SPECIAL REPORT

Prospective multicentre study on the prediction of relapse after antithyroid drug treatment in patients with Graves’ disease

H. Schleusener\(^1\), J. Schwander\(^1\), C. Fischer\(^2\), R. Holle\(^2\), G. Holl\(^1\), K. Badenhoop\(^3\), J. Hensen\(^1\), R. Finke\(^1\), U. Bogner\(^1\), W. R. Mayr\(^4\), G. Schernthaner\(^5\), H. Schatz\(^6\), C. R. Pickardt\(^7\) and P. Kotulla\(^1\)

Endocrine Department\(^1\), Medical Clinic, Klinikum Steglitz, Free University of Berlin, Berlin, FRG; Statistics and Data Centre\(^2\), University of Heidelberg, Heidelberg, FRG; IIInd Medical Clinic\(^3\), Klinikum Mannheim, University of Heidelberg, Mannheim, FRG; Department of Transfusion Medicine\(^4\), Medical Faculty, RWTH Aachen, FRG; IIInd Medical Clinic\(^5\), University of Vienna, Vienna, Austria; Medical Clinic\(^6\), Justus Liebig University, Giessen FRG and Medical Clinic\(^7\), University of Munich, Munich, FRG

Abstract. Graves’ disease is an autoimmune disease characterized by a course of remission and relapse. Since the introduction of antithyroid drug treatment, various parameters have been tested for their ability to predict the clinical course of a patient with Graves’ disease after drug withdrawal. Nearly all these studies were prospective and often yielded conflicting results. In a prospective multicentre study with a total of 451 patients, we investigated the significance of a variety of routine laboratory and clinical parameters for predicting a patient’s clinical course. Patients who had positive TSH receptor antibodies activity at the end of therapy showed a significantly higher relapse rate than those without \((P < 0.001)\). However, the individual clinical course cannot be predicted exactly (sensitivity 0.49, specificity 0.73, \(N = 391\)). The measurement of microsomal \((P = 0.99, \text{ sensitivity 0.37, specificity 0.63, } N = 275)\) or thyroglobulin antibodies \((P = 0.76, \text{ sensitivity 0.18, specificity 0.84, } N = 304)\) at the end of antithyroid drug therapy did not show a statistically significant difference in the antibody titre between the patients of the relapse and those of the remission group. Additionally, HLA-DR typing (HLA-DR3: \(P = 0.37, \text{ sensitivity 0.36, specificity 0.58, } N = 253\) was proven to be unsuitable for predicting a patient’s clinical course. Patients with abnormal suppression or an abnor-

In addition, the following colleagues participated in this multicentre study: G. Benker (Essen), J. Beyer (Mainz), W. Börner (Würzburg), K. Brandstetter (Mainz), R. Bretzel (Giessen), T. Droste (Tübingen), T. Fiek (Berlin), K. Gräf (Berlin), S. Grebe (Giessen), K. Hengst (Münster), U. Hopf (Berlin), M. Hübner (Heidelberg), K. Joseph (Marburg), T. Jungmann (Frankfurt), G. Kahaly (Mainz), K. Koppenhagen (Berlin), G. Müller-Eckardt (Giessen), D. Oppermann (Mainz), F. Raue (Heidelberg), Ch. Reiners (Essen), D. Reinwein (Essen), F. Schillerdecker (Frankfurt), F. Schulz (Frankfurt), P. Schumm-Draeger (Frankfurt), U. Schwedes (Mannheim), F. J. Seif (Tübingen), K. H. Usadel (Mannheim), M. Wdowinski (Mannheim), A. Witte (München), R. Ziegler (Heidelberg). Biometry: G. Maigatter (GSD, Berlin), R. Schlittgen (Essen).
Graves’ disease is thought to be an autoimmune disease in which antibodies against the TSH receptor stimulate thyroid function (McKenzie 1980; Burman & Baker 1985). Antithyroid drug therapy has been carried out with thioureylenes for more than 40 years in patients with Graves' thyrotoxicosis (Astwood 1944), though the treatment pattern varies widely in the different countries (Glinser et al. 1987). Many studies were undertaken to find parameters which can predict the clinical course of a patient with Graves’ disease after antithyroid drug withdrawal. These studies were performed with therapeutic regimens that varied with respect to the duration of therapy and follow-up, the dose of methimazole or propylthiouracil, and, in some cases, the criteria ‘relapse of hyperthyroidism’. The parameters investigated for their ability to predict a clinical course were the serum activity of TSH receptor antibodies, the HLA status and the suppression test (for review see Schleusener et al. 1987b,c). The studies also examined clinical parameters such as goitre size or the degree of an accompanying ophthalmopathy (Meng et al. 1982).

As to some of these tests, nearly all former investigations were performed in a retrospective design and yielded very conflicting results. We therefore undertook a prospective multicentre study from 1982 until 1986 in order to re-evaluate the significance of these parameters for predicting the clinical course in the individual patient with Graves’ disease.

**Patients and Methods**

**Eligibility**

Criteria for entry into the study were 1) status of hyperthyroidism (elevated T₃ and/or T₄ levels and lack of a TSH response to TRH) and 2) diffuse thyroidal uptake of technetium or radioiodine. Exclusion criteria were: 1) the patient wished to undergo surgical or radioiodine treatment; 2) a large goitre which required surgery because of local problems or 3) 131-I-iodine treatment was indicated because of other severe diseases. Patients were excluded from evaluation if they required surgery or radioiodine during the time of therapy or if they could not be followed up.

**Treatment and follow-up**

Antithyroid drug therapy generally lasted for 1 year, starting with 40 mg of methimazole. In most patients, we reduced the dosage to 5–10 mg of methimazole after they had become euthyroid. Some received 50 or 100 mg of propylthiouracil because they had shown allergic reactions to methimazole. In order to prevent hypothyroidism during antithyroid drug therapy, most patients additionally received 50 µg of L-T₄ daily; there was no difference in the results of patients with and without T₄-substitution. The patients were followed up for at least 1 year after the end of drug treatment.

**Evaluation of subjects**

Clinical and laboratory evaluation included medical history of the patients, goitre size determined by palpation, degree of ophthalmopathy (WHO classification) and the following laboratory tests:

A radioligand receptor assay was used as previously described (Schleusener et al. 1978) to measure serum activity of TSH receptor antibodies (TBlab) at least 3 times: before or during the first 3 weeks of therapy, at the end of therapy, and at the end of the follow-up period. This assay, using thyroid cell membrane fragments, proved to be more sensitive than the commercially available kits (Schleusener et al. 1987a,b).

Microsomal and thyroglobulin antibodies (Mc-ab and Tg-ab, respectively) were measured at least twice, i.e. at the beginning and end of therapy, using Wellcome kits (Thymune M, Thymune T, Wellcome Diagnostics, Dartford, UK). TBlab and Tg-ab were measured centrally at Klinikum Steglitz and the Medical Clinic of Gießen, respectively. T₄, T₃, and TSH were determined radioimmunologically using commercially available kits.

The HLA type was identified at any given time during therapy with the standard NIH microlymphotoxicity test for A, B and C, and the two-colour fluorescence technique (van Rood et al. 1976) for DR. Well characterized sera were used from the 7th, 8th and 9th International Histocompatibility Workshops, including positive and negative controls. HLA-typing was performed centrally in Dr Mayr’s laboratory.

A suppression test was carried out within an interval extending from 2 weeks before to 1 week after drug discontinuation and 30 days after administration of a single dose of 3 mg L-T₄. In 213 healthy controls, concomitant TRH tests had proved that this pattern guaranteed thyroid suppression. The suppressibility was measured 2 h after the administration of iodine-123 and was defined as ‘normal’, when less than 15% iodine were taken up by the thyroid.

A TRH test was performed at least 3 times: at the beginning of treatment, 3 to 12 weeks after termination of the antithyroid drug regimen, and at the end of follow-up. TSH was measured basally and after IV (or sometimes nasal) application of TRH by radioimmunoassay with commercially available kits. A difference of 2.5–25 mU TSH/l between the basal and the stimulated value was taken to be normal.

The goitre size was determined by palpation and evaluated according to the WHO classification. At the time the study was started in 1982, ultrasound measure-
ment of the thyroid was not yet available in all participating clinics and was therefore not used to determine goitre size. The WHO classification was also applied to assessing ophthalmopathy. This proved to be the most convenient classification for all participating clinics. These clinical parameters were evaluated at least 3 times: at the beginning and end of drug therapy and at the end of the follow-up period.

Response criteria and statistical analysis

Response was defined as the recurrence of hyperthyroidism within the 1-year follow-up period and remission as a euthyroid state during this period. The primary question was whether TBIab and the suppression test, performed at the end of therapy are suitable parameters for predicting remission or relapse. Furthermore, all other clinical and laboratory parameters were examined as potential prognostic factors. The prognostic value of a parameter was assessed in terms of sensitivity and specificity. Sensitivity refers to the proportion of relapsed patients who had been predicted to relapse, and specificity indicates the proportion of patients in remission who had been predicted to stay in remission. The calculation of sample size was based on the estimation of 95% confidence intervals for sensitivity and specificity with a deviation of 10% simultaneously for the two main questions. The hypothesis of equal relapse rates in subgroups was checked with the chi-square test.

In order to investigate the relationship between the relapse rate and the age of the patients, we used the statistical method of moving relapse-rate estimation, where the relapse rate is successively estimated in overlapping age groups of 45 patients. The connecting curve between these groups, presented in Fig. 2, was smoothed by spline interpolation.

Multivariate analysis

There are several possibilities for constructing rules of prognosis based on more than one variable. The most simple non-parametric way is to take two dichotomous variable such as the suppression test and TBIab. From the univariate point of view, a relapse of hyperthyroidism would be predicted if the suppression test is abnormal or if TBIab are detectable. Combining these two parameters, results can be categorized in the following four subgroups:

- suppression abnormal and TBIab detectable (+, +)
- suppression abnormal and TBIab not detectable (+, −)
- suppression normal and TBIab detectable (−, +)
- suppression normal and TBIab not detectable (−, −)

An optimistic prognostic rule would predict a relapse only for patients in whom both parameters are abnormal (+, +). A pessimistic rule, however, predict a relapse for all patients with at least one abnormal parameter (+, +), (+, −), (−, +). Obviously, the pessimistic rule yields a higher sensitivity than the optimistic one and was therefore the rule we applied. Since multivariate analysis based on a logistic model did not lead to the formulation of a better rule for prognosis, we will not present the data obtained by this method.

![Medical Clinic](image)

**Fig. 1.**

Relapse rates of the various clinics (black bars) and their 95% confidence limits in patients with Graves' disease 1 year after antithyroid drug treatment.

44*

Downloaded from Bioscientifica.com at 12/01/2018 06:51AM via free access
Results

Patient population
From January 1982 until December 1984, 1096 patients were recruited from thyroid outpatient departments of 14 medical centres throughout the FRG. Among them, 155 patients (14%) underwent thyroid surgery or radioiodine treatment during the time of drug therapy; 277 (25%) could not be followed up, and 38 patients (3%) were excluded because they had received antithyroid drugs for less than 4 months. Of the 626 patients (58%) who completed antithyroid drug therapy, 175 were excluded from further evaluation because they dropped out during the follow-up period (N = 139) or underwent thyroid surgery or radioiodine therapy before suffering a relapse of hyperthyroidism (N = 36). There was no difference in the distribution of positive and negative test results between 451 patients evaluated and the 626 patients who completed antithyroid drug treatment.

Relapse rate
We evaluated 451 patients; 227 of them (50.3%) relapsed within 1 year after drug withdrawal, and 224 stayed in remission. At the end of the follow-up period, a TRH test was performed in 92 remission patients, 24 of whom (26%) showed no TSH response to TRH application ('latent hyperthyroidism'). Among the 127 patients followed for more than 2 years after the termination of therapy, 61 (48%) had relapsed within 1 year and 71 (56%) within 2 years after drug withdrawal. The relapse rate (50.3%) varied from 30 to 67% among patients in the 14 co-operating clinics. Fig. 1 shows the different relapse rates and their 95% confidence limits. Both Berlin and Heidelberg had two participating clinics. The relapse rates were 56 and 36% in the two Berlin clinics and 56 and 67% in Heidelberg.

Fig. 2 indicates that the relapse rate is dependent on the age of the patient: it is lowest in patients from 30 to 40 years of age, but begins to increase steadily thereafter. Seventy-seven patients were under 30 years of age, 78 between 30 and 40, and 296 over 40. The differences between the relapse rates of these groups are significant (<30 vs 30-40, P = 0.012; 30-40 vs >40 years, P < 0.001; chi-square); these P-values, however, have to be interpreted in an explorative sense.

No significant difference could be found between the relapse rates of female and male patients: 187 of 378 female patients (49%) and 37 of 73 male patients (51%) relapsed (P = 0.39).

Whether the patient had an initial occurrence or a relapse of hyperthyroidism could be determined at the beginning of treatment in 413 patients, 132 (47%) of 280 patients who had a first occurrence of hyperthyroidism relapsed within 1 year after drug discontinuation. Nearly the same relapse rate (50%) was found in 101 patients with a first recurrence of hyperthyroidism. In contrast, patients who had already had one or more relapses of hyperthyroidism before receiving this treatment had a significantly higher relapse rate of 75% (24/32, P = 0.003)

Measurement of thyroid antibodies
Evaluation of the different parameters is based on slightly varying collective analyses because data were partly missing in some patients. Before treatment, we found TBIab activity in 248 of 354 (70%) patients with toxic 'diffuse' goitre and in 90 of 107 (84%) patients with concomitant Graves' ophthalmos. Presumably, some of the patients without
ophthalmopathy suffered from disseminated thyroid autonomy. At the end of therapy, TBlab activity was found in only 38% of the patients, and the titres were lowered in most antibody-positive patients. Patients who had been antibody-positive or -negative at the beginning of therapy showed no differences in their relapse characteristics (P = 0.76); the values for sensitivity and specificity were 0.71 and 0.31, respectively (N = 354). TBlab, measured at the end of therapy, indicated a statistically significant relapse rate (P < 0.001); the sensitivity and specificity, however, were around 0.49 and 0.73 (Table 1). No differences were found by also considering HLA-DR3 antigen (N = 123, P < 0.001, sensitivity 0.32, specificity 0.78) or TBlab-positivity at the beginning of therapy and exophthalmos (N = 115, P < 0.001, sensitivity 0.70, specificity 0.79).

To investigate the value of the other thyroid antibodies in predicting the clinical course, we divided our patients into groups with high (Mc-ab > 1:1600, Tg-ab > 1:160) and low (Mc-ab ≤ 1:1600, Tg-ab ≤ 1:160) antibody titres. At the beginning of drug treatment neither the titre of microsomal antibodies (N = 312, P = 0.46, sensitivity 0.53, specificity 0.43), nor that of thyroglobulin antibodies (N = 381, P = 0.153, sensitivity 0.15, specificity 0.79) showed significant values for P, sensitivity, or specificity. At the end of drug therapy, a comparable result was found (Tables 2 and 3).

**Table 1.**
Relapse rate of patients with thyrotoxicosis 1 year after antithyroid drug treatment according to TBlab at the end of therapy.

<table>
<thead>
<tr>
<th>TBlab</th>
<th>No relapse</th>
<th>Relapse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>147</td>
<td>96</td>
</tr>
<tr>
<td>Positive</td>
<td>55</td>
<td>93</td>
</tr>
<tr>
<td><strong>P &lt; 0.001</strong></td>
<td><strong>N = 391</strong></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.49</td>
<td>(95% CI 0.42-0.56)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.73</td>
<td>(95% CI 0.66-0.79)</td>
</tr>
</tbody>
</table>

**Table 2.**
Clinical status of patients with Graves' disease 1 year after antithyroid drug treatment according to the titre of microsomal antibodies (Mc-ab) at the end of therapy.

<table>
<thead>
<tr>
<th>Mc-ab</th>
<th>≤ 1:1600</th>
<th>&gt; 1:1600</th>
</tr>
</thead>
<tbody>
<tr>
<td>No relapse</td>
<td>88</td>
<td>51</td>
</tr>
<tr>
<td>Relapse</td>
<td>86</td>
<td>50</td>
</tr>
<tr>
<td><strong>P = 0.99</strong></td>
<td><strong>N = 275</strong></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.37</td>
<td>(95% CI 0.29-0.45)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.63</td>
<td>(95% CI 0.55-0.71)</td>
</tr>
</tbody>
</table>

**Table 3.**
Clinical status of patients with Graves' disease 1 year after antithyroid drug treatment according to the titre of thyroglobulin-antibodies (Tg-ab) at the end of therapy.

<table>
<thead>
<tr>
<th>Tg-ab</th>
<th>≤ 1:160</th>
<th>&gt; 1:160</th>
</tr>
</thead>
<tbody>
<tr>
<td>No relapse</td>
<td>127</td>
<td>25</td>
</tr>
<tr>
<td>Relapse</td>
<td>125</td>
<td>27</td>
</tr>
<tr>
<td><strong>P = 0.76</strong></td>
<td><strong>N = 304</strong></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.18</td>
<td>(95% CI 0.12-0.76)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.84</td>
<td>(95% CI 0.76-0.89)</td>
</tr>
</tbody>
</table>

**Table 4.**
Relapse and remission in patients with Graves' disease 1 year after antithyroid drug treatment according to HLA-DR3 locus.

<table>
<thead>
<tr>
<th>HLA</th>
<th>Others</th>
<th>DR3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remission</td>
<td>72</td>
<td>52</td>
</tr>
<tr>
<td>Relapse</td>
<td>82</td>
<td>47</td>
</tr>
<tr>
<td><strong>P = 0.36</strong></td>
<td><strong>N = 253</strong></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.36</td>
<td>(95% CI 0.28-0.47)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.58</td>
<td>(95% CI 0.49-0.67)</td>
</tr>
</tbody>
</table>
cantly less often than those without (P < 0.049). Sensitivity and specificity values for HLA-DR5, however, were unsatisfactory (N = 253, sensitivity 0.18, specificity 0.72).

Suppression test
We measured thyroid suppressibility at the end of antithyroid drug therapy in 259 patients. Suppression was normal in 134 patients and abnormal in 125. The relapse rate for the two groups was significantly different (62% in patients with abnormal vs 31% in those with normal suppression, P < 0.001). The value for both sensitivity and specificity was 0.66 (Table 5).

TRH test
The relapse rate varied significantly (P < 0.001, N = 228), being 25 vs 53% in patients with a normal and an abnormal TRH test, respectively, 3 to 12 weeks after the termination of therapy. The sensitivity was rather high, 0.89, but the specificity was only 0.30 (Table 6).

Table 5.
Clinical status of patients with Graves’ disease 1 year after antithyroid drug treatment according to suppression test at the end of therapy.

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Abnormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>No relapse</td>
<td>93</td>
<td>47</td>
</tr>
<tr>
<td>Relapse</td>
<td>41</td>
<td>78</td>
</tr>
<tr>
<td>P &lt; 0.001</td>
<td>N = 259</td>
<td></td>
</tr>
<tr>
<td>Sensitivity 0.66</td>
<td>(95% confidence intervals 0.56–0.74)</td>
<td></td>
</tr>
<tr>
<td>Specificity 0.66</td>
<td>(95% confidence intervals 0.58–0.74)</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.
Clinical course of patients with Graves’ disease 1 year after antithyroid drug treatment according to the TRH test within 6 weeks after the end of therapy.

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Abnormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>No relapse</td>
<td>36</td>
<td>84</td>
</tr>
<tr>
<td>Relapse</td>
<td>12</td>
<td>96</td>
</tr>
<tr>
<td>P &lt; 0.001</td>
<td>N = 228</td>
<td></td>
</tr>
<tr>
<td>Sensitivity 0.89</td>
<td>(95% confidence intervals 0.81–0.94)</td>
<td></td>
</tr>
<tr>
<td>Specificity 0.30</td>
<td>(95% confidence intervals 0.22–0.39)</td>
<td></td>
</tr>
</tbody>
</table>

Goitre size
Goitre size was documented at the beginning of therapy in 396 patients. The patients fell into two categories: one group (N = 203) with a small goitre or no palpable thyroid enlargement (≤ WHO I b) and the other group (N = 193) with a large goitre (≥ WHO I I). These two groups had significantly different relapse rates: 43 and 54%, respectively (P < 0.03). The sensitivity was 0.54 and the specificity 0.57 (Table 7). Regarding goitre size at the end of therapy we found only a slight difference between the relapse rates of these two categories (48 vs 53%, P = 0.41, N = 339). Being 0.46 and 0.58, respectively, the sensitivity and specificity were similar to those at the beginning of therapy.

Ophthalmopathy
To evaluate whether the degree of Graves’ ophthalmopathy can be used as a parameter for predicting the clinical course of hyperthyroidism, we divided our patients into one group with exophthalmos (≥ WHO II I) and another group with minor eye signs (≤ WHO II). Neither at the beginning (Table 8), nor the end of therapy (N = 327, P = 0.06, sensitivity 0.38, specificity 0.72) could any significant result be found for this parameter.

Combinations of various parameters
Table 9 shows the sensitivity and specificity values for several combinations of two parameters. Some of these groups are much smaller than the total group of 451 patients, and selection effects must therefore be taken into account. A combination of two parameters always showed a higher sensitivity than the univariate parameters in the respective collective analysis. No combination, however, has
Clinical status of patients with Graves’ disease 1 year after antithyroid drug treatment according to degree of ophthalmopathy at the beginning of therapy.

<table>
<thead>
<tr>
<th>No relapse</th>
<th>WHO 0-II</th>
<th>WHO III-VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relapse</td>
<td>131</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>62</td>
</tr>
</tbody>
</table>

$P < 0.205$ \text{ (} N = 361 \text{)}

Sensitivity 0.35 \text{ (95% confidence intervals 0.28–0.42)}

Specificity 0.71 \text{ (95% confidence intervals 0.64–0.77)}

Table 8.

Better values of both sensitivity and specificity than the univariate parameter and thus does not improve the prediction of an individual’s course. A good example of this is the combination of the suppression test and TBIab measured at the end of treatment, two parameters with highly significant $P$-values. The combination of these parameters is also associated with a highly significant $P$-value ($<0.001$), but the sensitivity value is not markedly better than the univariate parameter, and the specificity value is even lower (Table 9).

Table 9.

Sensitivity and specificity values in combinations of 2 parameters.

<table>
<thead>
<tr>
<th>Combination of parameters</th>
<th>No. of patients</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supp + TBIab</td>
<td>247</td>
<td>0.80</td>
<td>0.49</td>
</tr>
<tr>
<td>Supp + TRH</td>
<td>153</td>
<td>0.94</td>
<td>0.13</td>
</tr>
<tr>
<td>Supp + EO</td>
<td>202</td>
<td>0.83</td>
<td>0.52</td>
</tr>
<tr>
<td>Supp + goitre</td>
<td>233</td>
<td>0.78</td>
<td>0.58</td>
</tr>
<tr>
<td>TRH + goitre</td>
<td>200</td>
<td>0.94</td>
<td>0.19</td>
</tr>
<tr>
<td>EO + goitre</td>
<td>259</td>
<td>0.65</td>
<td>0.45</td>
</tr>
<tr>
<td>TBIab + goitre</td>
<td>341</td>
<td>0.72</td>
<td>0.40</td>
</tr>
<tr>
<td>Age + TBIab</td>
<td>391</td>
<td>0.78</td>
<td>0.52</td>
</tr>
<tr>
<td>Age + supp</td>
<td>259</td>
<td>0.82</td>
<td>0.58</td>
</tr>
<tr>
<td>Age + EO</td>
<td>327</td>
<td>0.70</td>
<td>0.44</td>
</tr>
<tr>
<td>Age + goitre</td>
<td>396</td>
<td>0.78</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Supp: suppression test; TBIab: TSH receptor antibodies; TRH: TRH test; EO: ophthalmopathy in 2 stages (≤ II vs ≥ III); goitre: goitre size in 2 stages (≤ I vs ≥ II); age: patient age in 2 stages (≤ 50 vs > 50 years).

Discussion

Previous studies on antithyroid drug treatment have yielded widely divergent results on the relapse rate and the possibility of reliably predicting the clinical course in patients with Graves’ disease (Slingerland & Burrows 1979; McGregor et al. 1980; Meng et al. 1982; Romaldini et al. 1983). These discrepancies might be partially due to differences in the duration of therapy, the maintenance dose, the follow-up period or the iodine supplementation. Relapse rates of 30% (Meng et al. 1982), 45% (Thalassinos et al. 1974), 51% (Allanic et al. 1983), and 58% (Romaldini et al. 1983) were reported for patients who received treatment comparable to that applied in this multicentre study. These differences might be explained by the small number of patients in the individual studies. In this multicentre study, relapse rates obtained in the different clinics varied from 30 to 67% (Fig. 1). Great differences arose even in two clinics in the same town. The overlapping of 95% confidence limits at a 50% relapse rate indicates that the differing relapse rates are accidental and caused by the small number of patients in the individual centres.

It is obvious that the percentage of hyperthyroidism relapse increases with patient age (Fig. 2). This may be due to the development of autonomously functioning nodules, which might exist even in Graves’ goitres and would be especially prevalent in the iodine-deficient area of Germany (Studer 1986). The size of these autonomous nodules often increases with age, so that, especially in elderly people, a relapse of hyperthyroidism may be caused not only by antibody stimulation of the TSH receptor but also by uncontrolled production of thyroid hormones in an autonomously functioning nodule (Studer 1986). No difference could be found between the relapse rates of female and male patients. Obviously, patients with at least two former occurrences of hyperthyroidism relapse significantly more often after a further antithyroid drug treatment than patients with a first occurrence of thyrotoxicosis. Patients with a first or second occurrence, however, do not differ in their relapse rate.

Earlier studies yielded controversial results as to whether antithyroid treatment over more than 2 years results in a higher remission rate than short-term therapy. In contrast to Tamai et al. (1980),
who reported a significant difference in the relapse rate depending on the duration of therapy, Meng et al. (1982), Bouma et al. (1982) and Madec et al. (1984) denied that prolonged therapy yielded better results. Controversial is also the question concerning the best dosage of antithyroid drugs: Romaldini et al. (1983) found a difference in remission rates of patients given more than 40 mg and less than 25 mg of methimazole. Further information on this topic will be obtained in a current multicentre prospective study (Benker et al. 1987).

We retrospectively evaluated those few patients who, for unknown reasons, were treated for a total of less than 9 months or more than 16 months or who received permanent therapy with 40 mg of methimazole. No difference could be observed between these patients and the majority of those treated according to the study design.

Many of the previous investigations on the validity of predictive parameters used P-values to determine a significant difference in the clinical course of relapse and remission patients. This, however, is impracticable in daily clinical routine when a decision has to be made about an individual patient. We therefore calculated the values for sensitivity (percentage of hyperthyroid patients, who are correctly predicted to relapse by the investigated test system) and specificity (percentage of the euthyroid patients who are correctly predicted to stay in remission). High sensitivity and specificity values indicate a reliable prediction for an individual patient. Ideally, both values are 1.0, but lower values normally have to be accepted.

The utility of a prediction parameter in the daily clinical routine depends on the consequences of a wrong decision based on the prediction of remission or relapse. The two types of wrong decision have different consequences. The risk of continuing antithyroid drug therapy is very low (Krüsskemper 1960), whereas a relapse of hyperthyroidism may be more risky, especially in the elderly. Therefore it may be more dangerous wrongly to stop antithyroid drug therapy than to continue it unnecessarily.

We therefore think a prediction parameter should have a high sensitivity (of at least 0.90) in order to reduce the probability of wrongly stopping therapy. A lower specificity (at least of 0.70) on the other hand seems acceptable, since it only increases the probability of wrongly continuing drug treatment.

TSH receptor antibodies have most frequently been investigated as a parameter for clinical prediction in Graves' disease. These antibodies stimulate thyroid function, and thus their detection was thought to be a good parameter (Burman & Baker 1985). This hypothesis seemed to be confirmed in studies by Davies et al. (1977), O'Donnell et al. (1978), Schleusener et al. (1978), McGregor et al. (1980), Hardisty et al. (1981), Wilson et al. (1985), and Creagh et al. (1985). These studies, however, had a maximum of 65 patients (McGregor et al. 1980) and other authors (Docter et al. 1980; Weetman et al. 1986) could not confirm the results. It must also be noted that, up to now, all investigators calculated whether the relapse group had a statistically higher prevalence of TBlab activity than the remission group. The sensitivity and specificity of this method, however, which are important for predicting an individual clinical course, were not determined. In a very recently published review on TSH receptor antibodies (Smith et al. 1988), it was speculated that our TBlab assay using thyroid membrane fragments is less sensitive and specific than the methods applied by other groups (Wilson et al. 1985, 1986). In a comparison of the available TBlab assay systems, we could show that ours is at least as sensitive and specific as the others (Schleusener et al. 1987a).

By re-evaluating the prognostic validity of TBlab measurement in 391 patients, we found that those with detectable TBlab at the end of therapy have a significantly higher relapse rate. A reliable prognosis for an individual patient is not possible, however, owing to the low sensitivity and specificity values. Measurement of TBlab at the beginning of antithyroid drug therapy does not allow reliable prediction either.

Some investigations have indicated that patients can become hypothyroid many years after treatment for hyperthyroidism (Irvine et al. 1977) and that this development is often accompanied by a rise of microsomal and/or thyrogblobulin antibodies (Wood & Ingbar 1979; Lamberg et al. 1981; Hirota et al. 1986). In contrast to these studies, results were published to the effect that high titres of Mca-b (Hamada et al. 1987) or Tg-ab (Meng et al. 1981) indicate a relapse of hyperthyroidism after drug withdrawal. In our study, we could not find a significantly different relapse rate for patients with high and low antibody titres at any given time during therapy, nor were we able to achieve sufficient
reliability for an individual prognosis by measuring antibody titres. The follow-up time in this study, however, was too short for a development of immune thyroiditis (Lambert et al. 1981), so that Mc-ab or Tg-ab titres could not be expected to predict hypothyroidism. Furthermore, new findings by our group (Bogner et al. 1987) in sera of patients with Graves' disease show that the Mc-ab and Tg-ab titres do not correlate with the cytotoxic activity against human thyroid cells. This is in accordance with the results of study by Lazarus and co-workers (1981), who found no correlation between the prevalence of thyroid antibodies and hypothyroidism in elderly.

There are also some methodological arguments to explain our results: Mc-ab and Tg-ab are normally synthesized by intrathyroidal lymphocytes (McLachlan et al. 1983), and this is probably also true for TBlab. Therefore, the detection of antibody activity in serum may only indicate an ‘overflow’ of intrathyroidal synthesis. Secondly, the radioligand receptor assay for TBlab measures not only the stimulating but also the blocking TSH antibodies. Therefore, TBlab may also be detected in patients with Graves' disease who are in remission (Endo et al. 1978; Schleusener et al. 1978).

It is well known that Caucasians with Graves' disease show a relatively high prevalence of HLA antigens B8 and DR3 (Grumet et al. 1974; Farid 1981; Dahlberg et al. 1981). HLA-typing of 253 patients in this study yielded a prevalence of 34% HLA-B8 and 39% HLA-DR3 in patients with Graves' disease, which is significantly elevated compared with normals. We previously found a prevalence of 18% HLA-B8 and 20% HLA-DR3 in normal Caucasians in our region (Schleusener et al. 1983). In patients with Graves' disease prevalences of 44–53% HLA-B8 (Allanic et al. 1980; Dahlberg et al. 1981) and 64% (Farid 1981) or 51–55% (Allanic et al. 1980; Dahlberg et al. 1981) HLA-DR3 have been reported; these prevalences were significantly elevated compared with normals. The prevalence of HLA-DR5 in the patients with Graves' disease of this study was as high as in 963 normal Caucasians (23 vs 26.6%). We are aware that HLA-DR5 can nowadays be split into HLA-DRw11 and HLA-DRw12 (Baur et al. 1984). This is, however, not so important here, because DRw12 is extremely seldom in Caucasians (Mayr, personal communication).

Determination of HLA-status as a method for predicting the clinical course was proposed by Bech et al. (1977) and McGregor et al. (1980) who found significantly different frequencies of HLA-DR3 in relapse and remission patients. Other studies (Schernthaner et al. 1980; Dahlberg et al. 1981; McKenna et al. 1982; Allanic et al. 1983; Weetman et al. 1986), however, are in agreement with our results indicating that none of the three tested HLA-antigens B8, DR3 and DR5 permitted reliable prediction of the individual clinical course. The difference between the prevalence of HLA-DR5 in the relapse and remission group is of borderline significance (uncorrected P-value) and has to be ascertained by further investigations. As already stated in a recent publication (Wilson et al. 1986), the concomitant detection of TBlab and HLA-DR3 does not worsen an individual's prognosis for remission.

Obviously, normal suppression of radioiodine uptake has often been proposed as an indicator for termination of drug treatment, because patients with normal suppression were more frequently found to stay in remission than those with suppressed uptake (Cassidy 1965; Alexander et al. 1970; Lowry et al. 1971; Hackenberg et al. 1973; Meng et al. 1983; Yamamoto et al. 1983; Yamada et al. 1983). Observed in our patients as well, this difference was found to be highly significant statistically. The low sensitivity and specificity values, however, prevent a reliable prediction for the individual patient. Even the combination of two parameters with highly significant P-values (suppression test and TBlab measured at the end of therapy) cannot reliable indicate the individual clinical course.

The TRH test as a predictive marker was investigated by Meng et al. (1983) and Dahlberg et al. (1985) who could demonstrate that patients with an abnormal TRH test after treatment relapse significantly more often than those with a normal TSH response to TRH. This is also valid for our study. Of greater importance is the finding that 89% of the patients with a future relapse of hyperthyroidism showed no TSH response to TRH at the time antithyroid drug therapy was discontinued. This was the highest sensitivity value of all single parameters tested. The specificity value, however, was too low. Combining the TRH and the suppression test led to a slight improvement of sensitivity, but still to a very low specificity value. Since ultrasensitive TSH measurement was not yet com-
mercially available at the beginning of this investigation, we have no data on its predictive value.

Some authors have studied the significance of clinical parameters for the future course of Graves' disease. Meng et al. (1982) and Laurberg et al. (1986) found that patients with large goitres show a higher relapse rate than those with small goitres or normal thyroids. We were able to confirm these results in the present investigation. Our study was performed in Germany, a region well-known for iodine deficiency and thyroid enlargement (Gutekunst et al. 1986). Autonomously functioning areas are thought to develop in some of these enlarged thyroids (Gerber et al. 1983), possibly even in Graves' goitres (Studer 1986). Müller-Gärnter et al. (1986), for example, reported on nodules in thyroids of patients with Graves' disease which did not respond to TSH receptor stimulation; and, in iodine-deficient areas, serum TSH levels related to goitre size (Fenzi et al. 1985). The high statistical significance of the parameters of thyroid suppressibility in this study may be interpreted as the result of an influence exerted by autonomously functioning nodules on the clinical course in our patients. Thus, there may be two reasons for a relapse after antithyroid drug therapy in patients with Graves' disease: firstly, increased immunological stimulation of the TSH receptors and, secondly, autonomous production of thyroid hormones in autonomously functioning areas. The extent of such autonomous areas is probably subject to a high degree of interindividual variation, and this may explain, along with immunological, genetic and environmental factors, why reliable prediction for the individual person is impossible.

The degree of an endocrine ophthalmopathy was not found reliably to indicate relapse and remission in patients with Graves' disease either in earlier studies (Meng et al. 1982) or in our investigation. It is controversial whether ophthalmopathy and Graves' hyperthyroidism are two independent diseases that often occur concomitantly. The development of the oedema therefore might not be an indicator of the course of the other (Wall et al. 1981; Jacobson & Gorman 1984).

The fact that the course of Graves' disease cannot be reliably predicted may be explained by the complex mechanism involved: many factors can influence the development of an autoimmune disease (Wick 1987). Major histocompatibility complex (MHC) genes as well as non-MHC-linked genes are responsible for the disturbance of immunoregulation and the susceptibility of the target organ. External factors such as steroids, viruses, bacteria, iodide, and stress (Tardieu et al. 1984; McGregor et al. 1985; Wick 1987; Wenzel et al. 1987) can influence the manifestation of an autoimmune thyroid disease. Additionally, the recurrence of hyperthyroidism might be explained in some cases by the finding of Studer (1986) that autonomous areas develop in long-standing Graves' goitres. This complexity makes it unlikely that a single factor, like certain HLA alleles, which represent only part of the immunogenetic factors, or the measurement of TBlab or Mc-ab, which indicates the current immunological status, can be used as a predictive parameter for the course of Graves' disease.

Acknowledgments

This study was financially supported by the Bundesminister für Forschung und Technologie (01 ZR 8508).

References


Bogner U, Wall J R & Schleusener H (1987): Cellular and antibody mediated cytotoxicity in autoimmune thy-


Mckenna R, Kearns M, Sugrue D, Drury M I & McCarthy...


Accepted February 23rd, 1989.

Dr H. Schleusener,
Medizinische Klinik,
Endokrinologische Abteilung,
Klinikum Steglitz der Freien Universität,
Hindenburgdamm 30,
D-1000 Berlin 45, FRG.