Can we predict relapse in Graves’ disease?

Results from a Systematic Review and Meta-analysis

*Tristan Struja¹, Hannah Fehlberg¹, Alexander Kutz¹, Larissa Guebelin¹, Christian Degen¹, Beat Mueller¹,², Philipp Schuetz¹,²

*equally contributing first authors

1 Kantonsspital Aarau, Aarau, Switzerland, Medical University Department, Clinic for Endocrinology, Diabetes & Metabolism
2 Basel, Switzerland, Medical Faculty of the University of Basel

ABBREVIATED TITLE: Relapse Risk in GD before treatment

WORD COUNT: 3582; NUMBER OF FIGURES: 2; NUMBER OF TABLES: 2

KEY TERMS: Graves’ disease, systematic review, meta-analysis, relapse, recurrence, ATD

SUPPLEMENTARY MATERIAL: Appendix/supplemental file

AUTHOR LIST FOR INDEXING: T. Struja MD (tristan.struja@gmail.com) and H. Fehlberg (hannah.fehlberg@stud.unibas.ch), A. Kutz MD (alexander.kutz@ksa.ch), L. Guebelin (larissa.guebelin@gmx.ch), C. Degen MD (christian.degen@ksa.ch), Prof. B. Mueller MD (beat.mueller@ksa.ch), Prof. Ph. Schuetz MD, MPH (philipp.schuetz@ksa.ch)

Corresponding author and person to whom reprint requests should be addressed:

Dr. Tristan Mirko Struja, MD
University Department of Medicine, Kantonsspital Aarau
Tellstrasse 15, CH-5001 Aarau, Switzerland
tristan.struja@gmail.com (e-mail)
Abstract

Background: Identification of pre-treatment risk factors predicting relapse in patients with hyperthyroidism of Graves’ disease after stopping anti-thyroid drugs [ATD] is decisive to guide therapeutic options.

Purpose: We performed a systematic search and meta-analysis to study predictors for relapse after stopping ATD in patients with Graves’ disease.

Methods: Based on a pre-specified protocol, we searched PubMed, EMBASE and Cochrane in July 2015 for case-control, controlled, and randomized-controlled trials reporting risk factors for relapse after stopping ATD. The primary endpoint was relapse of disease until follow-up. PRISMA and SIGN statements were used for reviewing the data and assessing quality of included trials.

Results: We included 54 trials with a total of 7595 participants. Most trials were small with moderate to high risk for bias. 10 trials were assessed qualitatively only (2227 patients), genomic data was reported in 13 trials (2178 patients) and 31 trials (4346 patients) were assessed quantitatively. In total, there were 3696 relapses in 7595 patients (48.7%).

By using random-effects meta-analysis, orbitopathy, smoking, thyroid volume measured by sonography, goiter size, fT4, tT3, TRAb and TBII was significantly associated with relapse, whereas male vs. female sex, age, and initial tT4 level did not show significant associations.

Conclusions: This analysis found several risk factors to predict relapse in Graves’ disease which can be combined in a risk score. Prospective studies should evaluate the prognostic accuracy of such a score to guide treatment decisions.
Introduction

Graves’ disease is the most common cause of primary hyperthyroidism in iodine sufficient areas. The overall prevalence of Graves’ disease varies but is reported to be around 0.5% (1). Graves’ disease is characterized by the autoimmune production of thyrotropin related antibodies [TRAb] stimulating the thyroidal cells and causing an overproduction of thyroid hormones (2). Furthermore, up to 50% of patients develop Graves’ orbitopathy by co-expression of TSH-receptors by pre-adipocytes within the orbital cavities inducing proliferation and thus proptosis (3).

One standard treatment for Graves’ disease includes use of anti-thyroid drugs [ATD] for a total duration of 12 to 18 months. ATD may also have immuno-modulating effects and positive impact on the course of Graves’ disease (4). This approach is generally preferred in Europe and Asia it offers the possibility of disease resolution without residuals (i.e., post-interventional hypothyroidism, worsening of orbitopathy) (5). Yet, risk for relapse is high (around 50%) and patients may suffer from side effects such as exanthema, pruritus, urticaria, hepatitis and agranulocytosis, especially those treated with higher doses. Due to the high risk of relapse, an alternative first line treatment used (mostly in the United States) is radioactive-iodine [RAI] ablation, a usually ablative treatment modality (3, 6).

Several risk factors predicting relapse after stopping ATD have been reported, e.g. younger age, large goiters, smoking, male sex, severe biochemical disease, and higher levels of antibodies (4). Still, there is a lack of a relapse risk prediction rule which could be helpful when counseling patients about the best first treatment option (i.e. ATD, surgery, or RAI). Subjecting them to a potentially unnecessary and time consuming ATD treatment. Herein, we performed a systematic review and meta-analysis to study possible pretreatment risk factors of relapse of patients with hyperthyroidism due to Graves’ disease.
Methods

Objective

The objective was to study pretreatment risk factor of disease recurrence after stopping ATD in patients with a first episode of hyperthyroidism due to Graves’ disease.

Protocol and eligibility criteria

This review adheres to PRISMA guidelines (7). First, we generated a review protocol (see supplemental file), outlining the main hypothesis, outcomes and search strategy. We included case-control, observational, randomized-controlled and controlled trials that assigned patients with a first episode of Graves’ disease to a standard treatment with any ATD. The diagnosis of Graves’ disease had to be established by a low thyrotropin [TSH], a high free or total thyroxine [f/T4] or triiodothyronine [f/T3], and if available diffuse goiter and positive TRAb. Participants had to be adults of at least 16 years of age. There were no restrictions on type and dosage of ATD used (e.g. methimazole, carbimazole, propylthiouracil, etc.). Also, we had no restrictions regarding language of publication, publication type or date and publication status. If studies were not available in a public library, we contacted the corresponding author. We also did include trials assessing the addition of T4 or T3 to standard ATD treatment as a former systematic review has shown that this does not affect recurrence rate (6).

We excluded trials with follow-up of <12 months, as well as case reports, reviews and trials not reporting relapse or risk factors, our primary endpoints of interest.

Outcomes

The primary outcome was relapse of hyperthyroidism according to pretreatment risk factors within a follow-up of at least 12 months after stopping ATD. We did not define any secondary outcomes.

Search strategy and study selection

We searched the electronic databases PubMed/Medline, EMBASE and the Cochrane CENTRAL Library (all until end of July 2015). We also hand-searched the references of leading articles to identify further articles.

Our search strategy included following terms: Graves’ disease, relapse, recurrence, goiter, hyperthyroidism, Basedow (see supplemental file for further details). The articles found were transferred into EndNote Version
X7 (Thomson Reuters, Philadelphia, PA, USA) and then screened by title and abstract. Full articles were retrieved for further assessment if the information given indicated that the study included patients with their first episode of Graves' disease, reported pretreatment risk factors and had well defined criteria for diagnosis of Graves' disease and relapse. If there was any doubt about the relevance of the article from the information given in the title and abstract, the full article was consulted for clarification. All articles were reviewed by two independent authors (HF and TS). In case of disagreements a solution was obtained by consensus and if not possible by a third reviewer (PS) (see figure 1).

**Data collection**

For all included studies, a data extraction table was generated as suggested in the Cochrane Handbook for Systematic Reviews of Interventions (8). It was then tested on two pilot studies and refined afterwards. Data were extracted by two authors (HF and CD) and cross-checked by another researcher (TS). Two review authors independently assessed trial quality (HF and TS). In case of disagreements a solution was again obtained by consensus and if not possible by a third reviewer (PS). We suspected duplicate publications if the same authors published multiple studies on the same subject within two years. In November 2015, we tried to contact 12 trial authors (9–20) for suspected double publication of data and/or for missing data, respectively. Four authors could not be contacted as there were no recent email addresses (15, 17, 19, 20). Two authors refused to share their data with us (10, 16). Until end of February 2016, we received no answers from the other authors, although we contacted them twice.

**Data items**

From all included trials, we extracted the following data: 1) General information: first author, year of publication, country; 2) Trial characteristics: design, if eligibility criteria are reported, duration, mode of randomization, blinding (assessors, patients), intention to treat analysis; 3) Interventions: basis intervention with ATD (days and dose), pretreatment risk factors predicting outcome; 4) Patients: number of patients included, sex, age; 5) Outcomes: primary outcome, secondary outcome, effect size.

**Risk assessment of bias in individual studies**

All included studies were assessed by two reviewers according to the Scottish Intercollegiate Guidelines Network (SIGN) (21) (see supplemental file). Assessment was done by using the appropriate checklist accordingly (i.e. RCT, cohort or case-control study). The risk of bias was judged to be unacceptable, high or low. The following aspects of internal and external validity were also assessed and reported: 1) Are there any eligibility criteria reported? 2) Mode of randomization (e.g. random number permutation table); 3)
Participants blinding? 4) Assessor blinding? 5) Intention to treat analysis? 6) Outcome well defined and complete? 7) Mean follow-up in months; 8) Percentage of lost to follow-up?

**Summary measures, planned methods of analysis and risk of bias across studies**

Only risk factors reported in at least three studies were considered. Due to low number of studies, we omitted data on the following risk factors: urinary iodine, fT4I, fT3, thyroglobulin antibodies, scintigraphy, past history, and family history. As none of the authors that we contacted about original data responded, we had to impute means and standard deviations for several studies as recommended, as only median and range had been reported (22). To impute a standard error for the within-study variation, Poisson distribution was assumed. To calculate thyroid volume when only a weight was given, we assumed a density of 1g per ml (same density as water).

For dichotomous data risks ratios with 95% CI were calculated by a random-effects model (23). For continuous data we computed the standardized mean difference according to Hedges [g] (24) as effect size as measurements were on different scales. To facilitate clinical interpretation, we also calculated weighted raw mean differences [RMD]. As test of heterogeneity, we computed the variation in risk ratio [RR] across studies attributable to heterogeneity (I^2) (25). Additionally, we imputed mean values for all missing data on the relapse groups in all 31 studies using a multivariate truncated linear regression model (26). To check for internal validity, we performed univariate meta-regression and compared the results to the ones obtained by classical meta-analysis. Significant results from the univariate meta-regression were fitted into a multivariate meta-regression (27) while omitting co-linear variables. Categorical data were transformed into percentages before integration into regression models. Analyses for subgroup and sensitivity analyses were planned post-hoc due to the heterogeneity of studies. Gene data were assessed qualitatively and if sufficient data was reported we calculated odds ratios [OR].

As most data were available for the risk factor of age, we used these to check for small study effects. For each of these trials we plotted the effect by the inverse of its standard error (28). The symmetry of this “funnel plots” was assessed both visually and formally with Egger’s test, to see if the effect decreased with increasing sample size.

The statistical analysis was conducted using Stata software version 12.1 (Stata Corp., College Station, TX, USA). All significance tests were two-sided and P < 0.05 was considered to be statistically significant.
Results

Study selection and study characteristics

A total of 1859 articles were evaluated by the accordingly skilled members of our study group (see figure 1 according to suggestions of the PRISMA group (7)). Of these, 375 full-text articles were assessed for eligibility and 54 trials were included in the final data analysis (9–20, 29–70). There were 31 trials (of which 5 randomized-controlled and 18 controlled trials) with 2322 relapses in a total of 4346 participants (overall relapse rate 53.4%) for quantitative meta-analysis (see tables 1 and 2 in supplemental file). Of note, the number of patients does not add up to the grand total as data reported was both used for the genomic data table and the quantitative analysis. Trials reporting genomic data only were only evaluated qualitatively. Trials were published between 1977 and 2015. Mainly propylthiouracil and methimazole were used as ATD.

Risk of bias within studies

Trials were mostly small with diverse risk of bias (10 high, 15 medium, and 29 low risk of bias; see table 3 in supplemental file). Of the 31 trials included in the quantitative analysis, 12 were of low risk, 14 were of medium risk, and 5 were of high risk of bias.

Qualitative reporting of gene data

Gene data were available from 13 of 54 studies (24.1%) with a total of 2178 patients being genotyped. Number of subjects ranged from 30 to 451 per trial and trials had a moderate-to-low risk of bias (one high, five medium, and seven low risk of bias; see table 3 in supplemental file). Results are summarized in table 1 (9, 14, 20, 32, 35, 40–42, 51, 54, 57, 60, 70). The effect size for most markers were rather modest with OR around 1-2 in general. Many were below a size, which could be deemed clinically significant (i.e. > 2), but with a wide range from 0.1 to 16.1.

Out of 35 studies testing HLA-types, only 4 studies reported their result in detail. Mostly, negative or not significant associations were omitted, apparently with high risk of publication bias. De Bruin (14) showed, contradictorily to others, no increased relapse rate for DR3 positive patients.

There were also investigations on the composition of peripheral leucocytes. One study did not find a difference in peripheral T-, B- and NK-lymphocyte subsets, to sufficiently predict relapse, even not after combination of factors (66). Van Ouwerkerk (34) reported the amount of subsets of T-lymphocytes (helper and suppressor type) without any predictive value.
Quantitative synthesis of results

Random-effects meta-analysis did show a significant predictive value for orbitopathy RR 1.16 (1.08; 1.25), smoking RR 1.13 (1.02; 1.25), thyroid volume assessed by sonography g 0.54 (0.18; 0.90) / RMD 5.46ml (1.92; 9.00), goiter size according to WHO grades g 0.18 (0.07; 0.30), fT4 g 0.16 (0.05; 0.28) / RMD 3.98pM (0.53; 7.44), tT3 g 0.25 (0.04; 0.46) / RMD 0.59nM (0.10; 1.09), TRAb g 0.33 (0.04; 0.62) / RMD 16.93U/l (1.94; 31.93), TBII g 0.36 (0.09; 0.63) / RMD 8.27% (2.61; 13.92), TSAb g 0.26 (0.06; 0.45) / RMD 127.39% (18.01; 236.77), whereas male vs. female sex RR 0.91 (0.82; 1.00), age g -0.14 (-0.35; 0.07) / RMD -0.80y (-2.54; 0.94), and initial tT4 level g -0.04 (-0.31; 0.23) / RMD 2.11nM (-13.84; 18.06) did not show significant associations (see table 2).

As we computed standardized mean difference according to Hedges, values are comparable across different units of measurement. This gives a crude idea of effect size between different factors. A large drawback of this method is its difficult interpretation. Generally, effect sizes with g values around 0.1 are viewed as small, around 0.3 as medium, and above 0.5 as strong (24).

For better clinical comparison, we calculated raw mean differences. Thus, an increase of the units depicted above results in a 1% increase in relapse risk (e.g. thyroid volume +1% relapse risk for every 5.46ml increase).

Additionally, we depicted the two risk factors age and goiter stratified by the subgroups relapse vs. remission for age and WHO grades for goiter (see figure 2, additional figures in supplemental file).

In univariable meta-regression values for TBII, tT3, tT4, sonographic thyroid volume, and smoking status became non-significant, whereas tT4 turned significant. As this might be due to imputation, we refitted the models with the original data, but observed no difference.

To further check for internal validity, we refitted models with values for remission. These models showed similar results except for the gender of participants were there was a collinear increase of remission and relapse with both sexes. This might be explained by the low statistical power of meta-regression and the small effect of sex on relapse risk. Also, age, and tT3 showed contradictory effect sizes, which we interpreted due to small effect size and scarcity of data, respectively. These results and measurement of heterogeneity by $I^2$ fitted well with the classical meta-analysis, except the ones noted above.

In multivariate meta-regression, we first fitted a model with all significant covariates from univariate regression. As these results could be influenced by imputation, we refitted the models with the original data, but observed again no difference. Therefore, we concluded that influential, biased studies attributed to these observed effects.

As data were handled as percentages, table 2 shows a percentage increase, or decrease for every unit of risk factor, respectively. As an example, it can be depicted that a male with Graves’ orbitopathy and goiter grade II
has a 47% increased risk for relapse (cumulative risk = 14% + 7% + (2 x 13%) = 47%) compared to a female patient without orbitopathy and no goiter.

Risk of bias across studies

As there was strong evidence for heterogeneity for most risk factors, we computed a funnel plot (see figure 2 in supplemental file) for age as a risk factors showing slight asymmetry. We chose age as most data were available on this risk factor. Egger’s test for no small study effects did show a significant result, thus rejecting publication bias across studies (p = 0.006). Of course, this results may also be influenced by small study effects, poor design of the included studies and true heterogeneity considering the genetic differences in the studied populations (28).
Discussion

Summary of evidence

This is the first systematic review and meta-analysis focusing on pretreatment risk factors predicting relapse before the first treatment course with ATD in patients with hyperthyroidism due to Graves’ disease. The 54 included trials, had moderate to low trial quality and risk of bias. The pooled analysis showed a good prediction of relapse with several risk factors, but in most trials these factors by themselves had not enough power. Risk factors were similar in randomized trials compared to case-controlled and controlled trials.

Graves’ disease is an autoimmune disease caused by stimulation of antibodies directed against the thyrotropin receptor in most patients (4). The risk for disease occurrence depends by part on environmental factors such as smoking and psychological factors such as stressful life events, and by part on genetic factors as previously demonstrated in twin studies (4, 71). Whereas environmental factors for relapse could be altered, genetic factors could not be directly influenced. But it would be possible to modify the immune system’s reaction by immunosuppressive drugs (72) until the environmental co-factors subside.

Our analysis has two main clinical benefits. First, it offers clinicians an up-to-date qualitative overview of genetic factors and other risk factor associated with recurrence of Graves’ disease. Second, the quantitative analysis also allows to calculate the risk for relapse in an individual patient. Such numbers may be used when discussing therapeutic options with patients, i.e. whether medical therapy or more definite treatment options (radio-iodine, surgery) should be used. Still, the effect sizes for most risk factors assessed in our analysis were small and statistically not significant. Thus, calculation of relapse risk base on this analysis should be used with caution and only in conjunction of good clinical judgement. Also, a prospective study should validate our findings before more widespread use in clinical practice.

Strengths and limitations

The strengths of our study are the predefined study protocol adhering to PRISMA guidelines for systematic reviews, the inclusion from studies all around the globe without language restrictions and thus reflecting different ethnicities and genetic backgrounds (2), and the relevance of this topic in every day practice. As one might argue that the response to ATD depends on genetic background, our robust findings of the included trials from different countries suggest otherwise. This increases the clinical applicability and extern validity. Although, we do not provide insights into new risk factor and there have already been various systematic reviews assessing the different treatment options for Graves’ disease (1, 5, 6, 73), we provide the very first systematic review on pretreatment risk factors so far.

We are aware of several limitations in regard to this analysis. First, retrieved data was very heterogeneous and authors of eligible studies could either not be contacted or refused to share date with us. Thus, we had to rely on several imputation methods, which could have influenced our results. For our data set, we assumed that
data were missing at random. A complete-case analysis would introduce bias and lower the power of the analysis. By using multiple imputation, we assume that bias is minimal as the approach is generally considered to be conservative. Still, imputation has limitations particularly if data are not normally distributed.

Secondly, trials were mostly monocentric, heterogeneous, and a lot had a rather high risk for bias. Many of the included studies were not randomized or case-controlled trials. Thirdly, there were many different laboratory assay over a wide time span of 38 years, making comparability difficult, especially for the immunological antibody assays. In addition, only two studies had blinding of participants or assessors (placebo-controlled). No trial had an intention to treat analysis performed and adverse effects were rarely, if at all, systematically reported. Still, studies reported eligibility criteria and mean follow-up was at least one year after withdrawal of ATD, mainly 24 months. Notably, most relapses occur within the first two years. Still, studies reported eligibility criteria and mean follow-up was at least one year after withdrawal of ATD, mainly 24 months. Notably, most relapses occurred within the first two years. Heterogeneity of data did not allow for formal quantitative analysis (e.g. mean time to relapse). Hence, we qualitatively present this data without formally analyzing it (see table 2 in supplemental file). Roughly 2/3 of the relapsed occurred between months 6 to 18 after stopping ATD, being in line with current literature (1) Meta-regression is prone to over-interpretation of findings. But in our case results were similar to classical meta-analysis. Furthermore, by adjusting p-values for multiple testing we are certain that the results reflect true findings.

Overall, the markers identified by our systematic review are whether alone nor combined strong enough to predict the clinical outcome of a single patient. Ideally, fresh markers would fill this gap. In this regard, a very recent publication from the Netherlands might shed new light. The authors generated a risk score based on a multivariate analysis of 173 prospectively enrolled patients subjected to their first ATD course. They identified lower age, higher fT4, higher TBII, large goiters, PTPN22 C/T SNP and HLA types DQB1*02, DQA1*05, and DRB1*03 as independent risk factors (74). Still, these results need to externally validated and replicated in a larger cohort first. But this approach would lead to a new perspective and could be the first step to a form of individualized treatment in Graves’ disease.

Conclusions

In conclusion, this systematic review and meta-analysis found several pretreatment risk factors of relapse in patients with standard treatment of Graves’ disease. The occurrence of orbitopathy, smoking, larger thyroid volume, and biochemically more severe disease (fT4, tT3, TRAb, TBII, and TSAb levels) did show a higher risk of relapse.
Yet, the small predictive power of a single risk factor is insufficient to predict the outcome of a single patient. This calls for novel markers to be identified and tested in future trials.
Disclosure and funding statements and acknowledgments

TS, BM and PS designed the study, HF and CD analyzed data. HF and TS wrote the first draft of the manuscript and had primary responsibility for final content. All authors read and approved the final manuscript. All authors declare no conflicts of interest. The authors want to express their gratitude to Mr. Alexander Litke, MD, and Mrs. Merih Guglielmetti for their translation of Russian, or Turkish articles, respectively. This study was supported in part by the Swiss National Science Foundation (SNSF Professorship, PP00P3_150531 / 1) and the Research Council of the Kantonsspital Aarau (1410.000.044). Funders had no role in the design, analysis or writing of this article. All authors confirm that they do not have a conflict of interest associated with this manuscript.


14 de Bruin, T. W., Bolk JH, Bussemaker JK, Stijnen T, Schreuder GM, de Vries, R. R. & van der Heide D. 

15 Kimball LE, Kulinskaya E, Brown B, Johnston C & Farid NR. Does smoking increase relapse rates in 
(doi:10.1007/bf03343979)

16 Laurberg P, Buchholtz Hansen PE & Iversen E. Goitre size and outcome of medical treatment of Graves' 

and antithyroidal microsomal antibodies in patients with hyperthyroidism due to Graves' disease treated 
with antithyroidal drugs: J Clin Endocrinol Metab. The Journal of clinical endocrinology and 

18 Yamada T, Aizawa T, Koizumi Y, Komiya I, Ichikawa K & Hashizume K. Age-related therapeutic 
response to antithyroid drug in patients with hyperthyroid Graves’ disease: J Am Geriatr Soc. Journal of 
the American Geriatrics Society 1994 42 513–516.

19 Khanna CM, Shankar LR, Jaggi CB, Bansal JK & Chugh P. Predictor of outcome of hyperthyroidism 
due to Graves disease: serum triiodothyronine/thyroxine ratio: J Assoc Physicians India. The Journal of 

20 Wilson R, McKillop JH & Henderson N. The ability of the serum thyrotrophin receptor antibody 
(TRAβ) index and HLA status to predict long-term remission of thyrotoxicosis following medical 

16 October 2015).


(doi:10.1016/j.cct.2015.09.002)

24 Borenstein M, Hedges LV, Higgins JPT & Rothstein HR. Introduction to meta-analysis. Chichester: 


26 Raghunathan TE, Lepkowski JM, van Hoewyk J & Solenberger P. A multivariate technique for multiply 
imputing missing values using a sequence of regression models. Survey methodology 2001 27 85–96.

27 Thompson SG & Higgins JPT. How should meta-regression analyses be undertaken and interpreted? 

28 Egger M, Davey Smith G, Schneider M & Minder C. Bias in meta-analysis detected by a simple, 


42 Kotulla P. Do HLA-DR-typing and measurement of TSH-receptor antibodies help in the prediction of
43 the clinical course of Graves' thyrotoxicosis after antithyroid drug treatment?: Acta Endocrinol Suppl
45
46 Schifferdecker E, Kuhnl P, Schoffling K, Manfras B, Holzberger G, Spielmann W & Bohm BO.
49
50 Scanziga BR. [Drug therapy of Basedow's disease. Long term functional results (302 cases)]: Schweiz
52
53 Romaldini JH, Bromberg N, Werner RS, Tanaka LM, Rodrigues HF, Werner MC, Farah CS & Reis LC.
54 Comparison of effects of high and low dosage regimens of antithyroid drugs in the management of
56 (doi:10.1210/jcem-57-3-563)
57
58 Prakash R. Prediction of remission in Graves' disease treated with long-term carbimazole therapy:
59 evaluation of technetium-99m thyroid uptake and TSH concentrations as prognostic indicators: Eur J
61
62 Park SM, Cho YY, Joung JY, Sohn SY, Kim SW & Chung JH. Excessive iodine intake does not increase
63 the recurrence rate of Graves' disease after withdrawal of the antithyroid drug in an iodine-replete area.
64 European thyroid journal 2015 4 36–42. (doi:10.1159/000375261)
65
66 Orhan Y, Azezli A., Aral F., Tascioglu C., Molvalilar S. & Sencer E. The frequency of remission and
67 relapse in Graves' disease was investigated. ISTANB. TIP FAK. MECM. 1992 55 525–530.
68
69 Murakami M, Koizumi Y, Aizawa T, Yamada T, Takahashi Y, Watanabe T & Kamoi K. Studies of
70 thyroid function and immune parameters in patients with hyperthyroid Graves' disease in remission: J
72 (doi:10.1210/jcem-66-1-103)
73
74 Lucas A, Salinas I, Rius F, Pizarro E, Granada ML, Foz M & Sanmarti A. Medical therapy of Graves'
75 disease: does thyroxine prevent recurrence of hyperthyroidism?: J Clin Endocrinol Metab. The Journal
76 of clinical endocrinology and metabolism 1997 82 2410–2413. (doi:10.1210/jcem.82.8.4118)
77
78 Komiyi I, Yamada T, Sato A, Kouki T, Nishimori T & Takasu N. Remission and recurrence of
79 hyperthyroid Graves' disease during and after methimazole treatment when assessed by IgE and
80 interleukin 13: J Clin Endocrinol Metab. The Journal of clinical endocrinology and metabolism 2001 86
81 3540–3544. (doi:10.1210/jcem.86.8.7734)
82
83 Kim KW, Park YJ, Kim TY, Park DJ, Park KS & Cho BY. Susceptible alleles of the CD40 and CTLA-4
84 genes are not associated with the relapse after antithyroid withdrawal in Graves' disease: Thyroid.
85 Thyroid official journal of the American Thyroid Association 2007 17 1229–1234.
86
87 Kim TY, Park YJ, Park DJ, Chung H-K, Kim WB, Kohn LD & Cho BY. Epitope heterogeneity of
88 thyroid-stimulating antibodies predicts long-term outcome in Graves' patients treated with antithyroid


Guilhem I, Massart C, Poirier JY & Maugendre D. Differential evolution of thyroid peroxidase and thyrotropin receptor antibodies in Graves' disease: thyroid peroxidase antibody activity reverts to pretreatment level after carbimazole withdrawal: Thyroid. Thyroid official journal of the American Thyroid Association 2006 16 1041–1045. (doi:10.1089/thy.2006.16.1041)


### Table 1 Associated genes for relapse

<table>
<thead>
<tr>
<th>Gene locus</th>
<th>SNP polymorphism of gene and location</th>
<th>Receptor or transcript</th>
<th>Allele</th>
<th>Effect size (95% confidence interval) for risk of relapse</th>
<th>Reported by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ch 6p21.31</td>
<td>N/A</td>
<td>HLA-DR3</td>
<td>N/A</td>
<td>RR 0.29 (p 0.016) / OR 0.1 (0.0; 0.7)</td>
<td>de Bruin, T.W.,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLA-Cw7</td>
<td></td>
<td>RR 2.05 (p 0.009) / OR 5.0 (1.4; 17.7)</td>
<td>1988, The</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLA-A1</td>
<td>N/A</td>
<td>OR 0.5 (0.2; 1.3)</td>
<td>Netherlands</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLA-B8</td>
<td></td>
<td>OR 1.1 (0.4; 3.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLA-DR3</td>
<td></td>
<td>OR 1.8 (0.6; 4.9)</td>
<td></td>
</tr>
<tr>
<td>Ch 6p21.31</td>
<td>N/A</td>
<td>HLA-A1</td>
<td>N/A</td>
<td>OR 2.2 (0.8; 6.1)</td>
<td>Allanic, H.,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLA-B8</td>
<td></td>
<td>OR 3.3 (1.2; 9.6)</td>
<td>1983, France</td>
</tr>
<tr>
<td>Ch 6p21.31</td>
<td>N/A</td>
<td>HLA-B12</td>
<td>N/A</td>
<td>RR 0.23 (p 0.02) / OR 0.2 (0.0; 1.1)</td>
<td>Irvine, W. J.,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLA-DR5</td>
<td></td>
<td>RR 9.85 (p 0.03) / OR 9.9 (1.2; 438.0)</td>
<td>1977, UK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLA-DR7</td>
<td></td>
<td>RR 0.27 (p 0.05) / OR 0.3 (0.1; 1.3)</td>
<td></td>
</tr>
<tr>
<td>Ch 6p21.31</td>
<td>N/A</td>
<td>HLA-DR3</td>
<td>N/A</td>
<td>OR 1.2 (0.7; 2.4)</td>
<td>Schlesusener H.,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1987, Germany</td>
</tr>
<tr>
<td>Ch 6p21.31</td>
<td>N/A</td>
<td>HLA-DR3</td>
<td>N/A</td>
<td>OR 0.8 (0.5; 1.4)</td>
<td>Schlesusener H.,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1989, Germany</td>
</tr>
<tr>
<td>Ch 6p21.31</td>
<td>N/A</td>
<td>HLA-DR3</td>
<td>N/A</td>
<td>OR 2.2 (0.4; 12.3)</td>
<td>Wilson R., 1986,</td>
</tr>
<tr>
<td>Ch 20q12-q13.2</td>
<td>5' UTR</td>
<td>CD40</td>
<td>CC</td>
<td>OR 1.2 (0.5; 2.9)</td>
<td>Kim, K.W.,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CT</td>
<td>OR 0.6 (0.3; 1.4)</td>
<td>2007, South</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT</td>
<td>OR 3.7 (0.5; 166.2)</td>
<td>Korea</td>
</tr>
<tr>
<td>Chromosome</td>
<td>Gene</td>
<td>Position</td>
<td>SNP</td>
<td>Allele Comparison</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td>-----</td>
<td>------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>20q12-q13.2</td>
<td>Intron 1 (rs745307)</td>
<td>CD 40</td>
<td>C/T vs. C/C</td>
<td>OR 8.0 (1.0; 62.7)</td>
<td>Wang, P. W., 2013, Taiwan</td>
</tr>
<tr>
<td></td>
<td>Intron 1 (rs11569309)</td>
<td></td>
<td>C/T vs. T/T</td>
<td>OR 8.1 (1.0; 63.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intron 8 (rs3765457)</td>
<td></td>
<td>A/G vs. A/A</td>
<td>OR 2.6 (1.1; 6.3)</td>
<td></td>
</tr>
<tr>
<td>2p33</td>
<td>Position +49A/G in exon 1 (rs231775)</td>
<td>CTLA</td>
<td>G/G vs. A/G &amp; A/A</td>
<td>OR 2.0 (1.2; 3.3)</td>
<td>Wang, P. W., 2013, Taiwan</td>
</tr>
<tr>
<td>2p33</td>
<td>CT60 polymorphism in 3'UTR</td>
<td>CTLA-4</td>
<td>AA</td>
<td>OR 0.7 (0.0; 42.2)</td>
<td>Kim, K.W., 2007, South Korea</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AG</td>
<td>OR 1.1 (0.4; 3.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td>OR 0.9 (0.3; 2.5)</td>
<td></td>
</tr>
<tr>
<td>2p33</td>
<td>Promoter at position -318</td>
<td>CTLA-4</td>
<td>CC</td>
<td>OR 1.1 (0.4; 2.7)</td>
<td>Kim, K.W., 2007, South Korea</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CT</td>
<td>OR 0.9 (0.4; 2.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT</td>
<td>OR 0.4 (0.0; 31.9)</td>
<td></td>
</tr>
<tr>
<td>2p33</td>
<td>Position +49 in exon 1</td>
<td>CTLA-4</td>
<td>AA</td>
<td>OR 1.9 (0.2; 96.2)</td>
<td>Kim, K.W., 2007, South Korea</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AG</td>
<td>OR 1.5 (0.4; 5.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td>OR 0.5 (0.1; 1.6)</td>
<td></td>
</tr>
<tr>
<td>2p33</td>
<td>Position +49 in exon 1</td>
<td>CTLA-4</td>
<td>AA</td>
<td>OR 1.2 (0.2; 5.2)</td>
<td>Wang, P. W., 2004, Taiwan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AG</td>
<td>OR 0.3 (0.1; 0.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td>OR 4.0 (1.9; 8.6)</td>
<td></td>
</tr>
<tr>
<td>2p33</td>
<td>Position 49 in exon 1</td>
<td>CTLA-4</td>
<td>AA</td>
<td>OR 0.3 (0.1; 0.8)</td>
<td>Tanrikulu, S., 2006, Turkey</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AG</td>
<td>OR 6.2 (2.4; 16.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td>OR 3.6 (0.8; 5.7)</td>
<td></td>
</tr>
<tr>
<td>8q24</td>
<td>exon 33</td>
<td>Thyroglobulin</td>
<td>CC</td>
<td>OR 1.5 (0.6; 3.6)</td>
<td>Tanrikulu, S., 2006, Turkey</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CT</td>
<td>OR 1.4 (0.6; 3.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT</td>
<td>OR 2.1 (0.7; 6.5)</td>
<td></td>
</tr>
<tr>
<td>Chromosome Region</td>
<td>SNP/Gene Information</td>
<td>Genotype</td>
<td>Odds Ratio (95% CI)</td>
<td>Author(s)</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------</td>
<td>----------</td>
<td>--------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Ch 20q13.2-q13.3</td>
<td>T393C polymorphism</td>
<td>G alpha subunit of G protein (GNAS1)</td>
<td>CC</td>
<td>OR 1.3 (0.7; 2.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CT</td>
<td>OR 0.5 (0.3; 0.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT</td>
<td>OR 2.4 (1.0; 6.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Glowacka, D., 2009, Germany</td>
<td></td>
</tr>
<tr>
<td>Ch 9p24.1</td>
<td>Position 8923 in PD-L1 intron 4</td>
<td>Programmed cell death-1 ligand (PD-L1)</td>
<td>AA</td>
<td>OR 0.4 (0.1; 1.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AC</td>
<td>OR 1.5 (0.3; 5.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>OR 16.1 (0.8; 954.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hayashi, M., 2008, Japan</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2 Summary of results

#### Pretreatment risk factors for relapse

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Classical meta-analysis</th>
<th>Univariable meta-regression</th>
<th>Multivariable meta-regression</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RR or SMD/ Hedges’ g (95% CI)</td>
<td>Weighted raw mean difference (95% CI)</td>
<td>Heterogeneity ($I^2$)</td>
<td>% Risk (95% CI)</td>
</tr>
<tr>
<td>Male vs. female sex</td>
<td>RR 0.91 (0.82; 1.00)</td>
<td>N/A</td>
<td>40.3%</td>
<td>178.75 (100.80; 256.71)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>g -0.14 (-0.35; 0.07)</td>
<td>-0.80 (-2.54; 0.94)</td>
<td>85.3%</td>
<td>1.62 (0.04; 3.20)</td>
</tr>
<tr>
<td>Graves’ orbitopathy (y/n)</td>
<td>RR 1.16 (1.08; 1.25)*</td>
<td>N/A</td>
<td>0.0%</td>
<td>89.31 (56.67; 121.95)</td>
</tr>
<tr>
<td>Goiter (WHO grades)</td>
<td>g 0.18 (0.07; 0.30)*</td>
<td>N/A</td>
<td>0.0%</td>
<td>-64.85 (-90.33; -39.37)</td>
</tr>
<tr>
<td>Smoking (y/n)</td>
<td>RR 1.13 (1.02; 1.25)*</td>
<td>N/A</td>
<td>7.8%</td>
<td>51.19 (-176.54; 278.92)</td>
</tr>
<tr>
<td>Thyroid vol. (ml)</td>
<td>g 0.54 (0.18; 0.90)*</td>
<td>5.46 (1.92; 9.00)</td>
<td>75.6%</td>
<td>0.74 (0.03; 1.46)</td>
</tr>
<tr>
<td>fT4 (pmol/l)</td>
<td>g 0.16 (0.05; 0.28)*</td>
<td>3.98 (0.53; 7.44)</td>
<td>2.0%</td>
<td>0.69 (0.08; 1.31)</td>
</tr>
<tr>
<td>tT4 (nmol/l)</td>
<td>g -0.04 (-0.31; 0.23)</td>
<td>2.11 (-13.84; 18.06)</td>
<td>83.5%</td>
<td>-0.47 (-0.65; -0.28)</td>
</tr>
<tr>
<td>tT3 (nmol/l)</td>
<td>g 0.25 (0.04; 0.46)*</td>
<td>0.59 (0.10; 1.09)</td>
<td>66.7%</td>
<td>-1.87 (-5.73; 1.99)</td>
</tr>
<tr>
<td>TRAb (U/l)</td>
<td>g 0.33 (0.04; 0.62)*</td>
<td>16.93 (1.94; 31.93)</td>
<td>75.2%</td>
<td>0.41 (0.29; 0.53)</td>
</tr>
<tr>
<td>TBII (% activity)</td>
<td>g 0.36 (0.09; 0.63)*</td>
<td>8.27 (2.61; 13.92)</td>
<td>75.8%</td>
<td>0.08 (-0.43; 0.59)</td>
</tr>
<tr>
<td>TSAb (% activity)</td>
<td>g 0.26 (0.06; 0.45)*</td>
<td>127.39 (18.01; 236.77)</td>
<td>46.4%</td>
<td>-0.07 (-0.08; -0.06)</td>
</tr>
</tbody>
</table>

2 *p<0.05; TRAb (thyrotropin receptor antibodies), TBII (thyrotropin binding inhibiting immunoglobulins), TSAb (thyroid stimulating antibody), f/tT3/4 (free/total triiodothyronine/thyroxine), Thyroid volume (assessed by sonography)
Records identified through database
  Embase: 1604
  Pubmed: 1176
  Cochrane: 94
  (n = 2874)
Additional records identified through other sources
  (n = 1)
Records after duplicates removed
  (n = 1859)
Records screened
  (n = 1859)
Full-text articles assessed for eligibility
  (n = 375)
Studies included in qualitative synthesis
  (n = 54)
Studies included in quantitative synthesis (meta-analysis)
  (n = 31)
Records excluded (n = 1484)
  • Case report/review: 10
  • Only ophthalmopathy: 5
  • Pregnancy/children <16y: 15
  • Not meeting inclusion criteria/different subject: 1454
Full-text articles excluded, with reasons (n = 321)
  • Follow up <12 months: 13
  • Insufficient data: 25
  • No included intervention: 181
  • No CT: 58
  • No translation: 3
  • No individual data: 16
  • No original data: 3
  • Irretrievable: 22
Figure 2A Relapse risk according to age at diagnosis

279x153mm (300 x 300 DPI)
Figure 2B Relapse risk according to WHO goiter grade

381x285mm (300 x 300 DPI)