Screening for Thyroid Cancer: A Systematic Evidence Review for the U.S. Preventive Services Task Force

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. HHSA-290-2012-00015-I-EPC4, Task Order No. 4

Prepared by:
Kaiser Permanente Research Affiliates Evidence-based Practice Center
Kaiser Permanente Center for Health Research
Portland, OR

Investigators:
Jennifer S. Lin, MD, MCR
Erin J. Aiello Bowles, MPH
Selvi B. Williams, MD, MPH
Caitlin C. Morrison, MPH

AHRQ Publication No. 15-05221-EF-1
May 2017
This report is based on research conducted by the Kaiser Permanente Research Affiliates Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. HHSA-290-2012-00015-I-EPC4, Task Order No. 4). The findings and conclusions in this document are those of the authors, who are responsible for its contents, and do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information (i.e., in the context of available resources and circumstances presented by individual patients).

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: Jennifer Croswell, MD, MPH, formerly of AHRQ; current and former members of the U.S. Preventive Services Task Force who contributed to topic deliberations; Gunjan Tykodi, MD, for expert consultation; Todd Hannon, MLS, and Smyth Lai, MLS, for creating and conducting the literature searches; Elizabeth L. Hess, MS, ELS(D), for editorial assistance; Ning X. Smith, PhD, and Brittany Burda, MPH, for assistance with meta-analyses; and Louise Davies, MD, MS, Hyeong Sik Ahn, MD, PhD, Edward G. Grant, MD, and Mike Tuttle, MD, for their expert feedback on this report; and Barry Kramer, MD, MPH, Dawn Holman, MPH, and Monica C. Skarulis, MD, who provided expert review of this report.

Suggested Citation

Structured Abstract

Objective: We conducted this systematic review to support the U.S. Preventive Services Task Force in updating its recommendation on screening for thyroid cancer. Our review addresses the following Key Questions (KQs): 1) Compared with not screening, does screening adults for thyroid cancer lead to a reduced risk of thyroid-specific mortality or morbidity, reduced all-cause mortality, and/or improved quality of life? 2) What are the test performance characteristics of screening tests for detecting malignant thyroid nodules in adults? 3) What are the harms of screening for thyroid cancer in adults? 4) Does treatment of screen-detected thyroid cancer reduce thyroid-specific mortality or morbidity, reduce all-cause mortality, and/or improve quality of life? 5) What are the harms of treating screen-detected thyroid cancer?

Data sources: We searched MEDLINE, PubMed, and the Cochrane Central Register of Controlled Trials to locate relevant studies for all KQs. We searched for articles published from January 1966 to January 2016.

Study selection: We reviewed 10,424 abstracts and 707 articles against specified inclusion criteria. Eligible studies included those written in English and conducted in asymptomatic adult populations at general risk or with a prior personal history of radiation exposure.

Data analysis: We conducted dual independent critical appraisal of all included studies and extracted study details and outcomes from fair- or good-quality studies. We synthesized results by KQ and type of screening test (i.e., palpation or ultrasound). We used primarily qualitative synthesis. We used random-effects meta-analyses to pool surgical harms. We also summarized the overall strength of evidence for each KQ.

Results: We found no studies that met our inclusion criteria for KQ 1. Ten fair-quality studies were included for KQ 2. In two studies, neck palpation was not sensitive to detect thyroid nodules. Two methodologically limited studies that used selected sonographic features demonstrated that screening with ultrasound can be specific for detecting thyroid malignancy; one of these studies suggested that using a combination of high-risk sonographic features, such as microcalcification or irregular shape, can optimize both sensitivity and specificity. Three fair-quality studies met our inclusion criteria for KQ 3, none of which suggested any serious harms from screening or ultrasound-guided fine-needle aspiration. However, we found no screening studies that directly examined the risk of overdiagnosis. Two studies met our inclusion criteria for KQ 4, but neither was designed to determine if earlier or immediate treatment versus delayed or no surgical treatment improves the outcomes of patients with well-differentiated thyroid cancer. Fifty-two studies were included for KQ 5. Based on 36 studies, permanent surgical harms, hypoparathyroidism, and recurrent laryngeal nerve palsy are not uncommon. Best estimates of permanent hypoparathyroidism are from 2 to 6 events per 100 thyroidectomies and are more variable with lymph node dissection. The rate of recurrent laryngeal nerve palsy is estimated at 1 or 2 events per 100 surgeries. Based on 16 studies, treatment of differentiated thyroid cancer with radioactive iodine (RAI) treatment is associated with a small increase in second primary malignancies; RAI treatment is also associated with increased permanent adverse effects on the salivary gland, such as dry mouth.
**Limitations:** The vast majority of studies that evaluated the diagnostic accuracy of ultrasound to detect thyroid tumors are not in screening populations. High statistical heterogeneity for surgical harms of hypoparathyroidism could not be explained by known clinical heterogeneity across studies. Differences in study designs and variable reporting on radiation doses limits our understanding of the magnitude and precision around the excess risk for second primary malignancies due to RAI.

**Conclusions:** Although ultrasound of the neck using high-risk sonographic characteristics plus followup cytology from fine-needle aspiration can reasonably identify thyroid cancer, it is unclear if population-based or targeted screening can decrease mortality or improve important patient health outcomes. More importantly, screening results in the identification indolent thyroid cancer, and treatment of these cases of overdiagnosed cancer can pose real patient harms.
Table of Contents

Chapter 1. Introduction ............................................................................................................... 1
  Condition Definition ................................................................................................................ 1
  Prevalence and Burden of Disease ...................................................................................... 1
  Risk Factors .......................................................................................................................... 2
  Natural History ..................................................................................................................... 2
  Screening ............................................................................................................................... 3
  Treatment ............................................................................................................................... 4
  Current Clinical Practice ...................................................................................................... 5
  Previous U.S. Preventive Services Task Force Recommendation ...................................... 6

Chapter 2. Methods ...................................................................................................................... 7
  Scope and Purpose ................................................................................................................ 7
  Analytic Framework and Key Questions ............................................................................. 7
  Data Sources and Searches ................................................................................................. 7
  Study Selection .................................................................................................................... 7
  Quality Assessment and Data Abstraction ......................................................................... 9
  Data Synthesis ..................................................................................................................... 9
  Expert Review and Public Comment ................................................................................. 10
  USPSTF Involvement ......................................................................................................... 10

Chapter 3. Results ....................................................................................................................... 12
  KQ 1. Compared With Not Screening, Does Screening Adults for Thyroid Cancer Lead to a Reduced Risk of Thyroid-Specific Mortality or Morbidity, Reduced All-Cause Mortality, and/or Improved Quality of Life? ................................................................................................................................. 12
  KQ 2. What Are the Test Performance Characteristics of Screening Tests for Detecting Malignant Thyroid Nodules in Adults? ................................................................................................................................. 12
  Diagnostic Accuracy of Palpation to Detect Thyroid Nodules ...................................... 12
  Diagnostic Accuracy of Screening Ultrasound to Detect Thyroid Cancer .................. 13
  Yield of Screening for Thyroid Cancer ........................................................................... 14
  KQ 3. What Are the Harms of Screening for Thyroid Cancer in Adults? ......................... 15
  Limitations .......................................................................................................................... 15
  KQ 4. Does Treatment of Screen-Detected Thyroid Cancer Reduce Thyroid-Specific Mortality or Morbidity, Reduce All-Cause Mortality, and/or Improve Quality of Life? .............................................................. 16
  Limitations .......................................................................................................................... 16
  KQ 5. What Are the Harms of Treating Screen-Detected Thyroid Cancer? .................... 17
  Surgical Harms ................................................................................................................... 18
  RAI Harms .......................................................................................................................... 20

Chapter 4. Discussion ................................................................................................................. 23
  Summary of Evidence ........................................................................................................ 23
  Key Contextual Issues ....................................................................................................... 25
  Diagnostic Accuracy of Ultrasound and FNA ................................................................. 25
  Overdiagnosis .................................................................................................................... 27
  Limitations of the Review ................................................................................................. 29
  Evidence Gaps and Future Research Needs ..................................................................... 32
  Conclusions ......................................................................................................................... 32
Chapter 1. Introduction

Condition Definition

Thyroid cancer starts in the thyroid gland, which is located in the front of the neck. The histologic types of thyroid cancer can be categorized into three groups based on their cellular origin and characteristics: cancer derived from the thyroid epithelium, which includes papillary and follicular carcinomas, most commonly differentiated; cancer from parafollicular (C) cells, called medullary carcinomas; and anaplastic carcinomas, the most undifferentiated type.\(^1\) Papillary microcarcinoma refer to papillary cancer smaller than 1 cm.

Prevalence and Burden of Disease

Palpable thyroid nodules are common, occurring in approximately 5 percent of U.S. adults age 50 years and older when screened by palpation. The prevalence of thyroid nodules increases to 16 to 68 percent of adults when screened by ultrasound.\(^3\)\(^-\)\(^5\) In contrast, thyroid cancer is uncommon, and the lifetime risk of developing it is 1.1 percent.\(^6\)

Since 1975, the incidence of detected thyroid cancer cases has been rising in the United States for both men and women.\(^7\) Estimates published in 2014 showed the incidence of thyroid cancer increased from 4.9 cases per 100,000 persons in 1975 to 14.3 cases per 100,000 persons in 2009, representing an absolute increase of 9.4 cases per 100,000 persons (95% confidence interval [CI], 8.9 to 9.9).\(^7\) However, over time mortality rates have remained stable at about 0.5 deaths per 100,000 persons per year.\(^6\) According to the Surveillance, Epidemiology, and End Results Program (SEER), there are 13.5 new cases of thyroid cancer per 100,000 persons per year (age-adjusted and based on 2008–2012 cases).\(^6\) Thyroid cancer is 2 to 3 times more common in women than in men,\(^1\) and incidence is greater among whites and Asians than among blacks or American Indians/Alaska Natives.\(^6\) Overall, a total of 601,789 adults were living with thyroid cancer in the United States in 2012.\(^1\) The American Cancer Society estimated 62,450 new cases of thyroid cancer would be identified in 2015.\(^8\) One study estimated that if current trends continue, as many as 89,500 cases may be diagnosed in 2019, costing the United States $18.6 to $21.6 billion to treat cancer diagnosed from 2010 to 2019.\(^9\)

Differentiated thyroid cancer accounts for 90 percent of all cases.\(^10\) Within this category, papillary thyroid cancer accounts for about 70 to 80 percent of cases and follicular cancer accounts for 10 to 15 percent. Papillary microcarcinoma accounts for about 24 percent of thyroid cancer cases in the United States; it is typically found incidentally on imaging studies of the head or neck.\(^11\) Most cases of differentiated cancer occur in adults ages 30 to 50 years.\(^12\) Approximately 5 percent of cases of differentiated cancer occur in persons with a family history of thyroid cancer.\(^13\) Papillary cancer is 3 times more common in women than in men. Hürthle cell is a rare subtype of follicular thyroid cancer, accounting for less than 3 percent of cases, and is no longer classified separately from follicular cancer by the World Health Organization.\(^14\) Medullary and anaplastic thyroid cancer are much rarer forms, accounting for about 4 and 2 percent of cases of thyroid cancer, respectively. Medullary cancer most often occurs between the
ages of 50 and 60 years. Approximately 25 percent of cases of medullary cancer are inherited and 75 percent are sporadic.\textsuperscript{15, 16} Most anaplastic thyroid cancer cases are diagnosed after age 65 years, and the majority of cases occur in women.\textsuperscript{17}

### Risk Factors

Risk factors for thyroid cancer include exposure to radiation, family history of thyroid cancer, and inherited genetic syndromes. Other potential risk factors for the development of thyroid cancer include high iodine intake (either from the environment or diet), increased thyroid-stimulating hormone levels, obesity, and exposure to nitrate.\textsuperscript{18}

Radiation exposure during childhood up to late adolescence is a well-known risk factor for thyroid cancer. Radiation exposure may come from environmental nuclear disasters (e.g., the Chernobyl accident); medical examinations (e.g., computed tomography [CT] scans, medical or dental x-rays); and medical treatment, including radiation therapy for head and neck cancer and radioactive iodine (RAI) (also called I-131 or 131I) treatment (e.g., for conditions such as hyperthyroidism).\textsuperscript{18, 19, 20} A pooled analysis showed an excess relative risk of thyroid cancer of 7.7 (95% CI, 2.1 to 28.7) per Gy of radiation exposure.\textsuperscript{21}

For a better understanding of radiation exposure dose in the United States, the average person is exposed to approximately 3 mSv per year from naturally occurring materials in the environment (note the radiation dose received in Sv is equivalent to the radiation exposure in Gy).\textsuperscript{22} More than 80 million CT scans are performed per year in the United States alone, with an average dose of 2 to 20 mSv per examination.\textsuperscript{22} For solid tumors, including thyroid cancer, observational and molecular studies have noted a linear dose-response relationship for radiation exposure.\textsuperscript{21} There does not appear to be any critical threshold of radiation exposure as a risk factor for thyroid cancer.

Family history of thyroid cancer in a first-degree relative may increase one’s risk of nonmedullary thyroid cancer by as much as 5-fold;\textsuperscript{13, 23} inherited genetic syndromes may increase one’s risk of thyroid cancer 10-fold.\textsuperscript{24} The prevalence of persons with a family history of thyroid cancer is unknown. Most inherited genetic syndromes are rare. Familial adenomatous polyposis occurs in 1 per 10,000 to 15,000 persons, multiple endocrine neoplasia type 2 occurs in 1 per 30,000 to 50,000 persons, and Carney complex and Cowden syndrome are even rarer. Persons with familial adenomatous polyposis may be recommended for regular thyroid surveillance via palpation and/or ultrasound.\textsuperscript{25} Persons with multiple endocrine neoplasia type 2 are recommended for prophylactic thyroidectomy because they have a 70 to 100 percent chance of developing medullary thyroid cancer during their lifetime.\textsuperscript{26}

### Natural History

The majority of cases of thyroid cancer have a very favorable natural history. However, if aggressive thyroid nodules are left untreated, they will grow over time, metastasize (most often to the lymph nodes and lungs), and can ultimately result in death.\textsuperscript{27-29} The natural history varies
by the three types of thyroid cancer: differentiated, medullary, and anaplastic.

The 10-year overall survival rates for papillary and follicular thyroid cancer are 93 and 85 percent, respectively, for all stages of the disease.\textsuperscript{12} Approximately 12 percent of thyroid glands are found to be cancerous on autopsy, suggesting that a subset of differentiated cancer cases may be very slow growing and pose little to no risk to the patient.\textsuperscript{30}

Historically, one of the most important prognostic factors for differentiated cancer is age older than 45 years.\textsuperscript{10} Other important prognostic factors are larger tumor size and involvement of the lymph node.\textsuperscript{28, 31, 32} Patients younger than age 45 years have a very low likelihood of dying of papillary or follicular cancer, which is classified as stage I or II depending only on whether there are distant metastases.\textsuperscript{10} For patients older than age 45 years, staging for differentiated cancer also accounts for tumor size and lymph node metastases. Spread to lymph nodes in the neck is common for patients with papillary cancer, affecting approximately 50 to more than 75 percent of cases, but less common for patients with follicular cancer, affecting less than 10 percent of cases, as follicular cancer more commonly spreads to the lungs or bones.\textsuperscript{33}

The 10-year survival rate for medullary thyroid cancer is 75 percent.\textsuperscript{12} The primary feature of medullary cancer is the production of calcitonin, and levels of calcitonin production may be correlated with prognosis.\textsuperscript{34, 35} The most important prognostic factors include age (patients age >40 or 45 years have poorer prognosis), disease stage (later-stage disease portends a worse outcome), and extent of surgery (patients with a lobectomy have a worse prognosis than those with a total thyroidectomy; the choice is often driven by whether metastatic disease exists).\textsuperscript{16, 17, 36}

Anaplastic thyroid cancer is extremely rare but lethal. In 90 percent of incident cases, metastases to the lymph nodes or distant organs are present at the time of diagnosis.\textsuperscript{37-39} Anaplastic thyroid cancer is highly aggressive and has a 1-year survival rate of only 20 percent.\textsuperscript{12}

**Screening**

There are two primary methods to screen for thyroid cancer: 1) neck palpation during a physical examination, which can identify palpable nodules, and 2) ultrasound, which can identify both palpable and nonpalpable nodules, especially those smaller than 1 cm. Ultrasound can also identify characteristics of a thyroid nodule that help predict whether a nodule is benign or malignant.\textsuperscript{40} Screening with both palpation and ultrasound can also identify abnormal cervical lymph nodes that may represent metastatic thyroid cancer. Screening for thyroid cancer could result in early detection of malignant thyroid nodules that are easily treatable, before the cancer spreads beyond the thyroid gland. Early detection could make treatment more effective, with potentially less harm than if administered later. Potential harms of screening include false-positive results, which may lead to unnecessary diagnostic tests. Screening may also result in overdiagnosis because it can detect very small and/or indolent tumors that might never affect a person’s morbidity or mortality.\textsuperscript{41, 42} Overdiagnosis might also lead to overtreatment.\textsuperscript{43}

Diagnostic workup of a thyroid nodule typically includes measurement of serum thyroid-
stimulating hormone and diagnostic ultrasound of the thyroid and neck (i.e., cervical lymph nodes). Depending on the results of initial testing, additional laboratory testing and imaging may be conducted. Fine-needle aspiration (FNA), with or without ultrasound guidance, is the procedure of choice when evaluating thyroid nodules to obtain cytology. The American Thyroid Association (ATA) recommends FNA for nodules larger than 1 cm with intermediate or highly suspicious ultrasound features and for nodules larger than 1.5 cm with mildly suspicious ultrasound features. The ATA defines high, intermediate, and mild suspicion as follows:\(^5\)

- **High suspicion (malignancy risk of 70% to 90%):** solid hypoechoic nodule(s) or solid hypoechoic component in partially cystic nodule(s), with at least one of the following features:
  - Irregular margins
  - Microcalcifications
  - Taller than wide shape
  - Disrupted rim calcifications, with a small extrusive hypoechoic soft-tissue component
  - Evidence of extrathyroidal extension

- **Intermediate suspicion (malignancy risk of 10% to 20%):** hypoechoic solid nodule(s) with smooth margins but without:
  - Microcalcifications
  - Extrathyroidal extension
  - Taller than wide shape

- **Low suspicion (malignancy risk of 5% to 10%):** isoechoic or hyperechoic solid nodule(s), or partially cystic nodule(s) with eccentric uniformly solid areas but without:
  - Microcalcifications
  - Irregular margin
  - Extrathyroidal extension
  - Taller than wide shape

FNA cytology, using the Bethesda System for Reporting Thyroid Cytopathology, can be classified as: 1) nondiagnostic/unsatisfactory, 2) benign, 3) atypia of undetermined significance or follicular lesion of undetermined significance (AUS/FLUS), 4) follicular neoplasm or suspicious for follicular neoplasm, 5) suspicious for malignancy, or 6) malignant. ATA guidelines recommend that persons with nondiagnostic cytology have repeat FNA with ultrasound guidance. Depending on the ultrasound characteristics, persons with indeterminate results (i.e., AUS/FLUS) may undergo additional testing (e.g., molecular testing, repeat FNA) before diagnostic surgery is pursued. Persons with nodules with malignant or suspicious cytology features generally proceed to surgery. Given the low-false negative rate of ultrasound-guided FNA, persons with nodules with a highly suspicious ultrasound pattern and benign FNA cytology are recommended to have repeat ultrasound with FNA in 12 months, whereas those with an intermediately suspicious ultrasound pattern are recommended to have repeat ultrasound in 12 to 24 months.\(^5\)

### Treatment

Surgery is the main form of treatment for thyroid cancer. The type of surgery depends largely on
what proportion of the thyroid gland is involved. Surgical options include total thyroidectomy or partial thyroidectomy (the latter is also known as a lobectomy). A partial thyroidectomy may be sufficient as the initial treatment for low-risk differentiated thyroid cancer, and is recommended if surgery is chosen for low-risk papillary microcarcinoma restricted to one lobe of the thyroid. Complete thyroidectomy (removal of the entire thyroid gland following initial lobectomy) may be performed for patients whose final pathology result comes back as malignant and for whom this diagnosis was not known or suspected preoperatively. Lymphadenectomy (lymph node dissection) may be done to remove lymph nodes that may have signs of thyroid cancer involvement (local lymph node metastasis). Preoperative neck ultrasound for cervical lymph nodes and FNA of sonographically suspicious lymph nodes are recommended to determine the need for therapeutic lymph node dissection. The ATA recommends consideration of prophylactic lymph node dissection for clinically advanced papillary thyroid cancer without known nodal involvement, but this recommendation is based on low-quality evidence. The most common permanent serious surgical harms from thyroidectomy, with or without lymph node dissection, include hypoparathyroidism (hypocalcemia) and recurrent laryngeal nerve palsy (vocal cord paralysis).

Additional treatment recommendations generally differ by postoperative disease status, tumor stage, and type of thyroid cancer. RAI is taken orally and absorbed by the thyroid. Postoperative RAI may be used for remnant ablation to destroy any normal remnant thyroid tissue or microscopic foci of cancer that may remain after surgery, which facilitates detection of recurrent disease during followup. RAI may also be used at higher therapeutic doses as adjuvant therapy to treat residual local cancer or metastatic spread outside the neck to improve the disease-free survival or disease recurrence rate in higher-risk patients. The administered activity (dose) for ablation is lower than that for therapy. RAI is not routinely recommended for patients with low-risk differentiated thyroid cancer due to its adverse effects on, for example, salivary and gonadal function, as well as possible increased risk of second primary malignancies. However, RAI is routinely recommended for high-risk patients (the ATA defines “high risk” as macroscopic invasion of tumor into the perithyroidal soft tissues; incomplete tumor resection; distant metastases; postoperative serum thyroglobulin level suggestive of distant metastases; pathologic N1 disease with any metastatic lymph node ≥3 cm; or follicular thyroid cancer with extensive vascular invasion).

External-beam radiation therapy and chemotherapy are not typically used in persons with differentiated thyroid cancer unless they have late-stage disease (stage III or IV) that did not respond to RAI therapy. Radiation therapy and chemotherapy are more commonly administered to patients with anaplastic thyroid cancer.

**Current Clinical Practice**

We are unaware of any professional medical society that recommends population-based screening for thyroid cancer. Although 2015 ATA management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer do not address population-based screening, they do state that there is insufficient evidence to support screening in persons with a family history of follicular cell–derived, differentiated thyroid cancer. To date, South Korea appears to
be the only country in the world that regularly practices asymptomatic thyroid cancer screening using ultrasound; this practice happened opportunistically as an “add-on” option for persons undergoing sanctioned screening through an organized cancer screening program initiated in 1999.47

**Previous U.S. Preventive Services Task Force Recommendation**

In 1996, the U.S. Preventive Services Task Force (USPSTF) stated that screening asymptomatic adults or children for thyroid cancer by neck palpation or ultrasound is not recommended (“D” recommendation). It found a lack of evidence that early detection of thyroid cancer improves outcomes and noted that neck palpation had poor sensitivity for detecting lesions, which contributed to a large number of false-positive results and resulted in unnecessary invasive testing. However, for asymptomatic persons with a history of external upper body irradiation in infancy or childhood, the USPSTF had a separate recommendation (“C” recommendation): although it found that there is insufficient evidence to recommend for or against screening in these high-risk persons, recommendations for periodic palpation can be made on other grounds, including patient preference or anxiety regarding increased risk. Insufficient evidence would likely have resulted in an “I” statement using modern grading methods.
Chapter 2. Methods

Scope and Purpose

The USPSTF will use this evidence review to update its 1996 recommendation regarding the effectiveness of thyroid cancer screening in average- and high-risk adults. This review addresses the benefit and harms associated with thyroid cancer screening and treatment of early-stage thyroid cancer.

Analytic Framework and Key Questions

We developed an analytic framework with five Key Questions (KQs) based on the previous review and a scan of research conducted since the previous review (Figure 1).

1. Compared with not screening, does screening adults for thyroid cancer lead to a reduced risk of thyroid-specific mortality or morbidity, reduced all-cause mortality, and/or improved quality of life?
2. What are the test performance characteristics of screening tests for detecting malignant thyroid nodules in adults?
3. What are the harms of screening for thyroid cancer in adults?
4. Does treatment of screen-detected thyroid cancer reduce thyroid-specific mortality or morbidity, reduce all-cause mortality, and/or improve quality of life?
5. What are the harms of treating screen-detected thyroid cancer?

Data Sources and Searches

We worked with a research librarian to develop our search strategy, which was peer reviewed by a second research librarian. We searched Ovid® MEDLINE, the Cochrane Central Register of Controlled Trials, and the publisher supplied segment of PubMed to locate relevant studies for all KQs. We searched for articles published from January 1966 to January 2016. The search strategy is included in Appendix A Table 1. We supplemented our database searches by reviewing reference lists from recent and relevant systematic reviews. We also searched ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform for relevant ongoing trials (Appendix B).

Study Selection

Two reviewers independently reviewed 10,424 titles and abstracts, using an online platform (Abstrackr91), and 707 articles (Appendix A Figure 1) against specified inclusion criteria (Appendix A Table 1). We resolved discrepancies through consensus and consultation with a third investigator. We excluded articles that did not meet inclusion criteria or those we rated as poor quality. Appendix C lists all excluded trials.
For the screening questions (KQs 1–3), we included any studies of asymptomatic adult populations, either at general risk (e.g., unselected) or with prior personal history of radiation exposure. We excluded populations that were selected based on very high radiation exposure due to environmental disasters, inherited genetic syndromes associated with a high risk for developing thyroid cancer, or a personal history of thyroid cancer. We included any screening studies evaluating palpation, ultrasound, or both. Diagnostic accuracy studies of palpation or ultrasound had to include a reference standard (i.e., ultrasound for detection of nodules on palpation, histopathology results from FNA or surgery for detection of cancer on ultrasound). We required diagnostic accuracy studies to apply the chosen reference standard to both screen-positive and screen-negative persons (e.g., all or a random subset of screen-negative persons). Given the limited number of diagnostic accuracy studies, we also included screening studies that described the yield of cases of thyroid cancer. We excluded diagnostic procedures that included enhanced ultrasound-based techniques (e.g., elastography) or FNA. For screening effectiveness (KQ 1), we included any patient health outcome of reduced morbidity or mortality associated with thyroid cancer. For test performance (KQ 2), we included cancer detection rates and measures of diagnostic accuracy (e.g., sensitivity, specificity, positive and negative predictive values). For harms of screening (KQ 3), we included direct harms of palpation and ultrasound, subsequent harms of diagnostic FNA, and measures of overdiagnosis. For overdiagnosis, we looked for studies that compared screened versus unscreened groups. Studies that examined the rising incidence of thyroid cancer over time, the incidence and natural history of thyroid nodules and cancer, and autopsy studies were not included but are summarized in the Discussion section.

For treatment questions (KQs 4, 5), we included any studies of surgery (i.e., complete thyroidectomy, near-total thyroidectomy, lobectomy), with or without lymph node dissection or RAI ablation. We excluded studies of chemotherapy, external-beam radiation therapy, and other nonsurgical ablative treatments other than RAI. To approximate the treatment of screen-detected cancer, we excluded treatment studies among persons with metastatic disease or anaplastic thyroid cancer. We excluded surgical studies for thyroid conditions other than cancer (e.g., multinodular goiter, thyroiditis). For treatment benefit (KQ 4), we required studies to have a control group (e.g., untreated, surveillance, delayed treatment). The most commonly excluded study designs were comparative effectiveness trials or observational studies comparing one active treatment versus another (e.g., thyroidectomy vs. thyroidectomy plus lymph node dissection, robotic vs. conventional surgery). To assess the benefit of treatment, we considered the patient health outcomes of recurrence, mortality, and quality of life. For treatment harms (KQ 5), we did not require a control group for direct procedural harms (e.g., hypoparathyroidism, recurrent laryngeal nerve palsy) but did for other types of harms (e.g., second primary malignancies from RAI). The evolution of standard of care for the diagnostic workup (e.g., use of ultrasound-guided FNA) and treatment of thyroid cancer over time has resulted in a change in the case mix of patients having surgery, lymph node dissection, and/or RAI, as well as improvements in surgical techniques and RAI administered activity (doses) over time. To identify the most applicable evidence, we excluded studies conducted prior to 1990 and single-surgeon case series. We also excluded transient harms and surrogate measures (e.g., luteinizing hormone/follicle-stimulating hormone, salivary markers).

For the greatest applicability to U.S. practice, we focused on studies conducted in developed countries (“very high” development according to the United Nations Human Development
Because of resource constraints, we included only studies that published their results in English.

**Quality Assessment and Data Abstraction**

At least two reviewers critically appraised all articles that met the inclusion criteria based on the USPSTF’s design-specific quality criteria (Appendix A Table 3). We supplemented these criteria with the Newcastle-Ottawa Scale for cohort and case-control studies and Quality Assessment of Diagnostic Accuracy Studies I and II for studies of diagnostic accuracy (Appendix A Table 3). We rated articles as good, fair, or poor quality. In general, a good-quality study met all criteria. A fair-quality study did not meet, or it was unclear if it met, at least one criterion but had no known important limitations that could invalidate its results. A poor-quality study had a single fatal flaw or multiple important limitations. The most common fatal flaw for diagnostic studies included application of the reference standard to only those participants who screened positive (verification bias), because when missing data are not random or selective, analysis will generate biased estimates of diagnostic accuracy and generally lead to overestimation of both sensitivity and specificity. We also excluded diagnostic studies if they did not describe the followup of screen-negative persons. We excluded poor-quality studies from this review. Disagreements about critical appraisal were resolved by consensus and, if needed, in consultation with a third independent reviewer.

One reviewer extracted key elements of included studies into standardized evidence tables in Microsoft Excel® (Microsoft Corp., Redmond, WA). A second reviewer checked the data for accuracy. Evidence tables were tailored for each KQ and to specific study designs. Tables generally included details on study design and quality, setting and population (e.g., country, inclusion criteria, age, sex, race/ethnicity, risk factors for thyroid cancer), screening and treatment details, reference standard or comparator details (if applicable), length of followup, and outcomes (e.g., cancer yield, diagnostic accuracy, harms).

**Data Synthesis**

We synthesized results by KQ. We used a standardized summary of evidence table to summarize the overall strength of evidence for each KQ. This table included the number and design of included studies, summary of results, consistency or precision of results, reporting bias, summary of study quality, limitations of the body of evidence, and applicability of the findings.

Because of the limited number of studies and the clinical heterogeneity of the studies, we provided a narrative synthesis of results and used summary tables to allow for comparisons across different studies. For screening test performance (KQ 2) of palpation or ultrasound, we prioritized diagnostic accuracy (e.g., sensitivity, specificity) over yield (e.g., cancer incidence) outcomes. For harms of treatment (KQ 5), we stratified results by type of treatment (i.e., type of surgery, RAI). When possible, we conducted quantitative analyses for serious harms, including permanent hypoparathyroidism and permanent recurrent laryngeal nerve palsy. Quantitative analyses were not performed for other serious adverse events, as they were not commonly or
consistently reported or defined.

For quantitative analyses, we used random-effects models to estimate rates of serious adverse events. We applied the restricted maximum likelihood estimation method to estimate 95 percent CIs. In subgroup analysis, when the number of studies was less than five, we used the fixed-effects model for the analysis. Analyses were performed using R version 3.2.2 (The R Project for Statistical Computing, Vienna, Austria). We visually inspected plots stratified or ordered by key study characteristics accounting for clinical heterogeneity among studies to see if these characteristics affected rates of surgical complications. Key study characteristics included the type of surgery (e.g., partial thyroidectomy, total thyroidectomy, with or without lymph node dissection, type of lymph node dissection), case mix of patients (e.g., histology of thyroid cancer, average tumor size, average age), setting (e.g., country, year), and type and definition of outcome (e.g., criteria for permanent adverse effect). We were not able to evaluate if study quality (all were fair quality) or surgical experience (experience and surgical volume were not reported in individual studies) affected rates of surgical complications. If mean tumor size was not reported in a study, we calculated the weighted mean tumor size by reported T stage category or tumor size category (if either was available). We assumed that T stage categories were equivalent to the following mean tumor sizes: T1=1.0 cm, T2=3.0 cm, and T3=5.0 cm. To calculate weighted means using tumor size categories that had been chosen by the study authors, we used the sum of the percentage of subjects in each category and multiplied it by the midpoint of each category. The proportion of subjects with advanced disease was based on the highest proportion of study subjects in one of the following categories: stage III–IV, T3–T4, lymph node involvement, or metastasis. We also examined whether pooled effects were biased due to small, imprecise studies having larger than expected effect sizes. We performed tests of publication bias that examined whether the distribution of the effect sizes was symmetric with respect to effect precision using funnel plots and Egger’s linear regression method.

**Expert Review and Public Comment**

A draft research plan that included the analytic framework, KQs, and inclusion criteria was available for public comment in January 2015. We made no substantive changes to our review methods based on the comments received. A draft version of this report was reviewed by invited content experts and federal partners listed in the acknowledgements. Comments received during this process were presented to the USPSTF during its deliberation of the evidence and, subsequently, as appropriate, addressed in the final version of the report. Additionally, a draft of the full report was posed on the USPSTF Web site from November 22, 2016 through December 26, 2016. There were no changes made to the report based on these comments.

**USPSTF Involvement**

We worked with four USPSTF liaisons at key points throughout the review process to develop and refine the analytic framework and KQs and to resolve issues regarding the scope of the final evidence review. This research was funded by the Agency for Healthcare Research and Quality under a contract to support the work of the USPSTF. Agency staff provided oversight for the
project, assisted in external review of the draft report, and reviewed the draft report.
Chapter 3. Results

KQ 1. Compared With Not Screening, Does Screening Adults for Thyroid Cancer Lead to a Reduced Risk of Thyroid-Specific Mortality or Morbidity, Reduced All-Cause Mortality, and/or Improved Quality of Life?

We found no studies that met our inclusion criteria for KQ 1. We found no randomized, controlled trials (RCTs) or controlled clinical trials that evaluated the effects of thyroid cancer screening on patient morbidity or mortality compared with no screening. Two cohort studies that compared screened individuals versus a comparator group did not meet our inclusion criteria for KQ 1. One study (n=4,296) of high-risk persons who were previously exposed to radiation was not included because screening was conducted with older technology (i.e., a thyroid scan using 99mTc-pertechnetate and a pinhole collimator), and it used a historical control as the comparator group. One large Japanese cohort study (n=152,651) was excluded because the comparator group consisted of persons who presented with symptoms.

KQ 2. What Are the Test Performance Characteristics of Screening Tests for Detecting Malignant Thyroid Nodules in Adults?

Ten fair-quality studies met our inclusion criteria for KQ 2 (Table 1). Two studies reported on diagnostic accuracy of palpation to detect nodules, five on diagnostic accuracy of ultrasound to detect cancer, four on cancer yield from screening using palpation plus followup ultrasound, four on cancer yield from screening using ultrasound only, and two older studies on cancer yield from screening adults with a history of childhood irradiation.

Evidence to inform the true diagnostic accuracy of screening using neck palpation or ultrasound to detect thyroid cancer is limited. In two studies in Finland (n=354) using a single examiner, neck palpation was not sensitive enough to detect thyroid nodules. On the basis of results of two methodologically limited studies conducted in South Korea (n=243), using selected sonographic features when screening with ultrasound can be specific for detecting thyroid malignancy; one of these studies suggested that a combination of high-risk sonographic features, such as presence of microcalcifications or irregular shape, can optimize both sensitivity and specificity. Among studies that provided the screening yield, ultrasound-based screening detected a greater proportion of cancer cases than did palpation. The clinical significance (i.e., morbidity and mortality) of cancer detected through either screening method is unknown.

Diagnostic Accuracy of Palpation to Detect Thyroid Nodules

Two fair-quality prospective studies, both included in the previous review, examined the accuracy of neck palpation to detect thyroid nodules. These two studies (n=354) were
conducted by the same investigator in Finland in the late 1980s. Both studies used a single examiner to screen persons with neck palpation, which was immediately followed by ultrasound as the reference standard. However, only one study reported the diagnostic accuracy of palpation for all screened persons (regardless of the ultrasound results). In that study, screening by neck palpation was conducted among 253 randomly selected persons ages 20 to 49 years from the community. An abnormal result from neck palpation (i.e., thyroid nodule or diffuse enlargement of the thyroid) was found in 5.1 percent of subjects, whereas an abnormal result from ultrasound was found in 27.3 percent. The sensitivity and specificity of palpation to detect thyroid nodules (size not reported) were 11.6 (95% CI, 5.1 to 21.6) and 97.3 percent (95% CI, 93.8 to 99.1), respectively. In the other study, screening for thyroid nodules by neck palpation was conducted by a single examiner among 101 women ages 49 to 58 years who attended a mammogram screening program. The palpation results were reported for the 36 patients with abnormal ultrasound results. The sensitivity of palpation to detect nodules in persons with an abnormal ultrasound result was 27.8 percent. In this study, a second examiner palpated a subset of persons and examination findings were concordant with those of the first examiner in 18 of the 25 cases.

Limitations

Both studies are fair-quality prospective studies. However, these older, single-examiner studies were conducted outside the United States, and only one reported results for all persons who underwent examination of the thyroid by palpation. In addition, the number of participants in both studies was relatively small, and no cancer was detected.

Diagnostic Accuracy of Screening Ultrasound to Detect Thyroid Cancer

The vast majority of studies that examined the diagnostic accuracy of ultrasound to detect thyroid cancer were not (or were not reported to be) conducted in screening populations. These studies were therefore not included for KQ 2 but are summarized in the Discussion section. We identified two fair-quality studies that examined the diagnostic accuracy of screening ultrasound to detect thyroid cancer, neither of which was included in the previous review. Both studies (n=243) were conducted by the same investigators in South Korea from 2004 to 2007 but had different study designs.

The better quality study prospectively examined the diagnostic accuracy of screening for thyroid cancer by ultrasound in 113 of 2,079 screened women ages 15 to 77 years who attended a breast ultrasound clinic. Seventy-seven of the screened women (3.7%) revealed one or more malignant sonographic characteristics (e.g., presence of microcalcifications, irregular shape, ill-defined or microlobulated margin, marked hypoechogenicity, orientation of taller than wide). These 77 women and a small, nonrandomly selected subset of 36 screened women with probable benign ultrasound findings underwent ultrasound-guided FNA and either surgical confirmation or clinical followup at 2 years (depending on the FNA results). Among these 113 women who underwent FNA, 53 women were diagnosed with papillary thyroid cancer. The sensitivity and specificity of having one or more malignant features on screening ultrasound were 94.3 (95% CI, 84.3 to 98.8) and 55.0 percent (95% CI, 41.6 to 67.9), respectively (Table 2). Most screened women, however, did not undergo any further followup, including 756 women whose ultrasound
results revealed a nonsuspicious nodule or other abnormalities and 1,209 women whose results showed normal-appearing glands.

The other study was a retrospective analysis of 130 individuals from a series of 16,352 asymptomatic persons who referred themselves to thyroid cancer screening. A total of 1,009 screened persons had a sonographic lesion with one or more malignant features and were followed up with FNA. Malignant sonographic characteristics were defined as presence of microcalcifications, marked hypoechogenicity, well-defined spiculated margins, or solid rather than cystic nodule. The study sample included 58 of 150 lesions classified as malignant (all papillary thyroid cancer) and 82 of 823 lesions classified as benign by FNA results, for a total of 140 nodules in 130 persons. The 352 lesions with indeterminate or nondiagnostic FNA results were excluded from further analysis. The malignant lesions were diagnostically confirmed after surgical pathology, and the included benign lesions were all followed and confirmed to have no change in size by ultrasound or repeat FNA after at least 2 years. Among these 140 nodules, the sensitivity and specificity of having two or more malignant features on screening ultrasound were 94.8 and 86.6 percent, respectively (95% CI values could not be calculated) (Table 2). The authors found that the solid sonographic characteristic was sensitive (93.1%) but not specific (51.2%) for detecting malignancy, as opposed to other sonographic characteristics (i.e., microcalcifications, marked hypoechogenecity, and spiculated margins) that were specific, but less sensitive, for detecting malignancy. A total of 15,343 persons with no lesions or benign-appearing lesions had no followup after the initial screening with ultrasound.

Limitations

Both of these studies were fair-quality diagnostic accuracy studies conducted outside the United States. Both studies reported accuracy only among patients who had at least one study-defined malignant ultrasound characteristic and did not follow up on the vast majority (n=18,188) of screened individuals without these characteristics. Due to this study design, the potential amount of false-negative cases is generally unknown and estimates of sensitivity are likely overestimated. Despite the inherent limitations of the study design, we included these studies because they represent the only evidence of the accuracy of screening ultrasound in explicitly asymptomatic populations.

Yield of Screening for Thyroid Cancer

We found 10 studies that reported the yield of thyroid cancer cases from screening: four from palpation plus followup ultrasound, four from ultrasound only, and two older studies from palpation plus followup thyroid imaging using iodine or technetium in adults with a history of childhood irradiation. (Table 3). Four fair-quality studies, two in Japan (n=199,084) and two in Finland (n=352), reported the yield of thyroid cancer cases from screening palpation followed by ultrasound (if applicable) from 1980 to 2005. Overall, from 0 to 4.3 thyroid cancer cases were detected per 1,000 persons, and 90 percent of cases (372/415) were revealed to be papillary thyroid cancer. Three fair-quality prospective studies and one fair-quality retrospective study (n=20,521) on ultrasound screening for thyroid cancer were conducted in South Korea from 1997 to 2007, three of which were exclusively among women who presented for breast cancer screening or followup. In these four studies, 9.2 to
30.3 thyroid cancer cases were detected per 1,000 persons, and all but one case was diagnosed as papillary thyroid cancer. Studies were inconsistent in whether they reported the size and stage of screen-detected cancer. One study reported that 43 percent of ultrasound screen-detected patients had extrathyroidal extension. All four studies reported that 33 to 63 percent of patients with screen-detected thyroid cancer had lymph node metastases, with the highest percent detected by palpation and the lowest percent detected by ultrasound.

Two fair-quality prospective studies, one conducted in the United States (n=1,500) and the other in Israel (n=443), reported the yield of thyroid cancer in persons who had been treated with irradiation in childhood. Most of the subjects had been exposed to irradiation to the head and neck for benign conditions (which is no longer practiced). Initial screening in these two cohorts was by palpation; patients with palpable nodules then received thyroid imaging using iodine or technetium radioactive tracer with blood testing, including thyroid function tests. No cancer cases were detected in the smaller study, and 11.3 thyroid cancer cases per 1,000 persons were detected in the larger U.S. study. The histology of the cancer cases was not reported.

Limitations

The majority of these studies allow for calculation of cancer yield but not test performance. Across the studies, there was some variation in how cancer was identified and defined using histology from FNA, surgery, or both. Cancer yield varied by screening modality; however, differences in population characteristics (e.g., age, sex, personal history of irradiation) may have contributed to this variation.

KQ 3. What Are the Harms of Screening for Thyroid Cancer in Adults?

We found three studies that met our inclusion criteria for KQ 3 (Table 4). We found no studies that examined the harms of thyroid cancer screening with palpation or ultrasound and no studies that directly examined the effects of overdiagnosis in a screened versus unscreened group. A number of other study designs may indirectly inform the clinical importance and magnitude of overdiagnosis in thyroid cancer screening; these studies are summarized in the Discussion section of this report. Overall, there is very limited evidence to evaluate the potential harms of screening for thyroid cancer, including harms of diagnostic followup with FNA. One small U.S. study found that about a quarter of persons who had undergone FNA of a thyroid nodule did not meet the Society of Radiologists in Ultrasound (SRU) recommendation for FNA. Nonetheless, we found no evidence to suggest serious harms to patients from ultrasound-guided FNA.

Two fair-quality retrospective studies evaluated the harms of FNA of thyroid nodules. One large study at an institution in Japan reported the number of cases of needle tract implantation of papillary thyroid cancer from FNA. This study found 7 cases of tumor on the line of needle insertion in a total of 4,912 persons who had undergone ultrasound-guided FNA from 1990 to 2002. Subjects with needle tract implantation were older (age ≥50 years) and had a high incidence of poorly differentiated cancer and extrathyroidal extension of the tumor. It is unclear what effect, if any, this complication had on patient morbidity or mortality. A second study at a
single institution in the United States evaluated if bleeding complications due to FNA were related to antiplatelet or anticoagulant therapy in a cohort of 582 patients. About a quarter of the patients were taking antiplatelet or anticoagulant medications; five of them developed a hematoma after ultrasound-guided FNA, as detected by “immediate” postprocedural ultrasound. There was no statistically significant difference in the incidence of hematoma between persons taking or not taking antiplatelet or anticoagulant medications, albeit the number of outcomes was small. This study found no major bleeding complications (e.g., bleeding requiring hospitalization) in the cohort who had undergone ultrasound-guided FNA for a thyroid mass.

Another fair-quality retrospective study was designed to determine the proportion of FNAs of thyroid nodules that did not meet clinical guideline recommendations and therefore would be considered “unnecessary” diagnostic workup. This U.S. study included 400 consecutive subjects with thyroid nodules who subsequently received ultrasound-guided FNA at a single institution. Ninety-six (24%) subjects had FNA of a nodule that did not meet the SRU recommendation published in 2005. The SRU recommends FNA for nodules that have a maximum diameter of 1.0 cm or larger and have microcalcifications, nodules 1.5 cm or larger and are solid or have coarse calcifications, nodules 2.0 cm or larger and are mixed solid and cystic, and nodules with substantial growth since a prior assessment using ultrasound.

Limitations

Although they met our review’s inclusion criteria, these fair-quality retrospective studies did not report serious harms to patients (e.g., health outcomes resulting from needle tract implantation, hematoma, or “overuse” of FNA). The SRU recommendation varies slightly from the most recent guidelines set by the ATA. It is unclear what proportion of FNAs would be considered “unnecessary” using the ATA’s current recommendations.

KQ 4. Does Treatment of Screen-Detected Thyroid Cancer Reduce Thyroid-Specific Mortality or Morbidity, Reduce All-Cause Mortality, and/or Improve Quality of Life?

We found two unique studies (reported in five articles) that met our inclusion criteria for KQ 4 (Table 5). We found no trials designed to evaluate if earlier treatment or treatment of screen-detected, well-differentiated thyroid cancer results in better patient outcomes compared with observation (i.e., delayed treatment) or treatment of symptomatic, well-differentiated thyroid cancer. Due to major limitations in the study designs (e.g., lack of adjustment for confounders), it is uncertain if earlier or immediate treatment versus delayed or no surgical treatment improves patient outcomes for papillary carcinoma or papillary microcarcinoma.

One fair-quality retrospective observational study using SEER data from 1973 to 2005 compared survival rates of persons treated versus not treated for papillary thyroid cancer. Treatment included partial or total thyroidectomy, with or without postoperative RAI ablation. In total, 35,663 persons were analyzed; only 440 (1.2%) had not been treated. Compared with treated patients, untreated patients were older (mean age, 51 vs. 46 years), had a shorter length of
followup (mean, 5.9 vs. 7.6 years), had more missing data on tumor size (46% vs. 16% undocumented size), and had a smaller proportion of small tumors (≤1 cm) (13% vs. 40%). Overall, untreated persons had a slightly worse 20-year survival rate compared with treated persons (97% vs. 99%; p<0.001). These results were not adjusted for age, sex, tumor size, or any other potential confounders between treated and untreated persons, which limited our ability to compare the effect of treatment (vs. a case mix of patients) on patient outcomes. In a subgroup analysis limited to the 381 untreated patients from 1983 to 2005, patients who were recommended for treatment had a slightly lower 10-year survival rate than those who were not recommended for treatment (98.1% vs. 99.3%; p<0.001). In comparison, treated patients had a 10-year survival rate of 99.5 percent. Survival rates were not adjusted for other important confounders.

One fair-quality prospective study (reported in four separate articles) in Japan conducted from 1993 to 2013 examined the recurrence of disease in and the survival rate of persons with papillary microcarcinoma who opted for immediate surgery versus observation or active surveillance.70-73 From 1993 to 2004, 1,395 persons were analyzed, 340 of whom opted for observation with ultrasound once or twice per year.70 Thirty-two percent (n=109) of those who opted for observation ultimately had surgery. After approximately 6 years of followup, two persons in the immediate surgery group and no persons in the observation group had died. Three percent (32/1,055) of subjects in the immediate surgery group experienced disease recurrence, and no subjects in the observation group developed distant metastases. An additional 2,153 persons were diagnosed with papillary microcarcinoma from 2005 to 2013, 1,179 of whom opted for active surveillance and 974 of whom opted for immediate surgery.73 Only 8 percent (n=94) of those who opted for observation ultimately had surgery. After approximately 4 years of followup, no patients in either group developed distant metastases or died from thyroid cancer. Again, patients self-selected into one of two groups; therefore, the observation group had several statistically significant differences compared with the immediate surgery group at baseline, and outcomes were not adjusted for any confounders.

Limitations

Although these fair-quality observational studies met our review’s inclusion criteria, neither was adequately designed to evaluate the benefit of early surgical treatment versus observation (i.e., delayed or no surgery). Both studies had minimal or no adjustment for potential confounders that could have affected the decision for treatment versus observation of thyroid cancer.

KQ 5. What Are the Harms of Treating Screen-Detected Thyroid Cancer?

Fifty-two studies met our inclusion criteria for KQ 5. Thirty-six studies reported on surgical harms (i.e., total or partial thyroidectomy [including lobectomy], with or without lymph node dissection) (Table 6) and 16 studies reported on RAI harms (Table 7). Due to changes in the dose of RAI and the case mix of persons receiving RAI over time, we excluded older studies that addressed the long-term sequelae of RAI but summarize these studies in the Discussion section of this report. Overall, permanent surgical harms, hypoparathyroidism, and recurrent laryngeal
nerve palsy are not uncommon. The rate of permanent hypoparathyroidism varies widely; best estimates are from 2 to 6 events per 100 thyroidectomies and are more variable with lymph node dissection. The rate of permanent recurrent laryngeal nerve palsy is less variable and estimated at 1 or 2 events per 100 surgeries (including with lymph node dissection). Other serious surgical harms include death, adverse cardiopulmonary events, airway injury, wound complications, and infection. Having thyroid cancer is associated with an increased risk of second primary malignancies; however, best evidence suggests that treatment of differentiated thyroid cancer with RAI is independently associated with a small increase in second primary malignancies, although differences in study designs and variable reporting on radiation doses limit our understanding of the magnitude and precision of this small excess risk. Nonetheless, studies demonstrate that common clinically used doses of RAI are associated with an increased risk of both second solid and hematologic malignancies. RAI treatment is also associated with increased permanent adverse effects to the salivary gland, such as dry mouth.

### Surgical Harms

We found 36 studies that reported surgical harms, 32 studies of permanent hypoparathyroidism (hypocalcemia), 73-104 28 studies of permanent recurrent laryngeal nerve palsy (vocal cord paralysis), 73, 74, 76, 78-84, 86-103, 105, 106 two studies of surgical mortality, 107, 108 and 15 studies of other major surgical harms 73, 78, 79, 82, 85, 86, 88, 98-100, 102, 103, 107-109 (Table 8).

Studies reporting permanent or serious surgical harms were quite varied. The majority of studies were retrospective observational studies, although we also included three trials.77, 88, 94 Cohort size ranged from 76 to 13,854 persons. Only seven studies were conducted in the United States.92, 96, 99, 101, 105, 107, 108 Most of the studies included persons who had undergone surgery for thyroid cancer in the 1990s and 2000s. The average age of subjects was in the mid-40s to early 50s, with a predominance of women. The main surgeries evaluated were total or partial thyroidectomy, with or without lymph node dissection. Two studies reported surgical harms from partial thyroidectomy alone.75, 81 Lymph node dissection could be unilateral, bilateral, or not specified, as well as prophylactic, therapeutic, or not specified. Overall, there were 64 study arms in 36 studies. There was some variation in how permanent harm was defined, but it was generally defined as the adverse outcome persisting beyond 6 months.

Overall, there was large variation in the rate of permanent hypoparathyroidism due to total or partial thyroidectomy without lymph node dissection (15 study arms); the 95% CI of the pooled estimate ranged from 2 to 6 events per 100 surgeries ($I^2=73\%$) (Figure 2). The rate of permanent hypoparathyroidism from thyroidectomy with lymph node dissection was even more varied ($I^2=73\%$); the 95% CI for unilateral neck dissection (10 study arms) ranged from 1 to 4 events per 100 surgeries and, for bilateral neck dissection (9 study arms), from 1 to 10 events per 100 surgeries ($I^2=91\%$) (Figure 3). Given the very high statistical heterogeneity, it may be misleading to quantitatively pool rates of hypoparathyroidism across studies. A study by Viola et al,77 an outlier for hypoparathyroidism from total thyroidectomy with bilateral lymph node dissection, was an RCT that used the “most sensitive methods available” to determine hypocalcemia. However, other studies with comparatively high estimates of permanent hypoparathyroidism from lymph node dissection did not differ notably from the other included studies in study design, population, setting, tumor, intervention, or outcome characteristics.83, 86,
The rate of hypoparathyroidism did not seem to vary by year, setting, country, study-level proxy for more advanced tumors (average age, tumor size, histology), indication for lymph node dissection (prophylactic vs. therapeutic), or definition of “permanent” outcomes. In these pooled analyses, we excluded studies that did not distinguish between permanent and temporary harms. Both the funnel plots and Egger’s test suggested biased estimates due to smaller studies. Smaller studies in general appeared to report fewer events.

In contrast, there was little variation in the rates of permanent recurrent laryngeal nerve palsy due to thyroidectomy, with or without lymph node dissection. The 95% CI for recurrent laryngeal nerve palsy from thyroidectomy without lymph node dissection (14 study arms) was 1 to 2 events per 100 surgeries ($I^2=13\%$) (Figure 4). Estimates were similar for thyroidectomy with lymph node dissection (33 study arms) (Figure 5).

One fair-quality prospective study (n=2,153) by Oda et al. evaluated the differential surgical harms between persons who received immediate surgery (n=974) versus active surveillance (n=1,179) for papillary microcarcinoma. Persons who received active surveillance self-selected to be followed by serial neck ultrasound and laboratory testing. Ultimately, 94 of the 1,179 persons in the active surveillance group had surgery. Median followup was 47 months (range, 12 to 116 months). Permanent hypoparathyroidism was more common in the immediate surgery group (16/974) than in the active surveillance group (1/1,179) (p<0.0001). Recurrent laryngeal nerve palsy was uncommon and therefore not different between the two groups (2 vs. 0 events in the immediate surgery vs. active surveillance groups). The difference in surgical harms resulted from the lower number of persons in the active surveillance group going on to surgery.

Two fair-quality studies reported on surgical mortality. A large prospective observational cohort study in the United States of 5,584 persons with thyroid cancer who had undergone surgical treatment in 1996 found 15 deaths (0.3%) within 30 days, five of which were persons with undifferentiated or anaplastic cancer. Surgical mortality did not differ by type of surgical procedure, albeit the number of deaths was too low to make any meaningful conclusions. In a large retrospective study in the United States of 13,854 persons with thyroid cancer who had undergone surgery in 1999 to 2003, the same-stay mortality rate was 0.2 percent among those who had undergone lobectomy (9/4,238 patients) and 0.1 percent among those who had undergone total thyroidectomy (12/9,616 patients) (p=0.22). The study did not report mortality by type or stage of cancer. However, it did report other serious perioperative harms, including myocardial infarction, cerebrovascular accident, pulmonary embolus, pneumonia, airway injury (including tracheal injury), chyle leak, bleeding or hematoma requiring reoperation, and wound complications or infections. Because these harms were not consistently reported or defined, we do not discuss them further here.

Limitations

No studies reported serious adverse events not necessarily related to surgery (e.g., death, cardiopulmonary events) in an untreated control group. Clinical and statistical heterogeneity limited confidence in pooled estimates of permanent hypoparathyroidism. The studies did not allow us to evaluate if surgical experience or surgical volume influenced the rate of permanent surgical complications. Statistical tests for publication bias for pooled analyses of
hypoparathyroidism suggested biased estimates due to small studies. Pooled estimates may underestimate complications of hypoparathyroidism.

**RAI Harms**

We found 16 studies that reported harms of RAI. Eight studies addressed the risk of second primary malignancies, six addressed permanent adverse effects on salivary glands, one focused on hyperparathyroidism, and one on reproductive harms (Table 9).

Eight fair-quality retrospective studies (n=320,912) examined the incidence of second primary malignancies in persons with differentiated thyroid cancer being treated or not treated with RAI. Three of the studies were conducted using SEER data, none of which reported the indication for, nor the dose of, radiation from RAI. The largest of these three studies (n=37,176) included persons with papillary thyroid cancer and used SEER data from 13 registries with data from 1973 to 2006. Second primary malignancy was defined as a solid or hematologic cancer diagnosed more than 6 months after the index thyroid cancer was diagnosed. With an average of 11 years of followup (408,750 person-years), patients who received RAI experienced an excess absolute risk of 11.9 cancer cases per 10,000 person-years compared with a reference cohort. The standardized incidence ratio (SIR), the ratio of observed to expected second primary malignancies, was 1.18 (95% CI, 1.10 to 1.25) among persons who received RAI compared with a reference population of identical age, sex, race/ethnicity, and time. For persons who did not receive RAI, the SIR was 1.02 (95% CI, 0.98 to 1.06) compared with the same reference population. The second study (n=28,286) included persons with papillary or follicular thyroid cancer using SEER data from 1973 to 2002. Second primary malignancy was defined as cancer diagnosed more than 2 months after the index thyroid cancer. With an average of 10 years of followup (292,490 person-years), patients who received RAI had an excess absolute risk of 13.3 cancer cases per 10,000 person-years compared with a reference cohort (the general U.S. population). The SIR for second primary malignancies at any site in persons who received RAI was 1.21 (95% CI, 1.12 to 1.31) compared with the general U.S. population. Among persons who did not receive RAI, the SIR was 1.05 (95% CI, 1.00 to 1.10) compared with the general U.S. population. The third SEER study (n=29,456) included persons with thyroid cancer of any histology diagnosed from 1973 to 2000. Second primary malignancy was defined as any cancer diagnosed at least 2 months after the index thyroid cancer, which included newly diagnosed thyroid cancer. On average, there was about 8 years of followup (280,580 person-years). This study did not report the number of excess cancer cases by RAI exposure status. In a subgroup of persons from 1988 to 2000, the SIR for second primary malignancies at any site in those who received RAI was 1.14 (95% CI not reported) compared with a reference population of identical age, sex, race/ethnicity, and time. The SIR for second primary malignancies appeared to be similar in persons who did not receive RAI (1.19 [95% CI not reported]), although the statistical significance between groups was not reported. It is unclear what accounts for the difference in findings in this study compared with the previous two SEER studies. However, the primary aim of the study was not to determine the excess risk of second primary malignancy from RAI, and it differs from the other two SEER studies in three main ways: 1) it was not limited to papillary cancer, 2) it included thyroid cancer as a second primary malignancy, and 3) it had shorter followup for assessment of second primary malignancy.
Three smaller studies (n=4,273), from South Korea, Finland, and Hong Kong also examined the incidence of second primary malignancies in persons with differentiated thyroid cancer being treated or not treated with RAI. These studies generally reported the cumulative radiation doses in gigabecquerel (GBq) units (1 GBq=27.03 mCi). Radiation doses in clinical practice vary and generally correspond to the indication for RAI, such that lower doses (1.11 GBq) are used for ablation and higher doses (≤5.5 GBq) are used for adjuvant therapy for known or suspected residual disease. The largest study (n=2,468) included persons with differentiated thyroid cancer with at least 1 year of followup after thyroidectomy at a national university hospital in South Korea from 1976 to 2010. Second primary malignancy was defined as a nonsynchronous, nonthyroidal malignancy diagnosed at least 12 months after the index thyroid cancer diagnosis or RAI treatment. With an average of 7 years of followup (total person-years not reported), patients who received the highest cumulative dose of RAI (≥37 GBq) had an excess risk of 101.4 cancer cases per 10,000 person-years compared with the general Korean population. The excess risk decreased with lower cumulative doses (22.3 to 36.9 GBq; 24.6 cancer cases per 10,000 person-years), and no excess risk was observed at cumulative doses below 22.2 GBq. The study from Finland (n=910) included persons treated for differentiated thyroid cancer at one of two university hospitals from 1981 to 2002. Cases were matched to five controls with no prior thyroid cancer (selected from a national population register) on age, sex, and place of residence. Second primary malignancy was defined as an invasive cancer diagnosed at least 12 months after the index thyroid cancer was diagnosed. With an average of 16 years of followup (14,104 person-years) for all cases, patients who received more than the median cumulative dose of RAI (>3.7 GBq) had an excess risk of 25.3 cancer cases per 10,000 person-years compared with the controls. However, persons who did not receive any RAI also had an excess risk of 29.2 cancer cases per 10,000 person-years compared with controls. The third study from Hong Kong (n=895) included persons with papillary or follicular thyroid cancer using data from a single hospital in Hong Kong from 1971 to 2009. This study used the “standard” ablative dose (3 GBq), but higher doses were considered in the presence of more extensive disease. Second primary malignancy was defined as cancer diagnosed more than 12 months after diagnosis of the index thyroid cancer. With an average of 7.8 years of followup, 8.7 percent of patients who received RAI developed a second primary malignancy versus 3.2 percent of patients who did not receive RAI (p=0.004).

Two additional studies examined the incidence of specific subtypes of second primary malignancies—breast and leukemia diagnoses—in persons with thyroid cancer. The study examining the risk of developing breast cancer (n=10,361) included persons with thyroid cancer of any histology from Taiwan who were diagnosed from 2000 to 2008. Cases were frequency matched to four controls each from national health insurance data based on year of index diagnosis and age. Breast cancer outcomes were included if they were diagnosed after the thyroid index date or completion of RAI treatment, if RAI was received. With a median 6.5 years of followup (69,554 person-years), the excess risk of breast cancer among persons who received a cumulative RAI dose greater than 4.44 GBq was 2.7 cancer cases per 10,000 person-years compared with controls. The excess risk was 7.9 cancer cases per 10,000 person-years among persons who received a cumulative RAI dose of 4.44 GBq or less, and 4.6 cancer cases per 10,000 person-years among persons who did not receive RAI, compared with the same controls. The other study, examining the risk of developing leukemia (n=211,360), included persons with thyroid cancer of any histologic type as reported by the South Korean national health insurance.
Leukemia diagnoses were included if they occurred after the index thyroid cancer or after completion of RAI treatment, if RAI was received. With a median 2.4 years of followup (542,845 person-years), the incidence of leukemia was elevated among persons who received the highest cumulative doses of RAI (2.1 cancer cases per 10,000 person-years among those who received >5.5 GBq and 3.0 cancer cases per 10,000 person-years among those who received 3.7 to 5.5 GBq); incidence was 1.0 case per 10,000 person-years among persons who did not receive RAI (p <0.001 for trend).

One fair-quality retrospective and five fair-quality prospective studies (n=830) assessed the permanent harms of RAI on the salivary glands. The studies were generally small and had an average followup of 1 to 8.4 years. The mean radiation dose from RAI ranged from 1.1 to 5.3 GBq. The most common adverse effect of RAI to the salivary glands was xerostomia (dry mouth), which ranged from 2.3 to 35 percent. Dry mouth can adversely affect quality of life and vocal function and increase the risk of dental disease. One good-quality retrospective study (n=8,946), which used national data from Taiwan from 1997 to 2008, found no notable difference in the incidence of hyperparathyroidism between persons with papillary or follicular thyroid cancer who had received RAI and those who did not over an average followup of about 5 years. One fair-quality retrospective study (n=18,850) including a U.S. cohort of persons with papillary or follicular thyroid cancer found no notable difference in the birthrate between women who had received RAI and those who had not.

**Limitations**

Studies evaluating the harms of RAI using SEER data prior to 1987 may include radiation from other modalities, such as brachytherapy. Studies using SEER data did not account for the dose of radiation exposure. Studies reporting dose of radiation exposure varied in study design and ranges/thresholds of radiation doses, limiting direct comparisons. Studies that evaluated the harms of RAI on second primary malignancy and on salivary glands used different study methods, including how adverse outcomes were defined and adjustment (if any) for important known confounders. In addition, most of the studies that evaluated harms to salivary glands did not use a comparator arm (i.e., persons not exposed to RAI).
Chapter 4. Discussion

Summary of Evidence

In theory, screening for thyroid cancer could result in early detection of malignant nodules that can be treated more effectively, and with less harm, than if detected later or with symptomatic identification of thyroid cancer. However, we found no direct evidence to support this logic. To date there are no trials evaluating the (net) benefit of thyroid cancer screening (KQ 1) and, because most cases of thyroid cancer have a long latency period, screening trials of the benefit on patient health outcomes (i.e., morbidity and mortality) are unlikely to be conducted.

Well-designed studies evaluating the diagnostic accuracy of palpation or ultrasound in screening relevant populations (e.g., unselected or asymptomatic persons) are extremely limited (KQ 2). Two older Finnish studies, which were included in the prior review to support the 1996 USPSTF recommendation, demonstrated very low sensitivity of the neck examination to detect thyroid nodules, but these studies were limited to a single examiner.55, 56 Two small South Korean studies evaluated the diagnostic accuracy of various sonographic features with ultrasound in a screening population.57, 58 In those studies, using any one of several malignant sonographic characteristics (i.e., microcalcification, taller rather than wider, irregular shape, ill-defined or spiculated margin, a solid component with marked hypoechogenicity) could have high sensitivity (94.3%) but not specificity (and thus false-positive results) to detect cancer, whereas using a combination of two or more of these characteristics could have both high sensitivity (94.8%) and specificity (86.6%). These two studies, however, did not provide followup on most patients without malignant sonographic characteristics and thus likely overestimate the true sensitivity. Many other studies with similar study designs were excluded because they were not conducted or not reported to be conducted in a screening population. These two studies confirmed that these sonographic characteristics are most predictive of thyroid cancer but that the precision regarding the diagnostic accuracy of each characteristic varies (see “Diagnostic Accuracy of Ultrasound and FNA” below).

Potential harms of thyroid cancer screening are due primarily to subsequent diagnostic procedures and treatment. The major concerns for harms in thyroid cancer screening are “unnecessary” diagnostic procedures and treatments resulting from a false-positive finding and, more important, overdiagnosis (i.e., persons diagnosed with indolent cancer that would have never caused any suffering or death). We found very limited evidence to evaluate the potential harms of screening or FNA (KQ 3). One study suggested that, in practice, many persons (24%) with abnormal ultrasound findings go on to receive unnecessary FNA (i.e., do not meet current-day criteria and thus would not be expected to have a high risk of malignancy). Two studies suggested that there are no major harms of FNA, even though the procedure can result in localized hematomas or, very rarely, implantation of cancer along the needle tract.66, 67 Nonetheless, FNA is generally regarded as a safe procedure when performed by an experienced clinician;33-35, 37 the main potential harm is diagnostic inaccuracy or nondiagnostic samples leading to repeat procedures or unnecessary surgery (see “Diagnostic Accuracy of Ultrasound and FNA” below). We found no direct evidence (i.e., studies comparing screening with no screening) to evaluate the magnitude or effect of thyroid cancer screening on overdiagnosis.
However, because this is such an important issue for thyroid cancer screening—and for many, the main reason in the argument against thyroid cancer screening—we have included a discussion of the supporting evidence below (see “Overdiagnosis”).

We found no studies that evaluated if treatment of earlier-stage or screen-detected cancer compared with symptomatic cancer improves patient health outcomes (KQ 4). It is still unclear if immediate surgery, versus observation, improves patient health outcomes for small, well-differentiated thyroid cancer. We found one large study using SEER data that found very good 20-year survival rates in both treated and untreated persons with papillary thyroid cancer, albeit higher in treated persons. In this study, the untreated group was a self-selected minority (1% of the studied population) that differed in measured (and likely unmeasured) potential confounders for which the study did not adjust. It is therefore unclear if differences in survival were due to treatment rather than a case mix of patients who self-selected to be treated (or not). Nonetheless, this study demonstrates that the overwhelming majority of persons diagnosed with papillary cancer in the United States will get surgery. One Japanese study found no deaths after an average of 6 years of followup among persons who opted for observation of papillary microcarcinoma compared with two deaths among persons who opted for immediate surgery. Ultimately, 56 out of 162 persons in the observation group had surgical treatment. Again, this study did not adjust for confounders.

In contrast, we found a relatively large body of observational literature describing harms of treatment with surgery and RAI (KQ 5). These studies generally included a mixture of patients, likely not from screening alone. Limited data from included screening studies, however, suggest that a substantial proportion of screen-detected cancer cases include extrathyroidal extension or lymph node metastases at the time of detection, therefore warranting various interventions comparable with those in the treatment studies we identified. In addition, the rate of surgical harms did not appear to vary by study-level averages of proxies for prognosis (e.g., age, tumor size, tumor stage). We found that permanent surgical complications were not uncommon. Our pooled analysis showed that thyroidectomies were associated with 2 to 6 cases of permanent hypoparathyroidism and 1 to 2 cases of permanent recurrent laryngeal nerve palsy per 100 surgeries. The rate of hypoparathyroidism appears to be more variable with lymph node dissection. Our review is generally consistent with findings of existing systematic reviews. Jeannon et al reviewed 27 studies and estimated the incidence of permanent recurrent laryngeal nerve palsy at 2.3 percent, and Shan et al reviewed 16 studies and found no substantive increased risk of permanent hypoparathyroidism or recurrent laryngeal nerve palsy due to thyroidectomy with lymph node dissection compared with thyroidectomy alone. Limitations in the included primary literature did not allow for assessment of the effect of surgical volume on the variation of surgical harms across studies. However, evidence from nonincluded studies (not specific to thyroid cancer) suggests that surgeons with higher case volumes have lower rates of case complications, but even experienced surgeons have complication rates consistent with estimates in our review. Other serious harms may include death, airway injury, cardiopulmonary events, wound complications, and infection but are not commonly reported.

RAI is not routinely employed as treatment of thyroid cancer but is considered for persons with high-risk cancer. Radiation doses vary and generally correspond to the indication for RAI, such
that lower doses (1.11 GBq or 30 mCi) are used for ablation and higher doses (≤5.5 GBq or 150 mCi) are used for adjuvant therapy for known or suspected residual disease. Two studies using SEER data found an excess cancer risk of about 12 or 13 cancer cases per 10,000 person-years; however, neither reported the radiation dose. Smaller studies that reported radiation dose demonstrated an association with excess cancer risk at clinically used doses of RAI; however, differences in study designs and variable reporting on radiation dose limits our understanding of the magnitude and precision of this small excess risk. Our findings are consistent with older excluded literature. A 2009 systematic review by Sawka et al that estimated the risk of second primary malignancies after RAI treatment of thyroid cancer included two large studies, one by Brown et al using a U.S. cohort (included in our review) and another by Rubino et al using three European cohorts (n=6,841) (excluded because treatment dates were as early as 1934). The average radiation dose reported in the European cohorts was approximately 6 GBq (or 162 mCi), and the calculated excess absolute risk of secondary malignancies with approximately 1 GBq or 27 mCi was about 15 cancer cases (14.4 solid and 0.8 hematologic) per 100,000 person-years. Six studies showed that RAI (mean dose, 2.96 to 5.3 GBq or 80 to 142 mCi) is associated with permanent dry mouth (2.3% to 35%), which can adversely affect quality of life and increase the risk of dental disease. Although RAI can affect gonadal function, we found no evidence (including three older excluded studies) to suggest that lower doses of radiation from RAI result in male or female infertility. Our findings are consistent with existing systematic reviews that have examined the effects of RAI on gonadal function in men and women.

Key Contextual Issues

Diagnostic Accuracy of Ultrasound and FNA

Ultrasound

Our review was limited to diagnostic accuracy studies of screening the thyroid by ultrasound in unselected or asymptomatic persons, in which all or a random subset of screen-negative persons also received a reference standard (i.e., histology). Therefore, many studies evaluating the diagnostic accuracy of thyroid ultrasound (e.g., studies limited to persons with known thyroid nodules) and a rather large body of literature on the diagnostic accuracy of various ultrasound characteristics (e.g., nodule shape, margins, echogenicity, calcifications) for malignancy were excluded. Several studies have shown that certain ultrasound characteristics, in combination with nodule size, could help to determine the risk of malignancy and therefore potentially reduce unnecessary FNA testing.

Brito et al reviewed 31 studies on diagnostic accuracy (not in screening populations) published from 1985 to 2012 and showed that the three ultrasound characteristics most predictive of thyroid cancer malignancy were taller than wide shape, internal calcifications, and infiltrative margins. Taller than wide shape had a sensitivity of 0.53 (95% CI, 0.50 to 0.56) and specificity of 0.93 (95% CI, 0.91 to 0.94), internal calcifications had a sensitivity of 0.54 (95% CI, 0.52 to 0.56) and specificity of 0.81 (95% CI, 0.80 to 0.82), and infiltrative margins had a sensitivity of 0.56 (95% CI, 0.50 to 0.60) and specificity of 0.79 (95% CI, 0.77 to 0.80). Brito also showed that nodules with spongiform or cystic features (though present in only about 2% of
nodules) were most likely benign. Nodule size alone was not an accurate predictor of malignancy, and the review did not evaluate accuracy based on multiple characteristics. Smith-Bindman et al.\(^4\) conducted a retrospective case-control study of 11,618 thyroid ultrasounds done at the University of California, San Francisco from 2000 to 2005. In a multivariable model, only three characteristics remained statistically significant predictors of thyroid cancer: microcalcifications (odds ratio [OR], 8.1 [95% CI, 3.8 to 17.3]), nodule size 2 cm or larger (OR, 3.6 [95% CI, 1.7 to 7.6]), and entirely solid composition (OR, 4.0 [95% CI, 1.7 to 9.2]). Additional large studies out of Korea,\(^1\) Turkey,\(^2\) and Italy\(^3\) demonstrated that nodules with calcifications, taller than wide shape, irregular margins, and hypoechoic patterns are the most predictive of thyroid cancer malignancy. The sensitivity of each of these characteristics to characterize malignant nodules varied widely between studies, from 44 to 86 percent for microcalcifications, 40 to 76 percent for taller than wide shape, 48 to 90 percent for irregular margins, and 41 to 87 percent for hypoechoogenicity.\(^1\) Specificity ranged from 54 to 90 percent for microcalcifications, 60 to 91 percent for taller than wide shape, 81 to 92 percent for irregular margins, and 47 to 92 percent for hypoechoogenicity.\(^1\) Several, but not all, sensitivity and specificity results from our two included studies for KQ \(^2\) fall within these wide ranges, further emphasizing the wide variability in diagnostic accuracy for these characteristics.\(^1\)

Published in 2015, the ATA’s evidence-based guidelines on the management of thyroid nodules and differentiated thyroid cancer strongly recommends FNA for nodules larger than 1 cm with highly suspicious sonographic patterns, including a solid hypoechoic nodule or a solid hypoechoic component in a partially cystic nodule with either irregular margins, microcalcifications, taller than wide shape, or disrupted rim calcifications with small extrusive hypoechoic soft tissue component or extrathyroidal extension.\(^5\) The ATA guidelines also strongly recommend FNA for nodules at least 1 cm in size with intermediate suspicious sonographic patterns, including a hypoechoic solid nodule with a smooth regular margin and without microcalcifications, extrathyroidal extension, or taller than wide shape, although this latter recommendation was based on lower-quality evidence.

**FNA**

FNA is a quick, low-risk, and reliable procedure and currently the best method available for determining which thyroid nodules should be surgically removed or observed over time. Nonetheless, FNA is not perfect. The 2015 ATA guidelines recommend that FNA cytology be classified using the Bethesda System to reduce variability in reporting.\(^5\) The Bethesda System categories are 1) nondiagnostic or unsatisfactory, 2) benign, 3) AUS/FLUS, 4) follicular neoplasm or suspicious for follicular neoplasm, 6) suspicious for malignancy, or 6) malignant. Patients who have nodules with initial nondiagnostic results should have a repeat FNA with ultrasound guidance. Patients with malignant nodules should be recommended for thyroid surgery. Management of indeterminate nodules (AUS/FLUS, follicular neoplasm, or suspicious for malignancy) is more controversial and may involve molecular testing, repeat FNA, and/or surgical excision depending on the patient’s risk factors, the ultrasound characteristics of the nodules, and patient preference.

A 2012 systematic review of eight studies reported diagnostic accuracy among 25,445 FNAs of...
thyroid nodules. Overall, 6,362 (25%) FNAs resulted in surgery, with the proportion varying from 11.8 to 45.1 percent across the studies. The overall sensitivity and specificity of FNA for detecting malignancy were 97.0 and 50.7 percent, respectively, when considering the Bethesda System categories of follicular neoplasm, suspicious for malignancy, and malignant as positive. The sensitivity and specificity increased to 97.2 and 60.2 percent, respectively, when also including AUS/FLUS results (9.6% of all FNAs) as positive. The positive predictive value for AUS/FLUS alone was 15.9 percent, and 39.2 percent of these cases ultimately went on to surgery. The 2015 ATA guidelines state that only 7 percent of all thyroid FNAs are expected to have this result; however, Bongiovanni showed that the percentage of FNAs with AUS/FLUS results from eight studies ranged widely, from 0.8 to 27.2 percent. Large percentages of indeterminate FNA results as well as variability in the management of these indeterminate findings could ultimately result in unnecessary surgery and overtreatment of thyroid nodules. In the United States alone, 59,478 persons underwent thyroidectomy in an inpatient setting in 2009. While 18,008 (30.3%) of those operations were for thyroid cancer, the remainder were benign conditions, including nontoxic nodular goiter (36.0%) and benign neoplasms (11.2%).

**Overdiagnosis**

Overdiagnosis of thyroid cancer occurs when a thyroid malignancy is diagnosed but would not have caused symptoms or death during a patient’s lifetime. Overdiagnosis occurs because some thyroid tumors grow so slowly that the cancer never progresses (and sometimes regresses) or progresses at such a slow pace that the person dies of other causes before the cancer is symptomatic. Welch and Black proposed two prerequisites for overdiagnosis, both of which are met by thyroid cancer: 1) the existence of a disease reservoir or a substantial number of subclinical cancer cases, and 2) a method to detect these subclinical cancer cases via screening. One of the major harms from overdiagnosis is overtreatment, or the overuse of procedures that may result in treatment harms without (or only marginally) improving patient outcomes. According to SEER data, 98.8 percent of persons with thyroid cancer diagnosed from 1973 to 2005 underwent definitive treatment. Overdiagnosis may also lead to preventable harms, such as increased patient anxiety, potential complications or adverse effects from treatment (as we reported), and increased health care costs without benefit to the patient. However, at present, there is no clear way to determine which cases of thyroid cancer would actually require treatment to improve patient survival and which would not.

Although overdiagnosis is arguably the most important harm of thyroid cancer screening, it is not addressed by our review of harms of screening (KQ 3) due to limitations in the evidence base. To accurately estimate overdiagnosis, studies must compare screened and unscreened groups. We found no studies that compared screened and unscreened groups with incidence or overdiagnosis of thyroid cancer as an outcome. A recent review by Carter et al provided an overview of the types of study designs needed to accurately quantify overdiagnosis in cancer screening, none of which is available for thyroid cancer. The four types of study designs are modeling studies, pathological and imaging studies, ecological and cohort studies, and followup of RCTs. Modeling studies compare the way cancer would hypothetically occur with and without screening. Biases may limit the quality of modeling studies and include a lack of direct evidence to support modeling assumptions, validation analyses, or generalizability. Pathology and imaging studies determine the extent of overdiagnosis-based characteristics seen in imaging or pathology.
studies, such as tumor growth rate. These studies assume that pathology or imaging characteristics are strongly correlated with cancer morbidity or mortality, which may be a difficult assumption to apply to thyroid cancer because it has not been determined which tumor characteristics are more predictive of mortality than others, particularly for papillary carcinomas. Ecological and cohort studies typically follow persons through cancer screening programs and compare the cancer incidence with unscreened control groups. These studies are subject to selection bias and confounding from control group selection (studies typically use historical controls or controls from different geographic areas without screening programs) or lead-time bias from insufficient followup time. Followup from RCTs comparing screening with no screening may be the least-biased study design for assessing overdiagnosis; however, this type of study is rare, even for other types of cancer.

Incidence and Mortality Data

The best evidence we have to suggest that overdiagnosis is a problem with thyroid cancer comes from studies showing a rising incidence in thyroid cancer detection over time with no corresponding change in the mortality rate. Several studies by Davies and Welch have used SEER data to estimate the incidence of thyroid cancer and cancer-related mortality in the United States since 1973. The most recent estimates, published in 2014, showed that the incidence of thyroid cancer increased from 4.9 cases per 100,000 persons in 1975 to 14.3 cases per 100,000 persons in 2009, representing an absolute increase of 9.4 (95% CI, 8.9 to 9.9) cases per 100,000 persons. When only cases of papillary cancer were examined, the absolute increase over the same time period was 9.1 (95% CI, 8.6 to 9.6) cases per 100,000 persons; thus, cases of papillary cancer have accounted for the majority of the increased incidence among all cases of thyroid cancer. These increases were 3 to 4 times greater in women than in men. The size distribution of diagnosed tumors has shifted toward smaller lesions, with the proportion of tumors smaller than 1 cm that were diagnosed increasing from 25 percent in 1988 to 1989 to 39 percent in 2008 to 2009. During the same time period, the rate of thyroid cancer mortality remained stable (approximately 0.5 deaths per 100,000 persons). Ho et al conducted a similar analysis using SEER data and found nearly identical thyroid cancer incidence and mortality rates over time. They also noted that the 10-year disease-specific survival for patients diagnosed from 1983 to 1999 increased from 95.4 to 98.6 percent (p=0.002), which may reflect, in part, that small, asymptomatic cancer accounts for most new diagnoses. Neither of these studies reported mortality by histologic type, which is an important element to consider given that patients with medullary and anaplastic cancer have much higher mortality rates than patients with differentiated cancer.

Data from other countries have shown similar findings. A summary report using data from the Cancer Incidence in Five Continents database showed steady increases in thyroid cancer incidence in 12 selected countries from 1960 to 2007, which, again, was primarily driven by a rise in papillary carcinoma diagnoses. Mortality data from the World Health Organization showed that the mortality rate from 2000 to 2010 either stabilized around 0.2 deaths per 100,000 men and 0.6 deaths per 100,000 women or declined by 2 to 3 percent per year for men and 2 to 5 percent per year for women. The declines are likely related to changes in both risk factors (due to improvements in diet, reductions in iodine deficiency, and medical use of ionizing radiation over the last couple of decades in some countries) and cancer detection (resulting in overdiagnosis of
cancer with a favorable prognosis). The best example that illustrates the problem of overdiagnosis of thyroid cancer comes from South Korea, which has had an organized cancer screening program since 1999. Although the program did not officially include thyroid cancer screening, providers frequently offered thyroid screening with ultrasound for a small additional cost. In 2011, the rate of thyroid cancer diagnoses was 15 times the rate in 1993, while the rate of thyroid cancer mortality remained stable. In 2011 alone, more than 40,000 persons were diagnosed with thyroid cancer, whereas fewer than 400 died. Nearly every person diagnosed with thyroid cancer underwent surgical treatment. One study noted that the tumors excised decreased in size over time: the proportion of tumors excised that were smaller than 1 cm increased from 14 percent in 1995 to 56 percent in 2005. Increases in thyroid cancer incidence ranging from 3.2 to 6.2 percent per year in men and 3.5 to 8.1 percent in women were noted in France, Australia, and Canada from the early 1980s to the late 1990s, although none of these countries implemented thyroid cancer screening as did South Korea. These studies also noted that the majority of the increase in incidence was due to an increase in cases of papillary thyroid carcinoma or microcarcinoma.

Several studies have evaluated whether the increased incidence in thyroid cancer is a result of increased detection (e.g., through increased imaging) or a true increase in risk factors for thyroid cancer (e.g., due to exposure to ionizing radiation). Davies et al conducted a small (n=279) study to identify examinations that led to detecting cases of asymptomatic cancer. The results showed that 46 percent (n=44 of 95) of identified cases of cancer were first noted on surgical evaluation following detection of a nodule during a routine examination (i.e., asymptomatic), imaging for an unrelated cause, or diagnostic workup for other problems where the thyroid might be involved (e.g., patient complaining of fatigue). The majority of these cases of “asymptomatic” cancer were papillary (n=37), and the mean tumor size was 1.9 cm (range, 0.2 to 10 cm), whereas the mean tumor size of cases of symptomatic papillary cancer was 2.4 cm (range, 0.2 to 8 cm). These results are consistent with the epidemiologic studies described above that noted that the increased incidence of thyroid cancer was due to increased diagnoses of small papillary tumors.

However, a review by Pellegriti et al noted that while the largest increase in thyroid cancer incidence occurred among tumors smaller than 1 cm, smaller increases have occurred among larger tumors. For example, a study using SEER data showed that the incidence of thyroid tumors 2 to 4 cm in size increased 4.6 percent (95% CI, 3.5 to 5.7) per year from 1995 to 2006 and that of larger tumors increased 4.1 percent (95% CI, 3.4 to 4.8) per year from 1983 to 2006. A second SEER study showed that from 1983 to 2006, papillary thyroid tumors smaller than 1 cm, 1.1 to 2 cm, 2.1 to 5 cm, and larger than 5 cm increased by 19.3, 12.3, 10.3, and 12.0 percent per year, respectively. A third SEER study showed that from 1992 to 2005, approximately 50 percent of the overall increase in papillary cancer incidence was from tumors 1 cm or smaller, 30 percent from tumors 1.1 to 2 cm, and 20 percent from larger tumors.

Pellegriti et al also pointed out that an increase in incidence solely due to increased detection should have affected all histologic types and sexes equally. As we noted above, prior studies primarily noted increases among cases of papillary cancer and larger increases in incidence in women than men. Additional studies, such as the study by Chen et al, used SEER data to demonstrate that increases in incidence by tumor size differed between men and women.
From 1988 to 2005, the incidence of tumors smaller than 1 cm increased by 4 percent (95% CI, 0.8 to 7.3) per year in men and by 8.6 percent (95% CI, 7.8 to 9.5) per year in women. During the same time period, the incidence of thyroid tumors 1.0 to 2.9 cm in size increased by 5.5 percent (95% CI, 4.2 to 6.8) in men and by 0.4 percent (95% CI, -3.0 to 3.8) in women. Enewold et al.\textsuperscript{154} also using SEER data, demonstrated that the incidence rate of thyroid cancer varied by race/ethnicity, sex, and histology. Papillary cancer was the only histologic type to see a significant increase in the rate from 1980 to 2005, but the increase differed by race/ethnicity and sex (8.0%, 2.7%, 3.80%, and 1.16% per 100,000 person-years for white women, white men, black women, and black men, respectively).

In the United States, the increased incidence of thyroid cancer over time may also be related to increased access to care. This hypothesis is supported by a study that used SEER data from 1973 to 2009 linked to U.S. Census data from 2000 to demonstrate that thyroid cancer incidence was positively correlated with income level, education, and employment.\textsuperscript{156} Another study, which used data from 1999 to 2009 from the U.S. Cancer Statistics Report linked to Lifescript doctor review data and administrative patient claims data, showed that the incidence of thyroid cancer was significantly correlated with the density of endocrinologists.\textsuperscript{43} Incidence varied from 4.7 per 100,000 person-years in Oklahoma to 9.1 per 100,000 person-years in Rhode Island, and much of the variation in incidence by U.S. State could be explained by the density of endocrinologists and general surgeons as well as the use of neck ultrasounds.

Pellegriti et al.\textsuperscript{18} also reviewed whether changes in potential risk factors for thyroid carcinoma are related to thyroid cancer incidence, which could suggest that some of the increased incidence is real and not related to overdiagnosis. Pellegriti et al noted that in the United States, the use of medical imaging such as CT scans and x-ray has increased dramatically and that individual doses of ionizing radiation from these examinations doubled from 1980 to 2006. Additional elements of diet, lifestyle, and pollution (e.g., iodine intake, food or environmental contaminants) may influence the risk of thyroid cancer, but these aspects have not been studied carefully.\textsuperscript{18} Thus, limited evidence exists for an external cause of increased thyroid cancer incidence beyond increased detection and diagnosis.

**Autopsy Data**

Autopsy studies have provided additional evidence on overdiagnosis of thyroid cancer. A 2014 review by Lee et al.\textsuperscript{30} summarized 15 studies published from 1969 to 2005 on latent thyroid cancer discovered at autopsy. Of 8,619 thyroid glands obtained at autopsy, 989 (11.5%) were positive for papillary thyroid carcinoma. The proportion of cases of papillary thyroid cancer varied widely, from 1.0 to 35.6 percent. The majority of the tumors were tiny (diameter <1 to 3 mm), and women and men were equally likely to have papillary cancer diagnosed on autopsy. The authors compared the autopsy diagnoses with 1,355 papillary microcarcinomas diagnosed clinically at their institution; most patients diagnosed clinically (67.3%) had tumors larger than 0.5 cm, and women were 11 times more likely than men to be diagnosed. Comparisons between the latent cancer cases diagnosed on autopsy and the papillary microcarcinomas diagnosed clinically are likely subject to selection bias, but the autopsy studies clearly demonstrate that a proportion of thyroid cancer cases would likely never result in symptoms or mortality.
Natural History Data

Studies describing the natural history of thyroid nodules and malignancies also lend evidence to the problem of overdiagnosis in thyroid cancer. The 2015 ATA guidelines note that benign nodule growth has been variably defined across studies and that there is no good cutoff to use for percent change in size when determining whether to conduct repeat FNA on nodules previously diagnosed as benign.5 Durante et al157 describe 5-year followup of 992 patients with benign thyroid nodules (size, 0.4 to 4 cm). In 686 patients (69%), the size of the nodules remained stable; in 184 patients (18.5%), the size of one or more nodules decreased; and in 153 patients (15.4%), the size of one or more nodules increased by 20 percent or more (the groups were not mutually exclusive because some persons had more than one nodule). Ultimately, only five patients were diagnosed with thyroid malignancy. There are currently no studies with followup of benign nodules beyond 5 years. Studies with longer followup are needed to help determine whether indefinite followup of nodules is necessary.

Nodule growth was evaluated in persons with thyroid cancer by Ito et al.71 Among 162 persons with papillary microcarcinoma who opted for observation instead of immediate surgical treatment, within 1 year the tumor increased in size by 2 mm or more in 20 patients (15.3%), decreased in size by 2 mm or more in 18 patients (13.8%), or did not change in 92 patients (70.8%). After 5 years of followup, 72.3 percent of tumors did not increase in size. Ultimately, 56 patients went on to have surgery, only 13 of whom had an increase in tumor size of 2 mm or more.

Another study, of 2,070 patients with papillary microcarcinoma, looked at recurrence-free survival up to 35 years after diagnosis and found that the survival rate was 96.7 percent for patients with tumors 5 mm or smaller and 86.0 percent for patients with tumors 6 to 10 mm in size (p<0.0001).158 In a multivariable survival model, neither patient age nor sex was predictive of disease-free survival. A total of 73 patients experienced recurrence at a median time of 10.3 years. A large SEER study evaluated recurrence and mortality outcomes among 18,445 patients with papillary microcarcinoma; at 15 years, the overall survival rate was 90.7 percent and the disease-specific survival rate was 99.3 percent. In multivariable survival models, patient characteristics related to poorer overall survival included age older than 45 years (hazard ratio [HR], 6.18 [95% CI, 4.80 to 7.97]), male sex (HR, 1.74 [95% CI, 1.44 to 2.11]), and African American race (HR, 2.56 [95% CI, 1.88 to 3.47]).159 A total of 49 patients died of thyroid cancer. Using SEER data, Davies et al69 estimated the rate of papillary thyroid cancer–specific survival by whether the patients received definitive treatment. The 10-year survival rate was 99 percent among 29,789 persons who had received definitive treatment and 97 percent among 440 persons who did not. In the United States, data from single institutions have demonstrated that the overwhelming majority of thyroid cancer diagnoses are stage I papillary cancer, with 20-year survival approximating 100 percent.160 These data highlight the slow-growing nature of thyroid tumors and the low potential for recurrence or mortality due to thyroid cancer, particularly papillary tumors and microcarcinoma. However, data on the survival of patients who never receive treatment are very limited. As Ho et al147 pointed out, the high survival rates of patients with thyroid cancer may be a result of length bias as increasing numbers of subclinical thyroid cancer cases are diagnosed, thereby shifting survival curves toward longer survival.
Limitations of the Review

Our review was designed to support the USPSTF in making a recommendation regarding screening for thyroid cancer such that our inclusion criteria reflected decisions about identifying the most applicable evidence for our primary stakeholder. We did not include studies primarily focused on cohorts exposed to high doses of radiation due to environmental disasters. In addition, we did not review the diagnostic accuracy of ultrasound or ultrasound characteristics to detect thyroid cancer in nonscreening populations, primarily due to referral bias, although we provide a summary of this literature (see “Diagnostic Accuracy of Ultrasound”). For our review of overdiagnosis, our inclusion criteria required studies that compared screened versus unscreened persons. However, because these types of studies do not exist, we summarize the supporting literature (see “Overdiagnosis”).

Because our review focused on the benefit of treatment versus observation or the treatment of asymptomatic versus symptomatic disease, we excluded studies evaluating the comparative benefits and harms of treatment (i.e., the most effective or safest treatment). We also excluded harms not directly related to surgery or RAI (e.g., subsequent harms from suppressive doses of thyroxine). We excluded older studies that examined harms of RAI, as radiation doses have changed over time. Nonetheless, we acknowledge that, over time, surgical techniques, RAI doses, and the case mix of persons undergoing surgery and/or RAI have changed; such included studies still may not accurately reflect modern-day practice. Given our primary audience and resource limitations, we limited our review to evidence conducted in countries with the most appropriate applicability to U.S. practice and to articles published in English. We do not believe these criteria resulted in omission of any key evidence.

Due to the sparse data for most of the KQs, we were limited to nonquantitative analyses. Our meta-analyses to pool surgical harms had high statistical heterogeneity for outcomes of hypoparathyroidism, which we could not explain by using several study-level characteristics.

Evidence Gaps and Future Research Needs

Overall, there is very little evidence examining the benefit of screening for thyroid cancer. No professional society recommends population-based thyroid cancer screening. Additionally, there is little evidence to support screening in persons with an elevated risk of thyroid cancer. Previously, the USPSTF stated there was insufficient evidence to recommend screening in persons with a personal history of irradiation (and we found no new studies in this review), and the ATA stated in 2015 that there was insufficient evidence to support screening persons with a family history of differentiated thyroid cancer. Although population-based screening trials for thyroid cancer are unlikely, trials or well-designed observational studies to address the benefit of screening in higher-risk populations (e.g., those with a personal history of irradiation or a family history of differentiated thyroid cancer) would be helpful to understand if there is any role at all for screening for thyroid cancer. The use of radiation to treat benign conditions in childhood ended several decades ago, so questions about best practices for screening in this population may not be a priority, but there are smaller subpopulations who have received radiation for diagnostic (e.g., CT scans) or therapeutic (e.g., treatment of hematologic cancer) purposes in childhood,
adolescence, or early adulthood, for whom assessing best practices may be relevant.

Given the indolent nature of many cases of thyroid cancer and the risks associated with treatment, trials or well-designed observational studies examining the benefit of early treatment versus observation or surveillance for patients with (smaller) well-differentiated thyroid cancer are also needed. In addition, we need studies to determine which set of prognostic indicators predict aggressive versus indolent disease. Over the past decade, better understanding of the genetic mechanisms of thyroid cancer and the creation of molecular tests to aide in cancer diagnosis have made molecular markers a very promising area of research to help derive a prognosis of thyroid cancer.\(^5\)

**Conclusions**

Although an ultrasound of the neck using high-risk sonographic characteristics plus followup cytology from FNA can reasonably identify thyroid cancer, it is still unclear if population-based or targeted screening can decrease mortality rates or improve important patient health outcomes. Screening results in the identification of indolent thyroid cancer which would not have resulted in any morbidity or mortality in a person’s lifetime. Treatment of these cases of overdiagnosed cancer can pose real harms, including complications from surgical and RAI treatment. There is a lack of evidence to understand the true magnitude of overdiagnosis as well as the risk markers that predict indolent versus progressive thyroid cancer.
References


42. Hughes DT, Haymart MR, Miller BS, Gauger PG, Doherty GM. The most commonly occurring papillary thyroid cancer in the United States is now a microcarcinoma in a patient older than 45 years. Thyroid. 2011 Mar;21(3):231-6. PMID: 21268762.


152. Davies L, Ouellette M, Hunter M, Welch HG. The increasing incidence of small thyroid cancers: where are the cases coming from? Laryngoscope. 2010 Dec;120(12):2446-51. PMID: 21108428.


Figure 1. Analytic Framework and Key Questions

Key Questions
1. Compared with not screening, does screening adults for thyroid cancer lead to a reduced risk of thyroid-specific mortality or morbidity, reduced all-cause mortality, and/or improved quality of life?
2. What are the test performance characteristics of screening tests for detecting malignant thyroid nodules in adults?
3. What are the harms of screening adults for thyroid cancer?
4. Does treatment of screen-detected thyroid cancer reduce thyroid-specific mortality or morbidity, reduce all-cause mortality, and/or improve quality of life?
5. What are the harms of treating screen-detected thyroid cancer?
Figure 2. Key Question 5 Results: Permanent Hypoparathyroidism From Surgery, Stratified by Type of Thyroidectomy

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Tumor Size</th>
<th>Events</th>
<th>Total</th>
<th>Event Rate per 100 (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palestini 2008</td>
<td>not USA</td>
<td>&lt;1.0 cm</td>
<td>4</td>
<td>148</td>
<td>2.70 [0.74, 6.78]</td>
</tr>
<tr>
<td>Kwan 2015</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>1</td>
<td>51</td>
<td>1.96 [0.95, 10.45]</td>
</tr>
<tr>
<td>Viola 2015</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>7</td>
<td>88</td>
<td>7.96 [3.26, 16.70]</td>
</tr>
<tr>
<td>Alm 2014</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>13</td>
<td>291</td>
<td>4.47 [2.45, 7.52]</td>
</tr>
<tr>
<td>Corese 2014</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>4</td>
<td>390</td>
<td>1.03 [0.28, 2.61]</td>
</tr>
<tr>
<td>Tartaglia 2014</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>27</td>
<td>284</td>
<td>9.51 [6.36, 13.33]</td>
</tr>
<tr>
<td>Calo 2013</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>8</td>
<td>159</td>
<td>4.73 [2.07, 9.11]</td>
</tr>
<tr>
<td>Raffelli 2012</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>0</td>
<td>62</td>
<td>0.00 [0.00, 5.78]</td>
</tr>
<tr>
<td>Harl 2013</td>
<td>not USA</td>
<td>2.1-4.0 cm</td>
<td>6</td>
<td>51</td>
<td>6.59 [2.46, 13.80]</td>
</tr>
<tr>
<td>Syvok 2006</td>
<td>not USA</td>
<td>2.1-4.0 cm</td>
<td>2</td>
<td>391</td>
<td>0.51 [0.06, 1.84]</td>
</tr>
<tr>
<td>Chaplin 1999</td>
<td>not USA</td>
<td>2.1-4.0 cm</td>
<td>2</td>
<td>103</td>
<td>1.94 [0.94, 3.84]</td>
</tr>
<tr>
<td>Beute 2013</td>
<td>not USA</td>
<td>unknown</td>
<td>2</td>
<td>22</td>
<td>9.06 [1.12, 20.16]</td>
</tr>
<tr>
<td>Giordano 2012</td>
<td>not USA</td>
<td>unknown</td>
<td>25</td>
<td>394</td>
<td>6.35 [4.15, 9.22]</td>
</tr>
<tr>
<td>RE model for subgroup</td>
<td>USA</td>
<td>unknown</td>
<td>104</td>
<td>2184</td>
<td>3.65 [2.23, 6.66]</td>
</tr>
<tr>
<td>Total or partial thyroidectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yassa 2007</td>
<td>USA</td>
<td>2.1-4.0 cm</td>
<td>2</td>
<td>296</td>
<td>0.68 [0.08, 2.42]</td>
</tr>
<tr>
<td>Shindo 1995</td>
<td>USA</td>
<td>unknown</td>
<td>8</td>
<td>156</td>
<td>5.13 [2.24, 9.85]</td>
</tr>
<tr>
<td>RE model for subgroup</td>
<td>USA</td>
<td>unknown</td>
<td>10</td>
<td>452</td>
<td>3.40 [1.83, 6.21]</td>
</tr>
<tr>
<td>Total thyroidectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Tumor size=calculated mean tumor size.

Abbreviations: CI=confidence interval; RE=random effects; FE=fixed effects.
Figure 3. Key Question 5 Results: Permanent Hypoparathyroidism From Surgery, Stratified by Type of Lymph Node Dissection

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Tumor Size</th>
<th>Events</th>
<th>Total</th>
<th>Event Rate per 100</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unilateral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calgkan 2012</td>
<td>not USA</td>
<td>&lt;1.0cm</td>
<td>11</td>
<td>428</td>
<td>2.57 [1.29; 4.55]</td>
<td></td>
</tr>
<tr>
<td>Calgkan 2012</td>
<td>not USA</td>
<td>&lt;1.0cm</td>
<td>0</td>
<td>414</td>
<td>0.00 [0.00; 0.89]</td>
<td></td>
</tr>
<tr>
<td>Kim 2011</td>
<td>not USA</td>
<td>&lt;1.0cm</td>
<td>4</td>
<td>138</td>
<td>2.90 [0.80; 7.26]</td>
<td></td>
</tr>
<tr>
<td>Raffaelli 2012</td>
<td>not USA</td>
<td>1.1.2.0cm</td>
<td>0</td>
<td>62</td>
<td>0.00 [0.00; 5.78]</td>
<td></td>
</tr>
<tr>
<td>Lee 2010</td>
<td>not USA</td>
<td>1.1.2.0cm</td>
<td>3</td>
<td>513</td>
<td>0.58 [0.12; 1.70]</td>
<td></td>
</tr>
<tr>
<td>Palestini 2008</td>
<td>not USA</td>
<td>1.1.2.0cm</td>
<td>0</td>
<td>93</td>
<td>0.00 [0.00; 3.89]</td>
<td></td>
</tr>
<tr>
<td>Son 2005</td>
<td>not USA</td>
<td>1.1.2.0cm</td>
<td>1</td>
<td>56</td>
<td>1.79 [0.05; 9.55]</td>
<td></td>
</tr>
<tr>
<td>Sywak 2008</td>
<td>not USA</td>
<td>1.1.2.0cm</td>
<td>1</td>
<td>56</td>
<td>1.79 [0.05; 9.55]</td>
<td></td>
</tr>
<tr>
<td>Giordano 2012</td>
<td>not USA</td>
<td>unknown</td>
<td>27</td>
<td>385</td>
<td>7.01 [4.67; 10.04]</td>
<td></td>
</tr>
<tr>
<td>Moo 2009</td>
<td>USA</td>
<td>unknown</td>
<td>0</td>
<td>12</td>
<td>0.00 [0.00; 26.46]</td>
<td></td>
</tr>
<tr>
<td>RE model for subgroup</td>
<td></td>
<td></td>
<td>47</td>
<td>2167</td>
<td>1.86 [0.64; 4.04]</td>
<td></td>
</tr>
</tbody>
</table>

| Bilateral   |         |            |        |       |                    |            |
| So 2010     | not USA | <1.0cm     | 6      | 561   | 1.09 [0.40; 2.35]  |            |
| Viola 2015  | not USA | 1.1.2.0cm  | 18     | 93    | 19.35 [1.69; 28.65]|            |
| Conzo 2014  | not USA | 1.1.2.0cm  | 13     | 362   | 3.59 [1.93; 6.08]  |            |
| Raffaelli 2012| not USA | 1.1.2.0cm  | 1      | 62    | 1.61 [0.04; 0.80]  |            |
| Lee 2010    | not USA | 1.1.2.0cm  | 1      | 97    | 1.03 [0.03; 5.51]  |            |
| Moo 2009    | USA     | 1.1.2.0cm  | 0      | 104   | 0.00 [0.00; 3.48]  |            |
| Sort 2008   | not USA | 1.1.2.0cm  | 3      | 55    | 5.17 [1.08; 14.38]|            |
| Palestini 2008| not USA | 2.1.4.0cm  | 0      | 64    | 0.00 [0.00; 5.00]  |            |
| Giordano 2012| not USA | unknown    | 50     | 308   | 16.23 [12.30; 20.84]|            |
| RE model for subgroup |        |            | 92     | 1599  | 3.46 [1.20; 9.56]  |            |

| Laterality not specified |         |            |        |       |                    |            |
| Oda 2016    | not USA | <1.0cm     | 16     | 974   | 1.64 [0.94; 2.65]  |            |
| Chang 2015  | not USA | <1.0cm     | 2      | 713   | 0.39 [0.04; 1.17]  |            |
| Donnini 2015| not USA | <1.0cm     | 3      | 251   | 1.99 [0.05; 4.59]  |            |
| Konnini 2015| not USA | <1.0cm     | 0      | 69    | 0.00 [0.00; 5.21]  |            |
| Kim 2014    | not USA | <1.0cm     | 0      | 392   | 0.00 [0.00; 0.94]  |            |
| Lee 2010    | not USA | <1.0cm     | 0      | 383   | 0.00 [0.00; 0.58]  |            |
| Lee 2010    | not USA | <1.0cm     | 4      | 1390  | 0.29 [0.06; 0.74]  |            |
| Del Rio 2015| not USA | 1.1.2.0cm  | 0      | 165   | 0.00 [0.00; 3.45]  |            |
| Taraglia 2014| not USA | 1.1.2.0cm  | 10     | 63    | 25.40 [15.27; 37.94]|            |
| Cibb 2013   | not USA | 1.1.2.0cm  | 5      | 46    | 10.87 [3.62; 37.57]|            |
| Crocchi 2012| not USA | 1.1.2.0cm  | 3      | 120   | 1.67 [0.20; 5.99]  |            |
| Kwan 2015   | not USA | 2.1.4.0cm  | 2      | 53    | 3.77 [0.46; 12.98] |            |
| Ann 2014    | not USA | 2.1.4.0cm  | 8      | 70    | 11.43 [5.07; 21.28]|            |
| Harfi 2013  | not USA | 2.1.4.0cm  | 4      | 155   | 2.58 [0.71; 8.48]  |            |
| Speal 2008  | USA     | 2.1.4.0cm  | 2      | 81    | 2.47 [0.30; 8.64]  |            |
| Bone 2013   | not USA | unknown    | 4      | 61    | 6.56 [1.62; 15.90]|            |
| Raj 2010    | not USA | unknown    | 0      | 125   | 0.00 [0.00; 2.01]  |            |
| Shah 2000   | not USA | unknown    | 12     | 65    | 16.40 [5.82; 30.03]|            |
| Sim 1998    | not USA | unknown    | 4      | 141   | 2.84 [0.78; 7.10]  |            |
| RE model for subgroup |        |            | 88     | 6410  | 2.26 [1.62; 4.96]  |            |

RE model for all studies |        |            | 225    | 9266  | 2.38 [1.48; 3.81]  |            |

Note: Tumor size=calculated mean tumor size.

Abbreviations: CI=confidence interval; RE=random effects; FE=fixed effects.
Table 1. Study Characteristics and Key Question 5 Results: Permanent Recurrent Laryngeal Nerve Palsy From Surgery, Stratified by Type of Thyroidectomy

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Tumor Size</th>
<th>Events</th>
<th>Total</th>
<th>Event Rate per 100</th>
<th>[95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palestini 2008</td>
<td>not USA</td>
<td>&lt;1.0 cm</td>
<td>2</td>
<td>148</td>
<td>1.35 [ 0.18, 4.80 ]</td>
<td></td>
</tr>
<tr>
<td>Kwan 2015</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>1</td>
<td>51</td>
<td>1.96 [ 0.95, 10.45 ]</td>
<td></td>
</tr>
<tr>
<td>Ahn 2014</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>8</td>
<td>291</td>
<td>2.75 [ 1.19, 5.34 ]</td>
<td></td>
</tr>
<tr>
<td>Cenize 2014</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>3</td>
<td>399</td>
<td>0.77 [ 0.16, 3.23 ]</td>
<td></td>
</tr>
<tr>
<td>Tartaglia 2014</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>8</td>
<td>284</td>
<td>2.82 [ 1.22, 5.47 ]</td>
<td></td>
</tr>
<tr>
<td>Calo 2013</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>0</td>
<td>169</td>
<td>0.00 [ 0.00, 2.16 ]</td>
<td></td>
</tr>
<tr>
<td>Raffaelli 2012</td>
<td>not USA</td>
<td>1.1-2.0 cm</td>
<td>0</td>
<td>62</td>
<td>0.00 [ 0.00, 5.78 ]</td>
<td></td>
</tr>
<tr>
<td>Haril 2013</td>
<td>not USA</td>
<td>2.1-4.0 cm</td>
<td>2</td>
<td>91</td>
<td>2.20 [ 0.27, 7.71 ]</td>
<td></td>
</tr>
<tr>
<td>Sywak 2006</td>
<td>not USA</td>
<td>2.1-4.0 cm</td>
<td>4</td>
<td>391</td>
<td>1.02 [ 0.28, 4.60 ]</td>
<td></td>
</tr>
<tr>
<td>Chaplin 1999</td>
<td>not USA</td>
<td>2.1-4.0 cm</td>
<td>1</td>
<td>103</td>
<td>0.97 [ 0.02, 5.29 ]</td>
<td></td>
</tr>
<tr>
<td>Beute 2013</td>
<td>not USA</td>
<td>unknown</td>
<td>1</td>
<td>22</td>
<td>4.56 [ 0.12, 22.84 ]</td>
<td></td>
</tr>
<tr>
<td>Giordano 2012</td>
<td>not USA</td>
<td>unknown</td>
<td>4</td>
<td>394</td>
<td>1.02 [ 0.28, 3.58 ]</td>
<td></td>
</tr>
<tr>
<td>RE model for subgroup</td>
<td>34</td>
<td>2396</td>
<td>1.63 [ 1.11, 2.40 ]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Tumor size=calculated mean tumor size.

Abbreviations: CI=confidence interval; RE=random effects; FE=fixed effects.
Figure 5. Key Question 5 Results: Permanent Recurrent Laryngeal Nerve Palsy From Surgery, Stratified by Type of Lymph Node Dissection

Note: Tumor size=calculated mean tumor size.

Abbreviations: CI=confidence interval; RE=random effects; FE=fixed effects.
Table 1. Included Studies for Key Question 2: Test Performance Characteristics of Screening Tests for Detecting Malignant Thyroid Nodules in Adults

<table>
<thead>
<tr>
<th>Screening Method</th>
<th>Study, Year, Quality</th>
<th>Country, Recruitment Years</th>
<th>Study Design</th>
<th>N Screened</th>
<th>Mean Age (years)</th>
<th>Women (%)</th>
<th>Diagnostic Accuracy Cancer Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palpation</td>
<td>Suehiro, 2006&lt;sup&gt;49&lt;/sup&gt; Fair</td>
<td>Japan 1989-2005</td>
<td>Retrospective</td>
<td>46,433</td>
<td>49</td>
<td>45&lt;sup&gt;*&lt;/sup&gt;</td>
<td>X</td>
</tr>
<tr>
<td>Palpation</td>
<td>Brander, 1991&lt;sup&gt;50&lt;/sup&gt; Fair</td>
<td>Finland 1989-1990</td>
<td>Prospective</td>
<td>253</td>
<td>35</td>
<td>51</td>
<td>X</td>
</tr>
<tr>
<td>Palpation</td>
<td>Brander, 1989&lt;sup&gt;50&lt;/sup&gt; Fair</td>
<td>Finland 1988</td>
<td>Prospective</td>
<td>101</td>
<td>52</td>
<td>100</td>
<td>X</td>
</tr>
<tr>
<td>Palpation</td>
<td>Ishida, 1988&lt;sup&gt;41&lt;/sup&gt; Fair</td>
<td>Japan 1980-1986</td>
<td>Prospective</td>
<td>152,651</td>
<td>NR</td>
<td>100</td>
<td>X</td>
</tr>
<tr>
<td>US</td>
<td>Kim, 2010&lt;sup&gt;51&lt;/sup&gt; Fair</td>
<td>South Korea 2005-2007</td>
<td>Prospective</td>
<td>2,079</td>
<td>43</td>
<td>100</td>
<td>X</td>
</tr>
<tr>
<td>US</td>
<td>Kim, 2008&lt;sup&gt;36&lt;/sup&gt; Fair</td>
<td>South Korea 2004-2006</td>
<td>Retrospective</td>
<td>16,352</td>
<td>53</td>
<td>67</td>
<td>X</td>
</tr>
<tr>
<td>US</td>
<td>Lee, 2003&lt;sup&gt;51&lt;/sup&gt; Fair</td>
<td>South Korea 2003</td>
<td>Prospective</td>
<td>697</td>
<td>43</td>
<td>100</td>
<td>X</td>
</tr>
<tr>
<td>US</td>
<td>Chung, 2001&lt;sup&gt;50&lt;/sup&gt; Fair</td>
<td>South Korea 1997-1998</td>
<td>Prospective</td>
<td>1,401</td>
<td>47</td>
<td>100</td>
<td>X</td>
</tr>
<tr>
<td>Palpation + diagnostic followup&lt;sup&gt;†&lt;/sup&gt;</td>
<td>Ron, 1984&lt;sup&gt;52&lt;/sup&gt; Fair</td>
<td>Israel Years NR</td>
<td>Prospective</td>
<td>443</td>
<td>29</td>
<td>49</td>
<td>X</td>
</tr>
<tr>
<td>Palpation + diagnostic followup&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>Shimaoka, 1982&lt;sup&gt;53&lt;/sup&gt; Fair</td>
<td>USA 1977-1980</td>
<td>Prospective</td>
<td>1,500</td>
<td>39</td>
<td>64</td>
<td>X</td>
</tr>
</tbody>
</table>

<sup>*</sup> Percentage of exam visits that were women (calculated).
<sup>†</sup> Diagnostic followup consisted of technitium-<sup>99m</sup> thyroid scan and thyroid function tests.
<sup>‡</sup> Diagnostic followup consisted of iodine 123 thyroid scan, blood tests, and indirect laryngoscopy.

**Abbreviations:** KQ=key question; NR=not reported; X=reported; US=ultrasound.
<table>
<thead>
<tr>
<th>Study, Year, Quality</th>
<th>N Analyzed</th>
<th>≥1 characteristics:</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim, 2010* Fair</td>
<td>113 persons</td>
<td>Microcalcification</td>
<td>94.3% (55.0% - 136 persons</td>
<td>83.3% (71.5% - 91.7%)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Irregular shape</td>
<td>34.0% (21.5% - 48.3%)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taller-than-wide shape</td>
<td>88.7% (77.0% - 95.7%)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ill-defined or microlobulated margin</td>
<td>67.9% (53.7% - 80.1%)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marked hypochoegenicity</td>
<td>86.8% (74.7% - 94.5%)*</td>
<td></td>
</tr>
<tr>
<td>Kim, 2008* Fair</td>
<td>140 nodules</td>
<td>≥2 characteristics:</td>
<td>94.8% (86.6% - 140 persons</td>
<td>98.8% (93.4% - 99.8%)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Microcalcification</td>
<td>70.7% (57.3% - 81.9%)*</td>
<td>98.8% (93.4% - 99.8%)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taller-than-wide or irregular shape</td>
<td>55.2% (41.5% - 68.3%)*</td>
<td>89% (80.2% - 94.8%)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spiculated margin</td>
<td>48.3% (35-61.8)</td>
<td>97.6% (91.4-99.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marked hypochoegenicity</td>
<td>55.2% (41.5-68.3)</td>
<td>96.3% (89.7-99.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Solid</td>
<td>93.1% (83.3-98)</td>
<td>51.2% (39.9-62.4)</td>
</tr>
</tbody>
</table>

* Calculated confidence intervals.

Note: Both studies report accuracy only among patients who had at least one study-defined malignant ultrasound characteristic, providing no followup on the vast majority (n=18,188) of screened individuals who did not have these characteristics.

Abbreviations: KQ = key question; CI = confidence interval.
#### Table 3. Key Question 2 Results: Cancer Yield From Thyroid Cancer Screening

<table>
<thead>
<tr>
<th>Screening Method</th>
<th>Study, Year</th>
<th>Quality</th>
<th>N Analyzed</th>
<th>Cancer Yield per 1,000 Persons</th>
<th>Histology</th>
<th>Cancer Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palpation</td>
<td>Suehiro, 2006&lt;sup&gt;26&lt;/sup&gt; Fair</td>
<td>59</td>
<td>Fair</td>
<td>46,433</td>
<td>4.3</td>
<td>Papillary, follicular</td>
</tr>
<tr>
<td>Palpation</td>
<td>Brander, 1991&lt;sup&gt;35&lt;/sup&gt; Fair</td>
<td>253</td>
<td>Fair</td>
<td>0</td>
<td>NA</td>
<td>Based on FNA cytology</td>
</tr>
<tr>
<td>Palpation</td>
<td>Brander, 1989&lt;sup&gt;36&lt;/sup&gt; Fair</td>
<td>99</td>
<td>Fair</td>
<td>0</td>
<td>NA</td>
<td>Based on FNA cytology</td>
</tr>
<tr>
<td>Palpation</td>
<td>Ishida, 1988&lt;sup&gt;34&lt;/sup&gt; Fair</td>
<td>152,651</td>
<td>Fair</td>
<td>1.4</td>
<td>Papillary, follicular, and 1 medullary</td>
<td>Based on surgical pathology</td>
</tr>
<tr>
<td>US</td>
<td>Kim, 2010&lt;sup&gt;27&lt;/sup&gt; Fair</td>
<td>2,079</td>
<td>Fair</td>
<td>25.5</td>
<td>Papillary only</td>
<td>Based on FNA cytology</td>
</tr>
<tr>
<td>US</td>
<td>Kim, 2008&lt;sup&gt;35&lt;/sup&gt; Fair</td>
<td>16,352</td>
<td>Fair</td>
<td>9.2</td>
<td>Papillary only</td>
<td>Based on FNA cytology</td>
</tr>
<tr>
<td>US</td>
<td>Lee, 2003&lt;sup&gt;61&lt;/sup&gt; Fair</td>
<td>693</td>
<td>Fair</td>
<td>30.3</td>
<td>Papillary only</td>
<td>Based on surgical pathology</td>
</tr>
<tr>
<td>Palpation + diagnostic followup†</td>
<td>Chung, 2001&lt;sup&gt;60&lt;/sup&gt; Fair</td>
<td>1,397</td>
<td>Fair</td>
<td>26.5</td>
<td>Papillary and 1 insular</td>
<td>Malignant on FNA biopsy and confirmed with surgical pathology</td>
</tr>
<tr>
<td>Palpation + diagnostic followup*</td>
<td>Ron, 1984&lt;sup&gt;62&lt;/sup&gt; Fair</td>
<td>443</td>
<td>Fair</td>
<td>0</td>
<td>NA</td>
<td>Based on surgical pathology</td>
</tr>
<tr>
<td>Palpation + diagnostic followup†</td>
<td>Shimaoka, 1982&lt;sup&gt;27&lt;/sup&gt; Fair</td>
<td>1,500</td>
<td>Fair</td>
<td>11.3</td>
<td>NR</td>
<td>Based on surgical pathology</td>
</tr>
</tbody>
</table>

* Diagnostic followup consisted of thyroid scan with technetium-99m and thyroid function tests.  
† Diagnostic followup consisted of thyroid imaging by iodine 123 thyroid scan, blood tests, and indirect laryngoscopy.

**Abbreviations:** KQ=key question; US=ultrasonography; FNA=fine-needle aspiration; NA=not applicable; NR=not reported.
Table 4. Included Studies and Results for Key Question 3: Harms of Screening and Diagnostic Fine-Needle Aspiration

<table>
<thead>
<tr>
<th>Study, Year, Quality</th>
<th>Country, Years</th>
<th>Design</th>
<th>N</th>
<th>Mean Age (years)</th>
<th>Women (%)</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hobbs, 2014&lt;sup&gt;a&lt;/sup&gt;</td>
<td>USA 2010-2011</td>
<td>Retrospective</td>
<td>400</td>
<td>55</td>
<td>83</td>
<td>To determine the proportion of thyroid nodules undergoing ultrasound-guided FNA that do not meet Society of Radiologists in Ultrasound recommendations from 2005&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Persons undergoing FNA not meeting Society of Radiologists in Ultrasound recommendations: 96/400 (24.0%)</td>
</tr>
<tr>
<td>Abu-Yousef, 2011&lt;sup&gt;b&lt;/sup&gt;</td>
<td>USA 2006-2007</td>
<td>Retrospective</td>
<td>582&lt;sup&gt;c&lt;/sup&gt;</td>
<td>56</td>
<td>71</td>
<td>To determine whether there is a significantly increased incidence of bleeding complications from ultrasound-guided FNA of neck masses in patients on antithrombotic or anticoagulant therapy (compared to patients not on therapy)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Major complications (hospitalization or intervention required): 0/582 (0%) Post-procedural hematoma: 5/582&lt;sup&gt;‡&lt;/sup&gt; (0.9%)</td>
</tr>
<tr>
<td>Ito, 2005&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Japan 1990-2002</td>
<td>Retrospective</td>
<td>4,912</td>
<td>NR&lt;sup&gt;§&lt;/sup&gt;</td>
<td>NR&lt;sup&gt;§&lt;/sup&gt;</td>
<td>To investigate the relationship between needle tract implantation of papillary thyroid cancer and clinicopathological characteristics</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tumor implantation: 7/4,912 (0.14%)</td>
</tr>
</tbody>
</table>

<sup>a</sup> FNA is appropriate for nodules that have a maximum diameter of 1 cm or larger and have microcalcifications; nodules that are 1.5 cm or larger and are solid or have coarse calcifications; nodules that are 2 cm or larger and are mixed solid and cystic; and nodules with substantial growth since the prior ultrasound.

<sup>b</sup> n for thyroid masses only.

<sup>c</sup> Difference in incidence of hematomas between persons who were on antiplatelet or anticoagulant therapy versus persons not on therapy not statistically significant.

<sup>‡</sup> Data reported for 10 persons with outcomes: mean age 65 years and 90% female.

Abbreviations: FNA=fine-needle aspiration; KQ=key question; NR=not reported.
<table>
<thead>
<tr>
<th>Study, Year, Quality</th>
<th>Country, Years</th>
<th>Study Design, Histology</th>
<th>Group</th>
<th>N</th>
<th>Mean Age (years)</th>
<th>Women (%)</th>
<th>Length of Followup (mean range in years)</th>
<th>Thyroid Cancer-Specific Deaths (%)</th>
<th>Thyroid Cancer-Specific Survival (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davies, 2010 Fair</td>
<td>USA 1973-2005</td>
<td>Retrospective observational Papillary</td>
<td>IG</td>
<td>35,223</td>
<td>46</td>
<td>77</td>
<td>7.6 (0-32)</td>
<td>161/35,223 (0.46%)</td>
<td>20-year survival: 99% (93% to 100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>440</td>
<td>51</td>
<td>81</td>
<td>5.9 (0-31)</td>
<td>6/440 (1.4%)</td>
<td>20-year survival: 97% (96% to 100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*Recommended: 4/216 (1.9%)</td>
<td>10-year survival: *Recommended: 98.1% (95.9% to 100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*Not recommended: 1/165 (0.6%) (p=0.10)</td>
<td>*Recommended: 99.3% (97.8% to 100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oda, 2016 Ito, 2014 Ito, 2010 Ito, 2003 Fair</td>
<td>Japan 1993-2004</td>
<td>Prospective observational Papillary microcarcinomas</td>
<td>IG</td>
<td>1,055</td>
<td>52</td>
<td>91</td>
<td>6.3 (0.08-15.3)</td>
<td>2/1,055 (0.2%)</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>340</td>
<td>NR</td>
<td>92</td>
<td>6.2 (1.5-15.6)</td>
<td>0/340 (0%)</td>
<td>NR</td>
</tr>
<tr>
<td>Japan 2005-2013 Prospective observational Papillary microcarcinomas</td>
<td>IG</td>
<td>974</td>
<td>55†</td>
<td>88</td>
<td>3.9† (1.0-9.7)</td>
<td>0/974 (0%)</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>1179</td>
<td>57†</td>
<td>88</td>
<td>3.9† (1.0-9.7)</td>
<td>0/1179 (0%)</td>
<td>NR</td>
</tr>
</tbody>
</table>

* Subset of population (1988-2005) that had treatment recommendation as a variable and refers to SEER classification: Recommended=recommended to be treated; Not recommended (NR)=not recommended to be treated.
† Recurrence in the observation group is reported in this study.
‡ Median.

Abbreviations: CG=control or comparator group, no (immediate) surgical treatment; IG=intervention group, surgical treatment; KQ=key question.
<table>
<thead>
<tr>
<th>Author, Year, Country, Recruiting Years</th>
<th>Study Design</th>
<th>N</th>
<th>Mean Age (years)</th>
<th>Women (%)</th>
<th>Permanent Hypoparathyroidism</th>
<th>Permanent RLN Palsy</th>
<th>Other Serious Harms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oda, 2016†</td>
<td>Japan 2005-2013</td>
<td>Prospective observational</td>
<td>974*</td>
<td>56</td>
<td>88</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Chang, 2015††</td>
<td>South Korea 2002-2013</td>
<td>Retrospective observational</td>
<td>613</td>
<td>46</td>
<td>91</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Del Rio, 2015†</td>
<td>Italy 2005-2007</td>
<td>Prospective observational</td>
<td>105</td>
<td>59</td>
<td>82</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Donatini, 2015†</td>
<td>France 1991-2015</td>
<td>Prospective observational</td>
<td>880</td>
<td>48</td>
<td>81</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Kwan, 2015‡</td>
<td>Hong Kong 1995-2011</td>
<td>Retrospective observational</td>
<td>105</td>
<td>51</td>
<td>78</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Viola, 2015‡</td>
<td>Italy 2008-2010</td>
<td>RCT</td>
<td>181</td>
<td>45</td>
<td>75</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Ahn, 2014‡‡</td>
<td>South Korea 2000-2007</td>
<td>Retrospective observational</td>
<td>361</td>
<td>48</td>
<td>85</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Conzo, 2014‡‡</td>
<td>Italy 1998-2005</td>
<td>Retrospective observational</td>
<td>752</td>
<td>45</td>
<td>80</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Francis, 2014‡‡‡</td>
<td>USA 1991-2009</td>
<td>Retrospective observational</td>
<td>5,670</td>
<td>74</td>
<td>69</td>
<td>NR</td>
<td>X</td>
</tr>
<tr>
<td>Kim, 2014‡‡‡</td>
<td>South Korea 2011-2012</td>
<td>Retrospective observational</td>
<td>515</td>
<td>46</td>
<td>82</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Tartaglia, 2014‡‡</td>
<td>Italy 2000-2010</td>
<td>Retrospective observational</td>
<td>347</td>
<td>48</td>
<td>78</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Boute, 2013‡‡‡</td>
<td>France 1998-2009</td>
<td>Retrospective observational</td>
<td>83</td>
<td>51</td>
<td>55</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Calo, 2013‡‡‡</td>
<td>Italy 2002-2008</td>
<td>Retrospective observational</td>
<td>215</td>
<td>51</td>
<td>81</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Hartl, 2013‡‡‡</td>
<td>France 1995-2010</td>
<td>Retrospective observational</td>
<td>246</td>
<td>46</td>
<td>78</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Caliskan, 2012‡‡</td>
<td>South Korea 2000-2005</td>
<td>Retrospective observational</td>
<td>842</td>
<td>46</td>
<td>91</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Cirocchi, 2012‡‡</td>
<td>Italy 2009-2010</td>
<td>CCT</td>
<td>321</td>
<td>NR</td>
<td>55</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Giordano, 2012‡‡</td>
<td>Italy 1997-2010</td>
<td>Retrospective observational</td>
<td>1,087</td>
<td>NR</td>
<td>NR</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Hyun, 2012‡‡</td>
<td>South Korea 2002-2009</td>
<td>Retrospective observational</td>
<td>152</td>
<td>47</td>
<td>81</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Raffaelli, 2012‡‡</td>
<td>Italy 2008-2010</td>
<td>CCT</td>
<td>186</td>
<td>43</td>
<td>80</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Kim, 2011‡‡‡</td>
<td>South Korea 2008-2009</td>
<td>Retrospective observational</td>
<td>302</td>
<td>52</td>
<td>75</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lee, 2010‡‡</td>
<td>South Korea 2006-2007</td>
<td>Retrospective observational</td>
<td>2,636</td>
<td>47</td>
<td>86</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Table 6. Included Studies for Key Question 5: Harms of Surgical Treatment of Screen-Detected Thyroid Cancer

<table>
<thead>
<tr>
<th>Author, Year, Quality</th>
<th>Country, Recruiting Years</th>
<th>Study Design</th>
<th>N</th>
<th>Mean Age (years)</th>
<th>Women (%)</th>
<th>Permanent Hypoparathyroidism</th>
<th>Permanent RLN Palsy</th>
<th>Other Serious Harms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raj, 2010 Fair</td>
<td>Australia 1993-2008</td>
<td>Retrospective observational</td>
<td>125</td>
<td>48</td>
<td>76</td>
<td>X</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>So, 2010 Fair</td>
<td>South Korea 2005-2009</td>
<td>Retrospective observational</td>
<td>551</td>
<td>50</td>
<td>80</td>
<td>X</td>
<td>X</td>
<td>Hematoma requiring reoperation</td>
</tr>
<tr>
<td>Moo, 2009 Fair</td>
<td>USA 2003-2009</td>
<td>Prospective observational</td>
<td>116</td>
<td>47</td>
<td>76</td>
<td>X</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Zerey, 2009 Fair</td>
<td>USA 1999-2003</td>
<td>Retrospective observational</td>
<td>13,854</td>
<td>48</td>
<td>76</td>
<td>NR</td>
<td>NR</td>
<td>Wound infection, tracheal or laryngeal perforation, esophageal perforation, adverse cardiopulmonary event (i.e., MI, CVA, PE), pneumonia, renal failure, same-stay mortality</td>
</tr>
<tr>
<td>Palestini, 2008 Fair</td>
<td>Italy 2000-2006</td>
<td>Retrospective observational</td>
<td>305</td>
<td>47</td>
<td>73</td>
<td>X</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Son, 2008 Fair</td>
<td>South Korea 2003-2005</td>
<td>Retrospective observational</td>
<td>114</td>
<td>48</td>
<td>75</td>
<td>X</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Spear, 2008 Fair</td>
<td>USA 1996-2000</td>
<td>Retrospective observational</td>
<td>82</td>
<td>42†</td>
<td>70</td>
<td>X</td>
<td>X</td>
<td>Phrenic nerve injury</td>
</tr>
<tr>
<td>Yassa, 2007 Fair</td>
<td>USA 1995-2004</td>
<td>Retrospective observational</td>
<td>2,587</td>
<td>50</td>
<td>88</td>
<td>X</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Shah, 2006 Fair</td>
<td>Canada 2002-2003</td>
<td>Prospective observational</td>
<td>76</td>
<td>46</td>
<td>82</td>
<td>X</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Sywak, 2006 Fair</td>
<td>Australia 1995-2005</td>
<td>Retrospective observational</td>
<td>447</td>
<td>42</td>
<td>72</td>
<td>X</td>
<td>X</td>
<td>Wound infection</td>
</tr>
<tr>
<td>Holzer, 2000 Fair</td>
<td>Germany 1996</td>
<td>Prospective observational</td>
<td>2,376</td>
<td>51†</td>
<td>77</td>
<td>NR</td>
<td>NR</td>
<td>Wound infection</td>
</tr>
<tr>
<td>Hundahl, 2000 Fair</td>
<td>USA 1996</td>
<td>Prospective observational</td>
<td>5,584</td>
<td>45†</td>
<td>75</td>
<td>NR</td>
<td>NR</td>
<td>Wound infection, airway problem, postoperative death</td>
</tr>
<tr>
<td>Chaplin, 1999 Fair</td>
<td>Australia NR</td>
<td>Retrospective observational</td>
<td>175</td>
<td>45†</td>
<td>70</td>
<td>X</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Sim, 1998 Fair</td>
<td>Singapore 1988-1994</td>
<td>Retrospective observational</td>
<td>149</td>
<td>45†</td>
<td>75</td>
<td>X</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Shindo, 1995 Fair</td>
<td>USA 1989-1994</td>
<td>Retrospective observational</td>
<td>181</td>
<td>NR</td>
<td>90</td>
<td>X</td>
<td>X</td>
<td>NR</td>
</tr>
</tbody>
</table>

* N analyzed for persons evaluated for surgical harms; total study population was 2,153.
† Median or calculated based on median.
‡ 2002-2004 for surgical harms data.
§ Not included in meta-analysis because study does not report permanent hypoparathyroidism or RLN palsy, or it does not clearly define outcomes as permanent or temporary.

**Abbreviations:** CCT=controlled clinical trial; CVA=cerebrovascular accident; KQ=key question; NR=not reported; RCT=randomized, controlled trial; RLN=recurrent laryngeal nerve; MI=myocardial infarction; PE=pulmonary embolus; X=reported.
Table 7. Included Studies for Key Question 5: Harms of RAI Treatment of Screen-Detected Thyroid Cancer

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Quality</th>
<th>Country, Recruiting Years</th>
<th>Study Design</th>
<th>N</th>
<th>Mean Age (Years)</th>
<th>Women (%)</th>
<th>Second Primary Malignancy</th>
<th>Salivary Gland Harms</th>
<th>Other Serious Harms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hakala, 2015</td>
<td>Fair</td>
<td>Finland 1981-2002</td>
<td>Retrospective observational</td>
<td>910</td>
<td>49</td>
<td>82</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Khang, 2015</td>
<td>Fair</td>
<td>South Korea 1976-2010</td>
<td>Retrospective observational</td>
<td>2,468</td>
<td>46</td>
<td>84</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Lin, 2015</td>
<td>Fair</td>
<td>Taiwan 2000-2008</td>
<td>Prospective observational</td>
<td>10,361</td>
<td>46</td>
<td>100</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Seo, 2015</td>
<td>Fair</td>
<td>South Korea 2008-2013</td>
<td>Retrospective observational</td>
<td>211,360</td>
<td>48</td>
<td>82</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Lang, 2012</td>
<td>Fair</td>
<td>Hong Kong 1971-2009</td>
<td>Retrospective observational</td>
<td>895</td>
<td>47</td>
<td>81</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Iyer, 2011</td>
<td>Fair</td>
<td>USA 1973-2006</td>
<td>Retrospective observational</td>
<td>37,176</td>
<td>NR</td>
<td>NR</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Brown, 2008</td>
<td>Fair</td>
<td>USA 1973-2002</td>
<td>Retrospective observational</td>
<td>28,286</td>
<td>42†</td>
<td>76</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Ronckers, 2005</td>
<td>Fair</td>
<td>USA 1973-2000</td>
<td>Prospective observational</td>
<td>29,456</td>
<td>43†</td>
<td>75</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Ryu, 2015</td>
<td>Fair</td>
<td>South Korea 2010</td>
<td>Prospective observational</td>
<td>160</td>
<td>49</td>
<td>81</td>
<td>NR</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Jeong, 2013</td>
<td>Fair</td>
<td>South Korea 2003-2006</td>
<td>Prospective observational</td>
<td>213</td>
<td>47</td>
<td>91</td>
<td>NR</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Grewal, 2009</td>
<td>Fair</td>
<td>USA 1995-2003</td>
<td>Retrospective observational</td>
<td>262</td>
<td>45</td>
<td>66</td>
<td>NR</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Ish-Shalom, 2008</td>
<td>Fair</td>
<td>Israel NR 1995-2003</td>
<td>Prospective observational</td>
<td>40</td>
<td>48</td>
<td>100</td>
<td>NR</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Hyer, 2007</td>
<td>Fair</td>
<td>UK NR 1990-1995</td>
<td>Prospective observational</td>
<td>76</td>
<td>51</td>
<td>75</td>
<td>NR</td>
<td>X</td>
<td>NR</td>
</tr>
<tr>
<td>Solans, 2001</td>
<td>Fair</td>
<td>Spain 1990-1995</td>
<td>Prospective observational</td>
<td>79</td>
<td>46</td>
<td>86</td>
<td>NR</td>
<td>X*</td>
<td>NR</td>
</tr>
<tr>
<td>Wu, 2015</td>
<td>Fair</td>
<td>USA 1999-2008</td>
<td>Retrospective observational</td>
<td>18,850</td>
<td>47</td>
<td>100</td>
<td>NR</td>
<td>Reproductive outcomes†</td>
<td></td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>Good</td>
<td>Taiwan 1997-2008</td>
<td>Retrospective observational</td>
<td>8,946</td>
<td>44</td>
<td>81</td>
<td>NR</td>
<td>Hyperparathyroidism</td>
<td></td>
</tr>
</tbody>
</table>

* 28,286 is the total number included in the RAI analysis; total n for the study is 30,278.
† Median.
‡ Study also reported dry eyes.
§ Total cohort included 25,333 persons; the reproductive outcomes subset included 18,850 women.
‖ Birthrate and median time to first delivery.

**Abbreviations:** KQ=key question; NR=not reported; X=reported.
Table 8. Results for Key Question 5: Harms of Surgical Treatment of Screen-Detected Thyroid Cancer

<table>
<thead>
<tr>
<th>Study, Year, Quality</th>
<th>Treatment Arm</th>
<th>LN Indication</th>
<th>LN Side</th>
<th>Histology</th>
<th>Mean Tumor Size Category (cm)</th>
<th>Lymph Node Metastases (%)</th>
<th>N</th>
<th>Hypoparathyroidism</th>
<th>RLN Palsy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oda, 2016* Fair</td>
<td>All</td>
<td>NR</td>
<td>NR</td>
<td>Micropapillary</td>
<td>0-1.0</td>
<td>0</td>
<td>974</td>
<td>16 (2)</td>
<td>2 (0.2)</td>
</tr>
<tr>
<td>Chang, 2015* Fair</td>
<td>All</td>
<td>NR</td>
<td>Ipsilateral and bilateral</td>
<td>Micropapillary</td>
<td>0-1.0</td>
<td>39</td>
<td>613</td>
<td>2 (0)</td>
<td>2 (0)</td>
</tr>
<tr>
<td>Del Rio, 2015* Fair</td>
<td>All</td>
<td>NR</td>
<td>NR</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>105</td>
<td>0 (0)</td>
<td>NR</td>
</tr>
<tr>
<td>Donatini, 2015* Fair</td>
<td>Less than TT</td>
<td>NR</td>
<td>NR</td>
<td>Micropapillary</td>
<td>0-1.0</td>
<td>1</td>
<td>69</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Chang, 2015* Fair</td>
<td>TT</td>
<td>NR</td>
<td>NR</td>
<td>Micropapillary</td>
<td>0-1.0</td>
<td>14</td>
<td>251</td>
<td>5 (2)</td>
<td>3 (1)</td>
</tr>
<tr>
<td>Chang, 2015* Fair</td>
<td>TT+ LND</td>
<td>Prophylactic and bilateral</td>
<td>Papillary</td>
<td>1.1-2.0*</td>
<td>NR</td>
<td>51</td>
<td>2 (2)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Kwan, 2015* Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Papillary</td>
<td>1.1-2.0*</td>
<td>NR</td>
<td>78</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Kwan, 2015* Fair</td>
<td>TT+ LND</td>
<td>Prophylactic and bilateral</td>
<td>Papillary</td>
<td>2.1-4.0*</td>
<td>NR</td>
<td>53</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Viola, 2015* Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>88</td>
<td>7 (8)</td>
<td>NR</td>
</tr>
<tr>
<td>Viola, 2015* Fair</td>
<td>TT+ LND</td>
<td>Prophylactic and bilateral</td>
<td>Papillary</td>
<td>1.1-2.0*</td>
<td>NR</td>
<td>93</td>
<td>18 (19)</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Ahn, 2014* Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>291</td>
<td>13 (4)</td>
<td>8 (3)</td>
</tr>
<tr>
<td>Conzo, 2014* Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Differentiated</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>362</td>
<td>13 (4)</td>
<td>6 (2)</td>
</tr>
<tr>
<td>Conzo, 2014* Fair</td>
<td>TT+ LND</td>
<td>Prophylactic and bilateral</td>
<td>Differentiated</td>
<td>1.1-2.0*</td>
<td>NR</td>
<td>136</td>
<td>13 (4)</td>
<td>6 (2)</td>
<td></td>
</tr>
<tr>
<td>Francis, 2014* Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Mixed (all)</td>
<td>Unable to determine</td>
<td>NR</td>
<td>5,670*</td>
<td>8 (8)</td>
<td>NR</td>
</tr>
<tr>
<td>Francis, 2014* Fair</td>
<td>TT+ LND</td>
<td>Prophylactic and therapeutic and bilateral</td>
<td>Papillary</td>
<td>0-1.0</td>
<td>NR</td>
<td>42</td>
<td>392</td>
<td>0 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Tartaglia, 2014* Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>284</td>
<td>27 (10)</td>
<td>8 (3)</td>
</tr>
<tr>
<td>Tartaglia, 2014* Fair</td>
<td>TT+ LND</td>
<td>NR</td>
<td>NR</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>63</td>
<td>16 (25)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Boute, 2013* Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Differentiated</td>
<td>Unable to determine</td>
<td>NR</td>
<td>44</td>
<td>2 (9)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Boute, 2013* Fair</td>
<td>TT+ LND</td>
<td>NR</td>
<td>NR</td>
<td>Differentiated</td>
<td>Unable to determine</td>
<td>NR</td>
<td>44</td>
<td>4 (7)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Caliskan, 2012* Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Differentiated</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>91</td>
<td>6 (7)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Caliskan, 2012* Fair</td>
<td>TT+ LND</td>
<td>Prophylactic and bilateral</td>
<td>Papillary</td>
<td>2.1-4.0*</td>
<td>NR</td>
<td>52</td>
<td>155</td>
<td>4 (3)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Cirocchi, 2012* Fair</td>
<td>TT</td>
<td>NR</td>
<td>NR</td>
<td>Differentiated</td>
<td>1.1-2.0*</td>
<td>NR</td>
<td>120</td>
<td>2 (2)</td>
<td>2 (2)</td>
</tr>
</tbody>
</table>

Screening for Thyroid Cancer

Kaiser Permanente Research Affiliates EPC
Table 8. Results for Key Question 5: Harms of Surgical Treatment of Screen-Detected Thyroid Cancer

<table>
<thead>
<tr>
<th>Study, Year, Quality</th>
<th>Treatment Arm</th>
<th>LN Indication</th>
<th>LN Side</th>
<th>Histology</th>
<th>Mean Tumor Size Category (cm)</th>
<th>Lymph Node Metastases (%)</th>
<th>N</th>
<th>N (%) Hypoparathyroidism</th>
<th>N (%) RLN Palsy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giordano, 2012 Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Papillary</td>
<td>Unable to determine</td>
<td>NR</td>
<td>394</td>
<td>25 (6)</td>
<td>4 (1)</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Ipsilateral</td>
<td>Papillary</td>
<td>Unable to determine</td>
<td>NR</td>
<td>385</td>
<td>27 (7)</td>
<td>2 (1)</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Bilateral</td>
<td>Papillary</td>
<td>Unable to determine</td>
<td>NR</td>
<td>308</td>
<td>50 (16)</td>
<td>7 (2)</td>
</tr>
<tr>
<td>Hyun, 2012 Fair</td>
<td>Less than TT</td>
<td>NA</td>
<td>NA</td>
<td>Micropapillary</td>
<td>0-1.0</td>
<td>NR</td>
<td>87</td>
<td>0 (0)†</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Less than TT+ LND</td>
<td>Prophylactic</td>
<td>Ipsilateral</td>
<td>Micropapillary</td>
<td>0-1.0</td>
<td>29</td>
<td>65</td>
<td>0 (0)†</td>
<td>NR</td>
</tr>
<tr>
<td>Raffaelli, 2012 Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>10</td>
<td>62</td>
<td>0 (0)†</td>
<td>0 (0)†</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Ipsilateral</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>29</td>
<td>62</td>
<td>0 (0)†</td>
<td>1 (2)†</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Bilateral</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>42</td>
<td>62</td>
<td>1 (2)†</td>
<td>0 (0)†</td>
</tr>
<tr>
<td>Kim, 2011 Fair</td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Ipsilateral</td>
<td>Micropapillary</td>
<td>0-1.0</td>
<td>NR</td>
<td>138</td>
<td>4 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Lee, 2010 Fair</td>
<td>Less than TT+ LND</td>
<td>Prophylactic and therapeutic</td>
<td>NR</td>
<td>Mixed (all)</td>
<td>0-1.0</td>
<td>NR</td>
<td>636</td>
<td>0 (0)</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Prophylactic and therapeutic</td>
<td>NR</td>
<td>Mixed (all)</td>
<td>0-1.0</td>
<td>NR</td>
<td>1,390</td>
<td>4 (0)</td>
<td>3 (0)</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Therapeutic</td>
<td>Ipsilateral</td>
<td>Mixed (all)</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>513</td>
<td>3 (1)</td>
<td>1 (0)</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Therapeutic</td>
<td>Bilateral</td>
<td>Mixed (all)</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>97</td>
<td>1 (1)</td>
<td>NR</td>
</tr>
<tr>
<td>Raj, 2010 Fair</td>
<td>All</td>
<td>NR</td>
<td>NR</td>
<td>Mixed (all)</td>
<td>Unable to determine</td>
<td>1</td>
<td>125</td>
<td>0 (0)†</td>
<td>1 (1)†</td>
</tr>
<tr>
<td>So, 2010 Fair</td>
<td>All</td>
<td>Prophylactic</td>
<td>Bilateral</td>
<td>Micropapillary</td>
<td>0-1.0</td>
<td>NR</td>
<td>551</td>
<td>6 (1)</td>
<td>7 (1)</td>
</tr>
<tr>
<td>Moo, 2009 Fair</td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Ipsilateral</td>
<td>Papillary</td>
<td>Unable to determine</td>
<td>NR</td>
<td>12</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Bilateral</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>45</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Zerey, 2009 Fair</td>
<td>Less than TT</td>
<td>NA</td>
<td>NA</td>
<td>NR</td>
<td>Unable to determine</td>
<td>NR</td>
<td>4,238</td>
<td>NR†</td>
<td>NR†</td>
</tr>
<tr>
<td></td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>NR</td>
<td>Unable to determine</td>
<td>NR</td>
<td>9,616</td>
<td>NR†</td>
<td>NR†</td>
</tr>
<tr>
<td>Palestini, 2008 Fair</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Papillary</td>
<td>0-1.0</td>
<td>NR</td>
<td>148</td>
<td>4 (3)</td>
<td>2 (1)†</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Ipsilateral</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>22</td>
<td>93</td>
<td>0 (0)</td>
<td>0 (0)†</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Bilateral</td>
<td>Papillary</td>
<td>2.1-4.0</td>
<td>72</td>
<td>64</td>
<td>0 (0)</td>
<td>0 (0)†</td>
</tr>
<tr>
<td>Son, 2008 Fair</td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Ipsilateral</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>56</td>
<td>1 (2)</td>
<td>0 (0)</td>
</tr>
<tr>
<td></td>
<td>TT+ LND</td>
<td>Prophylactic</td>
<td>Bilateral</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>58</td>
<td>3 (5)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Spear, 2008 Fair</td>
<td>All</td>
<td>NR</td>
<td>NR</td>
<td>Mixed (all)</td>
<td>2.1-4.0</td>
<td>13</td>
<td>81</td>
<td>2 (2)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Yassa, 2007 Fair</td>
<td>Less than TT</td>
<td>NA</td>
<td>NA</td>
<td>NR</td>
<td>2.1-4.0</td>
<td>NR</td>
<td>296</td>
<td>2 (1)†</td>
<td>1 (0)†</td>
</tr>
</tbody>
</table>
Table 8. Results for Key Question 5: Harms of Surgical Treatment of Screen-Detected Thyroid Cancer

<table>
<thead>
<tr>
<th>Study, Year, Quality</th>
<th>Treatment Arm</th>
<th>LN Indication</th>
<th>LN Side</th>
<th>Histology</th>
<th>Mean Tumor Size Category (cm)</th>
<th>Lymph Node Metastases (%)</th>
<th>N</th>
<th>N (%) Hypoparathyroidism</th>
<th>N (%) RLN Palsy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shah, 2006*</td>
<td>All</td>
<td>NR</td>
<td>Ipsilateral and bilateral</td>
<td>Mixed (all)</td>
<td>Unable to determine</td>
<td>9</td>
<td>65</td>
<td>12 (18)*</td>
<td>NR</td>
</tr>
<tr>
<td>Sywak, 2006</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Papillary</td>
<td>2.1-4.0</td>
<td>NR</td>
<td>391</td>
<td>2 (1)</td>
<td>4 (1)*</td>
</tr>
<tr>
<td></td>
<td>TT+LND</td>
<td>Prophylactic</td>
<td>Ipsilateral</td>
<td>Papillary</td>
<td>1.1-2.0</td>
<td>NR</td>
<td>56</td>
<td>1 (2)</td>
<td>0 (0)*</td>
</tr>
<tr>
<td>Holzer, 2000</td>
<td>All</td>
<td>NR</td>
<td>NR</td>
<td>Differentiated</td>
<td>2.1-4.0*</td>
<td>NR</td>
<td>2,376</td>
<td>NR†</td>
<td>NR†</td>
</tr>
<tr>
<td>Hundahl, 2000</td>
<td>Less than TT (lobectomy)</td>
<td>NA</td>
<td>NA</td>
<td>Mixed (all)</td>
<td>1.1-2.0*</td>
<td>NR</td>
<td>903</td>
<td>NR†</td>
<td>NR†</td>
</tr>
<tr>
<td></td>
<td>Less than TT (near-total)</td>
<td>NA</td>
<td>NA</td>
<td>Mixed (all)</td>
<td>1.1-2.0*</td>
<td>NR</td>
<td>840</td>
<td>NR†</td>
<td>NR†</td>
</tr>
<tr>
<td></td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Mixed (all)</td>
<td>1.1-2.0*</td>
<td>NR</td>
<td>1,928</td>
<td>NR†</td>
<td>NR†</td>
</tr>
<tr>
<td></td>
<td>TT+LND</td>
<td>NR</td>
<td>NR</td>
<td>Mixed (all)</td>
<td>1.1-2.0*</td>
<td>NR</td>
<td>1,464</td>
<td>NR†</td>
<td>NR†</td>
</tr>
<tr>
<td>Chaplin, 1999</td>
<td>TT</td>
<td>NA</td>
<td>NA</td>
<td>Differentiated</td>
<td>2.1-4.0*</td>
<td>NR</td>
<td>103</td>
<td>2 (2)*</td>
<td>1 (1)*</td>
</tr>
<tr>
<td>Sim, 1998</td>
<td>All</td>
<td>NR</td>
<td>NR</td>
<td>Mixed (all)</td>
<td>Unable to determine</td>
<td>NR</td>
<td>141</td>
<td>4 (3)</td>
<td>4 (3)</td>
</tr>
<tr>
<td>Shindo, 1995</td>
<td>All</td>
<td>NA</td>
<td>NA</td>
<td>NR</td>
<td>Unable to determine</td>
<td>NR</td>
<td>156</td>
<td>8 (5)*</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>NA</td>
<td>NA</td>
<td>NR</td>
<td>Unable to determine</td>
<td>NR</td>
<td>326</td>
<td>NR</td>
<td>3 (1)*</td>
</tr>
</tbody>
</table>

* N analyzed for persons evaluated for surgical harms; total study population was 2,153.
† Followup not defined.
‡ Calculated.
§ Followup ≤6 months.
¶ Not included in meta-analysis because study does not report permanent hypoparathyroidism or RLN palsy, or it does not clearly define outcomes as permanent or temporary.

Note: Studies with “All” for treatment arm indicate mixed surgery types ranging from less than total thyroidectomy to total thyroidectomy with or without lymph node dissection, and outcomes were not reported separately by surgery type.

Abbreviations: KQ=key question; LND=lymph node dissection; NA=not applicable; NR=not reported; RLN=recurrent laryngeal nerve; TT=total thyroidectomy.
Table 9. Results for Key Question 5: Harms of RAI Treatment of Screen-Detected Thyroid Cancer

| Author, Year, Quality | Country, Recruiting Years | N    | Radiation Dose (mean or median) | Followup Duration (years) | Permanent Outcomes
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Second Primary Malignancy diagnosed at least 12 months after thyroid cancer diagnosis:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N outcomes/N population: incidence per 10,000 person-years; excess risk per 10,000 person-years compared with controls without thyroid cancer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>All:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>109/910; 77.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stratified by RAI dose:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;3.7 GBq RAI: 27/214; 93.8; 25.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤3.7 GBq RAI: 56/526; 67.4; -4.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: 26/170; 89.1; 29.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p=NR p=NR p=NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rate ratio for second primary malignancy at any site vs. controls† without thyroid cancer:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;3.7 GBq RAI: 1.37 (95% CI, 0.90 to 2.09)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤3.7 GBq RAI: 0.94 (95% CI, 0.70 to 1.25)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: 1.49 (95% CI, 0.96 to 2.30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p=NR p=NR p=NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nonthyroid second primary malignancy diagnosed ≥12 months after thyroid cancer diagnosis or RAI treatment:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>All:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61/2468; 27.9†</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stratified by RAI dose:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥37 GBq RAI: 11/69; 131.4; 101.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.3–36.9 GBq RAI: 2/44; 54.6; 24.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.56–22.2 GBq RAI: 6/302; 19.0; -11.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.1–5.55 GBq RAI: 18/981; 23.7; -6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: 24/1072; 30.9; 0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p=NR p=NR p=NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adjusted odds ratio‡ for second primary malignancy at any site:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥37 GBq RAI: 5.54 (95% CI, 2.64 to 11.63)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.3–36.9 GBq RAI: 2.04 (95% CI, 0.48 to 8.70)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.56–22.2 GBq RAI: 0.67 (95% CI, 0.27 to 1.66)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.1–5–55 GBq RAI: 0.87 (95% CI, 0.47 to 1.62)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: reference p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Breast cancer diagnosed after thyroid cancer diagnosis:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>All:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>129/10,361; 18.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stratified by RAI dose:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;4.44 GBq RAI: 30/2848; 15.8; 2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤4.44 GBq RAI: 61/4221; 21.0; 7.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: 38/3,292; 17.7; 4.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p=NR p=NR p=NR</td>
</tr>
</tbody>
</table>

Hakala, 201512
Fair
Finland
1981-2002
910* Median: 3.7 GBq (100 mCi)
Mean: 5.3 GBq (143.2 mCi) Mean: 16.2

Khan, 201513
Fair
South Korea
1976-2010
2468 Mean: 5.1 GBq (137.8 mCi) Median: 7.0

Lin, 201514
Fair
Taiwan
2000-2008
10,361 NR Median: 6.5

Screening for Thyroid Cancer

Kaiser Permanente Research Affiliates EPC
Table 9. Results for Key Question 5: Harms of RAI Treatment of Screen-Detected Thyroid Cancer

<table>
<thead>
<tr>
<th>Author, Year, Quality</th>
<th>Country, Recruiting Years</th>
<th>N</th>
<th>Radiation Dose (mean or median)</th>
<th>Followup Duration (years)</th>
<th>Permanent Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Radiosurgery Dose</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Followup Duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Incidence</td>
<td></td>
<td>Excess risk</td>
</tr>
</tbody>
</table>

**Seo, 2015**

- **Country, Recruiting Years:** South Korea 2008-2013
- **N:** 211,360
- **Radiation Dose:** Mean: 3.7 GBq (100 mCi)
- **Followup Duration:** Median: 2.4 years

Leukemia diagnosed after thyroid cancer surgery or RAI treatment:

<table>
<thead>
<tr>
<th>Radiation Dose</th>
<th>Incidence</th>
<th>Excess risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;4.44 GBq RAI</td>
<td>0.90 (95% CI, 0.56 to 1.46)</td>
<td></td>
</tr>
<tr>
<td>≤4.44 GBq RAI</td>
<td>1.18 (95% CI, 0.79 to 1.77)</td>
<td></td>
</tr>
<tr>
<td>No RAI</td>
<td>reference</td>
<td></td>
</tr>
</tbody>
</table>

Adjusted hazard ratio for breast cancer diagnosis:

- >4.44 GBq RAI: 0.90 (95% CI, 0.56 to 1.46)
- ≤4.44 GBq RAI: 1.18 (95% CI, 0.79 to 1.77)
- No RAI: reference

**Lang, 2012**

- **Country, Recruiting Years:** Hong Kong 1971-2009
- **N:** 895
- **Radiation Dose:** 3 GBq (80 mCi)
- **Followup Duration:** Median: 7.8 years

Nonthyroid second primary malignancy diagnosed ≥12 months after thyroid cancer diagnosis:

<table>
<thead>
<tr>
<th>Incidence</th>
<th>Excess risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>All: 64/895</td>
<td>61.5†</td>
</tr>
</tbody>
</table>
| Stratified by RAI dose:
  - >5.5 GBq RAI: 15/23,356 2.1
  - 3.7–5.5 GBq RAI: 21/28,441 3.0
  - 1.1–3.7 GBq RAI: 6/28,397 0.9
  - ≤1.1 GBq RAI: 4/23,547 0.6
  - No RAI: 26/107,619 1.0

Adjusted hazard ratio for leukemia diagnosis:

- >5.5 GBq RAI: 2.08 (95% CI, 1.09 to 3.94)
- 3.7–5.5 GBq RAI: 3.09 (95% CI, 1.74 to 5.51)
- 1.1–3.7 GBq RAI: 0.88 (95% CI, 0.36 to 2.14)
- ≤1.1 GBq RAI: 0.62 (95% CI, 0.22 to 1.77)
- No RAI: reference

**Iyer, 2011**

- **Country, Recruiting Years:** USA 1973-2006
- **N:** 37,176
- **Radiation Dose:** NR
- **Followup Duration:** Mean: 11 years

Second primary malignancy diagnosed ≥6 months after thyroid cancer diagnosis:

<table>
<thead>
<tr>
<th>Incidence</th>
<th>Excess risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>All: 3,223/37,176</td>
<td>67.0†</td>
</tr>
</tbody>
</table>
| Stratified by RAI use:
  - RAI: NR
  - No RAI: NR
<table>
<thead>
<tr>
<th>Author, Year, Quality</th>
<th>Country, Recruiting Years</th>
<th>N</th>
<th>Radiation Dose (mean or median)</th>
<th>Followup Duration (years)</th>
<th>Permanent Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown, 2008</td>
<td>USA 1973-2002</td>
<td>28,286</td>
<td>NR</td>
<td>Mean: 10</td>
<td>Nonthyroid second primary malignancy diagnosed ≥2 months after thyroid cancer diagnosis:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N Incidence Excess risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>All: 2,191/28,286 74.9†</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stratified by RAI:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RAI: 618/10,257 75.8 13.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: 1,573/18,029 74.6 3.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p=NR p=NR p=0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Standardized incidence ratio †† for second primary malignancy at any site vs. the general population:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RAI: 1.21 (95% CI, 1.12 to 1.31)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: 1.05 (95% CI, 1.00 to 1.10)</td>
</tr>
<tr>
<td>Ronckers, 2005</td>
<td>USA 1973-2000</td>
<td>29,456</td>
<td>NR</td>
<td>Median: 7.9</td>
<td>Second primary malignancy diagnosed ≥2 months after thyroid cancer diagnosis:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N Incidence Excess risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>All: 2,214/29,456 78.9† 7.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stratified by RAI from 1988–2000 only (N=17,055):</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RAI: 236/6,745 NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: 394/8,951 NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p&lt;0.05 p=NR p=NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Standardized incidence ratio ‡‡ for second primary malignancy at any site vs. the general population:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RAI: 1.14 (95% CI, NR)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: 1.19 (95% CI, NR but excludes 1)</td>
</tr>
<tr>
<td><strong>Salivary gland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ryu, 2015</td>
<td>South Korea 2010</td>
<td>160</td>
<td>Range: Low dose: 1.1-2.2 GBq (29.7-59.5 mCi) High dose: ≥3.7 GBq (≥100 mCi)</td>
<td>Minimum: 1</td>
<td>RAI treatment and dose had no effect on vocal function (as a result of salivary gland dysfunction).</td>
</tr>
<tr>
<td>Jeong, 2013</td>
<td>South Korea 2003-2006</td>
<td>213</td>
<td>Mean: 5.1 GBq (138 mCi)</td>
<td>Mean: 5</td>
<td>RAI: dry mouth 35/213 (16.4%) No RAI: NA</td>
</tr>
</tbody>
</table>
Table 9. Results for Key Question 5: Harms of RAI Treatment of Screen-Detected Thyroid Cancer

<table>
<thead>
<tr>
<th>Author, Year, Quality</th>
<th>Country, Recruiting Years</th>
<th>N</th>
<th>Radiation Dose (mean or median)</th>
<th>Followup Duration (years)</th>
<th>Permanent Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grewal, 2009 Fair</td>
<td>USA 1995-2003</td>
<td>262</td>
<td>Mean: 5.3 GBq (142 mCi)</td>
<td>Mean: 7</td>
<td>RAI: Any: 13/262 (15%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dry mouth: 6/262 (2%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Salivary gland swelling: 2/262 (1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alterations in taste: 3/262 (1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Salivary gland pain: 1/262 (0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tear-duct blockage: 2/262 (1%)§§</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: NA</td>
</tr>
<tr>
<td>Ish-Shalom, 2008 Fair</td>
<td>Israel NR</td>
<td>40</td>
<td>Mean: 4.0 GBq (109 mCi)</td>
<td>Mean: 8.4</td>
<td>RAI: Dry mouth complaints: 8/23 (35%) (p=0.21)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Difficulty swallowing: 5/23 (22%) (p=0.05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alterations in taste: 3/23 (13%) (p=0.18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: Dry mouth complaints: 3/17 (18%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Difficulty swallowing: 0/17 (0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alterations in taste: 0/17 (0%)</td>
</tr>
<tr>
<td>Hyer, 2007 Fair</td>
<td>UK NR</td>
<td>76</td>
<td>3 GBq (80 mCi)</td>
<td>Minimum: 2</td>
<td>RAI: Dry mouth: 16/76 (21%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: NA</td>
</tr>
<tr>
<td>Solans, 2001 Fair</td>
<td>Spain 1990-1995</td>
<td>79</td>
<td>Range: 925 MBq to 18.5 GBq (25-500 mCi)</td>
<td>3</td>
<td>RAI: Dry mouth: 12/79 (15%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dry eyes: 11/79 (14%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: NA</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Birthrate*: Did not differ overall between groups (p=0.81) by age (adjusted p-value reported) per 1,000 women-years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Median time to first delivery following initial presentation: RAI: 34.5 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: 26.1 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p&lt;0.0001</td>
</tr>
<tr>
<td>Wu, 2015 Fair</td>
<td>USA 1999-2008</td>
<td>18,850</td>
<td>NR (California Cancer Registry)</td>
<td>Median: 4</td>
<td>Primary hyperparathyroidism: RAI: 4/6,153 patients or 27,318 person-years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No RAI: 4/2,793 patients or 10,930 person-years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hazard ratio: 0.35 (95% CI, 0.09 to 1.42)</td>
</tr>
<tr>
<td>Lin, 2014 Good</td>
<td>Taiwan 1997-2008</td>
<td>8,946</td>
<td>Median: 3.7 GBq (100 mCi)</td>
<td>Mean: 5</td>
<td>Permanent Outcomes</td>
</tr>
</tbody>
</table>

* 910 analyzed; 920 originally included in study.
† Adjusted for age at thyroid cancer diagnosis, sex, years of diagnosis, pathology, and RAI dose.
‡ Adjusted for age, all comorbidities, hormone therapy, mammography, ultrasonography, radiotherapy, chemotherapy, and thyroxine supplement.
§ Adjusted for age and sex.
\ Standard radiation dose of initial treatment, subsequent therapy administered as needed.
** Adjusted for age, sex, race, and time.
†† Adjusted for age, sex, and race.
Table 9. Results for Key Question 5: Harms of RAI Treatment of Screen-Detected Thyroid Cancer

‡‡ Adjusted for ethnic group, sex, age, and time.
§§ Corrected with surgery at 2 and 4 years after surgery.
¶¶ Study also reported rates of nasolacrimal duct stenosis.

Abbreviations: KQ=key question; RAI=radioactive iodine ablation; NA=not applicable; NR=not reported.
Note: 1 GBq=27.03 mCi.
### Table 10. Summary of Evidence, by Key Question

<table>
<thead>
<tr>
<th>Test or intervention</th>
<th># Studies (k), Sample size (n), Design</th>
<th>Summary of Findings*</th>
<th>Body of Evidence Limitations†</th>
<th>Quality</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KQ 1: Effectiveness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>k=0</td>
<td>No trials have evaluated the impact of screening for thyroid cancer on patient morbidity or mortality.</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><strong>KQ 2: Diagnostic accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palpation</td>
<td>k=2</td>
<td>2 older Finnish studies found that neck palpation was not sensitive (11.6% to 27.8%) in detecting nodules compared with ultrasound. 4 studies found the yield of cancer ranged from 0 to 4.3 cases per 1,000 persons. 2 additional studies in adults with a history of irradiation found the yield of cancer ranged from 0 to 11.3 cases per 1,000 persons.</td>
<td>Only 2 small studies reported diagnostic accuracy; 1 study did not follow up all persons with neck palpations. No evidence of reporting bias.</td>
<td>Fair</td>
<td>Poor: diagnostic accuracy studies are old and use a single examiner</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>k=2</td>
<td>2 South Korean studies found that using any 1 of several malignant sonographic characteristics can be highly sensitive (94.3%) in detecting cancer and that using a combination (≥2) of these characteristics can be both highly sensitive (94.8%) and specific (86.6%). In 4 South Korean studies, the yield of cancer ranged from 9.2 to 30.3 cases per 1,000 persons.</td>
<td>Only 2 small studies reported diagnostic accuracy, neither of which followed up with the vast majority of screened individuals, such that the reported sensitivities are likely overestimates. No evidence of reporting bias.</td>
<td>Fair</td>
<td>Fair: both diagnostic accuracy studies conducted in South Korea by the same investigators, 1 of which included women only</td>
</tr>
<tr>
<td><strong>KQ 3: Screening harms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ultrasound</td>
<td>k=1</td>
<td>1 U.S. study found that 24% of persons underwent FNA of a nodule that did not meet the Society of Radiologists in Ultrasound criteria for FNA.</td>
<td>Only 1 study.</td>
<td>Fair</td>
<td>Poor: single-institution; standards for referral to FNA have changed</td>
</tr>
<tr>
<td>Ultrasound-guided FNA</td>
<td>k=2</td>
<td>1 Japanese study (n=4,912) observed 7 cases of needle tract implantation of papillary thyroid cancer with FNA. It is unclear what impact, if any, this had on patient outcomes. 1 U.S. study observed hematomas from FNA but no major bleeding complications requiring hospitalization.</td>
<td>One study for each type of harm. Possible reporting bias.</td>
<td>Fair</td>
<td>Fair: both single-institution studies</td>
</tr>
</tbody>
</table>
Table 10. Summary of Evidence, by Key Question

<table>
<thead>
<tr>
<th>Test or intervention</th>
<th># Studies (k), Sample size (n), Design</th>
<th>Summary of Findings*</th>
<th>Body of Evidence Limitations†</th>
<th>Quality</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>KQ 4: Treatment benefit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>k=2, n=39,211 Pro/retrospective cohort</td>
<td>1 U.S. study using SEER data found that, overall, untreated persons with papillary thyroid cancer had a slightly worse 20-year survival rate (97%) than did treated persons (99%) (p&lt;0.001). 1 Japanese study found no deaths in persons with papillary microcarcinoma who opted for ultrasound observation vs. 2 deaths in persons who opted for immediate surgery.</td>
<td>Studies were not designed to evaluate the comparative benefit of treatment vs. no or delayed treatment. Lack of adjustment for confounders such that it is unclear if differences in survival are due to differences in treatment vs. case mix of persons. No evidence of reporting bias.</td>
<td>Fair to poor</td>
<td>Fair: U.S. study includes persons treated in 1970s and 1980s; Japanese study includes persons with papillary microcarcinoma</td>
</tr>
<tr>
<td>KQ 5: Treatment harms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>k=36, n=43,295 Pro/retrospective cohort</td>
<td>The rate of permanent hypoparathyroidism varied widely; best estimates were 2 to 6 events per 100 thyroidectomies and were more variable with lymph node dissection. The rate of recurrent laryngeal nerve palsy was less variable, estimated at 1 to 2 events per 100 surgeries (with or without lymph node dissection).</td>
<td>Possible publication bias for hypoparathyroidism but not recurrent laryngeal nerve palsy outcomes. The driver of the wide variation of estimates is unclear.</td>
<td>Fair</td>
<td>Fair: indication for type of surgery and case mix of patients going on to surgery have changed over time</td>
</tr>
<tr>
<td>RAI</td>
<td>k=16, n=281,796 Pro/retrospective cohort</td>
<td>Treatment with RAI for differentiated thyroid cancer is associated with a small increase in primary second malignancies: approximately 12 to 13 excess cancers per 10,000 patients. Smaller studies demonstrate an association of excess cancers at clinically used doses. Other commonly reported permanent harms from RAI include dry mouth, ranging from 2.3 to 35% of persons.</td>
<td>Differences in study designs and variable reporting on radiation doses limits our understanding of the magnitude and precision around risk of second primary malignancies. No evidence-reporting bias for commonly reported adverse outcomes.</td>
<td>Fair</td>
<td>Fair: indication and radiation dose of RAI have changed over time</td>
</tr>
</tbody>
</table>

* Includes consistency and precision.  
† Includes reporting bias.  
‡ Calculated sample size includes only the largest study using SEER data so as to avoid double-counting studies with overlapping populations.  

**Abbreviations:** CI=confidence interval; FNA=fine-needle aspiration; k=number of studies; n=number; NA=not applicable; RAI=radioactive iodine.
Appendix A. Search Strategy

Databases searched:
OVID MEDLINE
PubMed, publisher-supplied
Cochrane Central Register of Controlled Trials (CENTRAL)

Key:
/ = MeSH subject heading
? = wildcard
ti = word in title
ab = word in abstract
pt = publication type
* = truncation
kw = keyword
adj# = adjacent within # number of words
fs = floating subheading
us = ultrasonography
dt = drug therapy
rt = radiotherapy
su = surgery
th = therapy

MEDLINE
Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1946 to Present>, Ovid MEDLINE(R) Daily Update <January 12, 2016>
Search Strategy:

1 Thyroid Neoplasms/
2 Thyroid Nodule/
3 Thyroid Carcinoma, Anaplastic/
4 Carcinoma/
5 Carcinoma, Papillary/
6 Carcinoma, Medullary/
7 Carcinoma, Papillary, Follicular/
8 Adenocarcinoma, Follicular/
9 Adenocarcinoma, Papillary/
10 Adenocarcinoma/
11 Thyroid Gland/
12 11 and (4 or 5 or 6 or 7 or 8 or 9 or 10)
13 (thyroid adj3 (cancer* or carcinoma* or adenoma* or nodule* or tumor?r* or neoplasm* or lymphoma* or adenocarcinoma* or sarcoma* or papillar* or follicular* or hurthle* or oxyphil* or medullar* or anaplast* or malignan*)).ti.
14 1 or 2 or 3 or 12 or 13
15 Ultrasonography/
16 Elasticity Imaging Techniques/
17 Elasticity/
18 Mass Screening/
19 Multiphasic Screening/
20 "Early Detection of Cancer"/
21 early diagnosis/
22 Palpation/
Appendix A. Search Strategy

23 Population Surveillance/
24 Sentinel Surveillance/
25 (screen* or surveil*).ti,ab.
26 (test* or exam* or detect* or predict* or identif* or discover* or diagnos*).ti.
27 case finding.ti,ab.
28 ultrasound* .ti,ab.
29 ultrasonograph*.ti,ab.
30 elastogra*.ti,ab.
31 echotomograph*.ti,ab.
32 echograph*.ti,ab.
33 ultrasonic*.ti,ab.
34 sonograph*.ti,ab.
35 sonogram*.ti,ab.
36 (palpat* or palpab*).ti,ab.
37 Thyroid Neoplasms/us [Ultrasonography]
38 Thyroid Nodule/us [Ultrasonography]
39 Carcinoma/us [Ultrasonography]
40 Carcinoma, Papillary/us [Ultrasonography]
41 Carcinoma, Medullary/us [Ultrasonography]
42 Carcinoma, Papillary, Follicular/us [Ultrasonography]
43 Adenocarcinoma, Follicular/us [Ultrasonography]
44 Adenocarcinoma, Papillary/us [Ultrasonography]
45 Adenocarcinoma/us [Ultrasonography]
46 37 or 38
47 11 and (39 or 40 or 41 or 42 or 43 or 44 or 45)
48 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or
32 or 33 or 34 or 35 or 36
49 Clinical Trials as Topic/
50 Controlled Clinical Trials as Topic/
51 Randomized Controlled Trials as Topic/
52 Meta-Analysis as Topic/
53 Control Groups/
54 Double-Blind Method/
55 Single-Blind Method/
56 (clinical trial or controlled clinical trial or meta analysis or randomized controlled trial).pt.
57 random*.ti,ab.
58 clinical trial*.ti,ab.
59 controlled trial*.ti,ab.
60 meta analy*.ti,ab.
61 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60
62 (14 and 48) or 46 or 47
63 61 and 62
64 remove duplicates from 63
65 "Sensitivity and Specificity"/
66 "Predictive Value of Tests"/
67 ROC Curve/
68 False Negative Reactions/
69 False Positive Reactions/
70 Diagnostic Errors/
71 "Reproducibility of Results"/
72 Reference Values/
Appendix A. Search Strategy

73 Reference Standards/
74 Observer Variation/
75 receiver operat*.ti,ab.
76 roc curv*.ti,ab.
77 sensitivit*.ti,ab.
78 specificit*.ti,ab.
79 predictive value.ti,ab.
80 accuracy.ti,ab.
81 false positive*.ti,ab.
82 false negative*.ti,ab.
83 miss rate*.ti,ab.
84 error rate*.ti,ab.
85 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84
86 (14 and 48) or 46 or 47
87 85 and 86
88 remove duplicates from 87
89 Mortality/
90 Morbidity/
91 Death/
92 "Drug-Related Side Effects and Adverse Reactions"
93 Fatal Outcome/
94 "Quality of Life"
95 Stress, Psychological/
96 Anxiety/
97 Reoperation/
98 Recurrence/
99 Neoplasm Recurrence, Local/
100 Hypocalcemia/
101 Voice Disorders/
102 Voice Quality/
103 Voice/ (6278)
104 Hypesthesia/
105 safety.ti,ab.
106 harm*.ti,ab.
107 mortality.ti,ab.
108 complication*.ti,ab.
109 (death or deaths or die or dying).ti,ab.
110 (adverse adj2 (interaction* or response* or effect* or event* or reaction* or outcome* or feature*)).ti,ab.
111 adverse effects.fs.
112 mortality.fs.
113 overdiagnos*.ti,ab.
114 over diagnos*.ti,ab.
115 unnecessary exam*.ti,ab.
116 unnecessary procedure*.ti,ab.
117 unnecessary test*.ti,ab.
118 unneeded exam*.ti,ab.
119 unneeded procedure*.ti,ab.
120 unneeded test*.ti,ab.
121 unneeded surger*.ti,ab.
Appendix A. Search Strategy

122 unnecessary surger*.ti,ab.
123 reoperation*.ti,ab.
124 recur*.ti,ab.
125 overtreat*.ti,ab.
126 over treat*.ti,ab.
127 (secondary adj3 malignan*).ti,ab.
128 psychosocial*.ti,ab.
129 (anxiet* or anxious* or distress* or nervous*).ti,ab.
130 (burden* or challenge*).ti,ab.
131 side effect*.ti,ab.
132 hypocalcem*.ti,ab.
133 hypocalcaem*.ti,ab.
134 voice.ti,ab.
135 numb*.ti,ab.
136 hypesthes*.ti,ab.
137 incidence/
138 Time Factors/
139 Prognosis/
140 Autopsy/
141 (incidence or prognos* or natural histor* or autopsy or autopsies).ti,ab.
142 ((time or temporal) adj3 trend*).ti,ab.
143 89 or 90 or 91 or 92 or 93 or 94 or 95 or 96 or 97 or 98 or 99 or 100 or 101 or 102 or 103 or 104 or 105 or 106 or 107 or 108 or 109 or 110 or 111 or 112 or 113 or 114 or 115 or 116 or 117 or 118 or 119 or 120 or 121 or 122 or 123 or 124 or 125 or 126 or 127 or 128 or 129 or 130 or 131 or 132 or 133 or 134 or 135 or 136 or 137 or 138 or 139 or 140 or 141 or 142
144 Clinical Trials as Topic/
145 Controlled Clinical Trials as Topic/
146 Randomized Controlled Trials as Topic/
147 Meta-Analysis as Topic/
148 Control Groups/
149 Double-Blind Method/
150 Single-Blind Method/
151 Cohort Studies/
152 Longitudinal Studies/
153 Follow-Up Studies/
154 Prospective Studies/
155 Retrospective Studies/
156 (clinical trial or controlled clinical trial or meta analysis or randomized controlled trial).pt.
157 random*.ti,ab.
158 clinical trial*.ti,ab.
159 controlled trial*.ti,ab.
160 meta analy*.ti,ab.
161 cohort.ti,ab.
162 longitudinal.ti,ab.
163 (follow up or followup).ti,ab.
164 case-control studies/
165 (case adj2 (control* or base* or comparison* or referrent or referent or compeer*)).ti,ab.
166 144 or 145 or 146 or 147 or 148 or 149 or 150 or 151 or 152 or 153 or 154 or 155 or 156 or 157 or 158 or 159 or 160 or 161 or 162 or 163 or 164 or 165
167 (14 and 48) or 46 or 47
168 143 and 166 and 167
Appendix A. Search Strategy

169 remove duplicates from 168
170 Thyroidectomy/
171 Lymph Node Excision/
172 Neck Dissection/
173 Iodine Radioisotopes/
174 Iodine Isotopes/
175 surgery.ti,ab.
176 surgical.ti,ab.
177 treat*.ti.
178 thyroidect*.ti,ab.
179 lobect*.ti,ab.
180 ((thyroid or neck or lymph) adj3 (remov* or dissect* or excis* or extract*)).ti,ab.
181 lymphadenect*.ti,ab.
182 radioactive iodin*.ti,ab.
183 radioidin*.ti,ab.
184 radio iodin*.ti,ab.
185 radio nuclide*.ti,ab.
186 radionuclide*.ti,ab.
187 (iodine adj2 ablat*).ti,ab.
188 i-131.ti,ab.
189 131-i.ti,ab.
190 Thyroid Neoplasms/dt, rt, su, th [Drug Therapy, Radiotherapy, Surgery, Therapy]
191 Thyroid Nodule/dt, rt, su, th [Drug Therapy, Radiotherapy, Surgery, Therapy]
192 Thyroid Carcinoma, Anaplastic/dt, rt, su, th [Drug Therapy, Radiotherapy, Surgery, Therapy]
193 Carcinoma/dt, rt, su, th [Drug Therapy, Radiotherapy, Surgery, Therapy]
194 Carcinoma, Papillary/dt, rt, su, th [Drug Therapy, Radiotherapy, Surgery, Therapy]
195 Carcinoma, Medullary/dt, rt, su, th [Drug Therapy, Radiotherapy, Surgery, Therapy]
196 Carcinoma, Papillary, Follicular/dt, rt, su, th [Drug Therapy, Radiotherapy, Surgery, Therapy]
197 Adenocarcinoma, Follicular/dt, rt, su, th [Drug Therapy, Radiotherapy, Surgery, Therapy]
198 Adenocarcinoma, Papillary/dt, rt, su, th [Drug Therapy, Radiotherapy, Surgery, Therapy]
199 Adenocarcinoma/dt, rt, su, th [Drug Therapy, Radiotherapy, Surgery, Therapy]
200 190 or 191 or 192
201 11 and (193 or 194 or 195 or 196 or 197 or 198 or 199)
202 170 or 171 or 172 or 173 or 174 or 175 or 176 or 177 or 178 or 179 or 180 or 181 or 182 or 183 or 184 or 185 or 186 or 187 or 188 or 189
203 144 or 145 or 146 or 147 or 148 or 149 or 150 or 151 or 152 or 153 or 154 or 155 or 156 or 157 or 158 or 159 or 160 or 161 or 162 or 163
204 (14 and 202) or 200 or 201
205 203 and 204
206 (14 and 202) or 200 or 201
207 143 and 166 and 206
208 64 or 88 or 169 or 205 or 207
209 limit 208 to yr="1966 -Current"
210 limit 209 to english language
211 animals/ not (humans/ and animals/)
212 210 not 211
Appendix A. Search Strategy

**PubMed** [publisher supplied references only]

<table>
<thead>
<tr>
<th>Search</th>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>#44</td>
<td>Search #41 AND #42 AND #43</td>
</tr>
<tr>
<td>#43</td>
<td>Search English[language]</td>
</tr>
<tr>
<td>#42</td>
<td>Search publisher[sb]</td>
</tr>
<tr>
<td>#41</td>
<td>Search #18 OR #40</td>
</tr>
<tr>
<td>#40</td>
<td>Search #3 AND #39</td>
</tr>
<tr>
<td>#39</td>
<td>Search #19 OR #20 OR #21 OR #22 OR #23 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38</td>
</tr>
<tr>
<td>#38</td>
<td>Search iodine radio isotope*[tiab]</td>
</tr>
<tr>
<td>#37</td>
<td>Search iodine radioisotope*[tiab]</td>
</tr>
<tr>
<td>#36</td>
<td>Search iodine isotope*[tiab]</td>
</tr>
<tr>
<td>#35</td>
<td>Search 131I[tiab]</td>
</tr>
<tr>
<td>#34</td>
<td>Search I 131[tiab]</td>
</tr>
<tr>
<td>#33</td>
<td>Search ablat*[ti]</td>
</tr>
<tr>
<td>#32</td>
<td>Search radionuclide*[tiab]</td>
</tr>
<tr>
<td>#31</td>
<td>Search radio nuclide*[tiab]</td>
</tr>
<tr>
<td>#30</td>
<td>Search radio iodin*[tiab]</td>
</tr>
<tr>
<td>#29</td>
<td>Search radiiodin*[tiab]</td>
</tr>
<tr>
<td>#28</td>
<td>Search radioactive iodin*[tiab]</td>
</tr>
<tr>
<td>#27</td>
<td>Search lymphadenect*[tiab]</td>
</tr>
<tr>
<td>#26</td>
<td>Search #24 AND #25</td>
</tr>
<tr>
<td>#25</td>
<td>Search remov*[ti] OR dissect*[ti] OR excis*[ti] OR extract*[ti]</td>
</tr>
<tr>
<td>#24</td>
<td>Search thyroid*[ti] OR neck*[ti] OR lymph*[ti]</td>
</tr>
<tr>
<td>#23</td>
<td>Search lobect*[tiab]</td>
</tr>
<tr>
<td>#22</td>
<td>Search thyroidect*[tiab]</td>
</tr>
<tr>
<td>#21</td>
<td>Search treat*[ti]</td>
</tr>
<tr>
<td>#20</td>
<td>Search surgical*[tiab]</td>
</tr>
<tr>
<td>#19</td>
<td>Search surger*[tiab]</td>
</tr>
<tr>
<td>#18</td>
<td>Search #3 AND #17</td>
</tr>
<tr>
<td>#17</td>
<td>Search #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16</td>
</tr>
<tr>
<td>#16</td>
<td>Search elastic imag*[tiab]</td>
</tr>
<tr>
<td>#15</td>
<td>Search palp*[tiab] or palpab*[tiab]</td>
</tr>
<tr>
<td>#14</td>
<td>Search sonogram*[tiab]</td>
</tr>
<tr>
<td>#13</td>
<td>Search sonograph*[tiab]</td>
</tr>
<tr>
<td>#12</td>
<td>Search ultrasonic*[tiab]</td>
</tr>
<tr>
<td>#11</td>
<td>Search echograph*[tiab]</td>
</tr>
<tr>
<td>#10</td>
<td>Search echotomograph*[tiab]</td>
</tr>
<tr>
<td>#9</td>
<td>Search elastogra*[tiab]</td>
</tr>
</tbody>
</table>
### Appendix A. Search Strategy

<table>
<thead>
<tr>
<th>Search</th>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>#8</td>
<td>Search ultrasonograph*[tiab]</td>
</tr>
<tr>
<td>#7</td>
<td>Search ultrasound*[tiab]</td>
</tr>
<tr>
<td>#6</td>
<td>Search case finding*[tiab]</td>
</tr>
<tr>
<td>#4</td>
<td>Search screen*[tiab] OR surveil*[tiab]</td>
</tr>
<tr>
<td>#3</td>
<td>Search #1 AND #2</td>
</tr>
<tr>
<td>#1</td>
<td>Search thyroid*[ti]</td>
</tr>
</tbody>
</table>

#### Cochrane Central Register of Controlled Clinical Trials (CENTRAL)

#1  (thyroid):ti,ab,kw near/3 (cancer* or carcinoma* or adenoma* or nodule* or tumo?r* or neoplasm* or lymphoma* or adenocarcinoma* or sarcoma* or papillar* or follicular* or hurthle* or oxyphil* or medullar* or anaplast* or malignant*):ti,ab,kw

#2  (screen* or surveil*):ti,ab,kw

#3  (test* or exam* or detect* or predict* or identif* or discover* or diagnos*):ti

#4  "case finding":ti,ab,kw

#5  ultrasound*:ti,ab,kw

#6  ultrasonograph*:ti,ab,kw

#7  elastogra*:ti,ab,kw

#8  echotomograph*:ti,ab,kw

#9  echograph*:ti,ab,kw

#10 ultrasonic*:ti,ab,kw

#11 sonograph*:ti,ab,kw

#12 sonogram*:ti,ab,kw

#13 (palpat* or palpab*):ti,ab,kw

#14 (elasti*):ti,ab,kw near/3 (image or imaging):ti,ab,kw

#15 #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14

#16 #1 and #15

#17 "receiver operat*":ti,ab,kw

#18 "roc curv*":ti,ab,kw

#19 sensitivit*:ti,ab,kw

#20 specificit*:ti,ab,kw

#21 "predictive value*":ti,ab,kw

#22 accuracy:ti,ab,kw

#23 "false positive*":ti,ab,kw

#24 "false negative*":ti,ab,kw

#25 "miss rate*":ti,ab,kw

#26 "error rate*":ti,ab,kw

#27 reference near/3 standard*:ti,ab,kw

#28 reference near/3 value*:ti,ab,kw

#29 "observer variation*":ti,ab,kw

#30 #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 58443

#31 #16 and #30

#32 safety:ti,ab,kw

#33 harm*:ti,ab,kw
Appendix A. Search Strategy

#34 mortality:ti,ab,kw
#35 complication*:ti,ab,kw
#36 (death or deaths or die or dying or fatal*):ti,ab,kw
#37 adverse:ti,ab,kw near/2 (interaction* or response* or effect* or event* or reaction* or outcome* or feature*):ti,ab,kw
#38 overdiagnos*:ti,ab,kw
#39 "over diagnos*":ti,ab,kw
#40 "unnecessary exam*":ti,ab,kw
#41 "unnecessary procedure*":ti,ab,kw
#42 "unnecessary test*":ti,ab,kw
#43 "unneeded exam*":ti,ab,kw
#44 "unneeded procedure*":ti,ab,kw
#45 "unneeded test*":ti,ab,kw
#46 "unneeded surger*":ti,ab,kw
#47 "unnecessary surger*":ti,ab,kw
#48 reoperation*:ti,ab,kw
#49 recur*:ti,ab,kw
#50 overtreat*:ti,ab,kw
#51 "over treat*":ti,ab,kw
#52 (secondary near/3 malignan*):ti,ab,kw
#53 psychosocial*:ti,ab,kw
#54 (anxiet* or anxious* or distress* or nervous*):ti,ab,kw
#55 (burden* or challenge*):ti,ab,kw
#56 "side effect*":ti,ab,kw
#57 hypocalcem*:ti,ab,kw
#58 hypocalcaem*:ti,ab,kw
#59 voice:ti,ab,kw
#60 numb*:ti,ab,kw
#61 hypesthes*:ti,ab,kw
#62 (incidence or prognos* or "natural histor*" or autopsy or autopsies):ti,ab,kw
#63 (time or temporal):ti,ab,kw near/3 (trend* or factor*):ti,ab,kw
#64 morbidit*:ti,ab,kw
#65 "quality of life":ti,ab,kw
#66 #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65
#67 #16 and #66
#68 (surger* or surgical):ti,ab,kw
#69 treat*:ti
#70 thyroidect*:ti,ab,kw
#71 lobect*:ti,ab,kw
#72 (thyroid or neck or lymph*):ti,ab,kw near/3 (remov* or dissect* or excis* or extract*):ti,ab,kw
#73 lymphadenect*:ti,ab,kw
#74 "radioactive iodin*":ti,ab,kw
#75 radioiodin*:ti,ab,kw
#76 "radio iodin*":ti,ab,kw
#77 iodin*:ti,ab,kw near/3 (isotope* or radioisotope* or "radio isotope"):ti,ab,kw
#78 "radio nuclide*":ti,ab,kw
#79 radionuclide*:ti,ab,kw
#80 (iodine near/2 ablat*):ti,ab,kw
Appendix A. Search Strategy

#81 "i-131":ti,ab,kw
#82 "131-i":ti,ab,kw
#83 #68 or #69 or #70 or #71 or #72 or #73 or #74 or #75 or #76 or #77 or #78 or #79 or #80 or #81 or #82
#84 #1 and #83
#85 #1 and #83 and #66
#86 #16 or #31 or #67 or #84 or #85 in Trials
### Appendix A Table 1. Quality Assessment Criteria

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Quality Criteria</th>
</tr>
</thead>
</table>
| Randomized controlled trials USPSTF methods | - Valid random assignment?  
- Was allocation concealed?  
- Was eligibility criteria specified?  
- Were groups similar at baseline?  
- Were measurements equal, valid, and reliable?  
- Was there intervention fidelity?  
- Was there adequate adherence to the intervention?  
- Were outcome assessors blinded?  
- Was there acceptable followup?  
- Were the statistical methods acceptable?  
- Was the handling of missing data appropriate?  
- Was there evidence of selective reporting of outcomes?  
- Was the device calibration and/or maintenance reported? |
| Observational studies (e.g., prospective cohort studies), adapted from the Newcastle-Ottawa Scale (NOS) | - Was the cohort systematically selected to avoid bias?  
- Was eligibility criteria specified?  
- Were groups similar at baseline?  
- Was the outcome of interest not present at baseline?  
- Were measurements equal, valid, and reliable?  
- Were outcome assessors blinded?  
- Was there acceptable followup?  
- Were the statistical methods acceptable?  
- Was the handling of missing data appropriate? |
| Diagnostic accuracy studies adapted from QUADAS I and II | - Screening test relevant, available for primary care, and adequately described  
- Study uses a credible reference standard performed regardless of test results  
- Reference standard interpreted independently of screening test  
- Handles indeterminate results in a reasonable manner  
- Spectrum of patients included in study  
- Sample size reported  
- Administration of reliable screening test |

**Abbreviations:** USPSTF=U.S. Preventive Services Task Force; QUADAS=Quality Assessment of Diagnostic Accuracy Studies.
# Appendix A Table 2. Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Populations</th>
<th>Include</th>
<th>Exclude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Asymptomatic adults age ≥18 years</td>
<td>• Persons who are already under surveillance for thyroid cancer because of previous thyroid cancer</td>
</tr>
<tr>
<td></td>
<td>• High-risk populations (those with a history of radiation exposure or family history of thyroid cancer)</td>
<td>• Persons who have symptoms that may lead to thyroid evaluation</td>
</tr>
<tr>
<td></td>
<td>• Persons who are already under surveillance for thyroid cancer because of previous thyroid cancer</td>
<td>• Persons with known inherited genetic syndromes, such as multiple endocrine neoplasia type II, as selection criteria for studies</td>
</tr>
<tr>
<td></td>
<td>• Persons who have symptoms that may lead to thyroid evaluation</td>
<td>• Persons with thyroid disease</td>
</tr>
<tr>
<td></td>
<td>• Persons with known inherited genetic syndromes, such as multiple endocrine neoplasia type II, as selection criteria for studies</td>
<td>• Children and adolescents</td>
</tr>
</tbody>
</table>

| Screening tests | KQs 1–3: Palpation or ultrasound of the neck conducted by primary care providers or specialists as part of a routine well care visit | KQs 2: Sensitivity, specificity, positive predictive value, false-positives, false-negatives, nodule detection rates, and cancer detection rates |
|                | • Enhanced ultrasound methods, such as elastography or ultrasound with contrast media | Incidentally identified thyroid nodules |
|                | • Diagnostic procedures (e.g., fine needle aspiration) will be excluded as screening tests but reviewed under harms of screening | |
|                | • Other imaging tests (e.g., magnetic resonance imaging, positron emission tomography) that incidentally identify thyroid nodules | |
|                | • Blood tests (e.g., calcitonin, thyroid-stimulating hormone) | |
|                | • Self-examination | |
|                | • Diagnostic accuracy studies in persons with known nodules | |

| Treatment interventions | KQs 4, 5: Surgery, including lobectomy, near-total thyroidectomy, total thyroidectomy, and lymphadenectomy; radioactive iodine ablation | Chemotherapy |
|                        | • Chemotherapy | External beam radiation therapy |
|                        | • External beam radiation therapy | Nonsurgical ablative treatment, such as thermal ablation, radiofrequency ablation, or ultrasound-guided percutaneous ethanol injection |
|                        | • Nonsurgical ablative treatment, such as thermal ablation, radiofrequency ablation, or ultrasound-guided percutaneous ethanol injection | Older treatment studies pre-1990 |

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>KQs 1–3: No screening</th>
<th>KQs 4, 5: No treatment</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>KQs 1, 4: Reduced morbidity associated with any thyroid cancer (including papillary, follicular, medullary, and anaplastic), including:</th>
<th>Incidentally identified thyroid nodules</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Improved quality of life</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Decreased thyroid cancer mortality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Decreased all-cause mortality</td>
<td></td>
</tr>
<tr>
<td>KQ 2:</td>
<td>Sensitivity, specificity, positive predictive value, false-positives, false-negatives, nodule detection rates, and cancer detection rates</td>
<td></td>
</tr>
<tr>
<td>KQs 3, 5:</td>
<td>Any harm from screening or treatment, including overdiagnosis,* diagnostic tests, overtreatment,** psychosocial harms, secondary malignancies, or procedure-related adverse events</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Settings</th>
<th>U.S. primary care settings</th>
<th>Nations with environmental disasters that lead to very high radiation exposure (e.g., Ukraine, Japan)</th>
</tr>
</thead>
</table>
## Appendix A Table 2. Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Study designs</th>
<th>Include</th>
<th>Exclude</th>
</tr>
</thead>
</table>
|               | Fair- to good-quality studies published between January 1, 1966 and March 31, 2015  
KQ 1: Randomized, controlled trials  
KQ 2: Diagnostic accuracy studies with a reference standard, systematic evidence reviews  
KQs 3, 5: Randomized, controlled trials; controlled clinical trials; cohort studies; case-control studies  
KQ 4: Randomized, controlled trials; controlled clinical trials; cohort studies | Poor-quality studies with a fatal flaw; studies outside of the publication window; case reports and case series; decision analyses |

*Diagnosis of nonpalpable nodules measuring <1 cm and/or fine needle aspiration of nodules not meeting revised 2009 American Thyroid Association criteria for fine needle aspiration.  
**Including treatment of an overdiagnosed nodule and extended followup of benign nodules.
Appendix A Figure 1. Literature Flow Diagram

Abbreviation: KQ = key question.
Appendix B. Ongoing Studies

We searched selected grey literature sources, including ClinicalTrials.gov and WHO International Clinical Trials Registry Platform (ICTRP), for ongoing trials. From these sources no screening focused studies were identified. Multiple studies found focus on the effectiveness of surgery in people with low risk papillary carcinomas. The majority of these focus on people with low-risk papillary thyroid cancer or microcarcinomas of the thyroid. There are two randomized trials addressing efficacy and safety of prophylactic central lymph node dissection, both of which are still in the recruiting phase; the South Korean based study is expected to be completed in 2022\(^{155}\) and the U.S. based study is expected to be completed in 2020.\(^{156}\) Outcomes will include harms of surgery. For papillary microcarcinomas there are three ongoing studies that would contribute to the evidence around overdiagnosis or harms of treatment. Two studies from South Korea on the comparative effectiveness of surgical treatment of papillary microcarcinomas may report the harms.\(^{157,158}\) Another study will evaluate why patients with papillary microcarcinoma choose treatment versus active surveillance.\(^{159}\)
Appendix C. Excluded Studies

Exclusion Codes

E Codes
E1. Study relevance
E1a. Primary aim technology improvements
E2. Study design
E2a. Case report or case series
E2b. Comparative effectiveness only, no control (untreated arm)
E2c. Diagnostic accuracy studies in persons with known nodules
E3. Setting
E3a. Not a very high HDI country
E3b. Nation with environmental disaster with very high radiation exposure
E4. Population
E4a. Previous thyroid cancer
E4b. Symptomatic
E4c. Inherited genetic syndromes
E4d. Thyroid disease
E4e. Children < 18
E5. No relevant outcomes or incomplete outcomes
E5a. No additional relevant data (primary article included)
E5b. Incidentally identified thyroid nodules
E6. Intervention (including outdated technology)
E6a. Imaging other than ultrasound
E6b. Blood tests
E6c. Self-exam
E6d. Chemotherapy
E6e. External beam radiation therapy
E6f. Non-surgical ablative treatment other than Radioactive Iodine Ablation
E6g. older treatment study – pre 1980
E6h. Single-surgeon practice
E7. Poor Study Quality
E8. Key existing SER with out of date MA
E9. Not in English
E10. Unable to retrieve

Appendix C. Excluded Studies


Appendix C. Excluded Studies


Appendix C. Excluded Studies


Appendix C. Excluded Studies


Appendix C. Excluded Studies


78. Chisholm EJ, Kulinskaya E and Tolley NS. Systematic review and meta-analysis of the adverse effects of thyroidectomy combined with central neck dissection as compared with thyroidectomy alone. Laryngoscope 2009. p. 1135-9. PMID: 19358241. KQ4e2b, KQ5e8


Appendix C. Excluded Studies


102. Creach KM, Siegel BA, Nussenbaum B and Grigsby PW. Radioactive iodine therapy decreases recurrence in thyroid papillary microcarcinoma. Isrn Endocrinology Print2012. p. 816386. PMID: 22462017. KQ4e2b, KQ5e6h


105. Davies L, Ouellette M, Hunter M and Welch HG. The increasing incidence of small thyroid cancers: where are the cases coming from? Laryngoscope 2010. p. 2446-51. PMID: 21108428. KQ3e5


107. Davis S. Screening For Thyroid Cancer after the Fukushima Disaster: What Do We Learn From Such An Effort? Epidemiology. 2015/10/07 ed2015. PMID: 26441344. KQ1E5


Appendix C. Excluded Studies


Appendix C. Excluded Studies

131. Elsayed NM and Elkhatib YA. Diagnostic Criteria and Accuracy of Malignant Thyroid Nodules by Ultrasonography and Ultrasound Elastography with Pathologic Correlation. Ultrasom Imaging. 2015/05/03 ed2015. PMID: 25933616. **KQ2E2C**


Appendix C. Excluded Studies


Appendix C. Excluded Studies


Appendix C. Excluded Studies


Appendix C. Excluded Studies

204. Hughes DT, Haymart MR, Miller BS, Gauger PG and Doherty GM. The most commonly occurring papillary thyroid cancer in the United States is now a microcarcinoma in a patient older than 45 years. Thyroid. 2011/01/28 ed2011. p. 231-6. PMID: 21268762. KQ4E1B


Appendix C. Excluded Studies


224. Jatzko GR, Lisborg PH, Muller MG and Wette VM. Recurrent nerve palsy after thyroid operations--principal nerve identification and a literature review. Surgery1994. p. 139-44. PMID: 8310401. KQ4e2b, KQ5e8


Appendix C. Excluded Studies


253. Kim C, Bi X, Pan D, Chen Y, Carling T, Ma S, Udelsman R and Zhang Y. The risk of second cancers after diagnosis of primary thyroid cancer is elevated in thyroid microcarcinomas. Thyroid 2013. p. 575-82. PMID: 23237308. KQ5e6h


Appendix C. Excluded Studies


263. Kim SK, Kang SY, Yoon HJ and Jung SH. Comparison of conventional thyroidectomy and endoscopic thyroidectomy via axillo-bilateral breast approach in papillary thyroid carcinoma patients. Surg Endosc. 2015/10/30 ed2015. PMID: 26511120. KQ5E6H


266. Ko SY, Kim EK, Moon HJ, Yoon HJ, Kim HY and Kwak JY. Application of Thyroid Imaging Reporting and Data System in the Ultrasound Assessment of Thyroid Nodules According to Physician Experience. Ultrasound Quarterly. 2015/08/19 ed2015. PMID: 26280583. KQ2E2C


Appendix C. Excluded Studies


Appendix C. Excluded Studies

292. Lee HS, Park HS, Kim SW, Choi G, Park HS, Hong JC, Lee SG, Baek SM and Lee KD. Clinical characteristics of papillary thyroid microcarcinoma less than or equal to 5 mm on ultrasonography. European archives of otorhinolaryngology 2013. p. 2969-74. PMID: 23873032. KQ5E2


297. Lee MJ, Hong SW, Chung WY, Kwak JY, Kim MJ and Kim EK. Cytological results of ultrasound-guided fine-needle aspiration cytology for thyroid nodules: emphasis on correlation with sonographic findings. Yonsei medical journal 2011. p. 838-44. PMID: 21786450. KQ2e2c


308. Li YJ, Wang YZ, Yi ZB, Chen LL and Zhou XD. Comparison of Completion Thyroidectomy and Primary Total Surgery for Differentiated Thyroid Cancer: A Meta-Analysis. Oncology Research and Treatment 2015. p. 528-31. PMID: 26451702. KQ5E3A
Appendix C. Excluded Studies


310. Lin JD, Chao TC, Huang BY, Chen ST, Chang HY and Hsueh C. Thyroid cancer in the thyroid nodules evaluated by ultrasonography and fine-needle aspiration cytology. Thyroid 2005. p. 708-17. PMID: 16053388. KQ2e2c, KQ2e2


312. Lin JD, Hsueh C and Chao TC. Long-Term Follow-Up of the Therapeutic Outcomes for Papillary Thyroid Carcinoma With Distant Metastasis. Medicine 2015. p. e1063. PMID: 26131826. KQ4E2B

313. Lin JD, Hsueh C, Chao TC, Weng HF and Huang BY. Thyroid follicular neoplasms diagnosed by high-resolution ultrasonography with fine needle aspiration cytology. Acta cytologica 1997. p. 687-91. PMID: 9167684. KQ2E2C, KQ3E4


Appendix C. Excluded Studies


335. Lyle MA and Dean DS. Ultrasound-Guided Fine-Needle Aspiration Biopsy of Thyroid Nodules in Patients Taking Novel Oral Anticoagulants. Thyroid. 2015/01/15 ed2015. PMID: 25584817. KQ3E5


Appendix C. Excluded Studies


348. Marti JL, Jain KS and Morris LG. Increased risk of second primary malignancy in pediatric and young adult patients treated with radioactive iodine for differentiated thyroid cancer. Thyroid2015. p. 681-7. PMID: 25851829. KQ5E4


Appendix C. Excluded Studies


377. Moon HJ, Lee HS, Kim HK, Ko SY, Seo JY, Park WJ, Park HY and Kwak JY. Thyroid nodules < 5 mm on ultrasonography: are they "leave me alone" lesions? Endocrine 2015. p. 735-44. PMID: 25600483.


Appendix C. Excluded Studies


383. Morris LG, Sikora AG, Tosteson TD and Davies L. The increasing incidence of thyroid cancer: the influence of access to care. Thyroid 2013. p. 885-91. PMID: 23517343. KQ3e5


Appendix C. Excluded Studies


Appendix C. Excluded Studies


425. Peiling Yang S, Bach AM, Michael Tuttle R and Fish SA. Frequent screening with serial neck ultrasounds is more likely to identify false positive abnormalities than clinically significant disease in the surveillance of intermediate risk papillary thyroid cancer patients without suspicious findings on follow-up ultrasound evaluation. Journal of Clinical Endocrinology & Metabolism. 2015/01/31 ed2015. p. jc20143651. PMID: 25632970. KQ2E5, KQ3E5


Appendix C. Excluded Studies


437. Qiu ZL, Shen CT and Luo QY. Clinical management and outcomes in patients with hyperfunctioning distant metastases from differentiated thyroid cancer after total thyroidectomy and radioactive iodine therapy. Thyroid2015. p. 229-37. PMID: 25331724. KQ5E3A


444. Renshaw AA. Focal features of papillary carcinoma of the thyroid in fine-needle aspiration material are strongly associated with papillary carcinoma at resection. American Journal of Clinical Pathology2002. p. 208-10. PMID: 12162679. KQ3e2a


450. Rosario PW, Barroso AL, Rezende LL, Padrao EL, Borges MA, Guimaraes VC and Purisch S. Testicular function after radioiodine therapy in patients with thyroid cancer. Thyroid. 2006/08/08 ed2006. p. 667-70. PMID: 16889490. KQALLE3A


Appendix C. Excluded Studies


Appendix C. Excluded Studies


Appendix C. Excluded Studies


Appendix C. Excluded Studies


517. Song HJ, Qiu ZL, Shen CT, Wei WJ and Luo QY. Pulmonary metastases in differentiated thyroid cancer: efficacy of radiiodine therapy and prognostic factors. European Journal of Endocrinology 2015. p. 399-408. PMID: 26104753. KQ5E3A


Appendix C. Excluded Studies


Appendix C. Excluded Studies


Appendix C. Excluded Studies


565. Ukrainski MB, Pribitkin EA and Miller JL. Increasing Incidence of Thyroid Nodules and Thyroid Cancer: Does Increased Detection of a Subclinical Reservoir Justify the Associated Anxiety and Treatment? Clinical Therapeutics. 2015/10/06 ed2015. PMID: 26434793. **KQALLE5**


Appendix C. Excluded Studies


Appendix C. Excluded Studies


605. Wolinski K, Szkudlarek M, Szczepanek-Parulski E and Ruchala M. Usefulness of different ultrason features of malignancy in predicting the type of thyroid lesion: a meta-analysis of prospective studies. Polskie Archiwum Medycyny Wewnetrznej 2014. p. 97-104. PMID: 24473342. KQ2e8


Appendix C. Excluded Studies


624. Yuan WH, Chiou HJ, Chou YH, Hsu HC, Tiu CM, Cheng CY and Lee CH. Gray-scale and color Doppler ultrasonographic manifestations of papillary thyroid carcinoma: analysis of 51 cases. Clinical imaging 2006. p. 394-401. PMID: 17101408. KQ2e2c

625. Ywata de Carvalho A, Chulam TC and Kowalski LP. Long-term Results of Observation vs Prophylactic Selective Level VI Neck Dissection for Papillary Thyroid Carcinoma at a Cancer Center. JAMA Otolaryngology-- Head & Neck Surgery 2015. p. 599-606. PMID: 25997016. KQ5E3A


Appendix C. Excluded Studies


634. Zhao ZH, Li FQ, Han JK and Li XJ. Effect of I 'clear residual thyroid tissue' after surgery on the function of parathyroid gland in differentiated thyroid cancer. Exp Ther Med. 2015/12/17 ed2015. p. 2079-2082. PMID: 26668598. KQ5E3A
