The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/81732

Please be advised that this information was generated on 2020-09-29 and may be subject to change.
Composite Reference Interval for Thyroid-Stimulating Hormone and Free Thyroxine, Comparison with Common Cutoff Values, and Reconsideration of Subclinical Thyroid Disease

H. Alec Ross,1,2* Martin den Heijer,2 Ad R.M.M. Hermus,2 and Fred C.G.J. Sweep1

BACKGROUND: Examination of the 2-dimensional probability distribution of thyroid-stimulating hormone (TSH) and free thyroxine (FT4) shows that the widths of the TSH and FT4 reference intervals derived from this bivariate distribution are mutually interdependent, an aspect commonly ignored when interpreting thyroid testing results with separate reference intervals for TSH and FT4. We desired to establish and critically evaluate a composite reference interval for TSH and FT4 to allow bivariate classification of biochemical thyroid conditions.

METHODS: FT4 and TSH results of 871 healthy individuals [361 women and 510 men, 18 – 40 years old, without history of thyroid-related disease or medication, negative for anti–thyroid peroxidase (anti-TPO) antibody] were transformed to standard normal variables by logarithmic transformation with correction for skewness and subsequent normalization. We established a 95% reference interval of the distance of each FT4/TSH pair of values to the center of the 2-dimensional probability distribution.

RESULTS: The bivariate 95% reference interval is enclosed by a circular profile with radius 2.45 SD. By contrast, conventional reference intervals comprise a square with the boundaries of −1.96 and +1.96 SD for both FT4 and TSH that enclose only 90% of all data. Compared with the ±1.96 SD square, the bivariate reference interval classified 4% fewer of 3651 healthy individuals older than 40 years as subclinically hyperthyroid and 14% fewer of 712 anti-TPO–positive healthy individuals as subclinically hypothyroid.

CONCLUSIONS: Conventional application of separate cutoff values for FT4 and TSH leads to overestimation of the incidence of subclinical thyroid disease. Application of a composite overall reference interval is recommended.
ence population will exceed this bivariate reference limit. Because the reference population must be suit-
able to identify cases of subclinical thyroid conditions, no signs of thyroid autonomy or failure should be con-
spicious. Within a population harboring some degree of thyroid autonomy, a negative correlation between 
FT₄ and TSH is expected due to negative feedback; conversely, in a group of individuals in which pituitary 
failure occurs, a positive correlation is anticipated. Therefore, if neither organ displays autonomy or fails 
by itself, no correlation should be observed between TSH and FT₄. Basal TSH and FT₄ concentrations de-
pend on the partially genetically determined pituitary-thyroid set point (4) that appears to result in entirely 
random combinations of FT₄ and TSH values among individuals. When this is the case, a bivariate reference 
limit is easy to calculate.

To investigate the effect of this approach on the definition of subclinical hyper- and hypothyroidism, 
we estimated the frequencies of these conditions by composite, bivariate, and conventional univariate ap-
proaches in a euthyroid older (compared to the refer-
ence group) age group without anti-TPO antibodies, 
and in an anti-TPO-positive group.

Study Participants and Methods

The individuals included in this study originated from 
the Nijmegen Biomedical study (5). Serum TSH was 
measured by immunoluminometric assay in an Archi-
tect random access assay system (Abbott Diagnostics). 
The functional detection limit (i.e., the concentration 
at which the interassay CV is 20%) was 0.007 mU/L. At 
higher concentrations, interassay CVs were as follows: 
3.3% at 0.250 mU/L, 3.6% at 1.72 mU/L, and 3.0% at 
9.86 mU/L. Serum FT₄ was estimated by a lumines-
cence enzyme immunoassay in a Vitros ECI random 
access assay system (Ortho Clinical Diagnostics). This 
assay uses a labeled anti-T₄ antibody in a medium that 
is essentially free of other extraneous T₄-binding pro-
teins. Serum samples may be diluted up to 8 times 
without significant effect on measurement results. In-
terassay CVs were as follows: 3.8% at 10.1 pmol/L, 
4.5% at 15.8 pmol/L, and 4.6% at 28.7 pmol/L.

The total population consisted of 6434 individuals 
age 18 years and older. After exclusion of all those with 
self-reported thyroid disease and/or thyroid surgery, 
those on thyromimetics and thyrostatics and on other 
medication known to affect thyroid function or thy-
roid function parameters, pregnant women and 
women on oral contraceptives, a group of 5235 healthy 
individuals remained. Individuals with increased anti-
TPO antibodies (n = 712) (aTPO⁺ group) were ex-
cluded, and of the remaining 4523, TSH and FT₄ 
concentrations were logarithmically transformed in-
cluding a correction for skewness. Based on the mean 
and SD of the transformed values, 14 outliers were

Fig. 1. TSH and FT₄ values as measured in the refer-
ence group.

(A), Untransformed values for data points, with reference 
limits obtained by reversal of the transformation and nor-
malization, with conventional 1.96 SD limits (dotted lines) 
and composite limits (solid lines). Sections outside the 
curved reference zone correspond to thyroid conditions 
outlined in the text. (B), Normalized, transformed TSH and 
FT₄ values for the reference group which formed the basis 
of the composite reference limits.
eliminated (see Appendix for transformation and outlier algorithms). In addition, 16 individuals with TSH below the functional sensitivity limit of 0.007 mIU/L were removed. Because we had observed a tendency toward lower TSH and higher FT4 concentrations with age (5), the reference group was restricted to people 40 years old and younger (n = 871). The distribution of transformed FT4 and TSH values in this reference group was not found to differ from a normal gaussian distribution. From these data, we made a 2-dimensional diagram of transformed TSH plotted against transformed FT4. A reference limit was represented by a circular profile centered at (0,0), which encloses a specified fraction of all data points. The remaining 3622 individuals were supplemented with results for the previously removed 13 of 14 outliers and for 16 values with TSH below the functional detection limit, all >40 years old, to form the older anti-TPO negative (>40, aTPO–) group (n = 3651).

The squared normalized distance, D2, of each data point to the center of the distribution for this reference population equals the sum of the squared normalized TSH and FT4 values and is given by the following formula:

\[
D^2 = [(\log(TSH + 0.303) - 0.243)^2/0.171]^2 + [(\log(FT4 + 11.2) - 1.39)^2/0.0324]^2. \tag{1}
\]

With the assumption that TSH and FT4 are uncorrelated, D2 is mathematically identical to the Mahalanobis distance measure traditionally used for multivariate reference regions (6) and follows a χ2 distribution with 2 degrees of freedom so that P0.95 is at 5.99. For the distance D, the critical value is √5.99 or 2.45. The constants 0.303 and 11.2 are corrections for skewness, 0.243 and 1.39 are the means of the transformed TSH and FT4 values, and 0.171 and 0.0324 are the corresponding SDs (see Appendix). Because all outliers except one were >40 years old, outlier removal had no effect on the parameters of the reference group.

The area outside the composite reference interval may be subdivided into 8 sections (Fig. 1, A and B), i.e., the number of combinations of low, normal, and high TSH and FT4 with the combination normal/normal excluded. The 4 corner sections I, III, VI, and VIII represent the zones in which both TSH and FT4 exceed the same absolute limit of 1.73 (2.45/√2) SD (or conventionally, ±1.96 SD). Section I represents overt hypothyroidism with high TSH and low FT4, whereas section VIII represents overt hyperthyroidism with high FT4 and low TSH. Sections II and VII represent subclinical hypo- and hyperthyroidism, respectively.

Using the χ2-test for population proportions, we compared observed frequencies in the sections statistically to each other and expected values (7); P values <0.05 were considered significant. For P values >0.25, the predicate “not different” was assigned. Distributions were tested for normality by Kolmogorov–Smirnov test (SPSS v. 16; SPSS Inc.).

Results

The distribution of transformed FT4 and TSH values in the reference group was not found by Kolmogorov–Smirnov testing to differ from a normal gaussian distribution. The separate 95% univariate reference intervals obtained after reversal of the transformation were 0.51–3.48 mIU/L for TSH and 9.8–16.9 pmol/L for FT4 (Fig. 1A). The circumference of the bivariate 95% reference region is given by Eq. 1, with the value 5.99 assigned to D2 (distance D = 2.45). For each FT4 value, a pair of TSH values was obtained, representing the TSH reference interval corresponding to that particular FT4 value. Thus, for an FT4 value of 13.1 pmol/L, the lower and upper limits for TSH are 0.36 and 4.28 mU/L, respectively. Conversely, if TSH is 1.45, the range for FT4 is 9.0–18.0 pmol/L. Furthermore, if FT4 is 9 pmol/L, the reference interval for TSH would be restricted to a single value of 1.45 mU/L.

When using the separate univariate reference intervals for TSH and FT4, 10.8% (n = 94) of data points in the reference group fall outside the area in which both criteria are met; by contrast, with the bivariate composite approach, 5.2% (n = 45) fall outside the constructed combined reference region. Both numbers fall close to the expected values of 9.75% and 5% (z-test for population proportions). The same holds after further subdivision into subclinical hypo- and hyperthyroidism (see Table 1): separate limits give 2.4% and 2.8% (expected 2.4% for both), and the bivariate composite limit yields 1.2% and 0.8% (expected 1.1% for both).

Fig. 2, A and B, shows the data from the group of healthier older (>40 years) individuals and healthy individuals with positive anti-TPO superimposed on the reference grid.

Table 1 presents the observed frequencies of subclinical hypo- and hyperthyroidism as assessed by the composite and conventional methods. In many instances, the differences between observed frequencies were significantly greater than expected from inclusion of a larger percentage of normal values (2.4% vs 1.1%) alone. This observation was particularly apparent in the frequency of subclinical hypothyroidism in the anti-TPO positive group. The estimate of subclinical hypothyroidism according to the conventional approach was almost 24%, whereas the composite approach indicated only 10%. In the elderly anti-TPO–negative group, the observed frequency of subclinical hypothyroidism was slightly higher than in the refer-
ence group if the composite limits were used (1.5% vs 1.2%), but the difference was somewhat more pronounced (3.3% vs 2.4%) when applying separate limits. For subclinical hyperthyroidism, the composite approach gave frequencies for the elderly anti-TPO–negative and anti-TPO–positive groups that were higher than expected (3.4% and 2.8% vs 0.8%) on the basis of the reference group, but not significantly different from each other. For the conventional method, frequencies for the elderly anti-TPO–negative and anti-TPO–positive groups were 7.3% and 5.1%, respectively, vs 2.8%, of which the former is significantly higher than the latter, whereas both are higher than in the composite approach.

**Discussion**

Plots of TSH vs FT₄ values from larger data sets show that the shape of the density distribution of points is egg-like (Fig. 1A) rather than rectangular. Whereas FT₄ approximates a normal distribution, the distribution of TSH is skewed, but approximates a normal distribution after log transformation. Fine-tuning of the transformation by correcting for residual skewness, applying log transformation for FT₄ as well with subsequent normalization, results in a circular-shaped 2-dimensional diagram for healthy individuals without any thyroid medication or sign of disease. This is a reflection that, if the negative feedback system is at equilibrium, TSH and FT₄ combine randomly among individuals. A negative correlation will be observed if there is a tendency of thyroid autonomy within the data set, in the form of either thyroid hyperactivity or failure. Conversely, a trend toward pituitary autonomy would lead to a positive correlation. If TSH and FT₄ do correlate, the density distribution will tend toward an elliptic shape with \( Y = X \) and \( Y = -X \) as main axes. A trivariate [TSH, FT₄ index, free triiodothyronine (FT₃) index] probability density distribution was presented by Kagedal et al. (8) for 3885 women 39–60 years of age. In this group, negative correlations were observed between logTSH and both FT₄ and FT₃ indices and a positive correlation between FT₄ and FT₃ indices. This resulted in an ellipsoidal frequency distribution and reference limit. We also observed a negative correlation between FT₄ and logTSH for individuals >40 years old, which was the reason to exclude those from the reference group since our aim was to obtain a reference group that is suitable for detecting subclinical thyroid disease. Although the univariate probability density distributions of FT₄ and TSH are mutually independent, the bivariate density distribution depends on both FT₄ and TSH, and therefore the reference limits derived from the bivariate distribution for each parameter are mutually dependent. So the derived reference limits for FT₄ are further apart if TSH values tend toward their average, than with TSH values toward the extremes, and vice versa.

Classification of thyroid conditions on the basis of a logTSH/FT₄ diagram has been shown before (2). Although the authors indicated in their graph that the normal reference region was elliptical (and would be circular after normalization) rather than rectangular, this aspect was not further explored.

The squared distance of each data point to the center of the distribution equals the sum of squared normalized TSH and FT₄ values. As a consequence of normalization, TSH and FT₄ each have a mean of 0 with variance 1. Thus, each squared transformed and normalized TSH and FT₄ value is an estimate of this variance. Therefore, the sum of squared TSH and FT₄ will be \( \chi^2 \) distributed with 2 degrees of freedom. The 95% reference interval corresponds to a \( \chi^2 \) value of 5.99, so it encloses a circle with a radius of 2.45, which is the square root of 5.99. Indeed, 94.8% of all data from the reference group fall within this region; in contrast, 89.2% fall within the separately applied limits, meaning that 5% of healthy individuals would be inappro-

---

**Table 1. Observed and expected incidence rates in subclinical hyper- and hypothyroid ranges.**

<table>
<thead>
<tr>
<th></th>
<th>Composite</th>
<th>Conventional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subclinical hypothyroid</td>
<td>Subclinical hypothyroid</td>
</tr>
<tr>
<td>Expected incidence</td>
<td>0.0108</td>
<td>0.0108</td>
</tr>
<tr>
<td>Observed incidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>0.0115</td>
<td>0.00804</td>
</tr>
<tr>
<td>&gt;40, aTPO−</td>
<td>0.0152a,b</td>
<td>0.0337a</td>
</tr>
<tr>
<td>aTPO+</td>
<td>0.101a,b</td>
<td>0.0281a</td>
</tr>
</tbody>
</table>

*a Significant difference (z-test for population proportions) with reference and expected.

*b Significant difference between aTPO− and >40, aTPO−.
appropriately classified as abnormal with the latter approach. Conversely, with the composite approach, the number of healthy individuals not requiring further investigation increases by 5%.

Subdivision of the area outside the reference interval (see Fig. 1B) into sections corresponding to various thyroid conditions is intuitive when using conventional separate 95% limits at ±1.96 SD. With the circular reference area it is somewhat less obvious. The dividing lines must intersect at ±1.73 SD, by which the corner sections will correspond to regions where not only TSH and FT$_4$ combined fall outside the reference interval, but also separately exceed the same limit. This limit is defined by the 4 points on the circle where the absolute values of normalized TSH and FT$_4$ are equal. This is at $2.45\sqrt{2} = 1.73$ SD.

For the reference population, the difference between conventional and bivariate composite classifications is almost exclusively located in the sections corresponding to the 4 possible combinations normal/abnormal, 2 of which represent subclinical hyper- and hypothyroidism.

Plotting the data from the healthy anti-TPO-negative participants older than 40 in the grid (Fig. 2, A and B), it appears that the data are shifted downward and to the right with respect to the reference group. This is a corollary of observations we made earlier (5). Interpretation of this shift depends on whether the composite reference limits or conventional reference limits are used. Table 1 shows that there was a slight but significant increase in frequency of subclinical hypothyroidism compared to the reference group, but the difference tended to be more pronounced with conventional classification. This was also true when considering the data from the anti-TPO-positive participants (Fig. 2B and Table 1). The composite approach shows clearly higher incidence of subclinical hypothyroidism with respect to both the older anti-TPO negative and reference groups, but no difference with the former for subclinical hyperthyroidism. Thus, the composite approach does not indicate a relation between anti-TPO status and the incidence of subclinical hyperthyroidism, whereas the conventional approach does. The higher incidence of subclinical hyperthyroidism compared to the younger reference group suggests an age effect instead.

Because the purpose of this report is presentation of the composite reference interval approach, the underlying causes for the age and anti-TPO-related differences in probability distributions from the reference group in these individuals are not discussed here. However, since it is questionable whether all patients classified as having subclinical thyroid disease by the usual definition benefit from treatment (3), the approach that is based on a composite reference region appears to be more conservative and thus may help to prevent unnecessary treatment. A prospective study, randomly assigning those classified as having subclinical disease
by univariate criteria but not by the bivariate approach to treatment or no treatment, will be required to test this supposition.

Practical Considerations

Because of assay method differences, each laboratory should establish its own composite reference interval, along the guidelines given in the Appendix. Patient results must be substituted into the resulting formula for the distance to the center of the distribution as a “thyroid balance index.” If the result is $<2.45$, there should be no need for reporting separate TSH and FT$_4$ values, although many physicians still may want to have these values. For example, a TSH value of 4.2 mU/L exceeds the 1.96 SD limit, but if FT$_4$ is close to the average of 13.1 pmol/L the data pair remains within the composite overall 95% reference interval. For ease of use to the physician, the calculation can be performed in the laboratory instead of supplying assay results with the formula. The physician will then receive a report with TSH and FT$_4$ results supplemented with either “Thyroid balance index within the reference interval” or “Thyroid balance index outside the reference interval.”

An alternative could be to provide a diagram, prepared by the laboratory, similar to Fig. 1A, showing only the egg-shaped profile and grid, and plotting the reported TSH and FT$_4$ results to see if the data pair falls within the reference region, and if not, what clinical category is likely to apply. If “screening by TSH” as a first step is common practice, the warning limits for ordering FT$_4$ may be kept at the univariate P2.5 and P97.5 for TSH. If these warning limits are exceeded, subsequent FT$_4$ measurement will reveal if the same holds with regard to the bivariate reference limit. Such a strategy limits the number of unnecessary FT$_4$ measurements and does not impair diagnostic sensitivity compared to a bivariate approach.

In conclusion, the proposed bivariate reference interval for TSH and FT$_4$ may lead to a better definition of subclinical thyroid conditions with fewer false positives. Its implementation can be readily included in existing diagnostic strategies.

Appendix

**TRANSFORMATION OF DATA**

In Excel, to each tabulated TSH and FT$_4$ value the same offset ($9$) was added and the logarithm of the sum was calculated. The skewness of the distribution of transformed values of the healthy group was expressed as $(\text{mean} - \text{median})/\text{SD}$, i.e., Pearson’s skewness index divided by $3$ ($7$), and the value of the offset was obtained by adjusting until a skewness index value of $<0.001$ times its original value had been reached. Finally, from each transformed value the mean was subtracted and the difference was divided by the SD to obtain values following a standard normal distribution. This was confirmed by Kolmogorov–Smirnov testing. Symmetry was further ascertained by comparing the numbers of individuals exceeding $\pm 1.96$ SD limits by the $z$-test for population proportions.

**REMOVAL OF OUTLIERS**

An outlier was defined as a value for which the probability not to be encountered at all in a sample of the same size as the population was 50%, which means that the probability of finding at least $1$ observation exceeding this value was 50% or less. For a sample size of 4523, this holds for values of $P < 0.00015$ or $>3.61$ SD from the mean, according to the following formula:

$$P = 1 - 0.5^{(1/4523)} = 0.000153.$$ Because the $–3.61$ interval for TSH extends beyond the detection limit of the assay, it seemed sensible to exclude all values below this limit.

**References**


2. Demers LM, Spencer CA, eds. Laboratory medicine practice guidelines: laboratory support for the diagnosis and monitoring of thyroid disease.


6. Boyd JC. Reference regions in two or more di-
Composite Reference Interval for TSH and FT$_4$

