

## Supplementary Online Content

Siegmann E-M, Müller HHO, Luecke C, Philipsen A, Kornhuber J, Grömer TW. Association of depression and anxiety disorders with autoimmune thyroiditis: a systematic review and meta-analysis. *JAMA Psychiatry*. Published online May 2, 2018. doi:10.1001/jamapsychiatry.2018.0190

**eAppendix 1.** Introduction

**eAppendix 2.** R Code

**eAppendix 3.** Discussion

**eTable 1.** Current Terminology Concerning AIT: Similarities and Differences

**eTable 2.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Checklist

**eTable 3.** Coding Protocol

**eTable 4.** The Newcastle-Ottawa Quality Assessment Scale for Case-Control Studies

**eTable 5.** Study Characteristics of all Included Studies

**eTable 6.** Slope ( $\beta$ ), 95% CI, and *P* Values for Meta-Regression Models

This supplementary material has been provided by the authors to give readers additional information about their work.

## eAppendix 1. Introduction

As mentioned in the article file, there is a lack of clarity concerning the various terminology used to describe different forms of (autoimmune) thyroiditis. A lot of studies employ the terms (*chronic autoimmune thyroiditis*, *Hashimoto's thyroiditis* and *focal lymphocytic thyroiditis*) as synonyms. Additionally, the terms *overt*, *subclinical* and *latent hypothyroidism* are used in association with AIT. eTable 1 shows the different definitions and criteria used to distinguish between all these forms of hypothyroidism. Despite the diverse terminology, almost every form of hypothyroidism can be traced back to AIT: Hashimoto's thyroiditis is a variety of AIT (besides atrophic thyroiditis) and explains for most of its cases<sup>1</sup>. Up to 80% of subclinical and overt hypothyroidism are induced by autoimmune disorder of the thyroid gland<sup>1</sup> which is the most common cause of decreased thyroid hormone production in patients with acquired hypothyroidism<sup>2</sup>.

The pathophysiology of depression is still a matter of ongoing research. As monoaminergic drugs are effective in the treatment of depression, it has been postulated that a lack of monoamines is the underlying cause of depressive disorders<sup>3</sup>. Other hypotheses include a dysfunction in the hypothalamic–pituitary–adrenal regulation, causing stress and depressive symptoms<sup>4</sup> as well as anxiety disorders, for which monoaminergic drugs are also effective<sup>3</sup>. In addition, inflammatory processes have been described to be related with symptoms of major depressive disorder, especially by increasing levels of circulating pro-inflammatory cytokines<sup>5</sup>. AIT is an autoimmune disease in which affected persons develop antibodies against targets in the thyroid gland which has been mainly linked to genetic factors<sup>6</sup>. Importantly, thyroid metabolism has been shown to modulate the brain serotonin system by reducing the sensitivity of 5-HT<sub>1A</sub> autoreceptors in the Raphé area, and by increasing 5-HT<sub>2</sub> receptor sensitivity<sup>7</sup>. Concerning anxiety, it is likely that the thyroid-catecholamine system interaction (via noradrenalin and T<sub>3</sub>) and its influence on sympathetic nervous system activity<sup>8</sup> affect symptoms such as high blood pressure<sup>2</sup> or sensible nightly tachycardia<sup>9,10</sup>.

## eAppendix 2. R Code

```
# Meta-Analysis: Autoimmune thyroiditis and depression/anxiety disorders
```

```
# dat = import ()
```

```
install.packages("metafor")  
library(metafor)  
library(Matrix)
```

```
# Conduct meta-analysis for anxiety only
```

```
# Anxiety dataset
```

```
dat.anx <- subset(dat, construct == 1)
```

```
dat.anx <- escalc(measure="OR", ai=tpos, bi=tneg, ci=cpos, di=cneg,  
  data=dat.anx)
```

```
# Calculate missing values
```

```
# Odds ratios
```

```
dat.anx$yi <- replmiss(dat.anx$yi, log(dat.anx$OR))
```

```
# Cohen's d
```

```
dat.anx$yi <- replmiss(dat.anx$yi, dat.anx$d*(pi/sqrt(3)))
```

```
# Confidence intervals
```

```
dat.anx$sei <- replmiss(dat.anx$sei, with(dat.anx, (log(ciup) -  
  log(cilow))/(2*1.96)))
```

```
dat.anx$vi <- replmiss(dat.anx$vi, dat.anx$sei^2)
```

```
# Mean and standard deviation
```

```
dat.anx <- escalc(measure="D2ORN", m1i=m1_anx, sd1=sd1_anx, n1i=n2,  
  m2i=m2_anx, sd2i=sd2_anx, n2i=n3, data=dat.anx,  
  replace=F)
```

```
# Aggregate dependent effect sizes
```

```
library(MAd)
```

```
dat.anx$sexpr <- dat.anx$sexn/dat.anx$n1
```

```
dat.anx.help <- aggregate(dat.anx, by=list(dat.anx$sno), FUN=mean)
```

```
dat.anx.agg <- agg(id=sno, es=yi, var=vi, n.1=n2, n.2=n3, cor=0.5,  
  data=dat.anx)
```

```
dat.anx.agg$study <- c(dat.anx$study[1], dat.anx$study[2],  
  dat.anx$study[3], dat.anx$study[4], dat.anx$study[6],  
  dat.anx$study[7], dat.anx$study[8], dat.anx$study[9],  
  dat.anx$study[10], dat.anx$study[13], dat.anx$study[14],  
  dat.anx$study[15], dat.anx$study[16])
```

```
dat.anx.agg$age <- dat.anx.help$age
```

```
dat.anx.agg$sexpr <- dat.anx.help$sexpr
```

```

dat.anx.agg$qual1 <- dat.anx.help$qual1
dat.anx.agg$qual2 <- dat.anx.help$qual2
dat.anx.agg$qual3 <- dat.anx.help$qual3
dat.anx.agg$measure <- dat.anx.help$measure
dat.anx.agg$ak <- dat.anx.help$ak

ma.anx <- rma.uni(es, var, data=dat.anx.agg)
summary(ma.anx)

## Random-Effects Model (k = 13; tau^2 estimator: REML)
##
##   logLik  deviance      AIC      BIC      AICc
## -16.3834  32.7669  36.7669  37.7367  38.1002
##
## tau^2 (estimated amount of total heterogeneity): 0.7067 (SE = 0.3521)
## tau (square root of estimated tau^2 value):      0.8406
## I^2 (total heterogeneity / total variability):   89.75%
## H^2 (total variability / sampling variability):   9.76
##
## Test for Heterogeneity:
## Q(df = 12) = 104.4089, p-val < .0001
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
## 0.8411  0.2586  3.2518  0.0011  0.3341  1.3480 **
##
##---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

predict(ma.anx, transf=exp, digits=2)

## pred ci.lb ci.ub cr.lb cr.ub
## 2.32 1.40 3.85 0.41 13.00

# Conduct meta-analysis for depression only
# Depression dataset

dat.depr <- subset(dat, construct == 2)

dat.depr <- escalc(measure="OR", ai=tpos, bi=tneg, ci=cpos, di=cneg,
  data=dat.depr)

# Calculate missing values
# Odds ratios
dat.depr$yi <- replmiss(dat.depr$yi, log(dat.depr$OR))

```

```

# Cohen's d
dat.depr$yi <- replmiss(dat.depr$yi, dat.depr$d*(pi/sqrt(3)))
# p-values
dat.depr$zi <- sign(dat.depr$yi) * qnorm(dat.depr$pval/2, lower.tail=FALSE)
dat.depr$sei <- dat.depr$yi / dat.depr$zi
# Confidence intervals
dat.depr$sei <- replmiss(dat.depr$sei, with(dat.depr, (log(ciup) -
  log(cilow))/(2*1.96)))
dat.depr$vi <- replmiss(dat.depr$vi, dat.depr$sei^2)
# Mean and standard deviation
dat.depr <- escalc(measure="D2ORN", m1i=m1_depr, sd1=sd1_depr, n1i=n2,
  m2i=m2_depr, sd2i=sd2_depr, n2i=n3, data=dat.depr,
  replace=FALSE)

# Aggregate dependent effect sizes

dat.depr$sexpr <- dat.depr$sexn/dat.depr$n1
dat.depr.agg <- agg(id=sno, es=yi, var=vi, n.1=n2, n.2=n3,
  cor=0.5, data=dat.depr)
dat.depr.help <- aggregate(dat.depr, by=list(dat.depr$sno), FUN=mean)
dat.depr.agg$ak <- dat.depr.help$ak
dat.depr.agg$age <- dat.depr.help$age
dat.depr.agg$sexpr <- dat.depr.help$sexpr
dat.depr.agg$qual1 <- dat.depr.help$qual1
dat.depr.agg$qual2 <- dat.depr.help$qual2
dat.depr.agg$qual3 <- dat.depr.help$qual3
dat.depr.agg$measure <- dat.depr.help$measure
dat.depr.agg$study <- c(dat.depr$study[1], dat.depr$study[2],
  dat.depr$study[4], dat.depr$study[5], dat.depr$study[6],
  dat.depr$study[7], dat.depr$study[9], dat.depr$study[12],
  dat.depr$study[13], dat.depr$study[14], dat.depr$study[15],
  dat.depr$study[16], dat.depr$study[17], dat.depr$study[18],
  dat.depr$study[19], dat.depr$study[22], dat.depr$study[23],
  dat.depr$study[24], dat.depr$study[25], dat.depr$study[26])

ma.depr <- rma.uni(es, var, data=dat.depr.agg)
summary(ma.depr)

## Random-Effects Model (k = 20; tau^2 estimator: REML)
##
##   logLik  deviance      AIC      BIC     AICc
## -30.1841  60.3683  64.3683  66.2572  65.1183
##

```

```

## tau^2 (estimated amount of total heterogeneity): 1.1567 (SE = 0.4393)
## tau (square root of estimated tau^2 value):      1.0755
## I^2 (total heterogeneity / total variability):   92.06%
## H^2 (total variability / sampling variability):  12.59
##
## Test for Heterogeneity:
## Q(df = 19) = 205.7831, p-val < .0001
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
##  1.2711  0.2608  4.8727  <.0001  0.7598  1.7823  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

predict(ma.depr, transf=exp, digits=2)

```

```

## pred ci.lb ci.ub cr.lb cr.ub
##  3.56  2.14  5.94  0.41 31.19

```

```

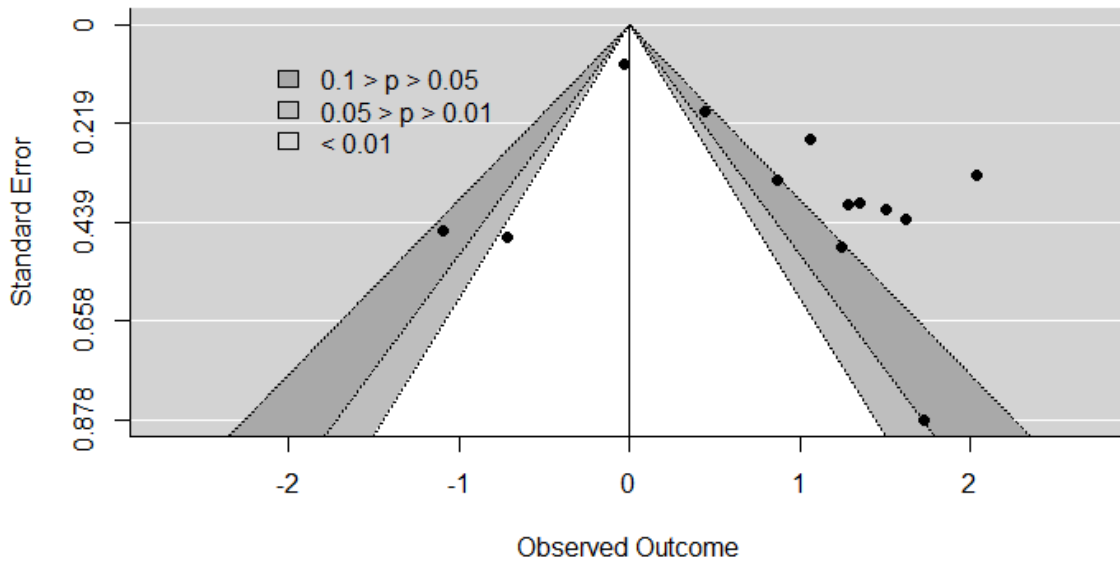
# Contour-enhanced funnel plot

```

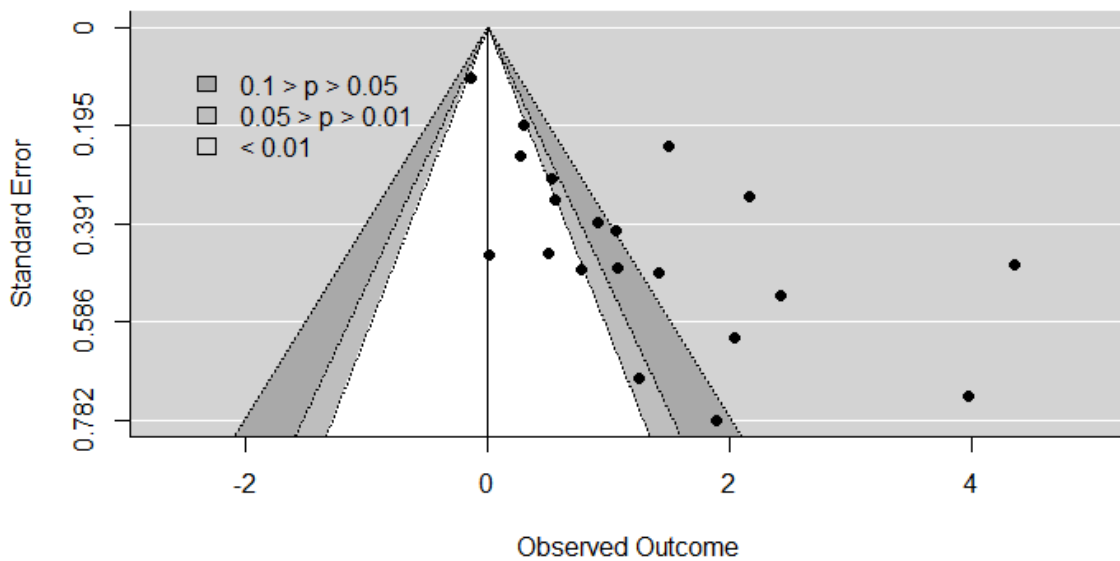
```

funnel(ma.anx, level=c(90, 95, 99), shade=c("white", "gray", "darkgray"),
       refline=0)
par(font=1)
legend(-2.2, 0.05, c("0.1 > p > 0.05", "0.05 > p > 0.01", "< 0.01"),
       fill=c("darkgray", "gray", "lightgray"), bty="n")

```



```
funnel(ma.depr, level=c(90, 95, 99), shade=c("white", "gray", "darkgray"),
       refline=0)
legend(-2.6, 0.05, c("0.1 > p > 0.05", "0.05 > p > 0.01", "< 0.01"),
       fill=c("darkgray", "gray", "lightgray"), bty="n")
```



```
# Egger's regression test
```

```
regtest.rma(ma.anx)
```

```
## Regression Test for Funnel Plot Asymmetry
```

```
##
```

```
## model:      mixed-effects meta-regression model
```

```
## predictor: standard error
```

```
##
```

```

## test for funnel plot asymmetry: z = 0.7292, p = 0.4659

regtest.rma(ma.depr)

## Regression Test for Funnel Plot Asymmetry
##
## model:      mixed-effects meta-regression model
## predictor: standard error
##
## test for funnel plot asymmetry: z = 2.8766, p = 0.0040

# Forest plots

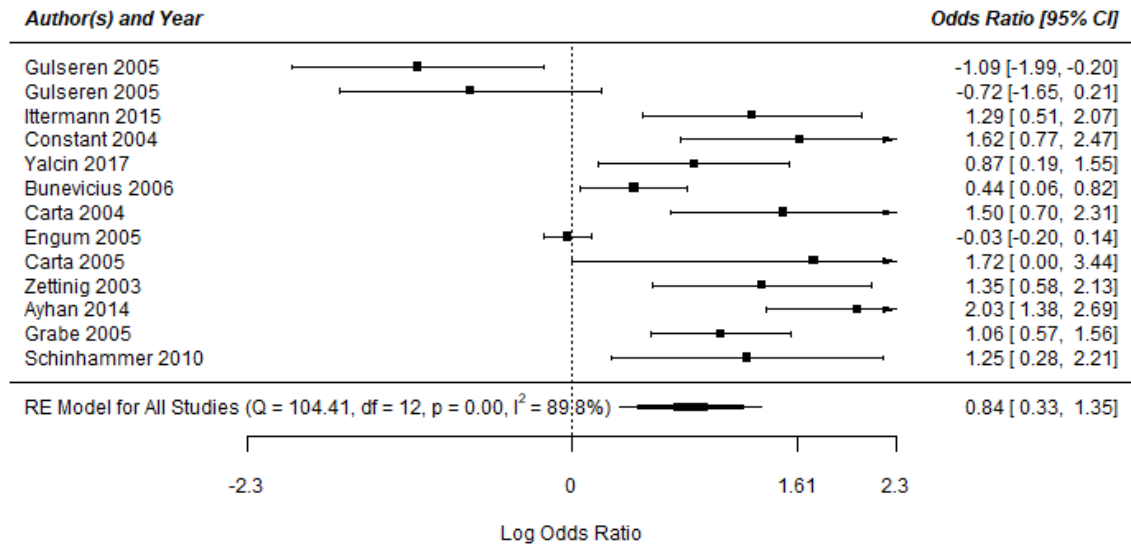
par(font=1)
forest(ma.anx,
      xlim=c(-4, 4),
      at=log(c(0.1, 1, 5, 10)),
      ilab.xpos=c(-9.5,-8,-6,-4.5),
      cex=0.75,
      slab = dat.anx.agg$study,
      xlab="Log Odds Ratio", mlab="",
      psize=1)

text(-4, -1, pos=4, cex=0.75, bquote(paste("RE Model for All Studies (Q =
      ",.(formatC(ma.anx$QE, digits=2, format="f")), ", df = ",.(ma.anx$k
      - ma.anx$p),"", p = ",.(formatC(ma.anx$QEp, digits=2, format="f")),
      ", ", I^2, " = ",.(formatC(ma.anx$I2, digits=1, format="f")), "%"))))

par(cex=0.75, font=4)
text(-4,15, "Author(s) and Year", pos=4)
text(4, 15, "Odds Ratio [95% CI]", pos=2)

```





```

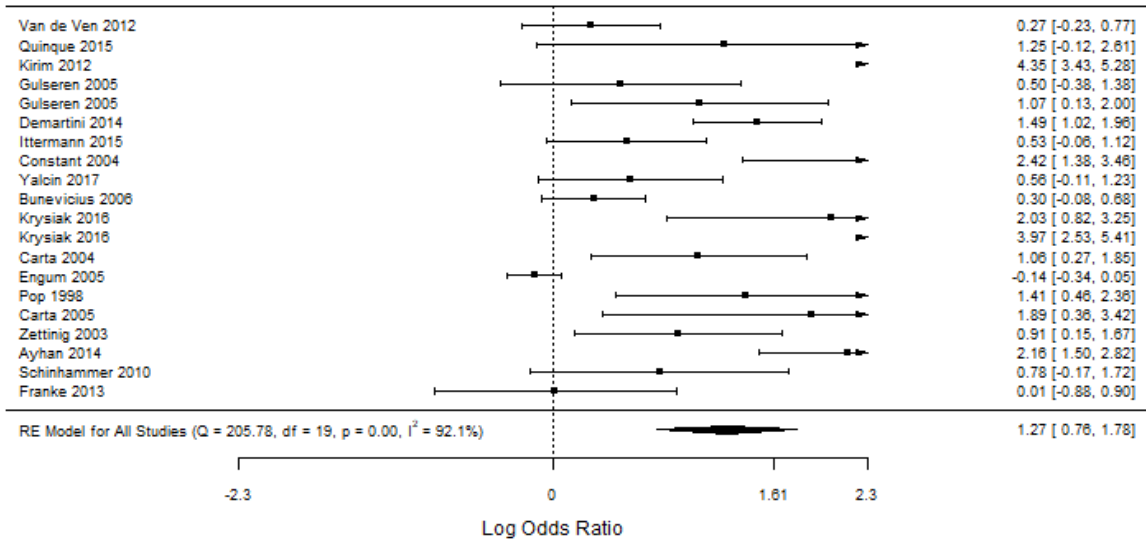
par(font=1)
forest(ma.depr,
      xlim=c(-4, 4),
      at=log(c(0.1, 1, 5, 10)),
      ilab.xpos=c(-9.5,-8,-6,-4.5),
      cex=0.75,
      slab = dat.depr.agg$study,
      xlab="Log Odds Ratio", mlab="",
      psize=1)
text(-4, -1, pos=4, cex=0.75, bquote(paste("RE Model for All Studies (Q =
",.(formatC(ma.depr$QE, digits=2, format="f")), ", df = ",(ma.depr$k
- ma.depr$p),", p = ", .(formatC(ma.depr$QEp, digits=2, format="f")),
", ", I^2, " = ",
.(formatC(ma.depr$I2, digits=1, format="f")), "%))))))

par(cex=0.75, font=4)
text(-4, 23, "Author(s) and Year", pos=4)
text(4.3, 23, "Odds Ratio [95% CI]", pos=2)

```

Author(s) and Year

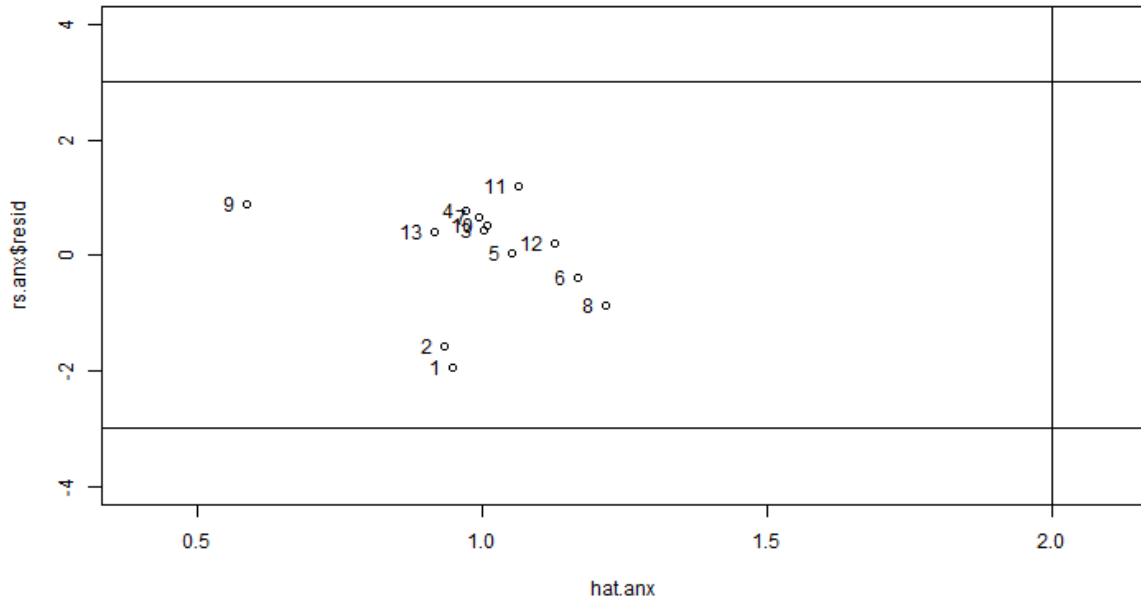
Odds Ratio [95% CI]



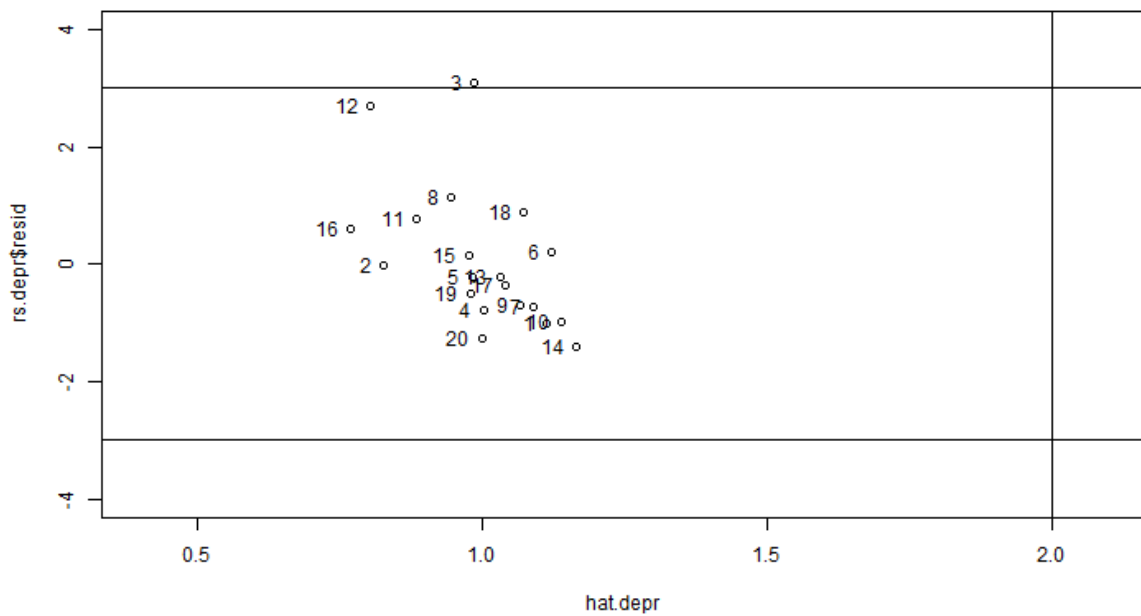
## # Sensitivity analysis

# RULE: If `rstandard > 3` AND `hatvalue > 2` times average of `hatvalues`,  
# run analysis with those cases deleted to test for sensitivity.

```
rs.anx = rstandard(ma.anx)
hat.anx = hatvalues(ma.anx)/mean(hatvalues(ma.anx))
plot(hat.anx, rs.anx$resid, ylim = c(-4,4), xlim=c(0.4,2.1))
par(font=1)
text(hat.anx, rs.anx$resid, cex= 1, pos = 2)
abline(h = -3)
abline(h = 3)
abline(v = 2)
```



```
rs.depr = rstandard(ma.depr)
hat.depr = hatvalues(ma.depr)/mean(hatvalues(ma.depr))
plot(hat.depr, rs.depr$resid, ylim = c(-4,4), xlim=c(0.4, 2.1))
text(hat.depr, rs.depr$resid, cex= 1, pos = 2)
abline(h = -3)
abline(h = 3)
abline( v = 2
```



# Meta-regression

```

dat.depr.agg$qual.all <-
  dat.depr.agg$qual1+dat.depr.agg$qual2+dat.depr.agg$qual3

ma.depr.mod1 = rma.uni(es,var,mods = ~qual.all, data = dat.depr.agg)

summary(ma.depr.mod1)

## Mixed-Effects Model (k = 20; tau^2 estimator: REML)
##
##   logLik  deviance      AIC      BIC      AICc
## -29.1008  58.2016  64.2016  66.8727  65.9159
##
## tau^2 (estimated amount of residual heterogeneity):      1.2350 (SE =
## 0.4787)
## tau (square root of estimated tau^2 value):              1.1113
## I^2 (residual heterogeneity / unaccounted variability):  92.29%
## H^2 (unaccounted variability / sampling variability):     12.97
## R^2 (amount of heterogeneity accounted for):              0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 18) = 192.2983, p-val < .0001
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 0.0151, p-val = 0.9021
##
## Model Results:
##
##           estimate      se      zval      pval      ci.lb      ci.ub
## intrcpt      1.4403  1.3739   1.0483  0.2945  -1.2525  4.1331
## qual.all    -0.0294  0.2394  -0.1230  0.9021  -0.4987  0.4398
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

dat.anx.agg$qual.all <-
  dat.anx.agg$qual1+dat.anx.agg$qual2+dat.anx.agg$qual3

ma.anx.mod1 = rma.uni(es,var,mods = ~qual.all, data = dat.anx.agg)

summary(ma.anx.mod1)

```

```

## Mixed-Effects Model (k = 13; tau^2 estimator: REML)
##
##   logLik  deviance      AIC      BIC      AICc
## -15.0540  30.1080  36.1080  37.3017  39.5366

```

```

##
## tau^2 (estimated amount of residual heterogeneity):      0.7279 (SE =
## 0.3765)
## tau (square root of estimated tau^2 value):             0.8532
## I^2 (residual heterogeneity / unaccounted variability): 89.87%
## H^2 (unaccounted variability / sampling variability):   9.87
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 11) = 104.4046, p-val < .0001
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 0.9038, p-val = 0.3418
##
## Model Results:
##
##           estimate      se      zval      pval      ci.lb      ci.ub
## intrcpt    -0.4555  1.3893  -0.3279  0.7430  -3.1785  2.2674
## qual.all     0.2262  0.2380   0.9507  0.3418  -0.2402  0.6927
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
ma.anx.mod2 = rma.uni(es,var,mods = ~sexpr, data = dat.anx.agg)
```

```

## Warning message:
## In rma.uni(es, var, mods = ~sexpr, data = dat.anx.agg) :
##   Studies with NAs omitted from model fitting.

```

```
summary(ma.anx.mod2)
```

```

## Mixed-Effects Model (k = 9; tau^2 estimator: REML)
##
##   logLik deviance      AIC      BIC      AICc
## -10.2622  20.5244  26.5244  26.3622  34.5244
##
## tau^2 (estimated amount of residual heterogeneity):      0.9248 (SE =
## 0.5678)
## tau (square root of estimated tau^2 value):             0.9617
## I^2 (residual heterogeneity / unaccounted variability): 89.11%
## H^2 (unaccounted variability / sampling variability):   9.19
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 7) = 48.7141, p-val < .0001

```

```

##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 0.9858, p-val = 0.3208
##
## Model Results:
##
##      estimate      se      zval      pval      ci.lb      ci.ub
## intrcpt      2.6605  1.8988   1.4012  0.1612  -1.0610  6.3821
## sexpr       -2.2085  2.2243  -0.9929  0.3208  -6.5681  2.1512
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

ma.depr.mod2 = rma.uni(es,var,mods = ~sexpr, data = dat.depr.agg)

```

```

## Warning message:
## In rma.uni(es, var, mods = ~sexpr, data = dat.depr.agg) :
##   Studies with NAs omitted from model fitting.

```

```

summary(ma.depr.mod2)

```

```

## Mixed-Effects Model (k = 15; tau^2 estimator: REML)
##
##      logLik deviance      AIC      BIC      AICc
## -21.4665  42.9330  48.9330  50.6278  51.5996
##
## tau^2 (estimated amount of residual heterogeneity):      1.3354 (SE =
## 0.6070)
## tau (square root of estimated tau^2 value):              1.1556
## I^2 (residual heterogeneity / unaccounted variability): 89.79%
## H^2 (unaccounted variability / sampling variability):     9.79
## R^2 (amount of heterogeneity accounted for):              0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 13) = 110.8479, p-val < .0001
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 0.4680, p-val = 0.4939
##
## Model Results:
##
##      estimate      se      zval      pval      ci.lb      ci.ub
## intrcpt      0.0881  2.1137  0.0417  0.9668  -4.0548  4.2309
## sexpr        1.6299  2.3825  0.6841  0.4939  -3.0397  6.2995
##
## ---

```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ma.anx.mod3 = rma.uni(es,var,mods = ~age, data = dat.anx.agg)
```

```
## Warning message:
```

```
## In rma.uni(es, var, mods = ~age, data = dat.anx.agg) :
```

```
##   Studies with NAs omitted from model fitting.
```

```
summary(ma.anx.mod3)
```

```
## Mixed-Effects Model (k = 10; tau^2 estimator: REML)
```

```
##
```

```
##   logLik  deviance      AIC      BIC      AICc
```

```
## -11.7085  23.4169  29.4169  29.6552  35.4169
```

```
##
```

```
## tau^2 (estimated amount of residual heterogeneity):      0.9095 (SE =  
## 0.5298)
```

```
## tau (square root of estimated tau^2 value):              0.9537
```

```
## I^2 (residual heterogeneity / unaccounted variability):  87.81%
```

```
## H^2 (unaccounted variability / sampling variability):     8.21
```

```
## R^2 (amount of heterogeneity accounted for):              0.00%
```

```
##
```

```
## Test for Residual Heterogeneity:
```

```
## QE(df = 8) = 50.9240, p-val < .0001
```

```
##
```

```
## Test of Moderators (coefficient(s) 2):
```

```
## QM(df = 1) = 0.0199, p-val = 0.8877
```

```
##
```

```
## Model Results:
```

```
##
```

```
##           estimate      se   zval   pval   ci.lb  ci.ub
```

```
## intrcpt    0.4281  2.8367  0.1509  0.8800  -5.1318  5.9880
```

```
## age        0.0088  0.0625  0.1412  0.8877  -0.1137  0.1313
```

```
##
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ma.depr.mod3 = rma.uni(es,var,mods = ~age, data = dat.depr.agg)
```

```
## Warning message:
```

```
## In rma.uni(es, var, mods = ~age, data = dat.depr.agg) :
```

```
##   Studies with NAs omitted from model fitting.
```

```
summary(ma.depr.mod3)
```

```

## Mixed-Effects Model (k = 16; tau^2 estimator: REML)
##
##   logLik  deviance      AIC      BIC      AICc
## -20.3970  40.7939  46.7939  48.7111  49.1939
##
## tau^2 (estimated amount of residual heterogeneity):      0.8266 (SE =
## 0.3891)
## tau (square root of estimated tau^2 value):              0.9092
## I^2 (residual heterogeneity / unaccounted variability):  84.36%
## H^2 (unaccounted variability / sampling variability):    6.39
## R^2 (amount of heterogeneity accounted for):              38.27%
##
## Test for Residual Heterogeneity:
## QE(df = 14) = 67.7139, p-val < .0001
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 8.2565, p-val = 0.0041
##
## Model Results:
##
##           estimate      se      zval      pval      ci.lb      ci.ub
## intrcpt      5.5713  1.4998  3.7147  0.0002  2.6317  8.5109  ***
## age          -0.0971  0.0338 -2.8734  0.0041 -0.1633 -0.0309  **
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

### # Subgroup analyses

#### # Subgroup 1: assessment of tpo-ab

```

ma.anx.sg1.1 = rma(es,var, data = dat.anx.agg, subset=ak=="1")
ma.anx.sg1.0 = rma(es,var, data = dat.anx.agg, subset=ak=="0")
dat.comp <- data.frame(estimate = c(coef(ma.anx.sg1.1),
  coef(ma.anx.sg1.0)), SE = c(ma.anx.sg1.1$se, ma.anx.sg1.0$se),
  sub = c("1","0"), tau2 = round(c(ma.anx.sg1.1$tau2,
  ma.anx.sg1.0$tau2),3))
rma(estimate, sei=SE, mods = ~ sub, method="FE", data=dat.comp, digits=3)

```

```

## Fixed-Effects with Moderators Model (k = 2)
##
## Test for Residual Heterogeneity:
## QE(df = 0) = 0.000, p-val = 1.000
##
## Test of Moderators (coefficient(s) 2):

```



```

## QM(df = 1) = 1.725, p-val = 0.189
##
## Model Results:
##
##          estimate      se   zval   pval   ci.lb  ci.ub
## intrcpt      0.458  0.436  1.050  0.294  -0.397  1.312
## sub1         0.698  0.532  1.314  0.189  -0.344  1.740
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ma.depr.sg1.1 = rma(es,var, data = dat.depr.agg, subset=ak=="1")
ma.depr.sg1.0 = rma(es,var, data = dat.depr.agg, subset=ak=="0")
dat.comp2 <- data.frame(estimate = c(coef(ma.depr.sg1.1),
  coef(ma.depr.sg1.0)), SE = c(ma.depr.sg1.1$se, ma.depr.sg1.0$se),
  sub = c("1","0"), tau2 = round(c(ma.depr.sg1.1$tau2,
  ma.depr.sg1.0$tau2),3))
rma(estimate, sei=SE, mods = ~ sub, method="FE", data=dat.comp2, digits=3)

```

```

## Fixed-Effects with Moderators Model (k = 2)
##
## Test for Residual Heterogeneity:
## QE(df = 0) = 0.000, p-val = 1.000
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 0.775, p-val = 0.379
##
## Model Results:
##
##          estimate      se   zval   pval   ci.lb  ci.ub
## intrcpt      0.989  0.304  3.252  0.001   0.393  1.586  **
## sub1         0.413  0.469  0.880  0.379  -0.506  1.332
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

#### # Subgroup 2: dimensional vs. categorical measures

```

ma.anx.sg2.1 = rma(es,var, data = dat.anx.agg, subset=measure=="1")
ma.anx.sg2.0 = rma(es,var, data = dat.anx.agg, subset=measure=="0")
dat.comp3 <- data.frame(estimate = c(coef(ma.anx.sg2.1),
  coef(ma.anx.sg2.0)), SE = c(ma.anx.sg2.1$se, ma.anx.sg2.0$se),
  sub = c("1","0"), tau2 = round(c(ma.anx.sg2.1$tau2,
  ma.anx.sg2.0$tau2),3))

```

```
rma(estimate, sei=SE, mods = ~ sub, method="FE", data=dat.comp3, digits=3)
```

```
## Fixed-Effects with Moderators Model (k = 2)
##
## Test for Residual Heterogeneity:
## QE(df = 0) = 0.000, p-val = 1.000
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 0.125, p-val = 0.724
##
## Model Results:
##
##          estimate      se    zval   pval   ci.lb  ci.ub
## intrcpt      0.974  0.448   2.173  0.030   0.095  1.852  *
## sub1        -0.198  0.561  -0.353  0.724  -1.296  0.901
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ma.depr.sg2.1 = rma(es,var, data = dat.depr.agg, subset=measure=="1")
ma.depr.sg2.0 = rma(es,var, data = dat.depr.agg, subset=measure=="0")
dat.comp4 <- data.frame(estimate = c(coef(ma.depr.sg2.1),
  coef(ma.depr.sg2.0)), SE = c(ma.depr.sg2.1$se, ma.depr.sg2.0$se),
  sub = c("1","0"), tau2 = round(c(ma.depr.sg2.1$tau2,
  ma.depr.sg2.0$tau2),3))
rma(estimate, sei=SE, mods = ~ sub, method="FE", data=dat.comp4, digits=3)
```

```
## Fixed-Effects with Moderators Model (k = 2)
##
## Test for Residual Heterogeneity:
## QE(df = 0) = 0.000, p-val = 1.000
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 2.327, p-val = 0.127
##
## Model Results:
##
##          estimate      se    zval   pval   ci.lb  ci.ub
## intrcpt      0.804  0.279   2.884  0.004   0.258  1.351  **
## sub1         0.710  0.465   1.526  0.127  -0.202  1.622
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### eAppendix 3. Discussion

Out of three moderators (study quality, proportion of females, mean age) only mean age altered the association between AIT and depression (For details, see eTable 6 of the supplement) on a Bonferroni corrected threshold. This result indicates that the association between AIT and depression decreases slightly when study participants show a higher averaged age. Since mean age is an aggregated variable, interpreting the moderator results as “the association of AIT and depression decreases with advancing age” must be treated with caution. Nevertheless, studies suggest that depression changes during the lifetime<sup>11</sup> with more age- and disease-related causal factors<sup>12</sup>, with elderly people being more vulnerable<sup>11</sup> and with a higher risk of relapse<sup>13</sup>. Additionally, several depression scales are loaded toward measuring somatic symptoms of depression, such as sleep disturbances or the decline of sexual function<sup>14</sup>. Among the elderly, these are prominent symptoms that even appear in non-depressed persons; hence, typical depression instruments overestimate the extent of depressive symptoms<sup>14</sup>. When applying rigorous, categorical diagnostic criteria, prevalence rates among elderly are much lower than when determining caseness by the level of depressive symptomatology<sup>15</sup>. Since we included both categorical and dimensional measures in our analysis, age as a moderating factor becomes plausible.

Both the other moderator variables and the subgroup analyses remained insignificant and can thus not explain for the high amount of heterogeneity (s. Table 1). We took into consideration post-hoc moderator analyses to reduce residual heterogeneity but due to a lack of data no further meta-regressions were conductible. Besides, there are possible explanations for the high  $I^2$  statistics in our meta-analysis. First, our study sample comprises investigations from nine different countries, each varying in its baseline prevalence of depression or anxiety disorders and in its handling of psychiatric illnesses. Second, the assessment of thyroid values is not always comparable between our studies since different methods were used (e.g. immunometric assays or chemiluminescent methods for measuring TSH). Third, psychiatric diagnoses were made employing either self-descriptive or observer-rated questionnaires. Even though most of these standardized instruments show high intercorrelations (convergent validity) they sometimes measure different aspects of one disease<sup>16–19</sup>. Both the method for assessing TSH values and the psychiatric instruments used were prespecified moderators that could not be tested because of little group sizes. We accepted a priori the possibility of high and significant heterogeneity in order to integrate a lot of data and to provide a broad overview of the existing evidence regarding AIT and psychiatric issues.

#### *Strengths and limitations*

We retrieved data from a profound and extensive literature search (s. Figure 1) including more databases than usual and searching psychological platforms, as well. Furthermore, the analysis was conducted entirely following the PRISMA statement<sup>20,21</sup> (eTable 2) and assessed study quality according to the Newcastle-Ottawa Quality Assessment Scale for case-control studies<sup>22</sup> (for details, see eTable 4 of the supplement). Three criteria, treating an adequate case definition, the definition of controls, and the application of the same method for determining exposure for both cases and controls, were fulfilled by every study we included. With a mean of 5.77 (SD = 1.17) for anxiety and a mean of 5.65 (SD = 1.14) for depression, our data are of sufficient quality. The meta-regression of study quality as a moderator remained nonsignificant (for both anxiety and depression), suggesting that our outcome measure is independent of the studies' risk of bias.

There are some limitations of this study. First, the high amount of heterogeneity could not be explained by the specified moderators. This indicates that our results show systematic variation not caused by the studies' sampling error. However, there are possible explanations for our high  $I^2$  statistics, as we explained above. Second, we detected asymmetry in the funnel plot concerning depression which potentially results in slightly overestimated effect sizes. Third, the diagnosis of an AIT depends on the sample examined and the criteria applied leading to insecurities about the comparability of our studies. Though, separate analyses for every form of hypothyroidism were not possible due to little group sizes. Forth, some of the prespecified analyses (particularly subgroup analyses) could not be conducted because of insufficient data thus leaving questions unanswered. This is one further reason contributing to the high amount of heterogeneity. Fifth, there are general problems with the methodology of case-control studies themselves. With the application of this study design, several points must be considered, including the masking of raters to reduce observation bias, an appropriate definition of caseness, an adequate control group and reporting of non-response rates. Nevertheless, a recent

review shows that many of the papers fail to include sufficient information to evaluate the impact of selection or information biases<sup>23</sup>. Hence, the generalizability of those findings is limited, and accurate replication is impeded. Another confounding influence may result from control self-selection via advertisements in which motivational factors related to personality traits or lifestyle are likely to play a role<sup>24</sup>. In case-control studies, the control group functions to represent the level of exposure within the general population from which the cases have been identified<sup>23</sup>. A typical problem for studies that assess episodes of depression and anxiety states that these diseases are somewhat common in the general population<sup>24</sup>. Screening out potential controls for such exposures may result in a sample of 'super-healthy' controls that would seriously inflate the case-control differences in the analysis and overestimate effect sizes<sup>23,24</sup>.

**eTable 1.** Current Terminology Concerning AIT: Similarities and Differences

<b>Term</b>	<b>Definition</b>
Hashimoto's thyroiditis (HT)	Elevated TSH values ft4 within normal range presence of thyroid peroxidase antibodies (tpo-ab)
Autoimmune thyroiditis (AIT)	Elevated TSH values ft4 within normal range presence of thyroid peroxidase antibodies (tpo-ab) comprising HT and atrophic thyroiditis
Subclinical hypothyroidism	Elevated TSH values ft4 within normal range
Latent hypothyroidism	Elevated TSH values ft4 within normal range
Overt hypothyroidism	Elevated TSH values

Abbreviations: ft4, serum free T4; TSH, thyroid stimulating hormone

**eTable 2.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)<sup>21,20</sup> Checklist

Section/topic	#	Checklist item	Reported on page #
<b>TITLE</b>			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
<b>ABSTRACT</b>			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	3, 4
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5, 6
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6
<b>METHODS</b>			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	8
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7, 8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	7
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7, Figure 1
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8

Section/topic	#	Checklist item	Reported on page #
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7, 8, eTable 3
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	9
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ ) for each meta-analysis.	8, 9, 10
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	10
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	9, 10
<b>RESULTS</b>			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	eTable 5
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	eAppendix 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figure 2, Figure 3
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	11, 12
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	12
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	12, 13, eAppendix 3
<b>DISCUSSION</b>			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	13ff

Section/topic	#	Checklist item	Reported on page #
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	eAppendix 3
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	14, 15, 16
<b>FUNDING</b>			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	See conflict of interest



### **eTable 3. Coding Protocol**

#### **Coding protocol**

*Hashimoto's Thyroiditis and anxiety or depression*

#### **Inclusion criteria:**

---

- a) Year of publication: 1992-2017
- b) Publication language: English, German
- c) Patients in the experimental group suffer from a diagnosed form of hypothyroidism
- d) Comparison to a control group free from any thyroid disease
- e) Anxiety disorder and depression are measured with standardized instruments
- f) Effect sizes or associated data to compute effect sizes are reported

#### **Exclusion criteria:**

---

- a) Abstracts or pilot data
  - b) Publication language other than English or German
  - c) No quantitative assessment of depression and anxiety disorders
  - d) The values for depression and anxiety in the experimental group were presented without comparing them to a healthy control group OR it was not possible to retrieve scores from a representative norm sample to use as control group
  - e) Study participants comprised pregnant women
-

### Coding procedure:

---

- (a) One line represents one effect size
- (b) If effect sizes are reported separately for the whole sample and for subgroups (e.g. men and women), only information concerning the whole sample will be extracted. If information is only reported for subgroups, every subgroup will be treated as distinct sample (variable sno).
- (c) If there are multiple effect sizes within one sample concerning different outcome variables (e.g. different subforms of anxiety), every effect size will be reported in its own line. These effect sizes are coded as dependent by allocating the same number for the variable sno.
- (d) No computations should be carried out while coding. Information is extracted directly without conversions.

Variable	Description	Code	Example
<b>General characteristics</b>			
study	Study name comprising lead author and year of publication.	Free specification	carlbring2007 carlbring2007a
pubyear	Year of publication	Range: [1992, 2017]	2007
incl	Effect size cannot be included in statistical analysis	0 = exclude 1 = include	1
sid	<u>Consecutive</u> number for every publication	Range: [1, ∞]	1
sno	<u>Consecutive</u> number for every sample	Range: [1, ∞]	1
colyear	Year of conduct	Range: [1992, 2017]	2007
centry	Country of conduct  If not reported, extract affiliation of lead author as ISO-CODE 2: <a href="http://en.wikipedia.org/wiki/ISO_3166-1_alpha-2">http://en.wikipedia.org/wiki/ISO_3166-1_alpha-2</a> or "XX" describing samples with participants originating from different countries	Free specification	DE
pubtype	Publication type	1 = Peer-reviewed Journal 2 = Book 3 = Thesis (Master / PhD) 4 = Poster 5 = Other	1

Variable	Description	Code	Example
<b>Sample</b>			
n1	Sample size <i>N</i>	Range: [2, ∞]	100
n2	Sample size of experimental group	Range: [2, ∞]	100
n3	Sample size of control group	Range: [2, ∞]	100
sample	Description of sample	Free specification	Adults
samtype	Description of sample (coded)	1 = students (undergraduates, college) 2 = Adults, mixed sample	2
sexn	Number of women	Range: [0, ∞]	40
sexpr	Proportion of women (in %) If number is not reported	Range: [0, 100]	40
age	Mean age (in years)	Range: [13, ∞]	16.86

Variable	Description	Code	Example
<b>Effect sizes</b>			
(1 = experimental group; 2 = control group)			
m1_anx	Mean anxiety score in the experimental group	Range: $[-\infty, \infty]$	7.00
sd1_anx	Standard deviation of anxiety score in the experimental group	Range: $[0, \infty]$	1.00
m1_depr	Mean depression score in the experimental group	Range: $[-\infty, \infty]$	7.00
sd1_depr	Standard deviation of depression score in the experimental group	Range: $[0, \infty]$	1.00
m2_anx	Mean anxiety score in the control group	Range: $[-\infty, \infty]$	7.00
sd2_anx	Standard deviation of anxiety score in the control group	Range: $[0, \infty]$	1.00
m2_depr	Mean depression score in the control group	Range: $[-\infty, \infty]$	7.00
sd2_depr	Standard deviation of depression score in the control group	Range: $[0, \infty]$	1.00
tpos	Proportion of psychiatric diagnoses in the experimental group	Range: $[0, 1]$	0.50
tneg	Converse probability of tpos	Range: $[0, 1]$	0.50
cpos	Proportion of psychiatric diagnoses in the control group	Range: $[0, 1]$	0.50
cneg	Converse probability of cpos	Range: $[0, 1]$	0.50
d	Cohen's d	Range: $[-\infty, \infty]$	1.50

Variable	Description	Code	Example
<b>Effect sizes</b>			
(1 = experimental group; 2 = control group)			
OR	Odds ratio	Range: [1, ∞]	1.50
pval	p-value reported for the odds ratio	Range: [0, 1]	0.05
ciup	Upper limit of the confidence interval reported for the odds ratio	Range: [-∞, ∞]	1.50
cilow	Lower limit of the confidence interval reported for the odds ratio	Range: [-∞, ∞]	1.50
sei	Standard error (sometimes reported instead of confidence intervals)	Range: [-∞, ∞]	1.50

Variable	Description	Code	Example
<b>Moderators</b>			
construct	Psychiatric disorder: Anxiety disorder or depression	1 = Anxiety 2 = Depression	1
ak	Assessment of thyroid peroxidase antibodies	0 = No 1 = Yes	1
measure	Assessment of psychiatric disorder via categorical or dimensional measure.	0 = Categorical 1 = Dimensional	1
qual1	Category Selection <ul style="list-style-type: none"> <li>- Is the case definition adequate?</li> <li>- Representativeness of cases</li> <li>- Selection of controls</li> <li>- Definition of controls</li> </ul>	0 = 0 Stars 1 = 1 Star 2 = 2 Stars 3 = 3 Stars 4 = 4 Stars	1
qual2	Category Comparability: Comparability of cases and controls on the basis of the design or analysis <ul style="list-style-type: none"> <li>- Study controls for thyroid disease</li> <li>- Study controls for any additional factor</li> </ul>	0 = 0 Stars 1 = 1 Star 2 = 2 Stars	1
qual3	Category Exposure <ul style="list-style-type: none"> <li>- Ascertainment of exposure</li> <li>- Same method of ascertainment for cases and controls</li> <li>- Non-response rate</li> </ul>	0 = 0 Stars 1 = 1 Star 2 = 2 Stars 3 = 3 Stars	1

**eTable 4.** The Newcastle-Ottawa Quality Assessment Scale<sup>22</sup> for Case-Control Studies

Category	Item	Answers
<b>Selection</b>	Is the case definition adequate?	<i>Yes, with independent validation</i> Yes, e.g. record linkage or based on self-reports No description
	Representativeness of cases	<i>Consecutive or obviously representative series of cases</i> Potential for selection bias or not stated
	Selection of controls	<i>Community controls</i> Hospital controls No description
	Definition of controls	<i>No history of disease (endpoint)</i> No description of source
<b>Comparability</b> (on basis of design)	Study controls for thyroid disease	Yes No
	Study controls for any additional factor	Yes No
<b>Exposure</b>	Ascertainment of exposure	<i>Secure record (e.g. surgical records)</i> <i>Structured interview where blind to case/control status</i> Interview not blinded to case/control status Written self-report or medical record only No description
	Same method of ascertainment for cases and controls	Yes No
	Non-response rate	<i>Same rate for both groups</i> Non-respondents described Rate different or no designation

Annotation: Answers which are awarded with one star are printed in cursive characters.



**eTable 5.** Study Characteristics of all Included Studies

<i>author(s), year</i>	No. of patients	mean age (years)	sex (% of females)	instrument(s)	country	prevalence rates for depression (in %)		prevalence rates for anxiety disorders (in %)	
						EG	CG	EG	CG
<i>van de Ven et al., 2016</i> <sup>25</sup>	906	56.8	NA	BDI	Netherlands	19.1	15.3	/	/
<i>Quinke, 2015</i> <sup>26</sup>	18	40.0	88.9	BDI, HAMD	Germany	NA	NA	/	/
<i>Kirim et al., 2012</i> <sup>27</sup>	201	37.7	98.0	HAMD	Turkey	91.5	12.1	/	/
<i>Gulseren et al., 2006</i> <sup>28</sup>	116	41.5	89.7	HARS, HAMD	Turkey	NA	NA	NA	NA
<i>Demartini et al., 2014</i> <sup>29</sup>	246	NA	92.7	HAMD, MADRS	Italy	63.8	28.5	/	/
<i>Ittermann et al., 2015</i> <sup>30</sup>	1714	51.2	NA	M-CIDI, BDI	Germany	16.7	13.4	8.2	6.4
<i>Constant et al., 2006</i> <sup>31</sup>	49	49.5	57.1	STAI, BDI	Belgium	NA	NA	NA	NA
<i>Yalcin et al., 2017</i> <sup>32</sup>	124	41.4	82.2	BAI, BDI	Turkey	17.3	3.3	58.2	41.3
<i>Bunevicius et al., 2007</i> <sup>33</sup>	348	52.0	100	HADS	Lithuania	20.5	12.5	39.0	25.5
<i>Krysiak et al., 2016</i> <sup>34</sup>	69	29.48	100	BDI	Poland	32.0	17.0	/	/
<i>Carta et al., 2004</i> <sup>35</sup>	222	NA	57.2	CIDIS	Italy	30.9	18.9	36.6	18.5
<i>Engum et al., 2005</i> <sup>36</sup>	30,175	NA	NA	HADS	Norway	11.6	13.2	16.3	16.7
<i>Pop et al., 1998</i> <sup>37</sup>	583	49.9	100	EDS	Netherlands	NA	NA	/	/
<i>Carta et al., 2005</i> <sup>38</sup>	95	NA	NA	CIDIS	Italy	NA	NA	NA	NA
<i>Zettinig et al., 2003</i> <sup>39</sup>	76	45.8	80.3	SAS, SDS	Austria	NA	NA	NA	NA
<i>Ayhan et al., 2014</i> <sup>40</sup>	119	34.4	95.0	BAI, BDI	Turkey	33.3	5.9	37.3	14.7

<i>author(s), year</i>	No. of patients	mean age (years)	sex (% of females)	instrument(s)	country	prevalence rates for depression (in %)		prevalence rates for anxiety disorders (in %)	
						EG	CG	EG	CG
<i>Grabe et al., 2005</i> <sup>41</sup>	1006	47.0	100	ZCS	Germany	/	/	NA	NA
<i>Schinhammer, 2010</i> <sup>42</sup>	50	45.9	NA	STAI, BDI	Germany	NA	NA	NA	NA
<i>Franke, 2013</i> <sup>43</sup>	57	44.2	86.0	BDI	Germany	NA	NA	/	/

Abbreviations: BAI, Beck's Anxiety Inventory BDI, Beck's Depression Inventory; CG, control group; CIDIS, Composite International Diagnostic Interview Simplified; EDS, Edinburgh Depression Scale; EG, experimental group; HADS, Hospital Anxiety and Depression Scale; HAMD, Hamilton Depression Rating Scale; HARS, Hamilton Anxiety Rating Scale; MADRS, Montgomery-Asberg Depression Rating Scale; M-CIDI, Munich-Composite International Diagnostic Interview; NA, not available; SAS, Zung's Self-Rating Anxiety Scale; SDS, Zung's Self-Rating Depression Scale; STAI, State-Trait Anxiety Inventory; ZCS, Zerssen's Complaints Scale;

**eTable 6.** Slope ( $\beta$ ), 95% CI, and *P* Values for Meta-Regression Models

Psychiatric disorder	study quality		proportion of females		mean age	
	$\beta$ ( $\pm$ CI)	p	$\beta$ ( $\pm$ CI)	p	$\beta$ ( $\pm$ CI)	p
depression	-0.029 (-0.499,0.440)	0.902	1.630 (-3.040,6.300)	0.517	<b>-0.097 (-0.163,-0.031)</b>	<b>0.004</b>
anxiety	0.226 (-0.240,0.693)	0.342	-2.209 (-6.568,2.151)	0.494	0.009 (-0.114,0.131)	0.888

Slopes with  $P < 0.008$  (Bonferroni corrected) are bolded.

Abbreviations: CI, confidence interval

## eReferences

1. Dayan C, Daniels GH. Chronic Autoimmune Thyroiditis. *N Engl J Med*. 1996;335(2):99-107. doi:10.1056/NEJM199607113350206.
2. Biondi B, Cooper DS. The Clinical Significance of Subclinical Thyroid Dysfunction. *Endocr Rev*. 2008;29(1):76-131. doi:10.1210/er.2006-0043.
3. Hirschfeld RM. History and Evolution of the Monoamine Theory of Depression. *J Clin Psychiatry*. 2000;61(suppl 6):4-6.
4. Pariante CM, Lightman SL. The HPA Axis in Major Depression: Classical Theories and New Developments. *Trends Neurosci*. 2008;31(9):464-468. doi:10.1016/J.TINS.2008.06.006.
5. Wohleb ES, Franklin T, Iwata M, Duman RS. Integrating Neuroimmune Systems in the Neurobiology of Depression. *Nat Rev Neurosci*. 2016;17(8):497-511. doi:10.1038/nrn.2016.69.
6. Chistiakov DA. Immunogenetics of Hashimoto's Thyroiditis. *J Autoimmune Dis*. 2005;2(1). doi:10.1186/1740-2557-2-1.
7. Bauer M, Heinz A, Whybrow PC. Thyroid Hormones, Serotonin and Mood: Of Synergy and Significance in the Adult Brain. *Mol Psychiatry*. 2002;7(2):140-156. doi:10.1038/sj/mp/4000963.
8. Bauer M, Whybrow PC. Thyroid Hormone, Neural Tissue and Mood Modulation. *World J Biol Psychiatry*. 2001;2(2):59-69.
9. Schenck JB, Rizvi AA, Lin T. Severe Primary Hypothyroidism Manifesting with Torsades De Pointes. *Am J Med Sci*. 2006;331(3):154-156.
10. Shojaie M, Eshraghian A. Primary Hypothyroidism Presenting with Torsades de Pointes Type Tachycardia: a Case Report. *Cases J*. 2008;1(1):298. doi:10.1186/1757-1626-1-298.

11. Alexopoulos GS. Depression in the Elderly. *Lancet*. 2005;365(9475):1961-1970. doi:10.1016/S0140-6736(05)66665-2.
12. Cole MG, Dendukuri N. Risk Factors for Depression among Elderly Community Subjects: a Systematic Review and Meta-analysis. *Am J Psychiatry*. 2003;160(6):1147-1156. doi:10.1176/appi.ajp.160.6.1147.
13. Mitchell AJ, Subramaniam H. Prognosis of Depression in Old Age Compared to Middle Age: a Systematic Review of Comparative Studies. *Am J Psychiatry*. 2005;162(9):1588-1601. doi:10.1176/appi.ajp.162.9.1588.
14. Yesavage JA, Brink TL, Rose TL, et al. Development and Validation of a Geriatric Depression Screening Scale: A Preliminary Report. *J Psychiatr Res*. 1983;17(1):37-49. doi:10.1016/0022-3956(82)90033-4.
15. Beekman ATF, Deeg DJH, Van Tilburg T, Smit JH, Hooijer C, Van Tilburg W. Major and Minor Depression in Later Life: a Study of Prevalence and Risk Factors. *J Affect Disord*. 1995;36(1-2):65-75.
16. Enns MW, Larsen DK, Cox BJ. Discrepancies Between Self and Observer Ratings of Depression: The Relationship to Demographic, Clinical and Personality Variables. *J Affect Disord*. 2000;60(1):33-41. doi:10.1016/S0165-0327(99)00156-1.
17. Hautzinger M, Bailer M, Worall H, Keller F. *Beck-Depressions- Inventar (BDI), Manual*, 2<sup>nd</sup> revised edition, Bern: Hans Huber; 2001.
18. Collegium Internationale Psychiatriae Salarum (CIPS), *Hamilton Depression Scale (HAMD)* in: Internationale Skalen für Psychiatrie., 5<sup>th</sup> revised edition, Göttingen: Hogrefe Verlag; 2005:261-266.
19. Collegium Internationale Psychiatriae Salarum (CIPS), *Hamilton Anxiety Scale (HAMA)* in: Internationale Skalen für Psychiatrie., 5<sup>th</sup> revised edition, Göttingen: Hogrefe Verlag; 2005:281-284.

20. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *BMJ*. 2009;339: b2700. doi:10.1136/bmj.b2700.
21. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6(7): e1000097. doi: 10.1371/journal.pmed.1000097.
22. Wells GA, Shea B, O'Connell D, Petersen J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomized Studies in Meta-Analyses. 2009. [www.ohri.ca/programs/clinical\\_epidemiology/oxford.htm](http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm)
23. Lee W, Bindman J, Ford T, et al. Bias in Psychiatric Case-Control Studies. *Br J Psychiatry*. 2007;190:204-209. doi:10.1192/bjp.bp.106.027250
24. Jablensky A. Research Methods in Psychiatric Epidemiology: an Overview. *Aust N Z J Psychiatry*. 2002;36(3):297-310.
25. van de Ven AC. *Towards an Optimal TSH Level: Different Goals for Different Outcomes and for Different Populations? [Dissertation]*. Nijmegen: Radboud Universiteit Nijmegen; 2016.
26. Quinke EM. *Brain, Mood and Cognition in Hypothyroidism [Dissertation]*. Leipzig: Max-Planck-Institut für Kognitions- und Neurowissenschaften; 2015.
27. Kırım S, Keşkek ŞÖ, Köksal F, Haydardedeoğlu FE, Bozkırlı E. Depression in Patients with Euthyroid Chronic Autoimmune Thyroiditis. *Endocr J*. 2012;59(8):705-708.
28. Gulseren S, Gulseren L, Hekimsoy Z, Cetinay P, Ozen C, Tokatlioglu B. Depression, Anxiety, Health-related Quality of Life, and Disability in Patients with Overt and Subclinical Thyroid Dysfunction. *Arch Med Res*. 2006;37(1):133-139. doi:10.1016/j.arcmed.2005.05.008.

29. Demartini B, Ranieri R, Masu A, Selle V, Scarone S, Gambini O. Depressive Symptoms and Major Depressive Disorder in Patients Affected by Subclinical Hypothyroidism. *J Nerv Ment Dis.* 2014;202(8):603-607. doi:10.1097/NMD.000000000000168.
30. Ittermann T, Völzke H, Baumeister SE, Appel K, Grabe HJ. Diagnosed Thyroid Disorders are Associated with Depression and Anxiety. *Soc Psychiatry Psychiatr Epidemiol.* 2015;50(9):1417-1425. doi:10.1007/s00127-015-1043-0.
31. Constant EL, Adam S, Seron X, Bruyer R, Seghers A, Daumerie C. Hypothyroidism and Major Depression: A Common Executive Dysfunction? *J Clin Exp Neuropsychol.* 2006;28(5):790-807. doi:10.1080/13803390591000990.
32. Yalcin MM, Altinova AE, Cavnar B, Bolayir B, Akturk M. Is Thyroid Autoimmunity Itself Associated with Psychological Well-Being in Euthyroid Hashimoto's Thyroiditis? *Endocr J.* 2017;64(4):425-429. doi:10.1507/endocrj.EJ16-0418.
33. Bunevicius R, Peceliuniene J, Mickuviene N, Bunevicius A, Pop VJ, Girdler SS. Mood and Thyroid Immunity Assessed by Ultrasonographic Imaging in a Primary Health Care. *J Affect Disord.* 2007;97(1-3):85-90. doi:10.1016/j.jad.2006.05.029.
34. Krysiak R, Droszol-Cop A, Skrzypulec-Plinta V, Okopien B. Sexual Function and Depressive Symptoms in Young Women with Thyroid Autoimmunity and Subclinical Hypothyroidism. *Clin Endocrinol (Oxf).* 2016;84(6):925-931. doi:10.1111/cen.12956.
35. Carta MG, Loviselli A, Hardoy MC, et al. The Link Between Thyroid Autoimmunity (Antithyroid Peroxidase Autoantibodies) with Anxiety and Mood Disorders in the Community: a Field of Interest for Public Health in the Future. *BMC Psychiatry.* 2004;5:1-5. doi:10.1186/1471-244X-4-25.
36. Engum A, Bjørø T, Mykletun A, Dahl AA. Thyroid Autoimmunity, Depression and Anxiety: are there any Connections? An Epidemiological Study of a Large Population. *J Psychosom Res.* 2005;59(5):263-268. doi:10.1016/j.jpsychores.2005.04.002.

37. Pop VJ, Maartens LH, Leusink G, et al. Are Autoimmune Thyroid Dysfunction and Depression Related? *J Clin Endocrinol Metab.* 1998;83(9):3194-3197. doi:10.1210/jcem.83.9.5131.
38. Carta MG, Hardoy MC, Carpiniello B, et al. A Case Control Study on Psychiatric Disorders in Hashimoto Disease and Euthyroid Goitre: not Only Depressive but also Anxiety Disorders are Associated with Thyroid Autoimmunity. *Clin Pract Epidemiol Ment Health.* 2005;1:23. doi:10.1186/1745-0179-1-23.
39. Zetting G, Asenbaum S, Barbara J, et al. Increased Prevalence of Subclinical Brain Perfusion Abnormalities in Patients with Autoimmune Thyroiditis: Evidence of Hashimoto's Encephalitis? *Anxiety.* 2003:637-643. doi:10.1046/j.1365-2265.2003.01901.x.
40. Ayhan MG, Uguz F, Askin R, Gonen MS. The Prevalence of Depression and Anxiety Disorders in Patients with Euthyroid Hashimoto's Thyroiditis: a Comparative Study. *Gen Hosp Psychiatry.* 2014;36(1):95-98. doi:10.1016/j.genhosppsy.2013.10.002.
41. Grabe HJ, Volzke H, Ludemann J, et al. Mental and Physical Complaints in Thyroid Disorders in the General Population. *Acta Psychiatr Scand.* 2005;112(4):286-293. doi:10.1111/j.1600-0447.2005.00586.x.
42. Schinhammer S. *Untersuchung Möglicher Zusammenhänge Zwischen Hashimoto-Thyreoiditis, Psychischen Störungen Und Genetischen Varianten Des Adenosinsystems [Dissertation].* Erlangen-Nürnberg: Friedrich-Alexander-Universität; 2010.
43. Franke JC. *Untersuchungen Zur Entwicklung von Depression Und Angststörungen Bei Hashimoto-Thyreoiditis [Dissertation].* Erlangen-Nürnberg: Friedrich-Alexander-Universität; 2013.