Chapter 3

Calcitonin testing for detection of medullary thyroid cancer in patients with thyroid nodules

Hans H.G. Verbeek, Jan Willem B. de Groot, Wim J. Sluiter, Anneke C. Muller Kobold
Edwin R. van den Heuvel, John T.M. Plukker, Thera P. Links

Protocol published in: Cochrane Database of Systematic Reviews 2012, Issue 10
Review submitted
Abstract

**Background** Thyroid nodules are very common. Calcitonin is a sensitive tumour marker for the detection of medullary thyroid carcinoma (MTC). Although the European Thyroid Association's guideline advocates calcitonin determination in patients with thyroid nodules, the role of routine calcitonin testing in patients with thyroid nodules is still debatable.

**Objectives** The objective of this review was to determine the diagnostic accuracy of calcitonin testing in the detection of MTC in patients with thyroid nodules.

**Search methods** We searched *The Cochrane Library*, MEDLINE, EMBASE and Web of Science from inception to March 2013.

**Selection criteria** We included all retrospective and prospective cohort studies in which all patients with thyroid nodules had undergone determination of basal calcitonin levels (and stimulated calcitonin, if performed).

**Data collection and analysis** Two review authors independently scanned all retrieved records. Data was extracted by using a standard data extraction form. We assessed risk of bias and applicability using the QUADAS-2 (quality assessment of diagnostic accuracy studies) tool. We obtained summary estimates of the expected operating points (sensitivity and specificity) for each threshold using the HSROC model.

**Main results** In 16 studies, 73052 patients with nodular thyroid disease were identified. Prevalence of MTC was 0.26% (n=187). Summary estimates of sensitivity and specificity for basal calcitonin testing were 99.2% (95% CI 96.4%-100%) and 98.7% (95% CI 97.5%-100%) respectively. For stimulated calcitonin testing sensitivity was slightly lower (98.5%; 95% CI 93.9%-100%) while specificity was higher (99.9%; 95% CI 99.7%-100%). The positive predictive value (PPV) of basal calcitonin testing was 7.5% and for stimulated calcitonin testing 72%.

**Authors’ conclusions** Both basal and stimulated calcitonin testing have a high sensitivity and specificity. The value of routine testing in patients with thyroid nodules remains questionable, due to the low PPV of basal calcitonin testing. Whether routine calcitonin testing improves prognosis in MTC patients remains unclear.
Background

Thyroid nodules are very common in the general population, and they can be found in 2.3% to 6.9% of all adults.¹-³ Ultrasound detects an even higher frequency of thyroid nodules (17% to 69%).⁴ Thyroid nodules are more prevalent in women than in men (1.5% to 2% vs. 6.4% to 10%) and the incidence increases with age.¹,⁵,⁶ Of all patients with thyroid nodules who undergo fine needle aspiration (FNA), approximately 7.7% to 12% have thyroid cancer and in 3.3% to 3.7% of these patients medullary thyroid cancer (MTC) will be diagnosed.⁷-¹¹

MTC is a neuro-endocrine tumour originating from the parafollicular C-cells. These C-cells secrete calcitonin, a 32-amino acid peptide, which can be used as a sensitive tumour marker. The 10-year survival for MTC is about 75%, but the prognosis depends on the extent of the primary tumour, the presence of nodal disease and distant metastases.¹² The primary treatment for MTC is surgery, consisting of a total thyroidectomy with central compartment dissection and even more extensive lymph node dissection depending on the extent of the disease. Some patients develop recurrent disease, which limits the therapeutic options. Patients with progressive disease may benefit from newly developed targeted therapies, although early diagnosis of MTC and adequate surgical treatment remain crucial for a favourable prognosis.¹³

Calcitonin is elevated in virtually all MTC patients and therefore a very sensitive tumour marker, although MTC does not always produce calcitonin.¹⁴,¹⁵ On the other hand hypercalcitoninaemia can also be caused by other conditions such as thyroiditis, sepsis, hypercalcaemia, hypergastrinaemia, other neuroendocrine tumours, chronic renal failure, chronic pulmonary disease, acute trauma, inhalation injury and pseudohypoparathyroidism.¹⁶-¹⁹

In the recent guidelines of the American Thyroid Association (ATA) the diagnostic work-up of a thyroid nodule consists, after history, physical examination and TSH determination, of a diagnostic ultrasound and FNA when a nodule is seen on ultrasound. The role of calcitonin testing in the work-up of thyroid nodules is unclear and there is no clinical consensus on calcitonin testing. While the ATA’s revised evidence-based guidelines for thyroid cancer do not recommend for or against calcitonin determination, the European Thyroid Association's consensus-based guideline advocates calcitonin determination in all patients with thyroid nodules.²⁰-²² Based on these guidelines and several studies, routine calcitonin testing is practiced in multiple centres, while the use remains disputed.
Despite the high sensitivity and specificity, only a small number of patients with elevated calcitonin levels have MTC. This is due to the low prevalence of MTC. Accordingly, the positive predictive value (PPV) in most studies is low, although some studies do report PPVs of up to 100%.\textsuperscript{23} Furthermore, the cut-off level of calcitonin has not yet been established and there are indications that different subgroups of patients need specific cut-off points, since there are gender specific cut-off levels.\textsuperscript{24} Perhaps only a subset of patients should undergo calcitonin testing. It is also unclear whether calcitonin testing can contribute to longer overall survival or will increase the quality of life of MTC patients. Finally, to determine its role in the evaluation of thyroid nodules the cost-effectiveness of calcitonin testing is also important.\textsuperscript{25,26}

**Role of calcitonin testing**

There are several potential roles for calcitonin testing in the diagnostic work-up of thyroid nodules (Figure 1). First it can be used as a screening tool. Screening, however, implies that the entire healthy population will undergo determination of calcitonin, which is currently not effective or clinically relevant. Therefore we focus only on calcitonin testing in patients with thyroid nodules, detected through palpation or ultrasound. It can be performed in all patients with thyroid nodules at an early stage and before FNA (Figure 1: I). In this case the supposed sensitivity is very high but a great number of patients will have false positive results which might lead to unnecessary surgery (resulting in life-long thyroid hormone supplementation and risk of recurrent nerve damage and hypoparathyroidism). As FNA is also commonly used for diagnosing other types of thyroid cancer which do not secrete calcitonin, calcitonin testing as a replacement for FNA is irrational and clinically not relevant.

Calcitonin testing can be used as an add-on test after FNA in patients with suspicious or indeterminate cytology (Figure 1: II). In this case the number of false positives will be lower, but some MTC patients might be missed (when cytology is benign) with the risk that MTC in these patients will be diagnosed at a later stage or not at all. Calcitonin testing can also be used as a preoperative test in all patients who will undergo thyroid surgery (Figure 1: III). In that case not all MTC patients will be detected but the risk of patients who undergo an operation receiving too restricted surgery decreases. This form of calcitonin testing will not be included in this review as it is more focused on preoperative assessment of tumour type than on screening.
This review will address the value of calcitonin testing for diagnosing MTC in patients with thyroid nodules for the triage and add-on roles of the calcitonin test. We want to give more insight into the different sensitivities and specificities for these different roles. By providing data on the diagnostic accuracy of the calcitonin test in light of the low prevalence of MTC in thyroid nodules we want to contribute to the discussion on the role of the calcitonin test in patients with thyroid nodules.

![Figure 1](image-url) Possible roles of calcitonin testing
Index tests

The available test for diagnosing MTC in thyroid nodules is the calcitonin assay. The former radioimmunoassays for calcitonin measurement recognised the monomeric and the dimeric form of calcitonin, as well as its precursors leading to false-positive results. The more recent and most commonly used immunometric assays mainly recognise the mature, monomeric form of calcitonin. They rely on a ‘sandwich’ formation by two monoclonal or polyclonal antibodies recognising different epitopes on calcitonin.\(^{27}\) However, limitations still exist in the calcitonin assays. If a one-step assay is applied, in case of an extremely high calcitonin concentration, all the antibodies including the signal antibodies are saturated with the antigen, preventing a sandwich formation. Then, the antigen concentration measured may be falsely low (also known as the ‘high dose hook’).\(^{28}\) Furthermore, also mainly in one-step assays, the presence of heterophilic antibodies may give erroneously high results of calcitonin by cross-linking the antibodies in the absence of calcitonin.\(^{29,30}\) Very rarely ‘blocking’ heterophilic antibodies are also able to produce false-negative results.\(^{31}\) Alternative methods for quantification, such as mass spectrometry may circumvent this problem, as was also shown for thyroglobulin.\(^{32}\) Furthermore, despite the World Health Organization international reference preparation for human calcitonin, differences exist between the same type of assays of different manufacturers, making it even more difficult to compare results from different studies and to establish an optimal cut-off value.\(^{27,33,34}\)

To improve the specificity of the calcitonin assay, calcitonin stimulation tests with pentagastrin or calcium are used.\(^{35}\) These stimulation tests can distinguish calcitonin secreted by MTC from other sources of calcitonin but there are some limitations.\(^{36}\) Stimulation with pentagastrin can induce unpleasant side effects, such as nausea, vomiting or skin rash.\(^{37}\) Furthermore, pentagastrin is not available in several countries. Calcium stimulation tests are better tolerated but are not routinely used although an increasing number of small studies have advocated the use of calcium.\(^{38-40}\) We planned to perform a heterogeneity analysis on whether basal calcitonin, stimulated calcitonin, or both, were determined and also the type of stimulation test used.

Alternative tests

The alternative test for diagnosing MTC in patients with thyroid nodules is fine needle aspiration cytology (FNAC) with eventually immunohistochemical examination in suspicious lesions. FNAC is an accurate and cost-effective method for evaluation of thyroid nodules, but
the sensitivity for diagnosis of MTC is not optimal, ranging from 63% to 89%.\textsuperscript{41-43} The outcome of the FNAC in these studies resulted in surgery in 91% to 100% of patients. Although a large proportion of the patients received surgery despite incorrect FNAC results, this might be an inadequate test as MTC requires a different surgical approach than differentiated thyroid cancer. Other techniques, such as measuring calcitonin levels in washout fluids of fine needle aspirates may improve accuracy but few studies have reported on this in limited numbers of patients.\textsuperscript{44-46}

**Rationale**

A number of studies and reviews on this topic advocate calcitonin testing for detection of MTC. These studies are hard to compare, however, since they have different inclusion criteria and different cut-off points for calcitonin levels. Moreover, there is no consensus between the American and European guidelines on thyroid nodules. Calcitonin testing in patients with thyroid nodules is associated with a high rate of false-positive results and a low PPV. It has not been established that calcitonin testing reduces MTC-related mortality in these patients. Cheung et al. stated that calcitonin testing in the US is cost-effective at the same level as mammography screening and advocates calcitonin testing in subgroups of patients such as young men with larger thyroid nodules, but this also remains a matter for debate.\textsuperscript{26}

**Objectives**

The objective of this review was to determine the diagnostic accuracy of calcitonin testing in the detection of MTC in patients with thyroid nodules.

**Investigation of sources of heterogeneity**

We planned to investigate several potential sources of heterogeneity, including differences in cut-off values, assay types and different verification methods. Possible factors that were evaluated as source for heterogeneity were:

- Age.
- Gender.
- Nodules detected by palpation or ultrasound.
- Nodule size.
- Number of nodules.
- Sonographic morphology of thyroid nodules.
- FNA procedures performed through ultrasound guidance versus palpation.
- Basal versus stimulated calcitonin testing.

**Methods**

**Criteria for considering studies for this review**

**Types of studies**
We included all retrospective and prospective cohort studies in which all patients with thyroid nodules had undergone determination of basal calcitonin levels (and stimulated calcitonin, if performed).

**Participants**
We included patients with nodular thyroid disease (defined as solitary thyroid disease (toxic/non-toxic), multinodular thyroid disease (toxic/non-toxic), autonomously functioning thyroid nodule) found by palpation or on ultrasound in whom calcitonin testing was performed. We distinguished between studies in which calcitonin testing was performed as a triage (before FNAC) or as an add-on test (after FNAC). We included patients with coexisting non-nodular disease such as autoimmune thyroid disease (Graves' disease or Hashimoto's thyroiditis) and subacute thyroiditis. We excluded patients with only non-nodular thyroid disease. If studies included both patients with nodular and non-nodular disease, we included them only if it was possible to separate the calcitonin levels and surgical outcomes of these patient groups or if fewer than 10% of patients had non-nodular disease. We excluded patients with known sporadic or familiar MTC (MEN2A/B, FMTC) prior to calcitonin screening. We also excluded studies that included these patients and did not describe them separately.

**Index tests**
The index tests for this review included all serum tests used to determine basal and stimulated serum calcitonin levels.

**Target conditions**
The target condition was MTC.
Reference standards
The optimal clinical reference standard for diagnosis of MTC was considered histopathological examination of the thyroid after surgery of all patients, even patients without elevated calcitonin levels. In all of the studies, however, the problem of differential verification was encountered and only patients with (markedly) elevated calcitonin levels or patients with suspicious cytology had histological verification (although some patients did undergo surgery for other reasons e.g. mechanical complaints due to a multinodular goitre). We planned therefore to make use of other reference standards such as clinical follow-up. A follow-up of at least three years was considered adequate as most clinically relevant MTCs will be identified at that time, while longer follow-up carries the risk that MTC patients are diagnosed while not having the disease at the time of calcitonin testing. To determine whether standard of verification significantly influences accuracy, we planned to include method of verification in the heterogeneity analysis.

Search methods for identification of studies
Electronic searches
We used the following sources for the identification of trials.

- *The Cochrane Library.*
- MEDLINE.
- EMBASE.
- Web of Science.

For detailed search strategies please see Appendix 1. The Editorial Base of the Cochrane Metabolic and Endocrine Disorders Group provided support for generating the optimal search strategy. We used PubMed's 'My NCBI' (National Centre for Biotechnology Information) email alert service for the identification of newly published studies using a basic search strategy (see Appendix 1). We included studies published in English language.

Searching other resources
We examined the references lists of relevant publications for additional studies. We searched in PubMed for related articles of relevant studies.
Data collection and analysis

Selection of studies
To determine the studies to be assessed further, two review authors (HHGV, JWBG) independently scanned the abstract, title or both sections of every record retrieved. All potentially relevant articles were investigated as full text. Any disagreements were resolved by a third reviewer (TPL). A PRISMA (preferred reporting items for systematic reviews and meta-analyses) flow-chart of study selection was made.\textsuperscript{47}

Data extraction and management
We extracted data on study design and study population using a standard data extraction form (Appendix 2), in which we included the following items:

- Study design.
- Included number of patients.
- Inclusion and exclusion criteria.
- General patient characteristics.
- Type of calcitonin assay and cut-off values.
- Number of patients with nodular thyroid disease.
- Number of patients with palpable nodules and/or nodules on ultrasound.
- Number of patients who had undergone calcitonin testing and number of positive patients.
- Number of patients operated and reason for operation.
- Number of patients with known follow-up and outcome of follow-up.
- Histological outcome of patients operated.
- Number of patients with MTC.

Assessment of methodological quality
We assessed risk of bias and applicability using the QUADAS-2 (quality assessment of diagnostic accuracy studies) tool. We rated each of the four key domains (patient selection, index test, reference standard, flow and timing) using the signalling questions as developed by the QUADAS-2 group.\textsuperscript{48} The criteria for each signalling question are provided in Appendix 3. We scored all items in the QUADAS-2 tool as ‘yes’, ‘no’ or ‘unclear’, and used graphs to present overall scores of risk of bias and applicability for each domain.
Statistical analysis and data synthesis
We incorporated true positives, false positives, true negatives and false negatives of each study in a 2x2 table and calculated test sensitivity and specificity with corresponding 95% confidence intervals. For extraction of data, we used pre-specified cut-offs based on previous literature with different cut-offs for basal and stimulated calcitonin levels. These cut-off values were 10, 15, 20, 30, 50 and 100 pg/ml for basal calcitonin levels and 100 pg/ml and 200 pg/ml for stimulated calcitonin levels. We entered the data into RevMan 5.2.3, to graphically present coupled forest plots, showing the pairs of sensitivity and specificity of each study, for each threshold.

Investigations of heterogeneity
We used SAS software for meta-analysis. We obtained summary estimates of the expected operating points (sensitivity and specificity) for each threshold using the HSROC model. Depending on the number of included studies and available data, covariates were added in this model, for investigation of possible sources of heterogeneity.

Sensitivity analyses
We performed sensitivity analyses on the different domains scored on the QUADAS-2 tool, in order to explore the influence of the quality of the included studies.

Results

Results of the search
A total of 2947 unique records were identified by our search in January 2012 and updated searches in June 2012 and March 2013. An additional two records were identified by examining references list of relevant publications. One other relevant publication was also included. Screening of all records resulted in 35 publications that were eligible for further evaluation. After assessment 19 articles were excluded. Eventually 16 studies were included in this review (Figure 2).
Included studies
Characteristics of the 16 included studies are shown in the table Characteristics of included studies (Appendix 4). A total of 73052 patients with nodular thyroid disease were included in these studies, of which 72368 underwent basal calcitonin testing with or without stimulated calcitonin testing as shown in Table 1. A total of 187 MTC patients were identified. Three studies performed only basal calcitonin testing, whereas in thirteen studies both basal and stimulated calcitonin testing was performed.

![Image of study flow diagram]

Figure 2 Study flow diagram.

Calcitonin assays
Two studies used an radio immunometric assay (RIA) for determination of calcitonin, including one study which during the study period switched from a RIA assay to an immunoradiometric assay (IRMA). Five other studies used also an IRMA assay.
## Table 1: Overview of study populations

<table>
<thead>
<tr>
<th>Study ID</th>
<th>[n] with nodular thyroid disease</th>
<th>[n] with calcitonin testing</th>
<th>[n] with positive basal calcitonin testing</th>
<th>[n] with stimulated calcitonin testing</th>
<th>[n] with positive stimulated calcitonin testing</th>
<th>[n] operated</th>
<th>[n] with follow-up</th>
<th>[n] with MTC</th>
<th>MTC prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rieu 1995</td>
<td>469</td>
<td>469</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>NR</td>
<td>4</td>
<td>0.85</td>
</tr>
<tr>
<td>Ozgen 1999</td>
<td>773</td>
<td>773</td>
<td>4</td>
<td>-</td>
<td>175</td>
<td>3</td>
<td>4</td>
<td></td>
<td>0.52</td>
</tr>
<tr>
<td>Hahm 2001</td>
<td>1448</td>
<td>1448</td>
<td>56</td>
<td>39</td>
<td>12</td>
<td>194</td>
<td>NR</td>
<td>10</td>
<td>0.69</td>
</tr>
<tr>
<td>Hatzl-Griesenhofer 2002</td>
<td>3899</td>
<td>3899</td>
<td>230</td>
<td>157</td>
<td>30</td>
<td>39</td>
<td>43</td>
<td>12</td>
<td>0.31</td>
</tr>
<tr>
<td>Elisei 2004</td>
<td>10864</td>
<td>10864</td>
<td>47</td>
<td>45</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>0.41</td>
</tr>
<tr>
<td>Karankas 2004</td>
<td>195</td>
<td>195</td>
<td>13</td>
<td>13</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.51</td>
</tr>
<tr>
<td>Vierhapper 2005</td>
<td>10292</td>
<td>10157</td>
<td>507</td>
<td>481</td>
<td>103</td>
<td>76</td>
<td>32</td>
<td>36</td>
<td>0.35</td>
</tr>
<tr>
<td>Papi 2006</td>
<td>1474</td>
<td>1425</td>
<td>23</td>
<td>19</td>
<td>6</td>
<td>315</td>
<td>NR</td>
<td>9</td>
<td>0.63</td>
</tr>
<tr>
<td>Scheutz 2006</td>
<td>105</td>
<td>105</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>NR</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Costante 2007</td>
<td>5817</td>
<td>5817</td>
<td>282</td>
<td>58</td>
<td>17</td>
<td>747</td>
<td>212</td>
<td>15</td>
<td>0.26</td>
</tr>
<tr>
<td>Rink 2009</td>
<td>21928</td>
<td>21928</td>
<td>885</td>
<td>218</td>
<td>62</td>
<td>214</td>
<td>28</td>
<td></td>
<td>0.13</td>
</tr>
<tr>
<td>Hasselgren 2010</td>
<td>959</td>
<td>702</td>
<td>39</td>
<td>np</td>
<td>-</td>
<td>246</td>
<td>702*</td>
<td>6</td>
<td>0.85</td>
</tr>
<tr>
<td>Hermann 2010</td>
<td>1007</td>
<td>1007</td>
<td>17</td>
<td>16</td>
<td>4</td>
<td>5</td>
<td>12</td>
<td>2</td>
<td>0.20</td>
</tr>
<tr>
<td>Schneider 2012</td>
<td>11270</td>
<td>11270</td>
<td>32</td>
<td>14</td>
<td>12</td>
<td>18</td>
<td>10</td>
<td>12</td>
<td>0.11</td>
</tr>
<tr>
<td>Giovanella 2012</td>
<td>1479</td>
<td>1236</td>
<td>14</td>
<td>14</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>0.16</td>
</tr>
<tr>
<td>Grani 2012</td>
<td>1073</td>
<td>1073</td>
<td>41</td>
<td>np</td>
<td>-</td>
<td>3</td>
<td>NR</td>
<td>2</td>
<td>0.19</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>73052</strong></td>
<td><strong>72368</strong></td>
<td><strong>2199</strong></td>
<td><strong>1083</strong></td>
<td><strong>300</strong></td>
<td><strong>2042</strong></td>
<td><strong>1280</strong></td>
<td><strong>187</strong></td>
<td><strong>0.26</strong></td>
</tr>
</tbody>
</table>

*“-” denotes not reported. * Cross linkage with Danish Thyroid Cancer Database. Abbreviations: MTC: medullary thyroid cancer, np: Not performed
Two of these five studies switched during the study period to a chemiluminescence assay (ICMA). The remaining nine studies used an ICMA assay. In conclusion, thirteen studies used only one calcitonin assay during their study period, while three studies used two assays. One of these three studies, that switched from an IRMA to an ICMA assay, used the ICMA assay only in 14 out of 702 patients, and was therefore in further analyses regarded as using an IRMA assay. The other two studies that switched from calcitonin assay were not included in the covariate analysis regarding assay type.

In total, calcitonin assays of nine different manufacturers were used (Appendix 4; Characteristics of included studies). Especially in the seven studies using a RIA or IRMA assay a large heterogeneity in manufacturers was present (n=7); some studies used during the study period assays from 2 different producers. Within the nine studies using a ICMA assay, one study did not report the manufacturer, while in the other studies an assay was used from one of two producers.

Verification method

Differential verification was present in all studies; all patients with a (highly) elevated basal and/or stimulated calcitonin underwent surgery, while only a subset of patients with negative calcitonin tests had surgery. We considered clinical follow-up of calcitonin negative patients as an appropriate alternative for detection of missed MTC patients. However, none of the included studies did report on clinical follow-up of all of their calcitonin negative patients. Only in the study of Hasselgren follow-up was performed that consisted of cross linkage with a national thyroid cancer database.

Calcitonin as triage or add-on test

None of the studies included, provided explicit information on the role of calcitonin testing in the diagnostic pathway of thyroid nodules. In nine studies FNA was described in the materials and methods section as part of the diagnostic protocol. Most of these studies stated that surgery was indicated if basal or stimulated calcitonin was clearly elevated (e.g. >100 pg/ml) regardless of the results of FNA. In these studies the role of calcitonin testing can be considered as a triage test in which calcitonin positive patients are subjected to surgery, while calcitonin negative patients require more diagnostic work-up in the form of FNA. In all studies in which FNA was not described in the diagnostic protocol, calcitonin testing was performed in all included patients, independent of another diagnostic procedure, and if
markedly elevated an indication to perform surgery. Therefore in these studies calcitonin testing was also regarded as a triage test.

**Patient and study characteristics**

Average and/or median age was described in twelve studies, but only one study reported the results of calcitonin testing specified in different age groups. Information on gender of the included patients was provided in 15 studies, although only in seven studies detailed information on outcome was given for both sexes. In nine studies information was available on whether thyroid nodules in the included patients were detected through palpation or US; four studies included patients with thyroid nodules found by US and five studies included patients with thyroid nodules detected through US or palpation. With regard to nodule size only one study provided information on summary measures of size for the included patients, although no detailed information was provided for patients with elevated calcitonin levels. No study presented information on number of nodules or US morphology of all patients. In four studies information was given on whether FNA procedures were performed through palpation or US; in one study both techniques were performed, in the three others US-guided FNA was performed.

**Excluded studies**

In the table Characteristics of excluded studies (Appendix 5) reasons for exclusion for the 19 excluded studies are shown. Of six studies only a meeting abstract was available and no full text article was published. Four articles were written in non-English language. Two studies used a study population that was also described in a later publication. Three studies did not specify the numbers of patients with thyroid nodules. Three studies reported on calcitonin testing in pre-operative patients, which is not the topic of this review. One study reported on calcitonin testing in isthmic thyroid nodules; because this patient group evaluated only nodules in a specific part of the thyroid, we excluded this study.

**Methodological quality of included studies**

In Figure 3 the overall quality of the 16 included studies is shown, with regard to the risk of bias and concerns about applicability scored according to the QUADAS 2 domains. In the domain Patient selection, one study scored high on the risk of bias as patients were included who showed evidence of growth during follow-up examinations. This might have increased
the rate of included patients with a malignancy. In all studies the risk of bias by the conduction or interpretation of the calcitonin test was scored low. The risk of bias with the conduct or interpretation of the reference standard was unclear in all studies, for the reference standard in calcitonin negative patients, was not described. Due to this lack of reference standard, resulting in a verification bias, the risk of bias with regard to flow and timing was in all studies expect one regarded as high. In the only study using a cross linkage with a national thyroid cancer database the risk was scored as unclear.61 No concerns of applicability existed in all studies. In Figure 4 the individual quality assessment of all studies can be found.

Figure 3  Risk of bias and applicability concerns graph: review authors’ judgements about each domain presented as percentages across included studies

Findings
The sensitivity of the reported basal calcitonin testing cut-off in the included studies ranged from 83% to 100%, while the specificity ranged from 94% to 100% (Figure 5 and Figure 6). In the study of Schuetz et al. no MTC patients with nodular thyroid disease were identified so sensitivity could not be calculated.57 The summary estimates of sensitivity and specificity were 99.2% and 98.7% respectively (95% Confidence intervals (CI) 96.4%-100% and 97.5%-100% respectively).

Effect of cut-off value
We extracted data for several cut-off values from the included studies, ranging from 10 pg/ml to 100 pg/ml. With the cut-off value of 10 pg/ml the sensitivity of the individual studies ranged from 92% to 100% while specificity ranged from 95% to 99%; summary estimates were 99.9% (95% CI: 99.1-100%) for sensitivity and 96.8% (95% CI: 95.5%-98.1%) for specificity (Table 2). With the highest cut-off value of 100 pg/ml sensitivities ranged from 42%-100% and specificity from 95%-100%. Summary estimates of sensitivity and specificity
for the different cut-off values of basal calcitonin >10 pg/ml could not be calculated due to limited number of studies.

**Figure 4** Risk of bias and applicability concerns summary: review authors’ judgements about each domain for each included study

**Effect of stimulated calcitonin**

Thirteen studies were included in the analysis of stimulated calcitonin. Sensitivity of individual studies ranged between 82% and 100%. Specificity ranged from 99% to 100% (Figure 7). Summary estimates were 98.5% (95% CI 93.9%-100%) for sensitivity and 99.9% (95% CI: 99.7%-100%) for specificity, respectively (Summary of findings Table 1, Figure 8).
Due to limited numbers of studies no summary estimates could be calculated for different cut-off values.

Table 1 Summary of findings

<table>
<thead>
<tr>
<th></th>
<th>Summary points (95% CI)</th>
<th>No. of patients (studies)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basal calcitonin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported cut-off value</td>
<td>Sensitivity 99.2% (96.4%-100%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Specificity 98.7% (97.5%-99.9%)</td>
<td></td>
</tr>
<tr>
<td>10 pg/ml</td>
<td>Sensitivity 99.9% (99.1%-100%)</td>
<td>72369 (16)</td>
</tr>
<tr>
<td></td>
<td>Specificity 96.8% (95.5%-98.1%)</td>
<td>44393 (10)</td>
</tr>
<tr>
<td><strong>Basal and stimulated reported calcitonin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported cut-off value</td>
<td>Sensitivity 98.5% (93.9%-100%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Specificity 99.9% (99.7%-100%)</td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup analysis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Sensitivity 95.9% (83.1%-100%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Specificity 99.0% (97.6%-100%)</td>
<td>14858 (6)</td>
</tr>
<tr>
<td>Male</td>
<td>Sensitivity 100% (NA-NA*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Specificity 98.2% (95.9%-100%)</td>
<td>4339 (6)</td>
</tr>
</tbody>
</table>

Figure 5 Forest plot of basal reported cut-off values.
Effect of gender

In Summary of findings Table 1 summary estimates for basal reported calcitonin for subgroups of females and males are reported. For females the sensitivity ranged between 96% and 100% and the specificity ranged between 96% and 100%. For males the sensitivity ranged between 82% and 100% and the specificity ranged and between 90% and 100% (Figure 9). Summary estimates for females were 95.9% (95% CI 83.1%-100%) for sensitivity and 99.0% (95% CI: 97.6%-100%) for specificity. Summary estimates for males were 100% (95% CI NA) and 98.2% (95% CI: 95.9%-100%) for sensitivity and specificity respectively. Only one study used gender specific basal calcitonin cut-off values for all included patients. Another study which used two assays during the study period also had a gender specific cut-off for the second assay, but this concerned only 14 patients. The study of Rink et al. used gender specific stimulated calcitonin cut-off values. Due to limited number of studies no summary estimates could be calculated for different cut-off values.
Effect of assay type, detection method of thyroid nodules and FNA method
Due to limited number of studies no summary estimates could be calculated for subgroups with respect to assay type or manufacturer, detection method of thyroid nodules and method of FNA.
Sensitivity analysis
We planned to perform a sensitivity analysis with regard to quality items scored with the QUADAS 2 tool. However, no large differences were seen between studies regarding quality items. Only two studies had aberrant scores with regard to the risk of bias on the domains of patient selection and flow and timing (Figure 4).

Figure 9  Forest plot of basal calcitonin reported cut-off values for females and males.
Discussion

Summary of main results
In this review we included 16 studies for determination of the diagnostic accuracy of calcitonin testing in patients with thyroid nodules. We found high summary estimates of sensitivity and specificity for the reported basal calcitonin cut-off value of all studies. For reported basal and stimulated calcitonin sensitivity was slightly lower and specificity slightly increased. In subgroup analysis, sensitivity in females was lower, while specificity was higher compared to males.

Strengths and weaknesses of the review
We evaluated the diagnostic accuracy of calcitonin testing for detection of medullary thyroid cancer in patients with thyroid nodules with a comprehensive search of literature, and performing a formal diagnostic meta-analysis.

One of the major limitations of this review is the lack of adequate reference standards for calcitonin negative patients in nearly all included studies. This increases the risk that false negative patients are missed and the reported sensitivities are overestimated. In two of the included studies false negative MTC patients were identified, although in the study of Vierhapper et al. two of the three false negative patients were not operated and a histological diagnosis was not obtained.56 In studies performing preoperative calcitonin testing, the rate of false negative MTC patients ranged between 4.3%-12.5% of all MTC patients identified.18,83 Because the prevalence of MTC is low in patients with thyroid nodules, even a small number of false negative patients can markedly affect sensitivity. The clinical relevance of these false negative MTC can be discussed as in most calcitonin testing studies this are patients with a micro MTC without nodal metastasis. However, reports also exist on MTC patients with more aggressive disease and undetectable calcitonin levels.85

Another limitation of this study is the small number of studies that could be included in final analyses. Due to this small number only for a few subgroups summary estimates could be calculated. Furthermore, no optimal cut point could be assessed due to the heterogeneity of the used cut-off levels in the different studies. This was further complicated by the large heterogeneity in assay types and manufacturers.
Other reports

Other reports have been published on the value of calcitonin testing in the detection of medullary thyroid carcinoma, although no systematic reviews were performed. Daniels provided an overview of 15 studies but included also patients with preoperative calcitonin testing and multiple studies of one study group. The author concluded that due to the large reservoir of undetected of medullary thyroid micro carcinoma's of uncertain malignant potential, and the unavailability of pentagastrin in the US, calcitonin testing is not indicated in the United States and Canada. Costante et al. evaluated 11 studies in their review and concluded that the question whether to routinely measure calcitonin remained unsolved because no evidence exists whether testing actually reduces MTC-related mortality.

Applicability of findings to the review question

This review provides summary estimates of sensitivity and specificity for basal and stimulated calcitonin. The role of calcitonin testing in the diagnostic evaluation in thyroid nodules remains unclear. The final purpose of calcitonin testing is to detect MTC patients in an early stage, in which the chance of biochemical cure improves and the prognosis of patients. The findings of this review indicate that calcitonin testing is a very sensitive and specific test, but this has to be interpreted bearing in mind the low prevalence of MTC. The positive predictive value of calcitonin testing is therefore low, especially with lower cut-off values. Although several conditions are known to cause increased calcitonin levels, still in a fairly large proportion of patients with elevated calcitonin levels MTC cannot be excluded. Repeated calcitonin testing and follow up in these patients is therefore warranted. A number of these patients will be operated without histological evidence of a MTC. Some patients will have C-Cell hyperplasia, but the clinical relevance of this finding and its malignant potential remain unclear.

Cost effectiveness in health care becomes more and more important. In this review no formal cost effectiveness analysis is performed, so no validated statements can be made. However, Cheung et al performed in 2008 a cost effectiveness analysis in which calcitonin testing was concluded to be cost effective similar to colonoscopy and mammography screening. In their hypothetical model, several parameters had an important influence on cost-effectiveness, such as specificity of the calcitonin test and prevalence of MTC. Cheung et al. used a cut-off value of 50 pg/ml with a specificity of 98 % in the base line model, almost similar to the 98.7% specificity we found of the basal calcitonin test. However, the MTC
prevalence established in our review was 0.26%; one third of the prevalence used by Cheung et al. With a three times lower prevalence, costs will also increase almost three times. In a review of autopsy studies the prevalence of occult MTC was estimated to be 0.14%. Although it is not known if all MTC's detected at autopsy are clinically irrelevant and will be detected through calcitonin screening, a proportion of these tumours will be, further lowering the prevalence of clinically relevant MTC's. Our findings with regard to sensitivity, specificity and MTC prevalence, applied to the cost-effectiveness model of Cheung et al. imply that basal calcitonin testing does not seem to be cost effective. The effects on cost-effectiveness of a combined basal and stimulated calcitonin on cost effectiveness is more difficult to estimate. In their model Cheung et al. give a sensitivity and specificity for this combined approach of 80% and 98% respectively while our summary estimates show both a higher sensitivity and specificity. However, it is likely that also this model is influenced by the prevalence of MTC, decreasing cost-effectiveness with lower prevalence.

**Implications for practice**

Calcitonin testing can be a sensitive and specific instrument for detecting MTC in thyroid nodules. However, due to the low prevalence of MTC its role as a screening tool remains unclear. If we apply our findings from a basal calcitonin test with a cut-off of 10 pg/ml to a population of 10000 patients with a MTC prevalence of 0.26% (mean prevalence of the included studies), 26 patients will have MTC. All these 26 patients will have an elevated basal calcitonin test, while 319 patients without MTC also will have an elevated basal calcitonin. The positive predictive value of the calcitonin test in this situation is 7.5%. Surely not all patients with an elevated basal calcitonin will be operated, but even if a cause of elevated calcitonin can be found in 90% of the false positive patients, more than 50% (n=32) of the patients will be operated unnecessary. Increasing cut-off values results in a higher positive predictive value but at the cost of missed MTC patients. An optimal cut-off value for is therefore difficult to generate. Also the variation between assays used in different studies makes it hard to establish a common cut-off value. Adding a stimulated calcitonin test increases specificity with a very little effect on sensitivity. Still all 26 patients with MTC will be detected and only 10 patients will have a false positive stimulated calcitonin test. In this scenario, the positive predictive value increases almost 10 times to 72%. However, the most commonly used stimulative, pentagastrin, is not available in several countries.

The major reason to perform calcitonin testing is to ultimately improve prognosis of MTC patients. The supposed value of calcitonin testing is the detection of MTC patients in an
earlier stage in which biochemical cure is still possible. However, to assess this, one has to know which of the MTC patients would not have been detected through regular examinations (US/FNA/optional calcitonin), and which of these detected patients by calcitonin testing have or will develop a clinically relevant MTC. Some studies demonstrate that survival was significantly improved after introduction of routine calcitonin testing in patients with thyroid nodules.\textsuperscript{54,89,90} However, these studies have made their comparison with a historical group, and other factors might also have contributed to improved survival. These factors include, for instance, improved surgical treatment strategies and use of ultrasound. Furthermore, it is also interesting to note that MTC was supposedly detected at an earlier stage in these studies, but the age of the screened MTC patients was not lower compared to the patients that were not screened.\textsuperscript{54,89} As MTC is considered to be a slow growing tumour which takes several years to become clinically evident, the fact that MTC patients in the screened groups are of equal or even higher age, might indicate that additional MTC patients have been detected who would have otherwise had an indolent course of their disease. Another indication that otherwise undetected and possibly indolent MTC patients are identified is the increased number of MTC patients detected in shorter periods in the calcitonin tested patients compared to the historical cohorts.\textsuperscript{54,90}

**Implications for research**

This review shows that the diagnostic accuracy of calcitonin testing in MTC is high. However, this conclusion is based on studies in which the MTC prevalence in calcitonin negative patients might have been underestimated. Future studies should therefore report more accurately on the follow-up of calcitonin negative patients, to ensure that no MTC patients are missed, and also to provide more information on the clinical behaviour of these tumours. Furthermore accurate reporting of assay type and manufacturer is crucial for establishing optimal cut-off points for diagnosis of MTC. Also the role of the calcitonin test, being a triage or add-on test next to FNA (with or without measurement of calcitonin in washout fluids) should be further evaluated. Furthermore reporting results for subgroups may also identify subgroups with a higher MTC prevalence in which calcitonin testing can be more cost effective.
References

Calcitonin for detection of MTC


