Clinical outcomes after estimated versus calculated activity of radioiodine for the treatment of hyperthyroidism: systematic review and meta-analysis

A de Rooij1,2, J P Vandenbroucke1, J W A Smit2, M P M Stokkel3 and O M Dekkers1,2

Departments of 1Clinical Epidemiology C7-R, 2Endocrinology and Metabolic Diseases and 3Nuclear Medicine, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands

(Correspondence should be addressed to O M Dekkers at Department of Clinical Epidemiology C7-R, Leiden University Medical Center; Email: o.m.dekkers@lumc.nl)

Abstract

Background: Despite the long experience with radioiodine for hyperthyroidism, controversy remains regarding the optimal method to determine the activity that is required to achieve long-term euthyroidism.

Objectives: To compare the effect of estimated versus calculated activity of radioiodine in hyperthyroidism.

Design: Systematic review and meta-analysis.

Methods: We searched the databases Medline, EMBASE, Web of Science, and Cochrane Library for randomized and nonrandomized studies, comparing the effect of activity estimation methods with dosimetry for hyperthyroidism. The main outcome measure was the frequency of treatment success, defined as persistent euthyroidism after radioiodine treatment at the end of follow-up in the dose estimated and calculated dosimetry group. Furthermore, we assessed the cure rates of hyperthyroidism.

Results: Three randomized and five nonrandomized studies, comparing the effect of estimated versus calculated activity of radioiodine on clinical outcomes for the treatment of hyperthyroidism, were included. The weighted mean relative frequency of successful treatment outcome (euthyroidism) was 1.03 (95% confidence interval (CI) 0.91–1.16) for estimated versus calculated activity; the weighted mean relative frequency of cure of hyperthyroidism (eu- or hypothyroidism) was 1.03 (95% CI 0.96–1.10). Subgroup analysis showed a relative frequency of euthyroidism of 1.03 (95% CI 0.96–1.19) for Graves’ disease and of 1.05 (95% CI 0.91–1.19) for toxic multinodular goiter.

Conclusion: The two main methods used to determine the activity in the treatment of hyperthyroidism with radioiodine, estimated and calculated, resulted in an equally successful treatment outcome. However, the heterogeneity of the included studies is a strong limitation that prevents a definitive conclusion from this meta-analysis.

European Journal of Endocrinology 161 771–777

Introduction

Hyperthyroidism is a common disorder, with an estimated incidence varying from 25 cases per 100,000 persons per year in areas with high iodine intake to 38.7 per 100,000 in iodine deficient areas (1, 2). The most common causes of hyperthyroidism are Graves’ disease and toxic multinodular goiter. Hyperthyroidism is associated with increased morbidity mainly due to cardiovascular complications, especially in older patients (3). A prompt and preferably definite treatment is therefore desirable. Three treatment options are available for hyperthyroidism: antithyroid drugs, radioiodine-131 (I-131), and thyroidectomy. Antithyroid drugs are highly effective in controlling hyperthyroidism, but are known to have relapse rates up to 70% after 1 year treatment in Graves’ disease, whereas toxic multinodular goiter rarely resolves with drug treatment (4). Radioiodine treatment has been widely used as a definitive treatment of hyperthyroidism for over 60 years and has proven to be safe and cost-effective (5). It is the therapy of choice for most adults with hyperthyroidism in the USA; in Europe and Japan, it is usually second-line treatment after failure of antithyroid drug therapy (1, 6, 7).

Despite the long experience with radioiodine for hyperthyroidism, controversy remains regarding the optimal method to determine the activity that is required to achieve an optimal radiation dose to the thyroid gland and a subsequent optimal clinical outcome (8, 9). Administration of high I-131 activities, aimed at a rapid cure of hyperthyroidism, may result in a high incidence
of hypothyroidism, necessitating life-long hormonal substitution. On the other hand, low I-131 activities may cause a delay in achieving euthyroidism, thus prolonging the period at risk of cardiovascular complications, and may lead to higher recurrence rates. Many different strategies have been used to approach the ideal activity of I-131 for hyperthyroidism. In general, two methods are available to determine the activity: i) standardized activity estimation, using a fixed activity for all patients or roughly adapting the applied activity to individual thyroid size (10); ii) dosimetric calculation, based on 4- and 24-h radioiodine uptake in combination with an estimation of the weight of the thyroid gland. Many studies have compared the treatment outcomes of both methods, but results are conflicting, and no consensus has been reached regarding the best method to determine the activity for radioiodine.

We conducted a systematic review and meta-analysis of studies, aiming to compare the effects of two different methods of activity determination in nonablative radioiodine treatment for hyperthyroidism: the estimated versus the calculated method. The main aim of the systematic review and meta-analysis was to answer the question, which method of assessing the activity of radioiodine is most successful in eliminating hyperthyroidism without inducing hypothyroidism, i.e. in achieving a euthyroid outcome. In addition, we compared cure rates of hyperthyroidism (i.e. achieving a euthyroid or hypothyroid state) between the estimated and calculated method.

Methods

Search strategy

We searched the electronic databases Medline, EMBASE, Web of Science, and Cochrane Library from their inception to October 2007 for randomized and nonrandomized studies, comparing the effect of activity estimation methods with dosimetry on outcome of radioiodine for hyperthyroidism. The following terms were used: ‘hyperthyroidism’ (majr) OR hyperthyroid” (ti) OR ‘Graves’ disease’ (ti) OR Graves’ disease (ti) OR thyrotoxicosis (ti) OR thyroid crisis (ti) OR ‘goiter’ (majr) OR goiter (ti) AND (radioiodine (ti) OR ‘iodine radioisotopes’ (majr) OR iodine radioisotopes (ti) OR I-131 (ti) OR radioactive iodine (ti) OR radioiodinated (ti)) AND (therapeutic use OR treatment OR therapy OR therapeutic) AND (treatment outcome OR ‘clinical outcome’ OR ‘clinical outcomes’ OR ‘clinical assessment’ OR outcome assessment OR follow-up OR follow-up studies OR euthyroid” OR relapse OR recurrence OR convive OR hypothyroid OR hypothyroid*) NOT (‘neoplasms’ (majr) OR leukemia OR leukemia). We restricted this review to published trials; published abstracts and reports presented only on scientific meetings were not included. Language was restricted to English. In addition, reference lists of all included articles and relevant reviews were screened for additional studies. Data extraction was performed by A R. Data extraction of study endpoints (eu-, hypo-, and hyperthyroidism) was independently performed by a second extractor (O M D).

Study selection

For inclusion, studies had to compare the results of radioiodine after activity estimation (fixed activity or activity selection based on estimated thyroid weight) with calculated dosimetry (based on radioiodine uptake measurement). Randomized clinical trials as well as observational studies were eligible for inclusion. The minimum follow-up after radioiodine had to be 6 months; clinical outcomes (eu-, hypo-, and hyperthyroidism) had to be reported for the entire study population. Definitions of thyroid status from the individual studies were maintained. Studies on animals, children, or adolescents were excluded.

Data review and analysis

We divided activity selection methods in two groups: estimated (fixed activity or standardized activity based on thyroid size) or calculated (using individualized dosimetry), regardless of the nomenclature used in the individual articles. The main outcome measure of the present analysis was the frequency of treatment success, defined as the proportion of persistent euthyroidism after radioiodine treatment at the end of follow-up in the dose estimated and calculated dosimetry group. Furthermore, we assessed the cure rates of hyperthyroidism, defined as the proportion of euthyroidism as well as hypothyroidism at the end of follow-up in all studies. Relative frequencies of treatment outcomes were calculated for each included study for estimated versus calculated activity selection methods. For studies reporting on endpoints at several time points during follow-up, the time-point with the maximum number of patients was chosen for extraction of the endpoints. We aimed to calculate a weighted mean relative risk (RR) for all treatment outcomes (euthyroidism, hypothyroidism, hyperthyroidism, and cure from hyperthyroidism). The primary outcome measure was the Mantel–Haenszel risk ratio. We planned to perform a fixed effect model in the absence of heterogeneity or a random effect model in the case of evidence of heterogeneity. Heterogeneity was quantified using the I 2 test (11). We performed a meta-regression to estimate the effect of follow-up duration on treatment outcome. In addition, we planned to perform a subgroup analysis for randomized controlled trials and an analysis per diagnosis (Graves’ disease and toxic multinodular goiter). SPSS (Chicago, IL, USA) and Review Manager (Cochrane IMS) 4.2 were used for graphical and statistical analysis.
Results

Literature search

We identified 906 studies by database search (PubMed, n = 585; Embase, n = 588; Cochrane library, n = 79; Web of Science, n = 153; Fig. 1). Studies on animals (n = 20) and children or adolescents (n = 40) were excluded. From the remaining 846 studies, 835 were excluded after screening of title and abstract. Eleven studies were retrieved for full paper view; two publications involved the same study results, only the most recent publication was included (12, 13). In two studies, RRs could not be calculated for important treatment outcomes; these studies were therefore not included (14, 15). No extra studies were found after searching the citations of the included trials. A total of eight studies were included in the present analysis. For all included trials, rates of treatment success (euthyroidism), cure of hyperthyroidism (euthyroidism or hypothyroidism), and treatment failure (persistence or recurrence of hyperthyroidism) after I-131 therapy were reported or could be calculated.

Study characteristics

Three randomized and five nonrandomized trials were included in the present analysis. Detailed characteristics of the studies, including methods used for dosimetric calculation, are described in Table 1. Patients were treated with I-131 between 1964 and 1996, with a follow-up period ranging from 6 to 94 months. The number of included patients ranged from 52 to 326.

Three studies included only patients with Graves’ disease (10, 13, 16), two were restricted to patients with toxic multinodular goiter (17, 18), and three included all causes of hyperthyroidism (19–21). Administered activities of I-131 ranged from 177 to 747 MBq. In five studies, the estimated activity exceeded the calculated activity; in three studies, the calculated activity was higher. The mean difference in applied activities between the two methods varied from 10 to 374 MBq.

Treatment outcome

Treatment outcomes of all included studies are shown in Table 2. Reported frequencies of euthyroidism at the end of follow-up ranged from 4 to 73%. The lowest reported success rate after radioiodine treatment was 4%; in this study, patients were treated with an estimated activity of radioiodine; the associated rate of recurrent hyperthyroidism was 27% and hypothyroidism occurred in 69% (10). The highest reported rate of euthyroidism (73%) was found in patients treated with estimated activities, with 21% associated hyperthyroidism and induction of hypothyroidism in only 5% (17).

For euthyroidism, RR for the estimated activity versus the calculated activity ranged from 0.48 (10) to 1.32 (16). In all studies, the confidence interval (CI) for relative frequency of euthyroidism was large and included 1.0. For hypo- and hyperthyroidism, two studies reported significant differences. A RR of 0.30 (95% CI 0.14–0.65) of hypothyroidism for estimated activity versus calculated activity was reported, at the expense of a RR of persistent hyperthyroidism of 1.98 (95% CI 1.03–3.82) (16). On the other hand, a RR of 2.01 (95% CI 1.34–3.00) for hypothyroidism was associated with a RR of 0.68 (95% CI 0.46–1.00) for hyperthyroidism (13).

Meta-analysis

We aimed to pool the relative frequency of euthyroidism for all included studies (estimated versus calculated activity). The weighted mean relative frequency of euthyroidism for all studies was 1.03 (95% CI 0.91–1.16), showing no difference in successful treatment outcome after estimated activity or calculated activity (Fig. 2). I^2 test showed a low heterogeneity (11.5%), allowing the use of a fixed effect model. The use of a random effect model did not markedly influence the point estimate: this resulted in a RR of 1.06 (95% CI 0.93–1.20). Restriction to randomized controlled trials showed a relative frequency of euthyroidism of 0.92.

Figure 1 Identification of included studies.
### Table 1 Characteristics of included studies.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year of treatment</th>
<th>Study design</th>
<th>Diagnosis</th>
<th>Number of patients (estimated/calculated)</th>
<th>Mean follow-up (months)</th>
<th>Mean activity (MBq; estimated/calculated)</th>
<th>Method of dosimetry methods, and thyroid volume estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calegaro (16)</td>
<td>ND</td>
<td>RCT</td>
<td>GD</td>
<td>149 (50/99)</td>
<td>12</td>
<td>314/346</td>
<td>ETW, DA × ETW × 100 RAU24, Nuclear imaging + Palpation</td>
</tr>
<tr>
<td>Jarlov (19)</td>
<td>1989–1993</td>
<td>RCT</td>
<td>TA</td>
<td>163 (85/78)</td>
<td>12</td>
<td>324/316</td>
<td>ETW, DA × ETW × 100 RAU24, Nuclear imaging + Palpation</td>
</tr>
<tr>
<td>Khanna (17)</td>
<td>1989</td>
<td>Follow-up</td>
<td>GD</td>
<td>30 (75/25)</td>
<td>12</td>
<td>46/58</td>
<td>ETW, DA × ETW × 100 RAU24, Nuclear imaging + Palpation</td>
</tr>
<tr>
<td>Leslie (10)</td>
<td>1992–1996</td>
<td>RCT</td>
<td>TMG, TA</td>
<td>88 (45/43)</td>
<td>12</td>
<td>46/58</td>
<td>ETW, DA × ETW × 100 RAU24, Nuclear imaging + Palpation</td>
</tr>
<tr>
<td>Peters (13)</td>
<td>ND</td>
<td>RCT</td>
<td>GD</td>
<td>205 (98/107)</td>
<td>6</td>
<td>80/85</td>
<td>FA × 100 RAU124, Nuclear imaging + Palpation</td>
</tr>
<tr>
<td>Sun (18)</td>
<td>1984–1990</td>
<td>Follow-up</td>
<td>TMG, TA</td>
<td>52 (21/31)</td>
<td>6</td>
<td>30/32</td>
<td>FA × 100 RAU124, Nuclear imaging + Palpation</td>
</tr>
<tr>
<td>Ustun (21)</td>
<td>1991–2003</td>
<td>Follow-up</td>
<td>GD, TMG, TA</td>
<td>148 (78/70)</td>
<td>6</td>
<td>28/30</td>
<td>ETW, DA × ETW × 100 RAU24, Nuclear imaging + Palpation</td>
</tr>
</tbody>
</table>

- **ETW**: based on estimated thyroid weight; **RAIU**: radioactive iodine uptake (%); **RAIU**: radioactive iodine uptake (%); **ND**: no data.

#### Discussion

In the present meta-analysis, the effects of the two main methods of nonablative I-131 activity determination for hyperthyroidism were compared: estimated versus individually calculated activity methods. Pooling of the results from eight studies showed that estimation and calculation of applied I-131 activities resulted in similar rates of successful treatment outcome (i.e. achieving persistent euthyroidism) and cure of hyperthyroidism (i.e. achieving euthyroidism or hypothyroidism). However, it should be emphasized that large clinical heterogeneity between the studies, the use of different methods to determine radioactive iodine activities as well as the small number of randomized trials prevent a definitive conclusion being reached.

We are aware of several limitations of the present meta-analysis. First, the analysis comprised studies with different designs, with different characteristics of included patients, and different treatment settings. Although we compared two methods, it has to be realized that different formulas were used for the I-131 activity calculation and that different methods to determine the estimated activities were reported. In fact, only the study by Peters et al. presents a real dosimetric approach (13). A second limitation might be that including different causes of hyperthyroidism can result in heterogeneity. Although Graves’ disease, toxic adenoma, and toxic multinodular goiter all can be treated effectively with I-131, the latter is known to require higher activities of radiation than both other (95% CI 0.72–1.17, fixed effect model). Restriction to the five studies that used ultrasonography or nuclear imaging methods to determine thyroid volume showed a relative frequency of euthyroidism of 1.04 (95% CI 0.89–1.20, fixed effect model). Restriction to the five studies with a follow-up period of maximally 12 months showed a relative frequency of treatment success of 1.08 (95% RR 0.94–1.25).

In six studies, only patients with Graves’ disease were included, or data on Graves’ disease could be separated. The relative frequency for euthyroidism in this subgroup was 1.03 (95% CI 0.84–1.26). Exclusive data on toxic multinodular goiter were reported in five studies. In this subgroup, the relative frequency of euthyroidism was 1.05 (95% CI 0.91–1.19). I² test revealed moderate and low heterogeneities of 29.2 and 0% respectively.

The weighted mean frequency of cure of hyperthyroidism (i.e. euthyroidism or hypothyroidism at the end of follow-up) was 1.03 (95% CI 0.96–1.10) for estimated versus calculated dose (Fig. 3). I² test showed a moderate heterogeneity of 27%. The use of a random effect model revealed similar results: weighted mean frequency of 1.01 (95% CI 0.94–1.09).

Meta-regression showed no effect of follow-up duration on the difference in euthyroidism between estimated and calculated activity.
diagnoses (22). This is mainly due to the larger volume and lower radioiodine uptake in patients with toxic multinodular goiter (1, 23). However, subgroup analysis of treatment success according to etiology of the hyperthyroidism did not show differences between estimated versus calculated activity determination for Graves’ disease (RR 1.03) as well as toxic multinodular goiter (RR 1.05). Thirdly, for the majority of included studies, the follow-up period was limited to 12 months. It can be presumed that prolonged follow-up would have resulted in higher proportions of hypothyroidism after radioiodine therapy is about 3% (23). Not all authors report the use of antithyroid drugs prior to I-131 therapy properly. A slightly higher rate of treatment failure due to antithyroid drug use has been shown (24). This factor is corrected by uptake measurements in calculated methods, but not in estimated methods. In one study, patients were assigned to different intervention groups according to prior antithyroid drug use, but those who were pretreated received higher activities of radioiodine treatment (16). However, in most studies, either all patients received antithyroid drugs (13, 17, 19, 20) or the distribution of drug use was comparable between treatment groups (10, 18). Another source of heterogeneity between the studies is the use of palpation as the sole method to determine thyroid volume in three studies. This method is known to be suboptimal for determination of thyroid volume (25). Excluding the three studies using palpation as the method to determine thyroid volume did not materially influence the study results.

Ideally, the rate of euthyroidism should correlate with the applied activities of radioiodine, while too low activities would result in persistent hyperthyroidism and overdosing would induce hypothyroidism, in between these extremes an ideal activity would be observed. However, thus far the ideal way to determine activity does not exist, mainly because treatment success depends on the absorbed dose and not on the applied activity. Individualized methods make important assumptions with respect to the kinetics of radioiodine. All included studies used the 24-h radioactive iodine uptake to determine activity. However, this uptake method may be inaccurate due to the fact that the biological half-life of radioiodine is variable and that it assumes a first-order kinetic model, which may be too simplistic (23). Moreover, dosimetric formulas take into account individualized iodine metabolism, but ignore several other factors also known to affect treatment outcome, such as patient age, gender and severity of hyperthyroidism. Also the individual thyroid metabolism seems to fluctuate overtime. Thus, the uptake of a test activity of radioiodine does not necessarily correlate with the uptake of the therapeutic activity several days to weeks later (26). Since there will always be some time between the application of the uptake measurement and the therapy with I-131, this problem cannot be

<table>
<thead>
<tr>
<th>Study</th>
<th>Estimated</th>
<th>Calculated</th>
<th>Relative risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calegaro (16)</td>
<td>30/50 (60)</td>
<td>45/98 (46)</td>
<td>0.67 (0.37–1.21)</td>
</tr>
<tr>
<td>Jarlov (19)</td>
<td>49/85 (58)</td>
<td>38/78 (49)</td>
<td>1.28 (0.67–2.43)</td>
</tr>
<tr>
<td>Khanna (17)</td>
<td>55/75 (73)</td>
<td>59/72 (78)</td>
<td>0.97 (0.58–1.62)</td>
</tr>
<tr>
<td>Kok (20)</td>
<td>59/128 (46)</td>
<td>51/198 (26)</td>
<td>1.22 (0.70–2.12)</td>
</tr>
<tr>
<td>Leslie (10)</td>
<td>2/45 (4)</td>
<td>13/21 (62)</td>
<td>0.08 (0.01–0.53)</td>
</tr>
<tr>
<td>Peters (13)</td>
<td>24/98 (25)</td>
<td>28/78 (36)</td>
<td>0.71 (0.40–1.26)</td>
</tr>
<tr>
<td>Sun (18)</td>
<td>13/21 (62)</td>
<td>15/31 (48)</td>
<td>0.88 (0.36–2.15)</td>
</tr>
<tr>
<td>Ustun (21)</td>
<td>15/78 (19)</td>
<td>13/72 (18)</td>
<td>0.92 (0.38–2.23)</td>
</tr>
</tbody>
</table>
overcome with dosimetry. These considerations make clear that also calculated methods to determine activity may not be optimal.

We expected the difference between treatment arms to be larger according to the difference in applied activity. Surprisingly, in two studies with a large difference in administered activities, no difference in treatment outcome was shown (RR 1.06 and 1.00 respectively) (17, 20). This finding emphasizes that methods for activity determination are intrinsically imperfect and that applied activities do not correlate well with the absorbed dose. A third study with a large difference in administered activities showed better treatment outcome in the group receiving the highest mean activity (13). The fact that in three studies the applied activities were rather similar (16, 19, 21), points out that the difference between the two methods does not only lead to similar treatment effects, but that it also does not translate invariably into different applied activities.

It is important to consider what the main goal of I-131 therapy should be: long-term euthyroidism or cure from hyperthyroidism. If the main goal is cure from hyperthyroidism (ablative procedures), higher activities may be applied in order to prevent a prolonged hyperthyreotic state. However, induction of hypothyroidism, necessitating lifelong suppletion therapy, may be accompanied by intrinsic imperfection of hormonal replacement therapy (27). The goal of nonablative procedures is the adequate treatment of hyperthyroidism, without induction of hypothyroidism in a large proportion of patients. It was not the aim of the paper to decide on the goal of I-131 treatment. Irrespective of the treatment goal, radioactive treatments should be performed according to the ALARA (‘as low as reasonable achievable’) principle. This means for a predefined therapeutic goal, the lower the administered activity the better the treatment. This principle protects from unnecessary high administered activities. Both goals should be taken into account when considering the goal for radioactive iodine therapy in hyperthyroidism: the best clinical effect with the lowest possible activity. In three studies, the estimated activity was highest (13, 18, 19), and in the other five studies, the calculated activity was highest. This meta-analysis therefore provides no clear evidence that one of both methods results in consistently higher applied activities.

What is preferable: dose estimation or dose calculation? The calculation method is a relatively time-consuming procedure compared with estimation methods, requiring at least one extra visit to an outpatient clinic, adding to the complexity and cost of the procedure. On the other hand, no optimal fixed dose exists. Moreover, the absorbed dose depends on
geographical conditions and patient characteristics, and is variable especially in Graves’ disease. Some of these factors that determine the absorbed dose can be adjusted for in calculated methods. From a theoretical point of view, calculated methods may therefore be preferable, especially in Graves’ disease, although this did not translate in better treatment outcomes.

The present meta-analysis provided no convincing evidence for superiority of estimated or calculated activity determination methods of I-131 activity in hyperthyroidism. However, the heterogeneity in included studies is a strong limitation that prevents a definitive conclusion from this meta-analysis. Therefore, the ultimate verdict for the optimal method to determine the activity for hyperthyroidism treatment would need a large randomized trial.

**Declaration of interest**

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

**Funding**

This research did not receive any specific grant from any funding agency in the public, commercial, or not-for-profit sector.

**References**


Received 6 August 2009
Accepted 9 August 2009

www.eje-online.org