Serum Thyroxine Change and Clinical Recovery in Psychiatric Inpatients

Steven Southwick, John W. Mason, Earl L. Giller, and Thomas R. Kosten

Serum free thyroxine (FT_4), total thyroxine (TT_4), and Brief Psychiatric Rating Scale (BPRS) measurements were obtained following hospital admission and at 2-week intervals during hospitalization in 80 male psychiatric inpatients with a variety of major psychotic and affective disorders. A strong correlation between the range values for BPRS sum and for FT_4 ($p < 0.005$) and TT_4 ($p < 0.001$) levels indicated that change in overall symptom severity was linked to change in thyroxine levels during clinical recovery. We found the relationship not to be a simple one, but to require definition of criteria for three patient subgroups for each hormone, taking into account the initial absolute thyroxine level, as well as the direction and magnitude of hormonal change during recovery. The hormonally defined "good recovery" subgroup included patients with high initial thyroxine levels that then fell substantially, patients with low initial thyroxine levels that then rose substantially, and patients with initial thyroxine levels in the middle range that subsequently changed substantially. The hormonally defined "poor recovery" subgroup included those patients not meeting these criteria. The degree of clinical improvement in the hormonally defined good recovery group was significantly greater by almost twofold than the poor recovery group both for FT_4 ($p < 0.04$) and TT_4 ($p < 0.02$). These findings suggest that a "normalizing" principle underlies the relationship between clinical recovery and thyroxine levels and that both FT_4 and TT_4 levels within the normal range appear to have clinical significance in either reflecting or contributing to the course of a variety of psychiatric disorders and possibly having a role in pathogenesis.

Introduction

Although there is a very long history of research on thyroid function in psychiatric disorders (Hoskins and Sleeper 1930; Reiss 1954; Mason 1968), the significance of thyroid hormonal levels in relation to psychiatric illnesses is still not well understood. The primary purpose of many psychoendocrine studies of thyroid function has been to search for clinical or subclinical thyroid gland abnormalities in psychiatric patients, particularly hypothyroidism, with the hope that thyroid hormone administration might

From the Department of Psychiatry, Yale University School of Medicine (S.S., J.W.M., E.L.G., T.R.K.), West Haven VA Medical Center (S.S., J.W.M., E.L.G.), and the Connecticut Mental Health Center (T.R.K.).
Address reprint requests to Dr. Steven Southwick, VA Medical Center/116A, West Haven, CT 06516.
Supported in part by Veterans Administration Research Funds and National Institute of Mental Health Research Scientist Award MH-00346 to J.W.M.
Received March 9, 1987; revised September 16, 1987.

0006-3223/89/$00.00

have therapeutic usefulness in certain disorders (Hoskins and Sleeper 1930; Bauer and Whybrow 1986). Other studies have focused on the significance of transient elevations in thyroxine seen in many psychiatric patients at the time of hospital admission screening for thyroid disease (Cohen and Swiger 1979; Levy et al. 1981). As yet, however, there is very little reported longitudinal data on correlations between changing thyroid hormone levels and symptom or clinical improvement measurements during the course of recovery from episodes of major psychiatric illnesses. Perhaps one of the reasons for the dearth of such data is the traditional tendency to focus on thyroid hormone abnormalities as defined by the laboratory criteria for diseases of the thyroid gland and to discount thyroid hormonal changes within the "normal range" as probably having no significant role in psychiatric disorders. However, there are reported observations indicating that the role of the hypothalamic–pituitary–thyroid (HPT) system in psychiatric illnesses is relatively subtle, complex, and may to a large extent involve fluctuations within the endocrinologically defined "normal range" without organic thyroid disease (Brody 1949; Reiss 1954; Bauer and Whybrow 1986). The purpose of the present report is to describe our findings concerning the relationship between thyroxine levels and symptom ratings during the longitudinal course of recovery in a large group of male psychiatric inpatients with major psychiatric disorders, investigated with the viewpoint that thyroxine fluctuations within the normal range may have a significant role in either reflecting or contributing to the course of psychiatric illnesses.

Methods

Eighty male inpatients, 18–68 years old, with diagnoses of undifferentiated schizophrenia (n = 6), paranoid schizophrenia (n = 9), schizoaffective disorder (n = 18), bipolar disorder (n = 15), major depressive disorder (n = 16), posttraumatic stress disorder (PTSD) (n = 9), and other disorders (n = 7) were studied during their hospital course at the West Haven VA Medical Center. Mean and SE values for our sample were as follows: age 38.6 ± 1.3 years, hospital stay 121 ± 12 days, weight 169.5 ± 2.7 lb, height 69.4 ± 0.3 inches. Informed consent was obtained after full explanation of our procedures. Research Diagnostic Criteria (Spitzer et al. 1978) were used for all diagnoses, except for PTSD, for which DSM-III criteria were used. None of our patients was experiencing his first episode of psychiatric illness, all having a chronic or subchronic course. Patients with drug or alcohol abuse in the past month, major medical illnesses, or organic brain syndromes were excluded, as were patients receiving thyroid hormone therapy. Following hospital admission and at 2-week intervals thereafter, 9:00 AM serum samples were obtained, and the Brief Psychiatric Rating Scale (BPRS) was used to assess clinical symptoms (Overall and Gorham 1962). The serum was frozen at −70°C until analyzed for free thyroxine (FT_4) and total thyroxine (TT_4) concentrations by a radioimmunoassay kit procedure developed by Clinical Assays, Inc., Cambridge, MA. The interassay coefficients of variation were 4.2% for FT_4 and 3.7% for TT_4 and the intrassay coefficients of variation were 3.2% for FT_4 and 2.2% for TT_4, all in the concentration range of our unknown samples.

Our statistical analyses included Pearson product–moment correlations, matched sample *t*-tests for within-subject comparisons of first versus last samples, unpaired *t*-tests for other group comparisons, and a two-way repeated measures Analysis of Variance (ANOVA) for comparison of Time, Group, and Time–Group interaction effects in the recovery pattern of hormonally defined "good" versus "poor" recovery groups.

Results

As an index of clinical recovery, we used the difference (delta) between the last minus first sample value for the BPRS sum, i.e., the change in overall symptomatology from the beginning to the end of the hospital course. In our total sample, the mean BPRS sum declined significantly ($t = 8.29$, $p < 0.0001$) from 24.3 ± 1.2 to 13.3 ± 1.1 (delta −11.0 ± 1.4) during the course of hospitalization. A similar overall downward trend during recovery was seen in FT_4 levels ($t = 3.63$, $p < 0.0005$) and TT_4 levels ($t = 4.58$, $p < 0.0001$). Correlational analyses, however, indicate that the relationship between symptom change and hormonal change is not a simple one. First, the BPRS sum delta (last minus first) does not correlate significantly with the first FT_4 level ($r = 0.12$, $p < 0.3$) or with the first TT_4 level ($r = 0.17$, $p < 0.2$), indicating that the absolute hormone level shortly after hospital admission by itself does not predict the degree of clinical improvement. Secondly, the BPRS sum delta also does not correlate significantly with the FT_4 delta ($r = 0.09$, $p < 0.4$) or with the TT_4 delta ($r = 0.16$, $p < 0.2$), indicating that the change in symptom levels and change in hormone levels are not related in a simple unidirectional way. Thus, the relationship is not simply that those patients whose thyroxine levels fall show improvement, whereas those whose thyroxine levels rise get worse. It is interesting, however, that there is a highly significant relationship between the range (maximum minus minimum value) of the BPRS sum and the range of FT_4 levels ($r = 0.49$, $p < 0.0005$) and range of TT_4 levels ($r = 0.28$, $p < 0.0001$), possibly indicating in view of the foregoing that clinical improvement may be associated with falling thyroxine levels in some patients and rising thyroxine levels in other patients.

Upon further inspection of our data, we found that although many patients improved clinically as high initial thyroxine levels fell, there was also a subgroup of patients with low initial FT_4 and TT_4 levels who subsequently showed marked clinical improvement as their hormonal levels rose toward the middle range of values for the overall group. Accordingly, we next tested a modified "normalization" hypothesis in which three factors—namely, the initial absolute hormonal level, the direction of subsequent hormonal change, as well as the magnitude of hormonal change—were all taken into consideration in defining hormonal criteria for predicted "good recovery" and "poor recovery" patient subgroups. On the basis of the initial hormonal values, patients were first classified into high, middle, or low categories by dividing the entire sample into thirds. In establishing criteria for a "substantial" degree of hormonal change for further defining the two groups, we initially took half of the difference between the cutoff points for the three groups, but we subsequently found that widely ranging change criteria (e.g., from as low as 0.5 up to 1.2 μg/dl for TT_4) have equivalent sensitivity and specificity in identifying "good recovery" patients. Table 1 presents the FT_4 and TT_4 criteria that were used to define "good recovery" versus "poor recovery" patient subgroups.

Table 2 shows that clinical improvement, as reflected in the BPRS sum delta, i.e., the decrease in BPRS sum during hospitalization, was almost twice as great in the hormonally predicted good recovery group compared to the poor recovery group, as defined either by FT_4 or TT_4 criteria. Using a two-way repeated measured ANOVA, we found a significant time effect [$F_{(78,1)} = 62.2$, $p < 0.0001$] for the TT_4 criteria recovery groups, indicating that the decline in BPRS sum during hospitalization for the total sample was highly significant. There was also a significant Group by Time interaction [$F_{(78,1)} = 5.7$, $p < 0.02$], indicating that the predicted good recovery group had a significantly larger decrease in symptoms during hospitalization than the poor recovery group. No significant

Table 1. Thyroxine Criteria for Good Recovery versus Poor Recovery Patient Subgroups

I. Good recovery subgroup: Patients meeting the following criteria:
  Free thyroxine (FT4) good recovery criteria (ng/dl)
    (1) First FT4 > 1.59 and FT4 delta = −0.15
    (2) First FT4 < 1.26 and FT4 delta = +0.15
    (3) First FT4 between 1.59 and 1.26 and FT4 range 0.15
  Total thyroxine (TT4) good recovery criteria (μg/dl)
    (1) First TT4 > 9.20 and FT4 delta = −0.6
    (2) First TT4 < 8.05 and TT4 delta = +0.6
    (3) First TT4 between 9.20 and 8.05 and TT4 range 0.6
II. "Poor recovery" subgroup: Patients not meeting the above criteria

baseline differences were observed between the two groups in their first BPRS sum scores [$F_{(78,1)} = 3.1$, $p < 0.1$]. Similar analysis of the FT_4 criteria recovery groups showed a significant time effect [$F_{(78,1)} = 5.9$, $p < 0.0001$], indicating a highly significant decrease in symptoms during hospitalization in the total sample. Again, there was a significant Group by Time interaction [$F_{(78,1)} = 4.3$, $p < 0.04$], indicating that the predicted good recovery group showed a significantly greater decrease in symptoms during hospitalization than the poor recovery group. No significant difference was observed in the baseline first BPRS scores between the two groups [$F_{(78,1)} = 0.01$, $p < 0.9$].

With regard to possible confounding factors, no significant correlations were found among age, hospital stay, weight, or height and any of the thyroxine parameters used in this study. Also, no significant differences were observed in age, hospital stay, weight, or height between the good versus poor recovery groups established by either the FT_4 or TT_4 criteria. No significant correlations were found between the dosage of lithium, antidepressant, or antipsychotic (chlorpromazine equivalents) medication and any of the thyroxine parameters used in the study. In addition, there were no significant differences in lithium, antidepressant, or antipsychotic medication either at the time of admission or discharge between the good versus poor recovery groups established by either the FT_4 or TT_4 criteria. Finally, no significant differences were observed in any of the thyroid parameters between the patients on versus off lithium, antidepressant, or antipsychotic medication in our total sample, either at the time of admission or discharge.

Finally, with regard to a possible relationship between thyroid recovery groups and diagnostic groups, chi-square tests were performed to determine if there were any significant differences in the proportions of schizophrenia, affective disorder, and schizoaffective patients between the good versus poor hormonally defined recovery groups. No significant differences were found in distribution of these three diagnoses between the good versus poor recovery groups using FT_4 criteria ($X^2 = 0.28$, df = 2, $p < 0.9$, NS) or TT_4 criteria ($X^2 = 0.32$, df = 2, $p < 0.9$, NS).

Discussion

The present study supports the hypothesis that there is a significant relationship between serum thyroxine levels in the normal range and clinical recovery in a wide diagnostic range of psychiatric inpatients. The relationship, however, is not simple or unidirectional, but involves the need to subgroup patients with regard to three inter-

Table 2. Hormonal and Clinical Data for Good Recovery versus Poor Recovery Patient Subgroups as Defined by Hormonal Criteria

	Good recovery			Poor recovery		
	First	Last	Delta	First	Last	Delta
Free thyroxine criteria subgroups[a]						
BPRS sum	24.7 ± 1.6	11.3 ± 1.1	−13.4 ± 1.6	24.4 ± 1.7	16.7 ± 2.1	−7.7 ± 2.3
Free T4 (ng/dl)	1.55 ± 0.04	1.37 ± 0.03	−0.18 ± 0.05	1.39 ± 0.05	1.37 ± 0.05	−0.02 ± 0.03
Total T4 (μg/dl)	8.97 ± 0.21	8.17 ± 0.19	−0.79 ± 0.21	8.98 ± 0.27	8.48 ± 0.31	−0.50 ± 0.19
Total thyroxine criteria subgroups[b]						
BPRS sum	26.3 ± 1.5	12.4 ± 1.2	−13.8 ± 1.7	22.1 ± 1.7	14.7 ± 1.9	−7.4 ± 2.1
Free T4 (ng/dl)	1.54 ± 0.04	1.38 ± 0.04	−0.16 ± 0.05	1.42 ± 0.05	1.36 ± 0.04	−0.06 ± 0.05
Total T4 (μg/dl)	9.35 ± 0.20	8.37 ± 0.19	−0.98 ± 0.23	8.40 ± 0.27	8.17 ± 0.30	−0.24 ± 0.10

[a]Good recovery, n = 50; poor recovery, n = 30.
[b]Good recovery, n = 48; poor recovery, n = 32.

related factors: initial thyroxine levels, direction of hormonal change, and magnitude of hormonal change. Although change appears to be the parameter of principal importance, it does not appear to be a simple matter of either decreasing versus increasing thyroxine levels correlating with clinical improvement, or simply a "dynamic" versus "nondynamic" group differentiation. For example, a patient with a low initial thyroxine level that then falls to a substantially lower level would not meet criteria for the good recovery group in spite of a large change value. Thus, direction of change in relation to the absolute initial value appears to be crucial, as well as the total magnitude of change. It appears that the patients most likely to improve fall within three good recovery subgroups: those with low initial thyroxine levels that then rise substantially, those with high initial thyroxine levels that then fall substantially, and those with initial thyroxine levels in the middle range that then change in either direction substantially. The patients not meeting the above criteria constitute the three poor recovery subgroups. The size of some of these resultant six subgroups in our sample is sufficiently small so that finer statistical analyses are precluded, especially with regard to diagnostic breakdown, but a few impressions from inspection of the data may be of interest in suggesting directions for further work along these lines.

The subgroup that appears least likely to improve clinically includes those patients with initial thyroxine levels in the low normal range that fail to increase or that decrease further during hospitalization. On the other hand, there appear to be some patients with initial thyroxine levels in the "high normal" range who show substantial clinical improvement, even though hormonal levels remain high and do not decline like the majority of those with high initial levels and good recovery. It would be of interest in future studies to obtain sufficiently large samples of these six subgroups so that possible connections with diagnostic or other clinical parameters might be evaluated. It was feasible, however, to examine the question of the diagnostic makeup of the two main recovery groups in our sample, and no significant differences were found between good versus poor recovery groups with regard to diagnosis of schizophrenia, affective disorder, or schizoaffective disorder for either FT_4 or TT_4 criteria. The question also arises concerning possible differences between the recovery groups with regard to symptomatology that might cut across diagnostic categories. However, we found no significant differences in mean values for any of the BPRS factors or individual items between good versus poor recovery groups using either FT_4 or TT_4 criteria. The evaluation of both primary diagnosis and BPRS symptomatology in our sample, therefore, appears not to indicate any strong link between these variables and the hormonal criteria, and the question of other possible clinical or psychological correlates of the thyroxine and recovery patterns remains for further exploration.

Our findings suggest that the concepts of "normalization" and "lability" of thyroid function may be particularly important to understanding relationships between thyroxine and recovery in a number of psychiatric disorders. This fits with the conclusions drawn from the early work of Reiss (1954) and with more recent findings of poor prognosis in depressed patients and in paranoid psychotic patients whose blunted TRH tests do not normalize after treatment with antidepressants or neuroleptics. (Kirkegaard et al. 1975, Langer et al. 1983). Our findings also bring to mind the therapeutic value of T_3 augmentation in some depressed patients who do not respond to tricyclic antidepressants (Goodwin et al. 1982). Whybrow and Prange (1981) have also observed that "adequate mobilization of thyroid function is associated with rapid clinical recovery from depression, whereas poor mobilization is associated with prolonged illness."

There appears to be a need for longitudinal studies with a battery of thyroid indices in large samples of depressed patients in order to search for more specific criteria for identifying those patients who may benefit from thyroid hormone administration. Our findings further suggest that thyroxine may have relevance not only to depression, but also to a broader range of psychiatric disorders, including schizophrenia, schizoaffective disorder, and bipolar disorder.

Our present study does not address the possible causes of the variations of thyroxine levels within the normal range in our patients. Although nutritional or other physical factors may be involved, it is also well established that thyroxine levels can respond sensitively in psychological stress situations (Mason 1968), so that initial thyroxine elevations in some patients may be primarily related to emotional arousal or "state" changes. With regard to initial low levels, little if anything is known as yet about whether or not certain "trait" psychological mechanisms might influence thyroid secretion in a manner similar, for example, to the lowering of chronic mean basal cortisol levels in human subjects who characteristically use denial as a major psychological defense during prolonged psychological stress (Mason 1975). Further work is needed on the psychological as well as the central neuroregulator mechanisms linked to the HPT system if we are to be able to explore the meaning of thyroid hormones in relation to psychiatric illness more fully and to generate more specific hypotheses concerning pathogenesis for testing in clinical research. The bidirectional nature of thyroid–brain interactions must also be considered, recognizing on the one hand the modulating influences of the brain on hormonal secretion, whereas on the other hand, considering that circulating thyroxine changes from any cause may act on the brain as a target organ and thereby potentially play a predisposing, precipitating, perpetuating, or ameliorating role in psychiatric illness.

Finally, a number of limitations of the present work should be emphasized. Our study involves only men with chronic or subchronic courses of illness and represents a survey of a broad range of diagnoses with a patient sample size large enough for analysis of primary diagnostic groups but not more specific diagnostic subgroups. Further longitudinal studies will be needed not only to see if our findings can be confirmed in other samples, but also to determine if they can be generalized to other populations and extended to useful clinical applications in particular diagnostic subgroups. The importance in future studies of using a more complete battery of HPT indices, including not only free and total thyroxine, but T_3, free thyroxine index, thyroid-binding globulin, TSH levels, and the TRH test when indicated, should be considered in order to provide much greater opportunities to determine whether hormonal differences or changes are based on secretory, binding, or conversion mechanisms and to help localize the functional level within the HPT axis that is determining the hormonal effect. Although psychotropic medications did not appear to play a significant role in the findings in our patient sample, the nature of medication and thyroxine interactions would appear to continue to deserve evaluation in future studies, perhaps particularly in the case of lithium. The preliminary nature of our findings should be emphasized, but it may be that the single most important practical implication of the present study is the attention that it draws to the apparent clinical psychiatric significance of thyroid hormonal levels that are within the normal range. It thereby reopens for fresh exploration the possibility that relatively fine-tuned study of the rather tightly regulated HPT system may reveal some practical applications to a variety of psychiatric disorders.

References

Bauer MS, Whybrow PC (1986): The effect of changing thyroid function on cyclic affective illness in a human subject. *Am J Psychiatry* 143:633–636.

Brody EB (1949): Psychologic tension and serum iodine levels in psychiatric patients without evidence of thyroid disease. *Psychosom Med* 11:70–73.

Cohen KL, Swigar ME (1979): Thyroid function screening in psychiatric patients. *JAMA* 242: 254–257.

Goodwin FK, Prange AJ, Post RM, et al (1982): Potentiation of antidepressant effects of L-triiodothyronine in tricyclic nonresponders. *Am J Psychiatry* 139:34–38.

Hoskins RG, Sleeper FH (1930): The thyroid factor in dementia praecox. *Am J Psychiatry* 10: 411–432.

Kirkegaard C, Norlem N, Lauridsen UB, et al (1975): Protirelin stimulation test and thyroid function during treatment for depression. *Arch Gen Psychiatry* 32:1115–1118.

Langer G, Aschauer H, Greta K, et al (1983): The TSH-response to TRH: A possible predictor of outcome to antidepressant and neuroleptic treatment. *Prog Neuro-Psychopharmacol Biol Psychiatry* 7:335–342.

Levy RP, Jensen JB, Laus VG, Agle DP, Engel IM (1981): Serum thyroid hormone abnormalities in psychiatric disease. *Metabolism* 30:1060–1064.

Mason JW (1968): A review of psychoendocrine research on the pituitary–thyroid system. *Psychosom Med* 30:666–681.

Overall JE, Gorham DR (1962): The Brief Psychiatric Rating Scale. *Psychol Rep* 10:799.

Reiss M (1954): Correlations between changes in mental states and thyroid activity after different forms of treatment. *J Ment Sci* 100:687–703.

Spitzer RL, Endicott J, Robins F (1978): Research Diagnostic Criteria. *Arch Gen Psychiatry* 35: 773–782.

Whybrow PC, Prange AJ (1981): A hypothesis of thyroid–catecholamine receptor interaction. *Arch Gen Psychiatry* 38:106–113.