8	1.3	1.0	1.7	2.1	2.7
no-PSF no-TOF	SC 30i8s	1.8	1.5	2.0	2.8	3.4	3.9

Table 1: DEPs for the phantom with background.

PET reconstruction method	SPECT reconstruction method	10 mm	13 mm	17 mm	22 mm	28 mm	37 mm
PSF TOF	SC 6i8s	1.1	2.3	3.3	4.8	5.9	7.0
no-PSF no-TOF	SC 6i8s	1.7	3.2	3.2	4.2	4.8	5.5
PSF TOF	no-SC 6i8s	0.8	1.4	2.0	2.9	4.1	4.3
no-PSF no-TOF	no-SC 6i8s	1.3	2.6	2.6	3.3	4.2	4.0
PSF TOF	SC 30i8s	0.9	1.1	1.4	1.6	1.7	1.8
no-PSF no-TOF	SC 30i8s	1.4	1.7	1.6	1.6	1.7	1.7

Table 2: DEPs for the phantom without background.

images (approximately 5 mm) generally is higher than the resolution of ^{131}I SPECT images (approximately 15 mm). Therefore, differences in the CNR are large in the range of sphere sizes between 5 and 15 mm.

Activities frequently used in clinical practice and studies are 74 MBq of ^{124}I and 7400 MBq of ^{131}I, thus the ratio of ^{124}I administered activity is approximately 1% of the ^{131}I administered activity. However, as different scan delay times are used after administration, it is necessary to correct for physical decay during the scan delay time. The ^{124}I PET and the ^{131}I SPECT acquisitions are usually performed 24 hours and 7 days after administration respectively, so that at scan time the ^{124}I activity concentration is approximately 1.5% of the ^{131}I activity concentration, if we correct for physical decay. Table 1 and Table 2 show that for some sphere sizes and reconstruction algorithms, the DEP is smaller than 1.5%. Therefore, for small spheres (≤10mm) on PSF TOF PET an activity of 74 MBq ^{124}I is sufficient to achieve similar detectability on the low activity ^{124}I PET/CT as on the high activity ^{131}I SPECT/CT. This shows that false negative low activity ^{124}I PSF TOF PET/CTs as compared with the post-therapy high activity ^{131}I SPECT/CTs are not likely occurring owing to differences in detectability for small lesions. A significant part of the lesions in clinical practice is expected to be smaller than 10 mm.[24] For larger lesions (>10 mm) and no-PSF no-TOF

Isotope	Reconstruction method	10 mm	13 mm	17 mm	22 mm	28 mm	37 mm
^{124}I	PSF TOF	1.0×10^{3}	4.1×10^{2}	1.6×10^{2}	8.3×10^{1}	4.4×10^{1}	2.1×10^{1}
^{124}I	no-PSF no-TOF	3.0×10^{3}	5.1×10^{2}	3.0×10^{2}	1.9×10^{2}	8.6×10^{1}	4.6×10^{1}
^{131}I	SC 6i8s	1.1×10^{5}	2.2×10^{4}	8.2×10^{3}	1.7×10^{3}	6.5×10^{2}	
^{131}I	no-SC 6i8s	9.6×10^{4}	2.5×10^{4}	8.9×10^{3}	1.4×10^{3}	4.2×10^{2}	
^{131}I	SC 30i8s	1.9×10^{5}	3.3×10^{4}	1.5×10^{4}	4.2×10^{3}	2.0×10^{3}	

Table 3: MDAs in Bq/ml for the phantom with background.

PET, the DEP generally is larger than 1.5% and 74 MBq of ^{124}I is not sufficient to achieve similar detectability on the low activity ^{124}I PET/CT as on the high activity ^{131}I SPECT/CT. Reported discrepancies, such as shown the example in Figure 1, may therefore be caused by differences in the detectability.

The administered activity of ^{124}I is a factor that should be taken into account when comparing our results with other studies. Several clinical studies have been performed with lower activities than 74 MBq of ^{124}I.[8,9,11] The probability of a false negative ^{124}I PET/CT as compared with the post-therapy ^{131}I SPECT/CTs considerably increases with lower ^{124}I dosages. When 25 MBq of ^{124}I and 7400 MBq of ^{131}I is administered, the ratio of ^{124}I and ^{131}I activity concentration is approximately 0.5%, when corrected for physical decay. Since this is lower than the reported DEPs, discrepancies are more likely to occur as a result of differences in detectability.

Activities of 90 MBq are sufficient to achieve similar detectability for lesion diameters up to 17 mm on PSF TOF PET, with DEPs up to 1.8%. Based on DEPs of 3.5% for lesion diameters up to 17mm on no-PSF no-TOF PET, activities as high as 170 MBq of ^{124}I may be warranted to obtain equal detectability. However, a limiting factor for high pre-therapy activities may be thyroid stunning, although stunning due to ^{124}I is controversial, and the origin and existence of stunning still are a hotly debated issue.[25]

Discrepancies between the ^{124}I and ^{131}I distributions in the tissue as a consequence of differences in uptake/washout may be influenced by additional factors, not assessed in this study. Firstly, unlike phantoms, thyroid tumors or thyroid cancer metastases are inhomogeneous structures consisting of cancer cells, blood vessels and connective tissue with different iodine concentrations, inducing partial volume effects which may affect the measured contrast. Secondly, the uptake of iodine may be influenced by the preparation of the patient, since patients can be prepared either by thyroid hormone withdrawal or recombinant human thyroid-stimulating hormone (rhTSH) stimulation to

stimulate iodine uptake. Thirdly, cell damage as a consequence of delivered dose may influence the retention time and may therefore cause differences in the ^{124}I and ^{131}I physiological washout properties. A faster washout of ^{131}I is expected for damaged cells.[24] Fourthly, the contrast of ^{131}I lesions may increase when delayed scanning is performed, because clearance of the background activity can occur at a faster rate than clearance of the activity in the tumor, so that contra-intuitively, the contrast in the images increases over time (10). One of the few practically adjustable parameters in clinical practice that influences the activity concentration is the delay time between the administration and acquisition. However, the timing of the ^{131}I SPECT/CT acquisition still is a matter of debate in literature. In fact, Salvatori et al. state that 'perfect timing' probably does not exist due to differences in ^{131}I kinetics in different patients and in different metastases.[26] A scan delay time of 7 days for this study was chosen to allow washout of background activity.[27] Similar to ^{131}I, the uptake of ^{124}I in metastases shows significant differences among metastases, although most of the metastases showed to have their peak uptake at approximately 24 hours after administration.[28,29] Therefore, a scan delay time of 24 hours was used for the ^{124}I PET/CT acquisition.

The purpose of this study was not to investigate the impact of physiological factors, but to quantitatively compare the detectability of lesions on ^{124}I PET/CT and ^{131}I SPECT/CT images. Assuming that physical decay is the only factor affecting the ratio of ^{124}I and ^{131}I activity concentration may not be accurate and the interpretation of the results depends on these assumptions. However, the measured DEPs do not depend on physiological factors. To our knowledge, phantom measurements that compare the detectability of ^{124}I and ^{131}I quantitatively have not been published before.

CNR values lower than 1 were not used for the regression analysis and calculation of the DEP. These values are inherently noisy due to the low number of counts. Furthermore, rounding errors occurred by conversion to the DICOM format when the mean voxel value in the background compartment was lower than 1. Consequently, for low activity concentrations in the phantom without background activity, the standard deviation in the background compartment σ_B was underestimated and the CNR overestimated. Therefore, these data points (mean background VOI<1) were not used to determine the DEPs and MDAs.

Diagnostic ^{131}I SPECT/CT with 37-150 MBq of ^{131}I can also be performed for pre-therapy imaging.[3] However, this study shows that false negative diagnostic scans as compared with the post-therapy scans may be ascribed to the difference in activity that is used, and that false negative results are likely, especially for the smaller lesions that have high MDAs.

False negative ^{124}I PET/CTs as compared with the ^{131}I SPECT/CTs occur when the ^{124}I activity concentration is below minimum detectable activity (MDA) and ^{131}I concentration is above MDA. The MDAs of ^{131}I and ^{124}I were therefore determined for the different reconstruction methods. The MDA of ^{131}I was higher for the SC 6i8s reconstruction method than for the no-SC 6i8s reconstruction method for some sphere sizes (10, 22 and 28 mm). MDAs were

not necessarily lower for images obtained with scatter correction, possibly due to the addition of noise introduced in the reconstruction by the noisy scatter projections. Unfortunately, no ^{131}I MDAs could be obtained for the 37 mm spheres because not enough data was available for low CNRs.

The DEPs were lower for the 30i8s than for the 6i8s SPECT reconstruction method, because generally lower CNRs were obtained for the 30i8s than for the 6i8s SPECT reconstruction method. This was caused by the fact that the background noise level in the images increased with the number of iterations, which lowered the CNR. Therefore, the ^{124}I activity concentration expressed as a percentage of the ^{131}I activity concentration required to achieve the same CNR (i.e. the DEP), was lower for the 30i8s than for the 6i8s SPECT reconstruction method.

The sizes of the lesions in Figure 1 are 5 mm, 7 mm (Figure 1B) and 12 mm (Figure 1D), based on the co-registered CT data. The DEP for the 10 mm sphere is between 1.3 and 3.4, depending of the lesion-to-background ratio (Table 1 and Table 2), so that a relatively low activity of 74 MBq ^{124}I (~1% of ^{131}I activity) may not be sufficient to achieve a similar detectability. The false negative ^{124}I PET/CTs as compared with the ^{131}I SPECT/CTs may therefore be ascribed to differences in detectability. Unfortunately, no DEPs were calculated for sphere diameters smaller than 10 mm. In general, discrepancies between ^{124}I PET/CTs and ^{131}I SPECT/CTs depend on many factors, and this study shows that the dosage of ^{124}I should be chosen carefully.

Conclusion

A relatively low activity of 74 MBq ^{124}I (~1% of ^{131}I activity) is sufficient to achieve a similar detectability of lesions on ^{124}I PSF TOF PET/CT and ^{131}I SPECT/CT for small spheres (≤10mm), since the reported DEPs are close to 1%. Activities of 90 MBq of ^{124}I are sufficient to achieve similar detectability for lesion diameters up to 17 mm on PSF TOF PET, with DEPs up to 1.8%. False negative ^{124}I PET/CTs as compared with the post-therapy high activity ^{131}I SPECT/CTs may be ascribed to differences in detectability for large lesions (>10mm) and for no-PSF no-TOF PET, since DEPs are larger than 1%. Based on DEPs of 3.5% for lesion diameters up to 17mm on no-PSF no-TOF PET, activities as high as 170 MBq of ^{124}I may be warranted to obtain equal detectability.

References

1. Luster M, Clarke SE, Dietlein M, et al. Guidelines for radioiodine therapy of differentiated thyroid cancer. Eur J Nucl Med Mol Imaging. 2008;35(10):1941–59.

2. Ma C, Xie J, Kuang A. Is empiric 131I therapy justified for patients with positive thyroglobulin and negative 131I whole-body scanning results? J Nucl Med 2005;46(7):1164–70.

3. Pacini F, Lippi F, Formica N, et al. Therapeutic doses of iodine-131 reveal undiagnosed metastases in thyroid cancer patients with detectable serum thyroglobulin levels. J Nucl Med 1987;28(12):1888–91.

4. Pacini F, Agate L, Elisei R, et al. Outcome of differentiated thyroid cancer with detectable serum Tg and negative diagnostic (131) I whole body scan: comparison of patients treated with high (131)I activities versus untreated patients. J Clin Endocrinol Metab 2001;86(9):4092–7.

5. Urhan M, Dadparvar S, Mavi A, et al. Iodine-123 as a diagnostic imaging agent in differentiated thyroid carcinoma: a comparison with iodine-131 post-treatment scanning and serum thyroglobulin measurement. Eur J Nucl Med Mol Imaging 2007;34(7):1012–7.

6. Erdi YE, Macapinlac H, Larson SM, et al. Radiation Dose Assessment for I-131 Therapy of Thyroid Cancer Using I-124 PET Imaging. Clin Positron Imaging 1999;2(1):41–6.

7. Freudenberg LS, Antoch G, Jentzen W, et al. Value of (124)I-PET/CT in staging of patients with differentiated thyroid cancer. Eur Radiol 2004;14(11):2092–8.

8. de Pont C, Halders S, Bucerius J, Mottaghy F, Brans B. ^{12} I PET/CT in the pretherapeutic staging of differentiated thyroid carcinoma: comparison with posttherapy ^{131}I SPECT/CT. Eur J Nucl Med Mol Imaging 2013;40(5):693–700.

9. Freudenberg LS, Jentzen W, Müller SP, Bockisch A. Disseminated iodine-avid lung metastases in differentiated thyroid cancer: a challenge to 124I PET. Eur J Nucl Med Mol Imaging 2008;35(3):502–8.

10. Khorjekar GR, Van Nostrand D, Garcia C, et al. Do negative 124I pretherapy positron emission tomography scans in patients with elevated serum thyroglobulin levels predict negative 131I posttherapy scans? Thyroid 2014;24(9):1394–9.

11. Lammers GK, Esser JP, Pasker P. Can I-124 PET/CT Predict Pathological Uptake of Therapeutic Dosages of Radioiodine (I-131) in Differentiated Thyroid Carcinoma? Advances in Molecular ... 2014;

12. Phan HTT, Jager PL, Paans AMJ, et al. The diagnostic value of 124I-PET in patients with differentiated thyroid cancer. Eur J Nucl Med Mol Imaging 2008;35(5):958–65.

13. Van Nostrand D, Moreau S, Bandaru VV, et al. (124)I positron emission tomography versus (131)I planar imaging in the identification of residual thyroid tissue and/or metastasis in patients who have well-differentiated thyroid cancer. Thyroid 2010;20(8):879–83.

14. Kist JW, de Keizer B, Stokkel MPM, Hoekstra OS, Vogel WV, THYROPET study group. Recurrent differentiated thyroid cancer: towards personalized treatment based on evaluation of tumor characteristics with PET (THYROPET Study): study protocol of a multicenter observational cohort study. BMC Cancer 2014;14(1):405.

15. Rubello D, Salvatori M, Ardito G, et al. Iodine-131 radio-guided surgery in differentiated thyroid cancer: outcome on 31 patients and review of the literature. Biomed Pharmacother 2007;61(8):477–81.

16. Preylowski V, Schlögl S, Schoenahl F, et al. Is the Image Quality of I-124-PET Impaired by an Automatic Correction of Prompt Gammas? PLoS ONE 2013;8(8):e71729.

17. Willowson KP, Tapner M, QUEST Investigator Team, Bailey DL. A multicentre comparison of quantitative (90)Y PET/CT for dosimetric purposes after radioembolization with resin microspheres : The QUEST Phantom Study. Eur J Nucl Med Mol Imaging 2015;42(8):1202–22.

18. Dewaraja YK, Frey EC, Sgouros G, et al. MIRD pamphlet No. 23: quantitative SPECT for patient-specific 3-dimensional dosimetry in internal radionuclide therapy. J Nucl Med 2012;53(8):1310–25.

19. Dewaraja YK, Ljungberg M, Green AJ, et al. MIRD pamphlet No. 24: Guidelines for quantitative 131I SPECT in dosimetry applications. J. Nucl. Med. 2013;54(12):2182–8.

20. Duda RO, Hart PE. Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM 1972;15(1):11–5.

21. Rose A. The sensitivity performance of the human eye on an absolute scale. J Opt Soc Am 1948;38(2):196–208.

22. Bright DS, Newbury DE, Steel EB. Visibility of objects in computer simulations of noisy micrographs. J Microsc 1998;189(Pt 1):25–42.

23. Carlier T, Eugène T, Bodet-Milin C, et al. Assessment of acquisition protocols for routine imaging of Y-90 using PET/CT. EJNMMI Res 2013;3(1):11.

24. Jentzen W, Hoppenbrouwers J, van Leeuwen P, et al. Assessment of Lesion Response in the Initial Radioiodine Treatment of Differentiated Thyroid Cancer Using 124I PET Imaging. J Nucl Med 2014;55(11):1759–65.

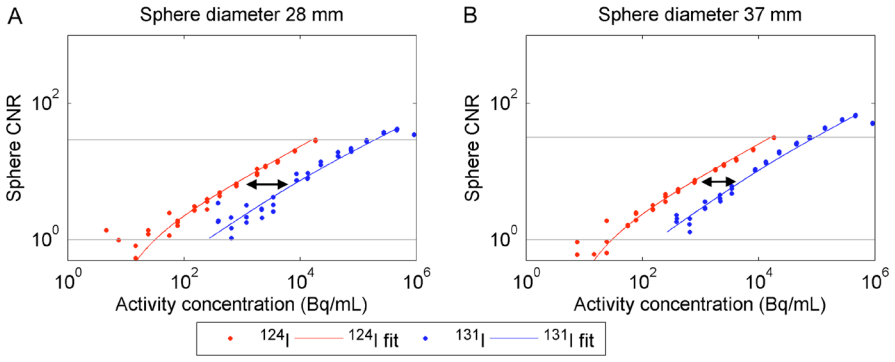
25. McDougall IR, Iagaru A. Thyroid stunning: fact or fiction? Semin Nucl Med 2011;41(2):105–12.

26. Salvatori M, Perotti G, Villani MF, et al. Determining the appropriate time of execution of an I-131 post-therapy whole-body scan: comparison between early and late imaging. Nucl Med Commun 2013;34(9):900–8.

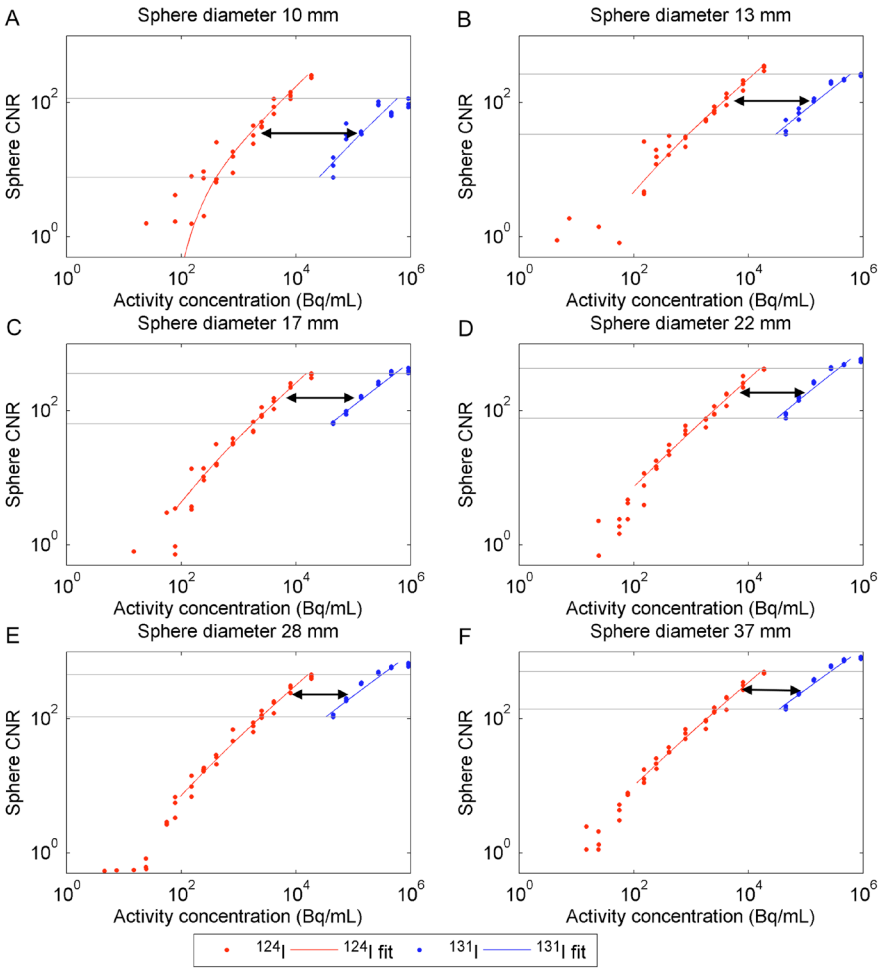
27. Chong A, Song H-C, Min J-J, et al. Improved detection of lung or bone metastases with an I-131 whole body scan on the 7th day after high-dose I-131 therapy in patients with thyroid cancer. Nucl Med Mol Imaging 2010;44(4):273–81.

28. Jentzen W, Freudenberg LS, Eising EG, Sonnenschein W, Knust J, Bockisch A. Optimized 124I PET Dosimetry Protocol for Radioiodine Therapy of Differentiated Thyroid Cancer. J Nucl Med 2008;49(6):1017–23.

29. Pettinato C, Spezi E, Nanni C, et al. Pretherapeutic dosimetry in patients affected by metastatic thyroid cancer using 124I PET/CT sequential scans for 131I treatment planning. Clin Nucl Med 2014;39(8):e367–74.



Supplemental Figure 1 : The sphere CNR as a function of activity concentration for the 124I (red) and 131I (blue) phantom with background for spheres with a diameter of (A) 28 mm and (B) 37 mm. The DEP is determined by calculating the ratio of the curves, which is graphically represented as the horizontal shift of the curves (black arrow). The gray lines denote the interval where the average ratio between the curves was determined. PET images were reconstructed using the PSF TOF method and SPECT images using the SC 6i8s method



Supplemental Figure 2: The sphere CNR as a function of activity concentration for the 124I (red) and 131I (blue) phantom without background for spheres with a diameter of (A) 10 mm, (B) 13 mm, (C) 17 mm, (D) 22 mm, (E) 28 mm and (F) 37 mm. The DEP is determined by calculating the ratio of the curves, which is graphically represented as the horizontal shift of the curves (black arrow). The gray lines denote the interval where the average ratio between the curves was determined. PET images were reconstructed using the PSF TOF method and SPECT images using the SC 6i8s method.

Isotope	Initial activity concentration spheres (Bq/ml)	Initial activity concentration background (Bq/ml)	Ratio
^{124}I	$1.8x10^4$	0	-
^{124}I	$1.8x10^4$	$1.8x10^3$	10
^{131}I	$9.1x10^5$	0	-
^{131}I	$9.1x10^5$	$9.0x10^4$	10

Supplemental table 1: Initial activity concentrations of the phantoms at time of the first experiment.

PART 3

General discussion and summary

GENERAL DISCUSSION AND FUTURE DIRECTIONS

CHAPTER 10

Jakob W. Kist

The major subjects discussed in this thesis relate to two current questions in contemporary clinical practice: firstly, how to characterize the numerous thyroid nodules that are (incidentally) diagnosed; and secondly, how to approach previously treated DTC patients with biochemical suspicion of recurrence? In this chapter, the studies presented in this thesis are discussed in light of recent literature with the aim to answer these questions. Furthermore, this chapter discusses perspectives on the directions of future research.

Diagnosis of differentiated thyroid cancer

The role of real-time elastography in diagnosing thyroid cancer

The increasing of different imaging modalities results in the sharp increase of detected thyroid nodules.[1] Of these, the majority is small and do not threaten patients' health or survival.[1] However, if a nodule has been detected the clinician and the patient are highly likely to feel the need to further evaluate its nature. The malignant potential of thyroid nodules partially justifies this. However, health resources to obtain a final diagnosis should be used as effectively as possible. In this light, we showed that elastography is an effective tool to identify benignity of thyroid nodules. The results presented in **chapter 2** are clear: real-time elastography (RTE) has a negative predictive value (NPV) of 97% and can thus play an important role in reducing the number of FNAs performed for thyroid nodules. The major disadvantage of elastography is its low positive predictive value (PPV). Any non-soft nodule still requires further work-up with FNA. Other disadvantages of elastography are its technical limitations: it is difficult to use in multinodular disease, obese patients and for nodules located more posterior or inferior in the thyroid.[2] Several studies studied the interobserver variability of elastography, but the results are conflicting.[3,4] Main argument in favor of existence of any interobserver variability is that RTE is a non-quantitative method, and thus its outcome depends on the technique of the operator. More advanced elastography methods, like the semi-quantitative strain-ratio elastography (SRE) and quantitative shear wave elastography (SWE), may aid to overcome some of these limitations. In SRE the ratio between two regions-of-interest (ROI) is calculated: one drawn over the nodule and one over healthy thyroid tissue. SRE is considered semi-quantitative because the result represents a relative quantitation and it therefore still depends on the pressure applied by the operator. The latter is the main difference with SWE, in which the pressure is standardized by an acoustic force emitted by the probe. The transverse displacement (the shear waves) can be measured by the device and represents the degree of elasticity of the ROI.

Both SRE and SWE were analyzed in recent meta-analyses.[5-7] Interestingly, none of the meta-analyses proved either technique to be superior to RTE. The main strength of RTE - diagnosing benignity - was similar to that of SWE in one study[6], while it was not reproduced in another study.[7] A higher PPV for SWE

in comparison to RTE was found by all meta-analyses, meaning that fewer patients need to undergo additional diagnostic work-up for a benign nodule. However, this advantage seems irrelevant because the lower NPV found in two of these studies means that a larger proportion of malignant nodules is falsely classified as benign.[5,7] Notably, the number of nodules included in these studies was significantly lower in comparison to our study, emphasizing the limited available evidence supporting the use of these techniques.

Thus adequately powered, prospective studies with clear eligibility criteria are warranted to determine the definite role of these techniques, preferably conducted with a blind comparison of the different elastography techniques. However, the data acquired in our study are convincing enough to omit an FNA in purely soft nodules. Close monitoring during routine is warranted to evaluate its diagnostic performance in daily practice.

In another meta-analysis by Trimboli et al. RTE was applied on an interesting different target population: patients with an indeterminate FNA result (in this study Bethesda category III and IV).[8] Remarkably, the high NPV (97%) from our own meta-analysis presented in **chapter 2** was not reproduced in this study (82%).[8] Although the higher incidence of malignant nodules in this subpopulation (~20%) lowered the NPV per definition, it does not fully explain this significant difference in NPV. The results of the meta-analysis were probably muddled by the methodological limitations of the included studies (e.g. small sample size and/or inhomogeneous study population) and the known significant inter-observer variability for thyroid FNA cytology interpretation.[9] Future studies focusing on indeterminate nodules - sufficiently powered and preferably with centrally reviewed FNA samples - are needed to confirm or invalidate the accuracy of RTE in this target population.

The Dutch and American guidelines on thyroid nodules and DTC published in 2015 do not yet advocate the routine use of any form of elastography for thyroid nodule evaluation.[2,10] Reasons behind this are likely the afore mentioned disadvantages of the technique, the relative low-quality of the available evidence, the limited availability of elastography, and the investment required for the purchase of the computational module and training of staff. However, the investment costs might be outweighed by cost savings by reducing the number of FNAs and thyroid lobectomies for benign nodules. Future cost-effectiveness studies should help to determine the applicability of elastography in daily practice.

Fine needle aspiration of thyroid nodules

Despite the development of the different elastography techniques, FNA remains the cornerstone of thyroid nodule evaluation. Nonetheless, for two categories of FNA results the next appropriate diagnostic step is not clear-cut: nondiagnostic or unsatisfactory cytology (Bethesda category I) and indeterminate cytology (mostly referred to as only Bethesda category III and IV, although some studies also include category V).

Nondiagnostic fine needle aspirations

Nondiagnostic FNAs results lead to repeat FNAs and (futile) diagnostic lobectomies, both causing patient anxiety.[11] The authors of the Bethesda System for Reporting Thyroid Cytopathology state in their guideline that ideally a maximum of 10% of FNAs should be nondiagnostic.[12] This guideline sharply contrasts with reported nondiagnostic rates in literature (up to 34%; average 12-17%)[13], a previous study at our institution (54%)[14] and the more recent study presented in chapter 4 (35%). On top of that, 39% of the respondents of a recently published survey among members of The Endocrine Society and American Thyroid Association (ATA) reported nondiagnostic rates of over 10%.[15] Published adequacy rates are thus likely to be subject to publication bias, leaving studies showing higher rates unpublished or even unnoticed by centers that do not adequately evaluate their own performance.

Nevertheless, the nondiagnostic rate at our center is too high according to professional standards and efforts have been made to reduce it. As discussed in chapter 4, several factors were associated with nondiagnostic FNA results on multivariate analysis: three or more clinic visits for FNA of the same thyroid nodule, rapid onsite adequacy assessment (ROSAA) and intranodular vascularization. The use of ROSAA significantly improved accuracy rates in our study. This might not only be due to the direct assessment of the FNA sample by a cytopathologist, but also to the increased number of needle passes.[11] Obtaining more aspirates per nodule is a less costly and less time consuming alternative to ROSAA. In the afore mentioned survey among endocrinologists this was the most common approach to increase the diagnostic yield.[15] The large number of different operators in our study (129 different operators performing 1381 FNAs) might also have contributed to the high nondiagnostic rate of 35% in our study. The consensus in literature is that accuracy of FNAs - in general and as well specifically for thyroid nodule FNA - is strongly operator dependent. The learning curve of thyroid nodule FNA is fairly short, limiting the effort to train physicians. More important seems that operators perform a significant number of FNAs on a regular basis.[15] Our center could relatively easily pursue this by reducing the number of different operators and by training specialized radiologists.

In conclusion, there are several different aspects of the FNA procedure that can be optimized. Every hospital should monitor their adequacy rate and adapt the FNA procedure if necessary. This should result in a higher accuracy rate, which should be confirmed by continuous evaluation.

Besides improvements in the FNA technique itself as discussed, two strategies to reduce the number of nondiagnostic results are of interest: biopsies performed with a screw needle or core needle (CNB) and molecular testing.

Recently, two meta-analyses were published on CNB versus repeat FNA after initial nondiagnostic FNA, which is a very relevant potential application of CNB in my opinion. Suh et al. reported a significantly different nondiagnostic rate of 1.4% for CNB versus 34.2% for repeat FNA.[16] The second meta-

analysis performed a similar analysis showing that the relative risk of gaining a nondiagnostic result after CNB was 0.22 (95%-CI 0.10-0.45), meaning that the likelihood of a nondiagnostic result after CNB are over four times lower compared to repeat FNA.[17] An analysis of the patients' comfort and tolerability of these techniques showed similar results as for FNA.[17] The reported complication rates of CNB of thyroid nodules are very low.[16] It should be noted that - like FNA - CNB cannot differentiate between a follicular adenoma and follicular carcinoma because the complete nodule has to be reviewed by the pathologist in order to detect invasion through the capsule.

In a survey members of The Endocrine Society and ATA suggested molecular testing as a potentially valuable diagnostic approach to nondiagnostic FNA.[15] However, it is questionable whether these tests on samples that do not meet diagnostic criteria (i.e. scant cellularity or bloody aspirates) will be able to accurately diagnose the nature of the nodule. Future studies on the role of CNB and molecular tests as additional diagnostic tools for non-diagnostic FNA results are needed to evaluate their definite value.

Indeterminate thyroid nodules

Mutation analysis for indeterminate nodules is gaining more and more attention in literature. Various mutations or combinations of mutations can be tested in order to correctly differentiate benign from malignant indeterminate nodules. In an interesting study by Kleiman et al. the influence of BRAF(V600E) testing of nodules with indeterminate FNA results (Bethesda category III, IV and V in this study) on the surgical strategy was evaluated.[18] In their study the BRAF(V600E) analysis was performed by a polymerase chain reaction (PCR) analysis. They concluded that it did not alter the surgical approach significantly. This conclusion does not stand when the Dutch approach had applied to their data as extensively discussed in **chapter 4**. Main difference is that Kleiman et al. propagated a total thyroidectomy for nodules with a suspicious for malignancy (Bethesda category V), while in the Netherlands a diagnostic lobectomy would be standard.[10] Due to this difference in approach BRAF(V600E) analysis of indeterminate nodules is cost-effective in the Netherlands by reducing the number of two-stage procedures and thus the costs and time until definite diagnosis.

In the future, BRAF(V600E) mutation analysis by immunocytochemical staining of direct smears of FNA samples might be an even cheaper alternative to PCR analysis and might thus increase the cost-effectiveness even further. Disadvantage of this technique is that nonspecific staining of colloid and macrophages might lead to false positive antibody trapping. A recent study on this subject reported, however, that with growing experience it was possible to discriminate true staining from false positive staining.[19]

BRAF(V600E) mutation analysis approach is not cost-effective when applied on nodules with a Bethesda III or IV FNAC result, as only around 2% of these nodules are BRAF(V600E) mutant.[20] Several commercial molecular tests have been proposed to guide decision making for these nodules, of which

a multigene expression classifier *(Afirma)*[21] and the molecular panels *ThyGenX Thyroid Oncogene Panel, ThyraMIR Thyroid miRNA Classifier*[22] and *ThyroSeq v2*[23] are the most renowned. Reports on their usefulness and cost-effectiveness are conflicting. Data on this matter applied in the Netherlands with the local treatment strategies, reimbursements and costs, are lacking. To our knowledge the current use in daily practice of these tests in the Netherlands is negligible.

Another potential diagnostic tool in the work-up of indeterminate nodules is ^{18}F-FDG PET/CT. In a meta-analysis by Vriens et al. ^{18}F-FDG PET/CT proved to have an excellent sensitivity, especially in nodules >1,5 cm (100%).[24] Another study from the same group used a Markov decision model to show that implementation of a PET-driven policy reduce the number of futile diagnostic surgeries for benign thyroid nodules in a cost-effective and oncologically safe manner.[25] This multifactorial hypothesis is currently being tested in a nationwide multicenter study.[26]

In conclusion, BRAF(V600E) testing on Bethesda category V FNA samples should be standard of care in the Netherlands as it effectively reduces the number of two-stage thyroidectomies. However, the search for the best diagnostic tool that can accurately determine whether a thyroid nodule with an indeterminate FNA result is benign or malignant is currently of great interest in literature and the future will tell which approaches will prevail.

Diagnosis of recurrent differentiated thyroid cancer

Determining the appropriate diagnostic and therapeutic strategy for patients with increased Tg-levels without proven locoregional recurrence is challenging. Diagnostic modalities such as ^{123}I and ^{131}I WBS have limited sensitivities, while blind or empiric ^{131}I therapy is futile in one third to half of the patients[27,28]. The Thyropet study - for which the protocol is described in chapter 6 - aimed to determine whether the combination of ^{124}I and ^{18}F-FDG PET/CT could prevent futile ^{131}I therapies.

Quantification of ^{124}I

A secondary aim of the Thyropet project was to create a calibration procedure enabling quantitative comparisons between participating centers. The procedure to realize this aim and its results is described in chapter 7. The reduction in relative error of the calibrated activity achieved, proved this procedure to be of added value. However, as the Thyropet study has been preliminary terminated **(chapter 8)**, regular repetition of the calibration procedure to maintain standardized quantification is currently not offered by any institution or organization in the Netherlands.

Despite the negative result of the Thyropet study, quantification of ^{124}I PET/CT scans might play a pivotal role in the future. One relevant application could be to use dosimetric analyses to determine optimal ^{131}I dosage. Currently the high dose blind therapies given in the Netherlands are all fixed dose (i.e.

5500 or 7400 MBq of ^{131}I). Multiple dosimetric analyses, however, showed that the delivered dose to DTC metastases varies widely. In one study average absorbed dose (AD) by the metastatic DTC lesions ranged from 0.22 to 657 Gray (Gy) per administered MBq of ^{131}I.[29]

In a recently published study ^{124}I PET/CT dosimetry was performed to determine the AD to metastases in a fixed dose setting.[30] The metastatic lesions with a complete response had significantly higher AD in comparison to metastases showing an incomplete response. These studies highlight the downsides of a fixed dose approach. The fixed dose ^{131}I therapies tend to deliver either too much or too little radiation to the target lesion. The first outcome is not in line with the 'As Low As Reasonably Achievable' (ALARA) principle, the latter lacks efficacy, while the patient is exposed to a treatment with associated morbidity and radiation. Hence, dosimetry should be pursued in patients with proven iodine-avid metastasis in the (near) future, requiring comparable and especially reliable quantification. The procedure described in **chapter 7** can be used to fulfill this need. It should be noted that because it requires specific and specialized knowledge and only relatively few patients per year need high dose ^{131}I treatments, centralization of this care to only a few specialized centers in the Netherlands is warranted.

Another potential use of dosimetric analyses of ^{124}I PET/CT scans is to determine therapy effect of targeted therapies, such as Selumetinib, a Tyrosine-kinase inhibitor.[31] Based on quantitative analysis of these ^{124}I PET/CT scans the study showed that the effect of the drug was significant, making it a potent new treatment strategy. Several other recently published studies on ^{124}I PET/CT in both benign and malignant thyroid disease used ^{124}I quantification.[30,32] To validate and expand the results of these studies, standardized ^{124}I scanning and comparable quantification are essential. Organizations like EATRIS and EANM should be encouraged to offer ^{124}I calibration, for which the results of our project described in chapter 7 can be the basis.

The Thyropet study – potential causes of false negative ^{124}I PET/CT scans

The main results of the Thyropet study, as presented in **chapter 8,** leave no room for doubt: ^{124}I PET/CT using 74 MBq ^{124}I after preparation with rhTSH is unable to predict ^{131}I WBS outcome reliably - bearing in mind the applied scan protocol used and quality of the scanners included. Every one of these aspects and variables should be investigated as a potential underlying cause and will be discussed in the following paragraphs.

The method of preparation on the scan and therapy: rhTSH versus thyroid hormone withdrawal

In our study the ^{124}I PET/CT scans were made after preparation with rhTSH whilst the blind ^{131}I therapy was given after thyroid hormone withdrawal (THW). The method of preparation with rhTSH might very well result in

lower iodine uptake in comparison THW. It is currently unclear how much this influence is per lesion and whether large differences per lesion can be expected. It is possible that this influence is negligible in some lesions, while in other lesions significantly increased uptake of iodine is seen after THW. New studies with intra-patient comparisons are warranted to elucidate this.

Another interesting aspect of this discussion is that it should be kept in mind that rhTSH was not developed for patients with metastatic DTC and thus with possibly reduced iodine avidity due to (early phase) dedifferentiation. Purely hypothetically, administering three or more doses of rhTSH instead of two doses of 0.9 mg rhTSH intramuscularly on consecutive days, could give similar - or at least more comparable - results to THW. Similarly, a higher dose per rhTSH administration may be warranted. Although higher or more doses of rhTSH might increase side effects and will definitely increase treatment costs it should be studied further, because accurate diagnostic scanning performed without the necessity of THW and with the potential to prevent futile ^{131}I therapies could be of great value.

Scan protocol and scanner improvements

Another aspect that could have contributed to the limited diagnostic accuracy of ^{124}I PET/CT in the Thyropet study are the scan protocol and the current generation of PET/CT scanners.

The scan protocol used in the Thyropet study was designed such that the maximum amount of time for a whole body PET-scan did not exceed 30 minutes. Longer acquisition times per bed position might result in better detection of lesions with low iodine uptake, reducing the number of false negative scans. However, this approach would result in longer total scan times; in daily practice the total scan time is preferably limited to 30 minutes to maximize the number of time slots per day. Although other studies used similar scan protocols for ^{124}I PET/CT scanning, these had a tendency towards longer scan time per bedposition.

In addition to optimization of the scan protocol, optimized reconstruction protocols for ^{124}I may improve diagnostic accuracy. In the Thyropet study we found that reconstruction artifacts occurred in some scanners in areas with high physiological uptake, like the stomach and the urinary bladder. These artifacts caused that no background activity, not even scatter, surrounded these areas (Fig. 1). Such artifacts can cause small tumor depositions in these locations to be missed. Reconstruction protocols specifically designed and adapted for ^{124}I are therefore necessary.

Besides such approaches to optimize acquisition of ^{124}I PET/CT two hardware innovations in PET scanning are of interest: PET/MRI and digital PET. PET/MRI is a hybrid technique combining PET images with magnetic resonance imaging (MRI) instead of CT, offering potential advantages above PET/CT, such as non-iodinated contrast agents and superior resolution. Binse et al. performed a head-to-head comparison between ^{124}I PET/MRI and ^{124}I PET/CT in patients at high risk of iodine-avid DTC metastases.[33] PET/MRI led

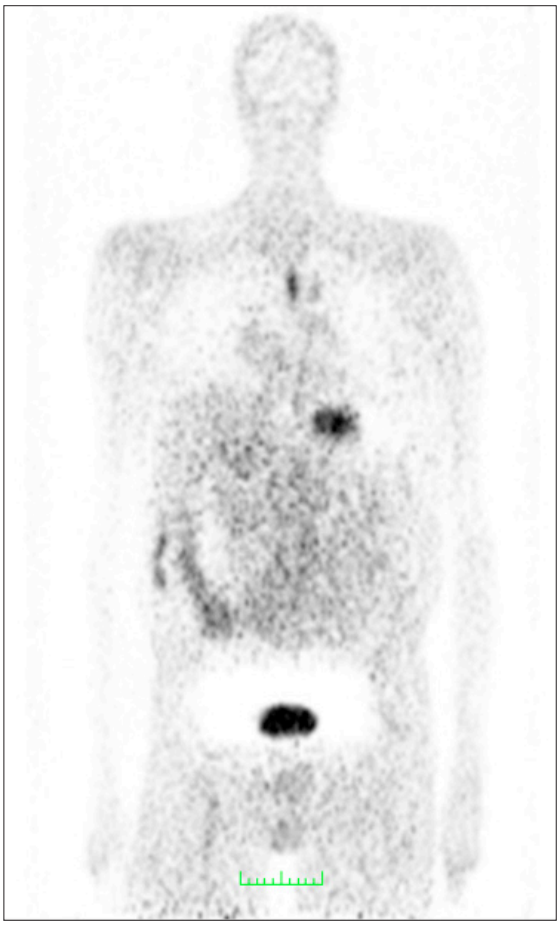


Figure 1. Coronal image of 124I PET showing physiological uptake in esophagus, stomach and bladder. Reconstruction artefacts around stomach and especially around bladder uptake.

to a significant higher number of detected iodine positive metastasis and thyroid remnants and also led to the detection of additional iodine negative metastasis. Although the results seem encouraging, others described several disadvantages of the technique: longer acquisition time, less cost-effective and less accurate attenuation correction.[34]

Digital PET is a second technical improvement using high-performance digital detectors instead of conventional photomultipliers, amongst other technical improvements. These improvements potentially result in higher spatial, energy and timing resolution, leading to improved PET images and quantification in comparison to the current generation of PET scanners.[35] The use of digital PET scanners is currently very limited and literature on clinically relevant benefits of this technique above conventional PET is scarce. To our knowledge no data is available using ^{124}I for DTC patients on digital PET, so it is too early to determine what role it could play in these patients, but it may (partially) overcome the current downsides of ^{124}I PET/CT in the future.

Administered activity of ^{124}I

A final potential cause of ^{124}I PET/CT to be false negative was studied in chapter 9: the difference in detectability between the ^{124}I PET/CT with a low activity (74 MBq ^{124}I; ~1% of ^{131}I) and the post-therapy ^{131}I SPECT/CT. This study showed that for larger lesions (>10 mm) with suboptimal PET reconstructions (i.e. non-TOF, non PSF) applied, the scouting dose ^{124}I is too low to result in a similar detectability in comparison to the post-therapy ^{131}I SPECT/CT. In the Thyropet study six of the seven false negative lesions were smaller than 10 mm, and the reconstruction protocols applied on these scans all were with PSF and TOF. The results of the study presented in chapter 9 did therefore not clarify why the ^{124}I scans were false negative in the Thyropet study. It did, however, show that higher doses of ^{124}I significantly increase detectability. In the previously discussed study by Ho et al. 222 MBq of ^{124}I was administered.[31] The protocol of that study referred to another study (not published) by the same investigators using the same dose of ^{124}I in which no toxicity was found by the higher dose. While this suggests it is safe to administer a higher ^{124}I dose, the costs will rise equally (around €2500 for 222 MBq of ^{124}I in the Netherlands) and the controversial issue on the so-called stunning effect will emerge.

The Thyropet study – the results in comparison to recent literature

In the previous paragraphs technical and physiological parameters that might have resulted in the ^{124}I PET/CT scans being false negative were discussed. In this part the results of the Thyropet study in light of other studies that have been recently published on this subject will be discussed.

In a large retrospective study by Ruhlmann et al. 137 DTC patients were included who underwent serial ^{124}I PET/CT using 24 MBq of ^{124}I and subsequently high dose radioiodine therapy.[36] The ^{124}I and ^{131}I scans were both derived using the same method of preparation for both scans. Of those 137 patients the majority (n=106) were high risk DTC patients at initial diagnosis and 31 were patients with increasing thyroglobulin levels or with indistinct findings during the first radioiodine therapy. Unfortunately, it was not possible to exclude the results of those 106 high-risk DTC patients to come to a more equal comparison with the Thyropet results. However, the results of their study were clear: of the 227 metastases detected by either ^{124}I PET/CT or ^{131}I SPECT/CT 97% were detected by both modalities. Only four lesions were negative on ^{124}I PET/CT and positive on ^{131}I SPECT/CT. As a potential explanation for the false negative ^{124}I results they suggest that these lesions have slow iodine kinetics leading to undetectable amounts of ^{124}I uptake, even after 5 days. The results of this study sharply contrast with our results. Although the study is a retrospective single center study including a heterogeneous cohort of patients, the number of patients included is sufficiently high and the presented sensitivity is that good that the results cannot be ignored. The most notable difference between this study and the Thyropet study is that the method of preparation for the ^{124}I PET/CT and the ^{131}I therapy were the same for each patient in their study, which might have led to more concordant scan results.

Another recently published study by Gulec et al. was prospectively designed and aimed to analyze ^{124}I PET/CT as a clinical imaging tool in DTC patients.[37] In this study a heterogeneous group of 15 patients was included, among which 7 patients with suspicion of metastatic disease similar to the Thyropet study population. The authors concluded on a lesion based analysis that ^{124}I is a valuable diagnostic modality with a reported sensitivity of 92.5% to detect iodine avid lesions. However, on a *patient* level in 3 out of 15 patients the ^{124}I PET/CT was false negative, resulting in a significant lower sensitivity of 80%. On top of that, when analyzing only the 7 patients with suspected recurrence of DTC, 4 out of 7 ^{124}I PET/CT scans were false negative (sensitivity 57%, 95% CI 18-90%).[38] This study underlines the importance of homogeneity of a study cohort and that clinically relevant outcomes have to be chosen in order to come to reliable conclusions. Similar to the study by Ruhlmann et al. patients were prepared with the same method, rhTSH or THW, for the ^{124}I PET/CT and ^{131}I SPECT/CT. This seems to contradict the suggestion that the consistency in preparation method is the most important factor for the high sensitivity found by Ruhlman et al.

From these studies and the results of the Thyropet study it is clear that more research is warranted on this matter. Specifically, large prospective diagnostic cohort studies with uniform scan protocols, strict inclusion criteria and clinically relevant endpoints are warranted. Furthermore, it would be very interesting to critically review our approach in detail with the group of Ruhlmann et al. to determine whether there are differences with their methods compared to ours, and if so, how ours could be optimized to come to comparable or at least more comparable scan results.

Optimal approach to patients suspected of recurrent DTC

While these issues regarding ^{124}I PET/CT should be further explored in future studies, the current clinical dilemma on the optimal approach of patients with suspicion of recurrence of DTC, remains. In short several aspects will be discussed that should be considered in the multidisciplinary meeting, in which these patients should be discussed.[39]

First, a representative US of the neck has to be performed and a FNA of any atypical lymph is warranted. In the Thyropet study 35% of the patients had metastatic DTC in a locoregional lymph node, even though all patients had a negative neck US upon inclusion of the study. This result is not unique; others found similar false negative rates.[40] This limitation of US should be kept in mind, and US should be performed by experienced and dedicated radiologists, aware of the characteristics of potentially pathological lymph nodes. If there is any suspicion of metastatic disease in a lymph node, FNA should be performed, optionally combined with measurement of Tg in the washout of the FNA needle.

The second point of consideration in patients with increased Tg-levels is the absolute level of the Tg. The 2015 Dutch guidelines on DTC treatment advocate considering ^{131}I treatment yet with Tg-levels of 1 ng/ml or higher.[39]

Most other guidelines advise to treat patients 'blind' with ^{131}I if Tg-levels are over 10 ng/ml.[2,41] The relatively low cut-off for ^{131}I therapy used in the Netherlands might be a cause of false negative post-therapy ^{131}I scans. The tumor volume and corresponding low iodine uptake in these patients might be below the detection limits of ^{131}I WBS and/or SPECT/CT. As patients are usually considered to have iodine refractory disease if a post-therapy ^{131}I scan is negative[2], the approach advocated in the Dutch guideline may result in patients being wrongfully classified as such, which should be prevented at all times.

The large difference between the Tg threshold in Dutch and other guidelines for ^{131}I therapy is remarkable, but both approaches are tenable. At low Tg levels^{131}I therapy may be more effective as the tumor load is small and the chances of dedifferentiation are limited. However, the disadvantage of this approach is that post therapy scans are more likely to yield false negative results because of the low tumor volume. On the other hand, using a higher Tg threshold, as advocated by most other guidelines, leads to fewer (false) negative post therapy scans, due to the larger tumor volume, but this approach increases the chances of dedifferentiation. Furthermore, in a fixed dose approach the chance that the AD is too low to result in a complete response is also increased.

Third, in the Thyropet study all patients underwent both a ^{124}I PET/CT scan and ^{18}F-FDG PET/CT scan. The results of these scans were not incorporated in **chapter 8,** because it would distract from the main message of that paper. Nonetheless, ^{18}F-FDG PET/CT remains an important diagnostic modality in patients with increased Tg-levels. If widespread FDG-avid DTC metastases are seen, the patient can be considered refractory to ^{131}I and other therapeutic strategies should be pursued.

With all the afore described considerations in mind, we believe the current strategy for patients with increased Tg-levels should remain 'blind' ^{131}I therapy, if 1) there has been a significant or continuing increase of Tg-level (thus a single Tg-level just above 1 ng/ml should be repeated in 3-6 months), 2) neck US - by an experienced radiologist - is negative and 3) ^{18}F-FDG PET has ruled out widespread iodine refractory metastases and surgically resectable metastases, with the latter being performed if the suspicion of dedifferentiation is high, based on the characteristics of the primary tumor, the absolute Tg-levels and its doubling time. In conclusion, an individualized approach is warranted and every patient should be discussed within a multidisciplinary team.

In conclusion, the results of the Thyropet study showed that treating patients with suspected recurrence blind with ^{131}I should be limited, as 47% of the post-therapy scans were negative. This underlines the desperate need for pre-therapy imaging correctly identifying those patients that are likely to benefit from the ^{131}I therapy. The second clear result of the Thyropet study is that ^{124}I PET/CT as applied in our study is not able to fulfill this need due to its high rate of false negativity. All factors potentially contributing to this

should be further studied to increase the diagnostic accuracy. Furthermore, collaborative efforts should be made to create a calibration network for ^{124}I PET/CT scanners, as quantification will become more important in the future.

References

1. Davies L, Welch HG. Current thyroid cancer trends in the United States. JAMA Otolaryngol Head Neck Surg 2014;140(4):317–22.

2. Haugen BR MD, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2015;:thy.2015.0020.

3. Park SH, Kim SJ, Kim E-K, Kim MJ, Son EJ, Kwak JY. Interobserver Agreement in Assessing the Sonographic and Elastographic Features of Malignant Thyroid Nodules. AJR Am J Roentgenol 2009;193(5):W416–23.

4. Lim D-J, Luo S, Kim M-H, Ko SH, Kim Y. Interobserver agreement and intraobserver reproducibility in thyroid ultrasound elastography. AJR Am J Roentgenol 2012;198(4):896–901.

5. Sun J, Cai J, Wang X. Real-time ultrasound elastography for differentiation of benign and malignant thyroid nodules: a meta-analysis. J Ultrasound Med 2014;33(3):495–502.

6. Veer V, Puttagunta S. The role of elastography in evaluating thyroid nodules: a literature review and meta-analysis. Eur Arch Otorhinolaryngol 2014;

7. Liu B-J, Li D-D, Xu H-X, et al. Quantitative Shear Wave Velocity Measurement on Acoustic Radiation Force Impulse Elastography for Differential Diagnosis between Benign and Malignant Thyroid Nodules: A Meta-analysis. Ultrasound Med Biol 2015;

8. Trimboli P, Treglia G, Sadeghi R, Romanelli F, Giovanella L. Reliability of real-time elastography to diagnose thyroid nodules previously read at FNAC as indeterminate: a meta-analysis. Endocr 2014;

9. Cibas ES, Baloch ZW, Fellegara G, et al. A prospective assessment defining the limitations of thyroid nodule pathologic evaluation. Ann Intern Med 2013;159(5):325–32.

10. Links TP, de Heide LJM, Janssen M, et al. [Guideline thyroid cancer including diagnostics of the nodule]. Ned Tijdschr Geneeskd 2015;159:A9413.

11. Redman R, Zalaznick H, Mazzaferri EL, Massoll NA. The impact of assessing specimen adequacy and number of needle passes for fine-needle aspiration biopsy of thyroid nodules. Thyroid 2006;16(1):55–60.

12. Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. Thyroid 2009;19(11):1159–65.

13. Witt BL, Schmidt RL. Rapid onsite evaluation improves the adequacy of fine-needle aspiration for thyroid lesions: a systematic review and meta-analysis. Thyroid 2013;23(4):428–35.

14. de Meer SGA, Schreinemakers JMJ, Zelissen PMJ, et al. Fine-needle aspiration of thyroid tumors: identifying factors associated with adequacy rate in a large academic center in the Netherlands. Diagn Cytopathol 2012;40 Suppl 1:E21–6.

15. Brito JP, Castro MR, Dean DS, Fatourechi V, Stan M. Survey of current approaches to non-diagnostic fine-needle aspiration from solid thyroid nodules. Endocr 2015;49(3):745–51.

16. Suh CH, Baek JH, Kim KW, et al. The role of core-needle biopsy for thyroid nodules with initially non-diagnostic, fine-needle aspiration results: a systematic review and meta-analysis. Endocrine Practice 2016;22(6):679–88.

17. Wolinski K, Stangierski A, Ruchala M. Comparison of diagnostic yield of core-needle and fine-needle aspiration biopsies of thyroid lesions: Systematic review and meta-analysis. Eur Radiol 2016;

18. Kleiman DA, Sporn MJ, Beninato T, et al. Preoperative BRAF(V600E) mutation screening is unlikely to alter initial surgical treatment of patients with indeterminate thyroid nodules: a prospective case series of 960 patients. Cancer 2013;119(8):1495–502.

19. Wobker SE, Kim LT, Hackman TG, Dodd LG. Use of BRAF v600e immunocytochemistry on FNA direct smears of papillary thyroid carcinoma. Cancer Cytopathology 2015;123(9):531–9.

20. Trimboli P, Treglia G, Condorelli E, et al. BRAF-mutated carcinomas among thyroid nodules with prior indeterminate FNA report: a systematic review and meta-analysis. Clin Endocrinol (Oxf) 2015;

21. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. N Engl J Med 2012;367(8):705–15.

22. Labourier E, Shifrin A, Busseniers AE, et al. Molecular Testing for miRNA, mRNA, and DNA on Fine-Needle Aspiration Improves the Preoperative Diagnosis of Thyroid Nodules With Indeterminate Cytology. J Clin Endocrinol Metab 2015;100(7):2743–50.

23. Nikiforov YE, Carty SE, Chiosea SI, et al. Impact of the Multi-Gene ThyroSeq Next-Generation Sequencing Assay on Cancer Diagnosis in Thyroid Nodules with Atypia of Undetermined Significance/Follicular Lesion of Undetermined Significance Cytology. Thyroid 2015;25(11):1217–23.

24. Vriens D, de Wilt JHW, van der Wilt GJ, Netea-Maier RT, Oyen WJG, de Geus-Oei L-F. The role of [18F]-2-fluoro-2-deoxy-d-glucose-positron emission tomography in thyroid nodules with indeterminate fine-needle aspiration biopsy: syst... - PubMed - NCBI. Cancer 2011;117(20):4582–94.

25. Vriens D, Adang EMM, Netea-Maier RT, et al. Cost-effectiveness of FDG-PET/CT for cytologically indeterminate thyroid nodules: a decision analytic approach. J Clin Endocrinol Metab 2014;99(9):3263–74.

26. Vriens D, de Geus-Oei LF. FDG-PET/CT in Evaluation of Cytological Indeterminate Thyroid Nodules to Prevent Unnecessary Surgery (EffECTS) [Internet]. [cited 2015 Nov 11];Available from: https://clinicaltrials.gov/ct2/show/NCT02208544

27. Ma C, Xie J, Kuang A. Is empiric 131I therapy justified for patients with positive thyroglobulin and negative 131I whole-body scanning results? J Nucl Med 2005;46(7):1164–70.

28. van Tol KM, Jager PL, de Vries EGE, et al. Outcome in patients with differentiated thyroid cancer with negative diagnostic whole-body scanning and detectable stimulated thyroglobulin. Eur J Endocrinol 2003;148(6):589–96.

29. Pettinato C, Spezi E, Nanni C, et al. Pretherapeutic dosimetry in patients affected by metastatic thyroid cancer using 124I PET/CT sequential scans for 131I treatment planning. Clin Nucl Med 2014;39(8):e367–74.

30. Wierts R, Brans B, Havekes B, et al. Dose-Response Relationship in Differentiated Thyroid Cancer Patients Undergoing Radioiodine Treatment Assessed by Means of 124I PET/CT. J Nucl Med 2016;57(7):1027–32.

31. Ho AL, Grewal RK, Leboeuf R, et al. Selumetinib-enhanced radioiodine uptake in advanced thyroid cancer. N Engl J Med 2013;368(7):623–32.

32. Gabler AS, Kühnel C, Winkens T, Freesmeyer M. Assessment of minimum activity of 124Iodine in pre-therapeutic uptake measurement prior to radioiodine therapy of benign thyroid diseases. J Nucl Med 2016;:jnumed.115.171066.

33. Binse I, Poeppel TD, Ruhlmann M, et al. Imaging with 124I in differentiated thyroid carcinoma: is PET/MRI superior to PET/CT? Eur J Nucl Med Mol Imaging 2015;

34. Dercle L, Deandreis D, Terroir M, Leboulleux S, Lumbroso J, Schlumberger M. Evaluation of (124)I PET/CT and (124)I PET/MRI in the management of patients with differentiated thyroid cancer. Eur J Nucl Med Mol Imaging 2016;

35. Nguyen NC, Vercher-Conejero JL, Sattar A, et al. Image Quality and Diagnostic Performance of a Digital PET Prototype in Patients with Oncologic Diseases: Initial Experience and Comparison with Analog PET. J Nucl Med 2015;56(9):1378–85.

36. Ruhlmann M, Jentzen W, Ruhlmann V, et al. High Level of Agreement between Pretherapeutic 124I PET and Intratherapeutic 131I Images in Detecting Iodine-positive Thyroid Cancer Metastases. J Nucl Med 2016;

37. Gulec SA, Kuker RA, Goryawala M, et al. (124)I PET/CT in Patients with Differentiated Thyroid Cancer: Clinical and Quantitative Image Analysis. Thyroid 2016;26(3):441–8.

38. Kist JW, de Keizer B, Vogel WV. Letter to the Editor regarding the article "I-124 PET/CT in patients with differentiated thyroid cancer: clinical and quantitative image analysis." Thyroid 2016;:thy.2016.0180.

39. Dutch thyroid cancer managment guidlines (2.0). 2015;1–188.

40. Wu L-M, Gu H-Y, Qu X-H, et al. The accuracy of ultrasonography in the preoperative diagnosis of cervical lymph node metastasis in patients with papillary thyroid carcinoma: A meta-analysis. Eur J Radiol 2012;81(8):1798–805.

41. Basu S, Dandekar M, Joshi A, D'Cruz A. Defining a rational step-care algorithm for managing thyroid carcinoma patients with elevated thyroglobulin and negative on radioiodine scintigraphy (TENIS): considerations and challenges towards developing an appropriate roadmap. Eur J Nucl Med Mol Imaging 2015;42(8):1167–71.

SUMMARY (IN ENGLISH)

CHAPTER 11

The aim of this thesis was twofold: first to improve the diagnosis of differentiated thyroid cancer (**part 1**) and second to improve the diagnosis of suspected recurrent differentiated thyroid cancer (**part 2**).

Part 1

Thyroid nodules are highly prevalent and may harbor a malignancy. It is important to find diagnostic methods that accurately discriminate between benign and malignant nodules. The current golden standard is US combined with FNA cytology, if indicated. As FNA is an invasive procedure, its use should be restricted to those cases in which non-invasive approaches fail to come to a definite diagnosis. In **chapter 2** the results of our systematic review and meta-analysis on the value of thyroid nodule elastography. The results of twenty studies, compromising 3973 nodules, were included. Using a standard cut-off between elastography scale (ES) 2 and scale 3 resulted in a pooled sensitivity and specificity of 85% (95% CI, 79–90%) and 80% (95% CI, 73–86%), respectively. The respective pooled NPV and PPV were 97% (95%CI, 94–98%) and 40% (95% CI, 34–48%). The sensitivity and NPV even further increased to 99% (95% CI, 96–100%) and 99% (95% CI, 97–100%) with a cut-off between ES1 and 2. From this we concluded that if a nodule appeared soft on elastography, the chance that it would be a malignant are negligible and FNA can be safely omitted. Both cut-offs can be used, and it should be determined per center what risk of missing a malignant thyroid nodule is accepted.

The use of elastography has been studied as well in different target populations. Trimboli et al. performed a meta-analysis of studies aiming to determine the role of elastography in patients that underwent FNA of their thyroid nodule and of which the FNA result was indeterminate (i.e. Bethesda III or IV). This analysis resulted in a disappointing pooled sensitivity of 69%, a specificity of 75%, and a respective NPV and PPV of 82 and 63%. In **chapter 3** we elaborated on the results of these meta-analyses and potential causes of the difference with the results in **chapter 2**. The higher incidence of malignant nodules might have partially contributed to the lower NPV in the study of Trimboli et al. However, this does not explain the significant difference entirely. Other potential causes are the low number of included studies and nodules and heterogeneity of the studies. At this point, however, we can only speculate on the causes of the modest accuracy of elastography in indeterminate nodules versus the high NPV in an unselected cohort of thyroid nodules.

In **chapter 4** we studied methods to decrease the percentage of nondiagnostic FNA results (i.e. Bethesda I). In a retrospective study we showed that rapid on-site adequacy assessment and three or more needle passes can improve adequacy rates.

Similar to the objective of using elastography in indeterminate nodules, BRAF mutation analysis on indeterminate FNA results, can be performed to determine whether a nodule is malignant. If a mutation in this gene is found,

the nodule is malignant as it is pathognomonic for DTC. We showed in **chapter 5** that this analysis on nodules with a Bethesda III or IV FNA result is cost-effective in the Netherlands.

Part 2

When a recurrence of DTC is suspected, based on increased Tg-levels, current guidelines advocate that a 'blind' therapy with ^{131}I should be given. However, up to 40% of the post-therapy scans is negative. This implicates that no benefit from the treatment can be expected, while the patient has been exposed to radiation and THW. The combination of ^{124}I and ^{18}F-FDG PET/CT is promising in determining those patients likely to benefit from high dose ^{131}I therapy. In **chapter 6** the rationale and design of the Thyropet study were presented. This study aimed to evaluate the value of the combination of ^{124}I and ^{18}F-FDG PET/CT in 100 patients suspected of recurrent DTC. Primary endpoint of the study was the accuracy of ^{124}I PET/CT to predict, at a patient level, the post-therapy ^{131}I WBS test result, as an operationalization of the impact of the implementation of ^{124}I PET/CT as a scouting procedure to set the indication for ^{131}I therapy. A stopping rule was defined such that if 3 or more ^{124}I PET/CT scans were false-negative the study would be preliminary stopped.

As the study was designed as a multicenter study there was a need to standardize ^{124}I PET/CT scanning to facilitate quantitative comparisons. In **chapter 7** we reported on our approach to calibrate 18 scanners of two different vendors in 16 different centers. A phantom containing six vials containing increasing activities of ^{124}I was scanned on every scanner twice within one week. The two measurements on each scanner did not differ, showing that the procedure was reproducible. A linear fit between measured (A_m) and reference activities (A_r) proved to be the better fit over a linear fit. The $_{cal}$ for scanners of Philips and Siemens ranged from 0.80 to 0.98 and 0.74 to 0.97, respectively. A Blant-Altman analysis showed that the relative errors of the calibrated activities was smaller the for the measured activities. This showed that our procedure improved the quantification per scanner and made the cohort of all scanners more accurate.

The main result of the Thyropet study was presented in **chapter 8**. After inclusion of 19 patients, of which two had to be excluded, the study was preliminary terminated, based on the predefined stopping rule. Eight out of 17 (47%) of the post-therapy ^{131}I WBS was negative, emphasizing the need to better select patients for ^{131}I therapy. All eight were correctly predicted by a negative ^{124}I PET/CT. Of the nine positive post-therapy WBS, four pre-therapeutic ^{124}I PET/CT scans were concordant, meaning that five were discordant. Implementation of ^{124}I PET in this setting would thus have led to 47% (8/17) less futile ^{131}I treatments, but 29% of patients (5/17) would have been denied the only potential curative therapy. We concluded that ^{124}I PET/CT as applied in the Thyropet study should not be implemented in standard care at this point.

The explicit results of the Thyropet study raised the question on what the underlying causes were of the ^{124}I PET/CT scans being false negative. A potential cause is the difference in administered activity of ^{124}I (usually around 74 MBq) versus the administered therapeutic activity ^{131}I (usually around 7400 MBq). On the other hand, PET has a higher sensitivity in comparison to SPECT. In **chapter 9** we aimed to determine whether the administered activity of ^{124}I is sufficient to achieve equal detectability. The NEMA-2007 image quality phantom was used to determine the contrast-to-noise ratio (CNR) for the different spheres at a range of activity concentrations. Subsequently, to determine the difference in detectability between ^{131}I SPECT/CT and ^{124}I PET/CT, the ^{124}I activity concentration was expressed as a percentage of the ^{131}I activity concentration required to achieve the same CNR. This metric was defined as the detectability equivalence percentage (DEP). Smaller spheres resulted in lower DEPs, so that a relatively low ^{124}I activity concentration is sufficient to achieve a similar detectability of lesions with ^{124}I PET/CT as with ^{131}I SPECT/CT. The DEP was 1.5, 1.9, 1.9, 4.4, 9.0 and 16.2% for the spheres with a diameter of 10, 13, 17, 18, 25 and 37 mm respectively, for attenuation and scatter corrected SPECT versus point spread function (PSF) modeled and time-of-flight (TOF) PET. When suboptimal PET reconstruction protocols (i.e. no-TOF and no-PSF) were used the DEP dropped significantly. From this we concluded that the relatively low activity used for ^{124}I PET/CT scanning, around 1% of ^{131}I activity, is sufficient for small spheres in optimal PET reconstruction settings. However, false-negative ^{124}I PET/CT as compared with the post-therapy ^{131}I SPECT/CT scans in larger lesions (>10 mm) and for no-PSF and no TOF PET might be caused by a difference in detectability.

NEDERLANDSE SAMENVATTING (VOOR NIET INGEWIJDEN)

CHAPTER 12

Samenvatting (voor niet ingewijden)

Het doel van dit proefschrift was enerzijds om de diagnostiek van schildklierkanker te verbeteren (**deel 1**) en anderzijds om de diagnostische methoden om uitzaaiingen te detecteren te verbeteren (**deel 2**).

Deel 1 – Het diagnosticeren van schildklierkanker

Veel mensen hebben een knobbel (nodus) in de schildklier. Een klein deel van deze nodi is kwaadaardig. Om een goedaardige nodus van een kwaadaardige te onderscheiden, dient de gebruikte diagnostische methode dit zo nauwkeurig mogelijk te doen. Dat wil zeggen dat de kansen dat deze methode een kwaadaardige nodus als goedaardig kwalificeert, of andersom, zo klein mogelijk zijn.

De gouden standaard voor de evaluatie van schildkliernodi is 1) echografie en 2) dunne naald aspiraat. Echografie is een radiologisch onderzoek dat door middel van geluidsgolven en de specifieke weerkaatsing van deze golven beelden van weefsels genereert. Bij een dunne naald aspiraat (FNA) brengt de radioloog tijdens de echografie een dunne holle naald in de nodus en zuigt daar enkele cellen uit. Vervolgens worden deze cellen onder de microscoop onderzocht om te bepalen of de nodus al dan niet kwaadaardig is. Omdat FNA een invasieve methode is, moet het gebruik daarvan beperkt worden tot die patiënten, waarbij een niet-invasieve methode niet in staat is om een definitieve diagnose te stellen.

In **hoofdstuk 2** hebben wij onderzoek gedaan naar het gebruik van elastografie in de diagnostiek van schildkliernodi. Elastografie is een relatief nieuwe speciale echografie techniek, waarbij ook de elasticiteit van een schildkliernodus kan worden bepaald. De mate van elasticiteit van de nodus wordt onderverdeeld in vier categorieën: helemaal zacht (ES 1), grotendeels zacht (ES 2), grotendeels hard (ES 3) en helemaal hard (ES 4). De rationale is dat hardere nodi een grotere kans hebben om kwaadaardig te zijn dan zachtere nodi.

In **hoofdstuk 2** hebben wij de resultaten van twintig onderzoeken, waarin in totaal 3973 nodi werden onderzocht, nader geanalyseerd. De kans dat een zachte of grotendeels zachte nodus (ES 1 of 2) goedaardig is, is 97%, terwijl de kans dat een harde nodus (ES 3 of 4) kwaadaardig is, 40% is. Als de afkapwaarde tussen goed- en kwaadaardig wordt verschoven naar tussen ES 1 en ES 2 worden de resultaten nog beter. De kans dat een zacht nodus (ES 1) dan goedaardig is 99%. Hieruit concludeerden wij dat in het geval van een zachte nodus, de FNA veilig achterwege kan worden gelaten, omdat de kans dat een zachte nodus toch kwaadaardig blijkt te zijn, heel erg klein is. Welke afkapwaarde gebruikt wordt, dient per ziekenhuis door het behandelteam bepaald te worden.

Elastografie wordt niet alleen toegepast bij patiënten bij wie een schildkliernodus recent is ontdekt, maar ook bij patiënten die al een FNA van een schildkliernodus hebben gehad. FNA is, zoals eerder besproken, zeer belangrijk om de aard van een schildkliernodus te bepalen. De uitkomst van een FNA wordt onderverdeeld in zes categorieën, de zogenaamde Bethesda classificatie. Categorie III of IV worden tezamen vaak in het Engels '*indeterminate*' genoemd, hetgeen 'onbepaald' betekent. Deze term wordt gebruikt omdat het onduidelijk is of deze nodi goed- of kwaadaardig zijn. Voor nodi waarvan de uitslag van de FNA *'indeterminate'* is, is daarom nader onderzoek nodig, omdat de kans dat deze nodi kwaadaardig zijn, reëel is (~25%). Daarom is onderzocht of elastografie ook voor deze '*indeterminate*' nodi een nauwkeurige diagnostische methode is.

In een recent onderzoek uitgevoerd door een onderzoeksgroep uit Italië werden alle eerder verschenen onderzoeken naar dit onderwerp samengevoegd en geanalyseerd. In **hoofdstuk 3** hebben we de verschillen tussen de resultaten van dit onderzoek en die van ons onderzoek bediscussieerd. De diagnostische waarde van elastografie voor deze specifieke nodi bleek in het onderzoek uit Italië een stuk lager dan dat wij in ons onderzoek hadden aangetoond. Met name het vermogen om goedaardigheid aan te tonen kwam in dit onderzoek niet naar voren. Waar in ons onderzoek de kans dat een zachte nodus inderdaad goedaardig was 97% bedroeg, bleek dat in het onderzoek uit Italië slechts 82% te bedragen`. Dat betekent dat de kans dat een zachte nodus toch kwaadaardig is bijna 1 op 5 zou zijn. Deze kans is te groot om elastografie voor deze patiëntengroep te gebruiken.

Het is niet met zekerheid te zeggen waarom de resultaten van beide onderzoeken zo verschilden. Het ging immers in beide onderzoeken om schildkliernodi, alleen in het onderzoek uit Italië betrof het een subgroep van deze nodi. Mogelijk worden de verschillen veroorzaakt doordat de onderzoeken die in de studie uit Italië werden geïncludeerd een lager aantal nodi hadden onderzocht en dat de opzet van de studies te verschillend waren. Het blijft speculeren en grotere onderzoeken met voldoende patiënten die zich specifiek richten op de groep patiënten met een '*indeterminate*' FNA resultaat, zullen duidelijkheid moeten verschaffen over de rol van deze techniek in de diagnostiek van schildklierknobbels.

Een ander veelvoorkomend probleem bij de diagnostiek van schildklierkanker is dat het vaak voorkomt dat er te weinig cellen worden verkregen met de FNA. Op dit beperkte materiaal kan dan geen diagnose gesteld worden en dit wordt daarom een non-diagnostisch FNA resultaat genoemd. De patiënt moet daarom opnieuw een FNA ondergaan of zelfs worden geopereerd om een definitieve diagnose te stellen. In **hoofdstuk 4** hebben we onderzocht hoe vaak een FNA resultaat non-diagnostisch was in het UMC Utrecht en welke interventies hebben geleid tot een lager percentage non-diagnostische FNA's.

Wij hebben aangetoond dat een directe analyse onder de microscoop van het met FNA verkregen materiaal helpt om dit percentage te verlagen. Tevens lieten onze resultaten zien dat de kans dat er een diagnose gesteld kan worden op het met FNA verkregen materiaal hoger wordt, indien er vaker in de nodus wordt geprikt.

Naast elastografie worden ook andere diagnostische methoden onderzocht om te bepalen of een nodus met een '*indeterminate*' FNA resultaat goed- of kwaadaardig is. Een van deze andere technieken is om te bepalen of de cellen verkregen middels FNA een mutatie in het zogenaamde BRAF-gen hebben. Bij alle patiënten met schildklierkanker is deze mutatie aanwezig in de kankercellen. Als op basis van de FNA met zekerheid gezegd kan worden dat een patiënt schildklierkanker heeft, wordt direct de hele schildklier operatief verwijderd. Indien het FNA resultaat '*indeterminate*' is, dan wordt eerst de halve schildklier verwijderd en als in het verwijderde deel schildklierkanker wordt aangetoond, vervolgens met een nieuwe operatie de andere helft van de schildklier verwijderd. Omdat BRAF-mutatie analyse een relatief dure methode is, wordt deze analyse nog niet standaard toegepast. In **hoofdstuk 5** hebben wij laten zien dat de extra kosten van het uitvoeren van deze test opwegen tegen de kosten die kunnen worden bespaard door in één keer de gehele schildklier te verwijderen in plaats van eerst de ene helft gevolgd door de andere.

Deel 2 – Het diagnosticeren van mogelijke uitzaaiingen van schildklierkanker

Als een patiënt is behandeld voor schildklierkanker is de gehele schildklier operatief verwijderd en zijn er geen schildkliercellen meer in het lichaam aanwezig. De functie van de schildklier, het produceren van schildklierhormoon, wordt opgevangen door medicijnen die dit hormoon bevatten. Doordat er geen schildkliercellen meer in het lichaam aanwezig zijn, is in het bloed Thyreoglobuline (Tg), een eiwit dat door schildkliercellen wordt gemaakt, onmeetbaar laag. Bij alle patiënten die schildklierkanker hebben gehad, wordt daarom de Tg-waarde herhaaldelijk bepaald. Een stijging daarvan kan namelijk wijzen op uitzaaiingen van schildklierkanker. Als deze waarde verhoogd is, wordt middels echografie onderzocht of er uitzaaiingen in de lymfeklieren van de hals te zien zijn. Deze lymfeklieren zijn namelijk de meest voorkomende plek van uitzaaiingen van schildklierkanker. Indien daar uitzaaiingen worden gevonden, wordt de patiënt verwezen naar de chirurg om deze operatief te verwijderen.

In het geval dat de Tg-waarde verhoogd is en er geen uitzaaiingen in de lymfeklieren in de hals worden gevonden, is er een verdenking op uitzaaiingen van schildklierkanker, maar is het niet duidelijk waar deze zich bevinden. Er zijn momenteel geen betrouwbare scantechnieken om de aanwezigheid van uitzaaiingen te bevestigen en de locatie ervan op te sporen. De huidige

behandelrichtlijnen adviseren in zo'n geval om een zogenaamde 'blinde' dosis Jodium-131 (^{131}I) te geven. Omdat jodium belangrijk is voor de productie van schildklierhormoon, wordt het door schildkliercellen opgenomen, en in principe ook door schildklierkankercellen. Andere cellen in het lichaam nemen niet of nauwelijks jodium op. ^{131}I is een radioactieve variant van jodium en kan de cellen die dit opnemen van binnenuit bestralen en daarmee vernietigen. De blinde behandeling heeft als doel om de teruggekeerde schildklierkankercellen dit ^{131}I op te laten nemen en daarmee te vernietigen. De behandeling wordt 'blind' genoemd, omdat de arts vooraf aan de behandeling niet weet waar de uitzaaiingen zitten en of deze uitzaaiingen zich nog als schildkliercellen gedragen en dus jodium opnemen, of dat de uitzaaiingen getransformeerd zijn naar cellen die dit niet meer doen. Na de behandeling met ^{131}I wordt vervolgens een zogenaamde SPECT/CT scan gemaakt om te evalueren of het ^{131}I inderdaad in de uitzaaiingen is opgenomen en of enig effect van de behandeling is te verwachten. Uit eerdere onderzoeken is gebleken dat in tot wel 40% van de gevallen de SPECT/CT geen opname van ^{131}I laat zien, waarmee de behandeling als niet zinvol kan worden beschouwd. Dit betekent dat er behoefte is om vooraf aan de blinde behandeling te weten welke patiënten wel en welke patiënten geen baat zullen hebben bij de behandeling met ^{131}I.

In **hoofdstuk 6** hebben we het studieprotocol van de Thyropet studie gepresenteerd. Deze studie had tot doel om met behulp van relatief nieuwe scanmethoden (^{18}F-FDG PET/CT en ^{124}I PET/CT) beter te voorspellen welke patiënten wel en welke patiënten geen baat zullen hebben bij de behandeling met ^{131}I.

^{18}F-FDG is een radioactief gelabelde vorm van suiker. Deze methode wordt gebruikt bij het opsporen van uitzaaiingen van schilklierkanker, waarvan de cellen dusdanig veranderd zijn dat ze geen jodium meer opnemen, maar wel relatief veel suiker gebruiken. Jodium-124 (^{124}I) is een andere variant van radioactief jodium met andere eigenschappen. Cellen die deze vorm van radioactief jodium opnemen worden niet bestraald, maar de straling van ^{124}I kan met een zogenaamde PET/CT scanner worden gedetecteerd en daarmee afgebeeld.

In figuur 1 in **hoofstuk 6** (pagina 90) is de belangrijkste gedachte achter de Thyropet studie aan de hand van twee patiënten weergegeven. Bij de bovenste patiënt is er wel opname van ^{124}I (fig. 1a) en niet van ^{18}F-FDG (fig. 1b) in de uitzaaiing in de borstkas. Bij deze patiënt is de verwachting dat een behandeling met ^{131}I effect zal hebben, omdat de uitzaaiing jodium opneemt. Bij de onderste patiënt daarentegen is te zien dat er in de kleine uitzaaiingen in de longen (fig. 1c) geen ^{124}I wordt opgenomen, terwijl deze wel ^{18}F-FDG opnemen (fig. 1d). In dit geval zou een behandeling met ^{131}I waarschijnlijk niet effectief zijn en achterwege gelaten worden.

De Thyropet studie was een landelijk onderzoek dat werd uitgevoerd in meerdere ziekenhuizen met verschillende typen PET/CT scanners. In

hoofdstuk 7 hebben wij de procedure beschreven die wij hebben ontworpen en uitgevoerd om de ^{124}I PET/CT scans vergelijkbaar te maken. Een vergelijkbare procedure bestond al voor ^{18}F-FDG PET/CT. Er waren twee belangrijke doelen van dit project: 1) op alle 18 PET/CT scanners moest op dezelfde wijze gescand gaan worden tijdens het onderzoek en 2) de opname van ^{124}I in uitzaaiingen moest kunnen worden gekwantificeerd op een manier dat deze meting vergelijkbaar was tussen de ziekenhuizen. Dit laatste hield in dat onafhankelijk van waar een patiënt gescand was, wij konden bepalen hoeveel^{124}I in een uitzaaiing was opgenomen en dat die bepaling overeen zou komen met de werkelijke hoeveelheid opgenomen ^{124}I. Om dit te bewerkstelligen hadden wij in een fantoom zes flesjes met oplopende hoeveelheden activiteit ^{124}I op al deze PET/CT scanners gescand. De hoeveelheid activiteit in elk flesje was bekend en konden we daarom vergelijken met de gemeten activiteit. Aan de hand van deze metingen konden we met een model per scanner berekenen hoe de metingen nauwkeuriger konden worden gemaakt. In een vergelijking tussen de gemeten en de gekalibreerde activiteiten bleek dat de relatieve meetfout van de gekalibreerde activiteit lager was en daarmee de kalibratie procedure succesvol.

Het streven was om 100 patiënten met een verhoogde Tg-waarde, zonder aanwijzingen voor uitzaaiingen op echografie van de hals en gepland voor een 'blinde' dosis met ^{131}I, te includeren in meerdere ziekenhuizen in Nederland. Alle patiënten zouden eerst de ^{18}F-FDG en ^{124}I PET/CT scan krijgen, gevolgd door de behandeling met ^{131}I. In de studie hadden we een zogenaamde stopregel opgenomen: indien tijdens de studie zou blijken dat de ^{124}I PET/CT scan van drie patiënten negatief was, terwijl de SPECT/CT na de ^{131}I behandeling van deze patiënten positief was, zou de studie voortijdig gestopt worden. In dat geval zouden patiënten die baat zouden hebben bij de ^{131}I behandeling, deze op basis van de ^{124}I PET/CT onterecht onthouden worden.

De belangrijkste resultaten van de Thyropet studie zijn gepresenteerd in **hoofdstuk 8**. Nadat er negentien patiënten in de studie waren geïncludeerd, werd de studie voortijdig beëindigd op basis van de vooraf gedefinieerde stopregel. Twee patiënten konden om verschillende redenen niet worden geanalyseerd, waardoor er 17 patiënten over bleven voor de analyses. Acht van de zeventien (47%) SPECT/CT scans na de ^{131}I lieten geen opname zien, hetgeen benadrukt dat de noodzaak om de selectie van patiënten voor de behandeling met ^{131}I moet worden verbeterd. Al deze acht negatieve scans werden correct voorspeld door een negatieve ^{124}I PET/CT scan. Er waren negen ^{131}I SPECT/CT scans positief voor uitzaaiingen, waarvan er slechts vier correct werden voorspeld door de ^{124}I PET/CT. Dit betekent dat er vijf scans fout negatief waren. Als ^{124}I PET/CT op deze manier zou worden gebruikt in de zorg zouden er 47% minder patiënten een ^{131}I behandeling krijgen, waarvan ze geen baat zouden hebben. Daartegenover staat echter dat 29% van de patiënten onterecht geen behandeling met ^{131}I zou krijgen. Deze kans is onacceptabel hoog en daarom kan ^{124}I PET/CT voor dit doel nog niet worden ingezet.

De resultaten van de Thyropet studie deden de vraag rijzen wat de oorzaak was dat de ^{124}I PET/CT negatief kon zijn, terwijl de ^{131}I SPECT/CT positief was. Er zijn vele verschillende oorzaken mogelijk, waarvan er een door ons werd onderzocht. De resultaten van dit onderzoek staan beschreven in **hoofdstuk 9**. Een belangrijk verschil tussen de ^{124}I PET/CT scan en de ^{131}I SPECT/CT scan is dat de toegediende activiteit van ^{131}I ongeveer 100 maal hoger is dan die van ^{124}I (7400 vs. 74 Megabecquerel (MBq)). Daar staat wel tegenover dat een PET/CT scanner over het algemeen als ongeveer 100 keer zo gevoelig wordt beschouwd als een SPECT/CT scanner.

In **hoofdstuk 9** hebben we de resultaten beschreven van een onderzoek dat als doel had om uit te zoeken hoe deze verhoudingen in de praktijk de detecteerbaarheid van een uitzaaiing beïnvloeden. Dit hebben we onderzocht door te bepalen hoeveel ^{124}I er nodig is om dezelfde detecteerbaarheid van ^{131}I te bewerkstelligen. In een fantoom waarin bollen van verschillende diameter zaten, werden eerst opeenvolgende metingen met ^{131}I op de SPECT/CT gemaakt, gevolgd door metingen met hetzelfde fantoom gevuld met ^{124}I. De resultaten toonden dat 74 MBq ^{124}I voldoende was om bij uitzaaiingen kleiner dan 10 mm een vergelijkbare detecteerbaarheid te genereren als optimale PET instellingen worden gebruikt. Echter, voor grotere laesies en bij het gebruik van suboptimale PET scanner instellingen, is de activiteit ^{124}I onvoldoende om een vergelijkbare detecteerbaarheid te genereren. Mogelijk dat hogere doses toegediend ^{124}I fout-negatieve ^{124}I PET/CT scans in die gevallen zou kunnen hebben voorkomen.

REVIEW COMMITTEE
LIST OF PUBLICATIONS
CURRICULUM VITAE
DANKWOORD

CHAPTER 13

Review committee

Prof. dr. G.D. Valk
Department of Internal Medicine, University Medical Center Utrecht, Utrecht, the Netherlands

Prof. dr. P. van Diest
Department of Pathology, University Medical Center Utrecht, Utrecht, the Netherlands

Prof. dr. L-F. de Geus-Oei
Department of Radiology and Nuclear Medicine, Leiden University Medical Center, Leiden, the Netherlands

Prof. dr. J.W.A. Smit
Department of Internal Medicine, Radboud University Medical Center, Nijmegen, the Netherlands

Prof. dr. M.R. Vriens
Department of Surgery, University Medical Center Utrecht, Utrecht, the Netherlands

List of publications

Kist JW, de Keizer B, Vogel WV. Letter to the Editor regarding the article "I-124 PET/CT in patients with differentiated thyroid cancer: clinical and quantitative image analysis." *Thyroid 2016; 26(8):1141–2*.

Kist JW, van der Vlies M, Hoekstra OS, Greuter HN, de Keizer B, Stokkel MP, Vogel WV, Huisman MC, van Lingen A. Calibration of PET/CT scanners for multicenter studies on differentiated thyroid cancer with (124)I. *EJNMMI Res. 2016;6(1):39*.

Lodewijk L, Vriens MR, Vorselaars WMCM, van der Meij MANP, Kist JW, Barentsz MW, Verkooijen HM, Borel Rinkes IH, Valk GD. Same-day fine neelde aspiration cytology diagnosis for thyroid nodules achieves rapid anxiety decrease and high diagnostic accuracy. *Endocr Pract. 2016;22(5):561-566*.

Lodewijk L, Kluijfhout WP, Kist JW, Stegeman I, Plukker JT, Nieveen van Dijkum EJ, Bonjer HJ, Bouvy ND, Schepers A, de Wilt JHW, Netea-Maier RT, van der Hage JA, Burger JWA, Ho G, Lee WS, Shen WT, Aronova A, Zarnegar R, Benay C, Mitmaker EJ, Sywak MS, Aniss AM, Kruijff S, James B, Grogan R, Brunaud L, Hoch G, Pandolfi C, Ruan DT, Jones MD, Guerrero MA, Valk GD, Borel Rinkes IH, Vriens MR. Characteristics of contralateral carcinomas in patients with differentiated thyroid cancer larger than 1 cm. *Langenbecks Arch Surg. 2016;401(3):365-373*.

Kist JW, de Keizer B, van der Vlies M, Brouwers AH, Huysmans D, van der Zant FM, Hermsen R, Stokkel MP, Hoekstra OS, Vogel WV, for the THYROPET study group. 124I PET/CT to predict the outcome of blind 131I treatment in patients with biochemical recurrence of differentiated thyroid cancer; results of a multicenter diagnostic cohort study (THYROPET). *J Nucl* Med. 2016;57(5):701–7.

de Koster EJ, Kist JW, Vriens MR, Borel Rinkes IHM, Valk GD, de Keizer B. Thyroid Ultrasound-Guided Fine-Needle Aspiration: The Positive Influence of On-Site Adequacy Assessment and Number of Needle Passes on Diagnostic Cytology Rate. *Acta Cytologica 2016;60(1):39–45*.

Beijst C, Kist JW, Elschot M, Viergever MA, Hoekstra OS, de Keizer B, de Jong HW. Quantitative Comparison of 124I PET/CT and 131I SPECT/CT Detectability. *J Nucl Med. 2016;57(1):103-108*.

Kist JW, Nell S, de Keizer B, Valk GD, Borel Rinkes IHM, Vriens MR. The role of qualitative elastography in thyroid nodule evaluation: exploring its target populations. *Endocr 2015;50(2):265–7*.

Kist JW, Nell S, Debray TP, de Keizer B, van Oostenbrugge TJ, Borel Rinkes IH, Valk GD, Vriens MR. Qualitative elastography can replace thyroid nodule fine-needle aspiration in patients with soft thyroid nodules. A systematic review and meta-analysis. *Eur J Radiol. 2015;84(4):652-661*

Lodewijk L, Bongers PJ, Kist JW, Conemans EB, de Laat JM, Pieterman CR, van der Horst-Schrivers AN, Jorna C, Hermus AR, Dekkers OM, de Herder WW, Drent ML, Bisschop PH, Havekes B, Rinkes IH, Vriens MR, Valk GD. Thyroid incidentalomas in patients with multiple endocrine neoplasia type 1. *Eur J Endocrinol. 2015;172(4):337-342*

Kist JW, de Keizer B, Stokkel MP, Hoekstra OS, Vogel WV. Recurrent differentiated thyroid cancer: towards personalized treatment based on evaluation of tumor characteristics with PET (THYROPET Study): study protocol of a multicenter observational cohort study. *BMC Cancer. 2014;14:405*

Lodewijk L, Kist JW, Valk GD, Vriens MR, Borel Rinkes IHM. Preoperative BRAF(V600E) mutation screening is unlikely to alter initial surgical treatment of patients with indeterminate thyroid nodules: a prospective case series of 960 patients. *Cancer. 2014;120:1083–1084*

Vriens MR, Kist JW, Lodewijk L, van Hillegersberg R, Borel Rinkes IHM. Robot-assisted transaxillary thyroid surgery. *Ned Tijdschr Geneeskd. 2013;157:A5841*

Lodewijk L, Prins AM, Kist JW, Valk GD, Kranenburg O, Rinkes IH, Vriens MR. The value of miRNA in diagnosing thyroid cancer: a systematic review. *Cancer Biomark. 2012;11(6):229-238*

Mulder MC, Kist JW, Consten ECJ, Verheijen PM. Gastrointestinal metastasis as the first presentation of lung carcinoma. *Int J Colorectal Dis. 2012;27:839–840*

Kist JW, THYROPET studie groep. Recidief gedifferentieerd schildkliercarcinoom: richting gepersonaliseerde behandeling gebaseerd op functionele tumorkarakteristieken met PET (THYROPET Studie). *Ned Tijdschr Nucl Geneeskd. 2012;1:855-858*

Hummel R, Watson DI, Smith C, Kist JW, Michael MZ, Haier J, Hussey DJ. MiR-148a improves sensitivity to cisplatin and 5-FU in sensitive and resistant esophageal adeno- and squamous cell carcinoma cell lines. *J of Gastrointest Surg. 2011;15(3):429-438*

Curriculum Vitae

Jakob Willem Kist was born in Amsterdam on the 9th of May 1984 in Amsterdam. After he graduated in 2002 from the Stedelijk Gymnasium in Haarlem, he entered medical school at the University of Groningen. During his studies Jakob did an extracurricular elective at the Mount Sinaï Hospital in New York in the United States at the department of hepatobiliary surgery. Later on in his studies Jakob participated in a research project at Flinders Medical Centre in Adelaide in Australia focusing on Barrett's esophagus. Following his clinical rotations at the Deventer Ziekenhuis, he completed an elective clinical rotation at the department of surgical oncology at the University Medical Center Utrecht (UMCU) under the supervision of prof. dr. I.H.M. Borel Rinkes.

After obtaining his medical degree in 2010, Jakob started working as a surgical resident (ANIOS) at the department of surgery at the Meander Medical Center in Amersfoort (dr. A.J. van Overbeeke) and subsequently at the department of surgery at the UMCU (prof. dr. I.H.M. Borel Rinkes). In 2011 he got the opportunity to start as a PhD-student, which eventually resulted in this thesis. He worked as a PhD-student at the department of nuclear medicine of the Netherlands Cancer Institute (dr. W.V. Vogel) and at the department of surgery of the UMCU (prof. dr. I.H.M. Borel Rinkes). His research was co-supervised by prof. dr. O.S. Hoekstra of the department of nuclear medicine at the VU University Medical Center and dr. B. de Keizer of the department of radiology and nuclear medicine at the UMCU.

Jakob is currently working as a resident in radiology at the Meander Medical Center in Amersfoort (dr. H.J. Baarslag). He lives with Jarmila van der Bilt and their two sons, Olivier and Mees, in Naarden.