European Journal of Endocrinology (2000) **143** 727–731 ISSN 0804-4643

CLINICAL STUDY

Thyroid ultrasound compared with World Health Organization 1960 and 1994 palpation criteria for determination of goiter prevalence in regions of mild and severe iodine deficiency

Michael Zimmermann, Amina Saad[1], Sonja Hess, Toni Torresani[2] and Noureddine Chaouki[1]

Laboratory for Human Nutrition, Swiss Federal Institute of Technology, Zürich, Switzerland, [1]The Division of Nontransmissable Diseases, The Ministry of Health, Rabat, Morocco and [2]The Children's Hospital, Zürich, Switzerland

(Correspondence should be addressed to M Zimmermann, Laboratory for Human Nutrition, Swiss Federal Institute of Technology, Zürich, Seestrasse 72/Postfach 474, CH-8803 Rüschlikon, Switzerland; Email: michael.zimmermann@ilw.agrl.ethz.ch)

Abstract

Objectives: In 1994, WHO/International Council for the Control of Iodine Deficiency Disorders recommended replacing the WHO 1960 four-grade goiter classification with a simplified two-grade system. The effect of this change in criteria on the estimation of goiter prevalence in field studies is unclear. In areas of mild iodine deficiency disorders (IDD) where goiters are small, ultrasound is preferable to palpation to estimate goiter prevalence. However, in areas of moderate to severe IDD, goiter screening by palpation may be an acceptable alternative to thyroid ultrasound. To address these two issues, we compared WHO 1960 and 1994 criteria with thyroid ultrasound for determination of goiter prevalence in areas of mild and severe IDD in Morocco.

Design: A cross-sectional study of 400 six- to 13-year-old children from two mountain villages (Aït M'hamed and Brikcha) in rural Morocco was carried out.

Methods: Urinary iodine concentration (UI), whole blood TSH and serum thyroxine were measured. Thyroid size was graded by inspection and palpation by two examiners using both WHO 1960 and 1994 criteria. Thyroid volume was determined by ultrasound. Variation between examiners and examination methods was assessed. Sensitivity and specificity of the two classification systems compared with ultrasound were calculated.

Results: Median UIs in Aït M'hamed and Brikcha were 183 and 24 μg/l respectively. In Aït M'hamed, using 1960 and 1994 criteria, goiter prevalence was 21 and 26% respectively, compared with 13% by ultrasound. In Brikcha, with 1960 and 1994 criteria, goiter prevalence was 64 and 67% respectively, compared with 64% by ultrasound. Agreement between observers was better with the 1994 criteria than with the 1960 criteria in Aït M'hamed ($\kappa = 0.53$ and 0.47 respectively), while in Brikcha observer agreement was similar with the two systems ($\kappa = 0.67$). Using either the 1994 or 1960 criteria, agreement with ultrasound was only moderate in Aït M'hamed ($\kappa = 0.41$–0.44), but good in Brikcha ($\kappa = 0.55$–0.64). Overall, compared with ultrasound, sensitivity increased 3–4% using 1994 criteria, while specificity decreased 4–5%.

Conclusions: The WHO 1994 criteria are simpler to use than the 1960 criteria and provide increased sensitivity with only a small reduction in specificity. Agreement between observers is better with the 1994 criteria than with the 1960 criteria, particularly in areas of mild IDD. Like the 1960 criteria, the 1994 criteria overestimate goiter prevalence in areas of mild IDD, compared with ultrasound. However, the 1994 palpation criteria provide an accurate estimate of goiter prevalence in areas of severe IDD, and may be an acceptable and affordable alternative to thyroid ultrasound in these areas.

European Journal of Endocrinology **143** 727–731

Introduction

Goiter prevalence in school-age children is an important indicator of iodine deficiency disorders (IDD) in a population (1). The 1960 World Health Organization (WHO) criteria for classification of thyroid size defined four grades of goiter; goiter was present if the lateral lobes of the thyroid were larger than the terminal phalanges of the individual's thumbs (2). In 1994, a two-grade classification system was proposed by WHO/International Council for the Control of Iodine Deficiency Disorders (ICCIDD) as more practicable for field screening by trained non-medical personnel, such as schoolteachers (1). It defined goiter as any enlarged thyroid that is palpable and/or visible, which meant in some cases, smaller thyroids than before were to be

graded as goiter. Although the WHO 1994 criteria are simpler to use, the effect of the change in criteria on the estimation of goiter prevalence in field studies is unclear. In the single previous study comparing the two classification systems in an area of severe IDD in Tanzania, the newer criteria for goiter sharply decreased specificity and resulted in an extra 20–33% of children being diagnosed as goitrous (3). The authors suggested a return to the WHO 1960 criteria as the method of choice for goiter screening. However, the sensitivity and specificity of goiter grading depends to a large extent on the examiners and the characteristics of the study population. Additional data comparing the WHO 1960 and 1994 criteria, particularly in populations with mild to moderate IDD, would be valuable.

Goiter prevalence can also be estimated using thyroid ultrasound (1). In areas of mild IDD where the prevalence of visible goiters is low, sensitivity and specificity of palpation are poor, misclassification can be as high as 40%, and ultrasound is more reliable (1, 4, 5). However, there are few available data comparing ultrasound to palpation in areas of moderate to severe IDD. Peterson *et al.* recently reported that thyroid ultrasound in an area of severe IDD had an unacceptably low precision, and that palpation compared favorably with ultrasound in estimating goiter prevalence (3). In areas of moderate to severe IDD, goiter screening by palpation could be an acceptable and more affordable alternative, as thyroid ultrasound requires technical proficiency and is relatively expensive.

To address these issues, we compared the WHO 1960 and 1994 criteria with thyroid ultrasound for determination of goiter prevalence in areas of mild and severe IDD in Morocco.

Subjects and methods

The study was done in May 2000 in two mountain villages of rural Morocco. The village of Aït M'hamed is located in the High Atlas Mountains of central Morocco. Although formerly an area of endemic goiter (6), iodized salt (60–80 mg I/kg salt) has been widely available since 1997. The village of Brikcha is in the Rif Mountains of northern Morocco, where a 1993 survey reported goiter rates of 50–65% (6). Salt in Brikcha is locally produced and non-iodized. The study protocol was approved by the Ministries of Health and Education in Morocco and the Ethics Committee of the Swiss Federal Institute of Technology in Zürich. Informed oral consent was obtained from the chief medical officers of the villages as well as the teachers and parents of the children in the study.

We sampled 200 children from primary schools in each village. The children were registered and their age, sex, height and weight were recorded (7). A spot urine sample was collected and 5 ml whole blood were collected by venipuncture into heparinized tubes. In each child, goiter was carefully classified by inspection and palpation from the front using both the WHO 1960 and 1994 criteria by two physicians. Examiner A was a Moroccan pediatrician who directs the national iodized salt program; examiner B was a Swiss endocrinologist. Both had had extensive experience in goiter surveys. Thyroid volume was measured in the children sitting upright with the neck extended using an Aloka SSD-500 echocamera (Aloka, Mure, Japan) with a high-resolution 7.5 MHz 4 cm linear transducer (8). The ultrasound examiner had had extensive experience using ultrasonography in goiter surveys in Europe and West Africa. During the study, the three examiners were unaware of each other's results.

Children were defined as goitrous by the 1960 criteria if their thyroid was grade 1A (palpable lobes larger than the terminal phalanges of the subject's thumbs), 1B (visible with the neck extended), 2 (visible with the head in normal position) or 3 (visible at a distance) (2). They were defined as goitrous by the 1994 criteria if their thyroid was palpable but not visible with the head in a normal position (grade 1) or if the thyroid was visible with the neck in the normal position (grade 2) (1). Thyroid volume was determined using the method of Brunn *et al.* (9). Maximal depth, width and height of each lobe were measured on transverse and longitudinal images. The volume of each lobe was calculated using the formula $\text{depth} \times \text{width} \times \text{height} \times 0.479$, and the lobe volumes were summed. In countries with a high prevalence of child growth retardation, thyroid volume is considered to be more directly a function of total body surface area (BSA) than of age (10). BSA was calculated using the formula $\text{weight}^{0.425} \times \text{height}^{0.725} \times 71.84 \times 10^{-4}$, with weight in kg and height in cm. Current WHO/ICCIDD upper limits for thyroid volume in children aged 6–12 years by sex (range 5.0–11.7 ml) and BSA (range 4.7–11.9 ml) were used to define the presence or absence of goiter (10). Urine samples were aliquoted and frozen at −20 °C until analysis. Urinary iodine concentration (UI) was measured using the Pino modification of the Sandell–Kolthoff reaction (11) with ammonium persulfate as the oxidizing reagent. Ammonium persulfate has significant advantages in that it is non-hazardous and non-explosive and results using this method correlate closely with the reference chloric acid method ($r = 0.994$) (11). Whole blood was spotted onto filter paper for measurement of whole blood thyrotropin (TSH) and serum thyroxine (T_4) using immunoassay (normal reference values: whole blood TSH, <3.5 mU/l; serum T_4, 65–165 nmol/l) (12).

Statistics

Data processing and statistics were done using SPLUS 2000 (Mathsoft, Seattle, WA, USA) and Excel (Microsoft, Seattle, WA, USA). The sensitivity and specificity of goiter diagnosis by palpation were calculated compared with the 'gold standard' of ultrasound. κ values were calculated to estimate agreement between

Table 1 Goiter prevalence by palpation by examiners A and B using the WHO 1960 and 1994 criteria compared with thyroid ultrasound. Values are numbers of children (percentages).

	WHO 1960			WHO 1994			
	Grade	A	B	Grade	A	B	**Ultrasound**
All ($n = 400$)	0	233 (58.3)	229 (57.3)	0	216 (54.0)	215 (53.8)	—
	1A	30 (7.5)	17 (4.3)	—	—	—	—
	1B	85 (21.3)	110 (27.5)	1	132 (33.0)	141 (35.3)	—
	2	52 (13.0)	44 (11.0)	2	52 (13.0)	44 (11.0)	—
	3	0	0	—	—	—	—
Total goiter		167 (41.8)	171 (42.8)		184 (46.0)	185 (46.3)	152 (38.0)
Aït M'hamed ($n = 200$)	0	159 (79.5)	157 (78.5)	0	149 (74.5)	148 (74.0)	—
	1A	10 (5.0)	6 (3.0)	—	—	—	—
	1B	23 (11.5)	34 (17.0)	1	43 (21.5)	49 (24.5)	—
	2	8 (4.0)	3 (1.5)	2	8 (4.0)	3 (1.5)	—
	3	0	0	—	—	—	—
Total goiter		41 (20.5)	43 (21.5)		51 (25.5)	52 (26.0)	25 (12.5)
Brikcha ($n = 200$)	0	74 (37.0)	72 (36.0)	0	67 (33.5)	66 (33.0)	—
	1A	20 (10.0)	11 (5.5)	—	—	—	—
	1B	62 (31.0)	76 (38.0)	1	89 (44.5)	93 (46.5)	—
	2	44 (22.0)	41 (20.5)	2	44 (22.0)	41 (20.5)	—
Total goiter		126 (63.0)	128 (64.0)	—	133 (66.5)	134 (67.0)	127 (63.5)

examiners and examination criteria. For κ values, agreement < 0.20 = poor; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.80 = good; and 0.81–1.00, very good. Normally distributed data were expressed as means (S.D.) and were compared by Student's *t*-test. Parameters not normally distributed (UI, TSH) were expressed as medians with ranges, and were compared by Mann–Whitney tests.

Results

Mean age (S.D.) in Aït M'hamed and Brikcha was 9.4(1.8) and 9.9(1.6) years respectively. Both village samples included 47–48% girls. The median UI (range) in Aït M'hamed was 183(15–1188) μg/l, with 13% of samples <50 μg/l. In contrast, in Brikcha the median UI (range) was 24(3–171) μg/l, with 77% of samples <50 μg/l. There were no significant differences in thyroid profile between the villages. Median TSH (range) was 0.8(0.2–8.6) mU/l in Aït M'hamed and 0.8(0.2–3.6) mU/l in Brikcha. Mean T_4 (S.D.) in Aït M'hamed and Brikcha was 75(24) and 66(23) nmol/l respectively. Table 1 shows the prevalence of goiter by palpation using the WHO 1960 and 1994 criteria and thyroid ultrasound. In Aït M'hamed, using the 1960 and 1994 criteria, the prevalence of goiter was 20–21% and 25–26% respectively, compared with a 12.5% prevalence of increased thyroid volume by ultrasound. The difference in goiter rates between the 1960 and 1994 criteria in Aït M'hamed was due to a larger number of children classified as grade 1 by both examiners using the 1994 criteria, compared with the number classified as 1A or 1B using the 1960 criteria.

As shown in Table 2, agreement between observers was better with the 1994 criteria than with the 1960 criteria in Aït M'hamed ($\kappa = 0.53$ and 0.47 respectively), while in Brikcha between-observer agreement was similar with the two systems ($\kappa = 0.67$). Table 3 compares the sensitivity and specificity of the 1960 and 1994 criteria in the diagnosis of goiter, compared with ultrasound. Overall, sensitivity increased 3–4% and specificity decreased 4–5% using 1994 criteria, compared with the 1960 criteria. Table 4 shows the percentage of discrepant data by BSA, when palpation is compared with ultrasound.

Table 2 κ values (S.E.) on agreement of goiter diagnosis between examiners A and B, between the WHO 1960 and 1994 grading systems, and between examiners and ultrasound.

	Criteria/examiner			
	WHO 1960 B	WHO 1994 A	WHO 1994 B	**Ultrasound**
Aït M'hamed				
WHO 1960				
A	0.47 (0.06)	0.94 (0.02)	0.49 (0.06)	0.41 (0.07)
B	—	0.50 (0.06)	0.94 (0.02)	0.42 (0.06)
WHO 1994				
A	—	—	0.53 (0.06)	0.42 (0.07)
B	—	—	—	0.44 (0.06)
Brikcha				
WHO 1960				
A	0.67 (0.05)	0.88 (0.03)	0.62 (0.06)	0.60 (0.05)
B	—	0.64 (0.05)	0.89 (0.03)	0.64 (0.05)
WHO 1994				
A	—	—	0.67 (0.05)	0.55 (0.06)
B	—	—	—	0.61 (0.05)

Table 3 Sensitivity and specificity of palpation by examiners A and B using the WHO 1960 and 1994 criteria in goiter diagnosis.

	Sensitivity (%) (S.E.)			**Specificity** (%) (S.E.)		
	Aït M'hamed	Brikcha	All	Aït M'hamed	Brikcha	All
WHO 1960						
A	56 (3)	77 (3)	74 (3)	84 (2)	61 (3)	78 (3)
B	48 (3)	82 (2)	76 (3)	82 (2)	67 (3)	77 (3)
WHO 1994						
A	60 (3)	80 (2)	77 (3)	79 (2)	58 (3)	73 (3)
B	60 (3)	84 (2)	80 (3)	78 (2)	64 (3)	74 (3)

Table 4 Palpation using the WHO 1960 and 1994 criteria compared with thyroid ultrasound for determination of goiter: discrepant data. Values indicate the percentage of discrepant data at each BSA.

Grading system	**WHO 1960**		**WHO 1994**	
Palpation*	−	+	−	+
Ultrasound†	+	−	+	−
BSA (m^2)				
0.7 ($n = 27$)	0.0	7.4	7.4	7.4
0.8 ($n = 86$)	9.3	30.2	8.1	31.4
0.9 ($n = 100$)	10.0	9.0	9.0	11.0
1.0 ($n = 75$)	8.0	8.0	8.0	13.3
1.1 ($n = 47$)	4.3	12.8	4.3	19.1
1.2 ($n = 38$)	5.3	5.3	5.3	10.5
>1.3 ($n = 27$)	7.4	3.7	7.4	11.1
Total ($n = 400$)	8.0	13.3	7.8	16.8

* +, goiter by palpation; −, no goiter by palpation (observer B).
† +, goiter by ultrasound; −, no goiter by ultrasound.

Discussion

Several studies in European children have pointed out the inadequacy of palpation in areas of mild IDD where goiters are mostly grade 1. In such areas, the sensitivity and specificity of palpation is low and misclassification can be as high as 40% (1). Gutekunst & Martin-Teichert (5), using WHO 1960 criteria in 6- to 11-year-old German children, found the correlation between palpation vs thyroid ultrasound was poor ($r = 0.48$) for goiters grade 1A and 1B but fairly good for larger goiters ($r = 0.74$). The percentage of children with goiter grades of 0, 1A, 1B and 2 by palpation that were confirmed by ultrasound was 60, 82, 88, 100% (5). In an area of mild IDD in Italy (median UI = 72 μg/l), Vitti *et al.* compared thyroid ultrasound with palpation using the 1960 criteria and reported discrepant data in 24% of children (4). However, in areas of moderate to severe IDD, limited data suggest goiter palpation becomes more reliable. In Malaysian schoolchildren, Foo *et al.* (13) compared palpation using 1994 criteria with local references for thyroid volume by ultrasound. In an area of severe IDD (median UI = 11 μg/l) palpation had a sensitivity of 82%, while in an area of moderate IDD (median UI = 54 μg/l) sensitivity fell to 55% (13). In a moderate IDD region in Senegal (median UI = 32 μg/g creatinine), sensitivity and specificity for palpation using the 1994 criteria compared with ultrasound were 52 and 92% respectively (14).

Our data suggest palpation using either the WHO 1960 or 1994 criteria provides accurate estimates of goiter prevalence compared with ultrasound in areas of severe IDD, but not in mild IDD. In Brikcha, an area of severe IDD (median UI 24 μg/l, goiter rate 63.5%), both criteria provided an accurate estimate of goiter prevalence, and agreement between both the 1960 and 1994 criteria and ultrasound was good ($\kappa = 0.67$). Also, the sensitivity (77–84%) and specificity (58–67%) of both the 1960 and 1994 criteria were acceptable. In contrast, in Aït M'hamed with mild IDD (median UI 183 μg/l, goiter rate 12%), both criteria overestimated the goiter prevalence compared with ultrasound. Agreement between palpation and ultrasound was only fair, and the sensitivities of both criteria were low (48–60%).

In the only previous comparative study, Peterson *et al.* evaluated the WHO 1960 and 1994 criteria in an area of severe IDD in Tanzania where the goiter prevalence in school-age children by ultrasound was 75% (3). Compared with the 1960 criteria, the 1994 criteria increased sensitivity from 56–60% to 77–80%, but decreased specificity from 69–76% to 29–39%. The new criteria resulted in an extra 20–33% of children being diagnosed as goitrous. The authors suggested a return to the WHO 1960 criteria as the method of choice for goiter screening. Our data differ from those of Peterson *et al.* (3). In severely iodine-deficient Brikcha, using the WHO 1960 and 1994 criteria, the prevalence of goiter was 63–64% and 66–67% respectively, compared with 63.5% by ultrasound. Comparing the WHO 1960 and 1994 criteria, sensitivity increased 2–3% with the 1994 criteria, while specificity decreased only 3%. Agreement between observers was better with the 1994 criteria than with the 1960 criteria in Aït M'hamed, while in Brikcha between-observer agreement was similar with the two systems. The sharp differences in results of these two studies indicate sensitivity and specificity of goiter grading by palpation depend to a large extent on the examiners and the characteristics of the study population.

Both the 1960 and the 1994 criteria generated approximately 8% false negatives (goiter by ultrasound but not by palpation). The 1994 criteria generated more false positives (goiter by palpation but not by ultrasound), particularly at BSAs >1.0 m^2, compared with the 1960 criteria. This may reflect the subtle lowering of the criterion for classifying a thyroid as grade 1 goiter in the new WHO 1994 criteria. Whereas with the 1960 criteria, goiter was only present if the lateral lobes of the thyroid were greater in volume than the terminal phalanges of the individual's thumbs, with the new criteria any enlarged thyroid that is palpable,

even if smaller than the terminal phalanges, is classified as goitrous (1). This means in some cases, smaller thyroids than before are graded as goiter. This is shown in Table 2, where the difference in goiter rates between the 1960 and 1994 criteria in Aït M'hamed (20–21% vs 25–26%) was due an increased number of children classified as grade 1 by both examiners using the 1994 criteria, compared with the number classified as 1A or 1B using the 1960 criteria. In Brikcha, where the goiters were larger and more common, this did not occur, and both classification systems gave equal estimates of the goiter prevalence.

Taken together, these data support the new WHO/ICCIDD 1994 criteria for goiter classification. The 1994 criteria are simpler to use than the 1960 criteria and provide increased sensitivity with only a small reduction in specificity. Agreement between observers is better with the 1994 criteria than with the 1960 criteria, particularly in areas of mild IDD. Both the 1960 and 1994 criteria overestimate goiter prevalence in areas of mild IDD, and in these areas ultrasound, if available, is the method of choice. However, palpation using the WHO 1994 criteria provides an accurate estimate of goiter prevalence in areas of moderate to severe IDD, and may be an acceptable and more affordable alternative to thyroid ultrasound in these areas.

Acknowledgements

We would like to thank Dr L Molinari of the Children's Hospital in Zürich for providing statistical advice, the schoolteachers and children in Morocco for their participation, and the Swiss Federal Institute of Technology Zürich for financial support.

References

1 WHO, UNICEF & ICCIDD. Indicators for assessing iodine deficiency disorders and their control through salt iodization. WHO/NUT/94.6. Geneva: WHO 1994.

2 Perez C, Scrimshaw S & Munoz A. Technique of endemic goiter surveys. In *Endemic Goiter*, pp. 369–383. Geneva: WHO, 1960.

3 Peterson S, Sanga A, Eklöf H, Bunga B, Taube A, Gebre-Medhin M *et al.* Classification of thyroid size by palpation and ultrasonography in field surveys. *Lancet* 2000 **355** 106–110.

4 Vitti P, Martino E, Aghini-Lombardi F, Rago T, Antonangeli L, Maccherini D *et al.* Thyroid volume measurement by ultrasound in children as a tool for the assessment of mild iodine deficiency. *Journal of Clinical Endocrinology and Metabolism* 1994 **79** 600–603.

5 Gutekunst R & Martin-Teichert H. Requirements for goiter surveys and the determination of thyroid size. In *Iodine Deficiency in Europe: a Continuing Concern*, pp. 109–118. Ed F Delange, JT Dunn & D Glinoer. New York: Plenum Press, 1993.

6 Chaouki N, Ottmani S, Saad A, El Hamdaoui M, Benabdejlil C, Kadiri A *et al.* The prevalence of iodine deficiency disorders in children 6–12 yrs old in Morocco. *Bulletin Epidemiologique Royaume du Maroc* 1996 **17–20** 2–23.

7 World Health Organization. *Physical Status: the Use and Interpretation of Anthropometry*, pp. 427–429. Geneva: WHO, 1995.

8 Delange F, Benker G, Caron P, Eber O, Ott W, Peter F *et al.* Thyroid volume and urinary iodine in European schoolchildren: standardization of values for assessment of iodine deficiency. *European Journal of Endocrinology* 1997 **136** 180–187.

9 Brunn J, Block U, Ruf G, Bos I, Kunze WP & Scriba PC. Volumetrie der Schilddrüsenlappen mittels Real-time-Sonographie. *Deutsch Medizinische Wochenschrift* 1981 **106** 1338–1340.

10 WHO/ICCIDD Recommended normative values for thyroid volume in children aged 6–15 years. *Bulletin of the World Health Organization* 1997 **75** 95–97.

11 Pino S, Fang SL & Braverman LE. Ammonium persulfate: a safe alternative oxidizing reagent for measuring urinary iodine. *Clinical Chemistry* 1996 **42** 239–243.

12 Torresani T & Scherz R. Thyroid screening of neonates without use of radioactivity: evaluation of time-resolved fluoroimmunoassay of thyrotropin. *Clinical Chemistry* 1986 **32** 1013–1016.

13 Foo LC, Zulfiqar A, Nafikudin M, Fadzil MT & Asmah ASA. Local versus WHO/ICCIDD-recommended thyroid volume reference in the assessment of iodine deficiency disorders. *European Journal of Endocrinology* 1999 **140** 491–497.

14 Smyth P, Darke C, Parkes AB, Smith DF, John R, Hetherton AM *et al.* Assessment of goiter in an area of endemic iodine deficiency. *Thyroid* 1999 **9** 895–901.

Received 3 July 2000
Accepted 24 August 2000