Supplementary Online Content


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 (β), 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.
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. Up to 80% of subclinical and overt hypothyroidism are induced by autoimmune disorder of the thyroid gland which is the most common cause of decreased thyroid hormone production in patients with acquired hypothyroidism.

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. Other hypotheses include a dysfunction in the hypothalamic–pituitary–adrenal regulation, causing stress and depressive symptoms as well as anxiety disorders, for which monoaminergic drugs are also effective. 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. 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. Importantly, thyroid metabolism has been shown to modulate the brain serotonin system by reducing the sensitivity of 5-HT1A autoreceptors in the Raphé area, and by increasing 5-HT2 receptor sensitivity. Concerning anxiety, it is likely that the thyroid-catecholamine system interaction (via noradrenalin and T3) and its influence on sympathetic nervous system activity affect symptoms such as high blood pressure or sensible nightly tachycardia.
**eAppendix 2. R Code**

```r
# 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$study[26],
    dat.depr$study[27])

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

## Random-Effects Model (k = 21; tau^2 estimator: REML)
##
## logLik  deviance      AIC       BIC      AICc
## -32.0228  64.0455  68.0455   70.0370   68.7514
##
## tau^2 (estimated amount of total heterogeneity): 1.1970 (SE = 0.4387)
## tau (square root of estimated tau^2 value): 1.0941
## I^2 (total heterogeneity / total variability): 92.62%
## H^2 (total variability / sampling variability): 13.55
##
## Test for Heterogeneity:
## Q(df = 20) = 214.1945, p-val < .0001
##
## Model Results:
##
## | estimate | se    | zval | pval   | ci.lb | ci.ub |           |
##|----------|-------|------|--------|-------|-------|-----------|
##| 1.1960   | 0.2576| 4.6420| <.0001 | 0.6910| 1.7009| ***       |
##
##
## Signif. codes:  '***' 0.001  '**' 0.01  '*' 0.05  '.' 0.1  '1'

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

```r
# pred  ci.lb  ci.ub  cr.lb  cr.ub
## 3.31  2.00  5.48  0.37 29.93
```

# Contour-enhanced funnel plot

```r
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(-2.9, 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 = 3.1844, p = 0.0015
# Forest plots

```r
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)
```

```r
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")), ",%))")))
```

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

---

![Forest plot image](forest_plot.png)
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)

# 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))
\begin{verbatim}
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
\end{verbatim}
# Meta-regression

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

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

\texttt{summary(ma.depr.mod1)}

## Mixed-Effects Model (k = 21; tau^2 estimator: REML)
##
## logLik  deviance       AIC       BIC      AICc
## -30.9108   61.8215   67.8215   70.6549   69.4215
##
## tau^2 (estimated amount of residual heterogeneity):     1.2689 (SE =
## 0.4744)
##
## tau (square root of estimated tau^2 value):             1.1265
##
## I^2 (residual heterogeneity / unaccounted variability): 92.84%
##
## H^2 (unaccounted variability / sampling variability):   13.98
##
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 19) = 198.8732, p-val < .0001
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 0.0452, p-val = 0.8316
##
## Model Results:
##
##| estimate | se   | zval  | pval  | ci.lb  | ci.ub  |
##|----------|------|-------|-------|--------|--------|
##| intrcpt  | 1.493| 1.390 | 1.071 | 0.2839 | -1.2347| 4.2138 |
##| qual.all | -0.052| 0.242 | -0.213| 0.8316 | -0.5250| 0.4222 |
##
## 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 = \texttt{rma.uni(es,var,mods = ~qual.all, data = dat.anx.agg)}

\texttt{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 < .001
##
## 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 |
## |----------|---------|-------|------|-------|-------|
## | intccept | -0.4555 | 1.3893 | 0.3279 | -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:

<table>
<thead>
<tr>
<th>estimate</th>
<th>se</th>
<th>zval</th>
<th>pval</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>intrcpt</td>
<td>2.6605</td>
<td>1.8988</td>
<td>1.4012</td>
<td>-1.0610</td>
<td>6.3821</td>
</tr>
<tr>
<td>sexpr</td>
<td>-2.2085</td>
<td>2.2243</td>
<td>-0.9929</td>
<td>0.3208</td>
<td>-6.5681</td>
</tr>
</tbody>
</table>

---
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 = 16; tau^2 estimator: REML)

<table>
<thead>
<tr>
<th>loglik</th>
<th>deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>-22.8478</td>
<td>45.6955</td>
<td>51.6955</td>
<td>53.6127</td>
<td>54.0955</td>
</tr>
</tbody>
</table>

tau^2 (estimated amount of residual heterogeneity): 1.2836 (SE = 0.5622)
tau (square root of estimated tau^2 value): 1.1330
I^2 (residual heterogeneity / unaccounted variability): 89.81%
H^2 (unaccounted variability / sampling variability): 9.81
R^2 (amount of heterogeneity accounted for): 6.13%

Test for Residual Heterogeneity:
QE(df = 14) = 123.2131, p-val < .0001

Test of Moderators (coefficient(s) 2):
QM(df = 1) = 1.8984, p-val = 0.1683

Model Results:

<table>
<thead>
<tr>
<th>estimate</th>
<th>se</th>
<th>zval</th>
<th>pval</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>intrcpt</td>
<td>-0.8577</td>
<td>1.6653</td>
<td>-0.5151</td>
<td>0.6065</td>
<td>-4.1216</td>
</tr>
<tr>
<td>sexpr</td>
<td>2.6473</td>
<td>1.9214</td>
<td>1.3778</td>
<td>0.1683</td>
<td>-1.1185</td>
</tr>
</tbody>
</table>

---
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:  '***' 0.001 '**' 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 = 17; tau^2 estimator: REML)
##
## | logLik  | deviance | AIC      | BIC      | AICc     |
## |---------|----------|----------|----------|----------|
## | -21.4167| 42.8334  | 48.8334  | 50.9575  | 51.0152  |
##
## tau^2 (estimated amount of residual heterogeneity): 0.7693 (SE = 0.3506)
## tau (square root of estimated tau^2 value): 0.8771
## I^2 (residual heterogeneity / unaccounted variability): 84.76%
## H^2 (unaccounted variability / sampling variability): 6.56
## R^2 (amount of heterogeneity accounted for): 44.72%
##
## Test for Residual Heterogeneity:
## QE(df = 15) = 71.6356, p-val < .0001
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 10.9071, p-val = 0.0010
##
## Model Results:
##
## | estimate | se     | zval  | pval        | ci.lb   | ci.ub |
## |----------|--------|-------|-------------|---------|-------|
## | intrcpt  | 5.7950 | 1.4067| 4.1195      | <.0001  | 3.0379| 8.5521| ***     |
## | age      | -0.1030| 0.0312| -3.3026     | 0.0010  | -0.1642| -0.0419| ***     |
##
## 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))
rm(a(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):
### Model Results:

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
<th>p-value</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.458</td>
<td>0.436</td>
<td>1.050</td>
<td>0.294</td>
<td>-0.397</td>
<td>1.312</td>
</tr>
<tr>
<td>sub1</td>
<td>0.698</td>
<td>0.532</td>
<td>1.314</td>
<td>0.189</td>
<td>-0.344</td>
<td>1.740</td>
</tr>
</tbody>
</table>

---

**Signif. codes:** 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```r
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) = 0.416, p-val = 0.519

---

**Signif. codes:** 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```r
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))
```

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

```r
# 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) = 1.582, p-val = 0.208
##
## Model Results:
##
##          estimate     se   zval   pval   ci.lb  ci.ub
## intrcpt     0.804  0.279  2.884  0.004   0.258  1.351  **
## sub1        0.579  0.460  1.258  0.208 -0.323  1.481
##
## Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Appendix 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\textsuperscript{11} with more age- and disease-related causal factors\textsuperscript{12}, with elderly people being more vulnerable\textsuperscript{13} and with a higher risk of relapse\textsuperscript{13}. Additionally, several depression scales are loaded toward measuring somatic symptoms of depression, such as sleep disturbances or the decline of sexual function\textsuperscript{14}. Among the elderly, these are prominent symptoms that even appear in non-depressed persons; hence, typical depression instruments overestimate the extent of depressive symptoms\textsuperscript{14}. When applying rigorous, categorical diagnostic criteria, prevalence rates among elderly are much lower than when determining caseness by the level of depressive symptomatology\textsuperscript{15}. 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\textsuperscript{16–19}. 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\textsuperscript{20,21} (eTable 2) and assessed study quality according to the Newcastle-Ottawa Quality Assessment Scale for case-control studies\textsuperscript{22} (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 5.67 (SD = 1.11) 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. 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. 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. A typical problem for studies that assess episodes of depression and anxiety states that these diseases are somewhat common in the general population. 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.
**eTable 1. Current Terminology Concerning AIT: Similarities and Differences**

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
</table>
| 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) Checklist

<table>
<thead>
<tr>
<th>Section/topic</th>
<th>#</th>
<th>Checklist item</th>
<th>Reported on page #</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TITLE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>1</td>
<td>Identify the report as a systematic review, meta-analysis, or both.</td>
<td></td>
</tr>
<tr>
<td><strong>ABSTRACT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structured summary</td>
<td>2</td>
<td>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.</td>
<td>3, 4</td>
</tr>
<tr>
<td><strong>INTRODUCTION</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>3</td>
<td>Describe the rationale for the review in the context of what is already known.</td>
<td>5, 6</td>
</tr>
<tr>
<td>Objectives</td>
<td>4</td>
<td>Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).</td>
<td>6</td>
</tr>
<tr>
<td><strong>METHODS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol and registration</td>
<td>5</td>
<td>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.</td>
<td>8</td>
</tr>
<tr>
<td>Eligibility criteria</td>
<td>6</td>
<td>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.</td>
<td>7, 8</td>
</tr>
<tr>
<td>Information sources</td>
<td>7</td>
<td>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.</td>
<td>7</td>
</tr>
<tr>
<td>Search</td>
<td>8</td>
<td>Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.</td>
<td>7</td>
</tr>
<tr>
<td>Study selection</td>
<td>9</td>
<td>State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).</td>
<td>7, Figure 1</td>
</tr>
<tr>
<td>Data collection process</td>
<td>10</td>
<td>Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.</td>
<td>8</td>
</tr>
<tr>
<td>Section/topic</td>
<td>#</td>
<td>Checklist Item</td>
<td>Reported on page #</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---</td>
<td>---------------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Data items</td>
<td>11</td>
<td>List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.</td>
<td>7, 8, eTable 3</td>
</tr>
<tr>
<td>Risk of bias in individual studies</td>
<td>12</td>
<td>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.</td>
<td>9</td>
</tr>
<tr>
<td>Summary measures</td>
<td>13</td>
<td>State the principal summary measures (e.g., risk ratio, difference in means).</td>
<td>9</td>
</tr>
<tr>
<td>Synthesis of results</td>
<td>14</td>
<td>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.</td>
<td>8, 9, 10</td>
</tr>
<tr>
<td>Risk of bias across studies</td>
<td>15</td>
<td>Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).</td>
<td>10</td>
</tr>
<tr>
<td>Additional analyses</td>
<td>16</td>
<td>Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.</td>
<td>9, 10</td>
</tr>
</tbody>
</table>

**RESULTS**

<table>
<thead>
<tr>
<th>Section/topic</th>
<th>#</th>
<th>Checklist Item</th>
<th>Reported on page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study selection</td>
<td>17</td>
<td>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.</td>
<td>Figure 1</td>
</tr>
<tr>
<td>Study characteristics</td>
<td>18</td>
<td>For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.</td>
<td>eTable 5</td>
</tr>
<tr>
<td>Risk of bias within studies</td>
<td>19</td>
<td>Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).</td>
<td>eAppendix 3</td>
</tr>
<tr>
<td>Results of individual studies</td>
<td>20</td>
<td>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.</td>
<td>Figure 2, Figure 3</td>
</tr>
<tr>
<td>Synthesis of results</td>
<td>21</td>
<td>Present results of each meta-analysis done, including confidence intervals and measures of consistency.</td>
<td>11, 12</td>
</tr>
<tr>
<td>Risk of bias across studies</td>
<td>22</td>
<td>Present results of any assessment of risk of bias across studies (see Item 15).</td>
<td>12</td>
</tr>
<tr>
<td>Additional analysis</td>
<td>23</td>
<td>Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).</td>
<td>12, 13, eAppendix 3</td>
</tr>
</tbody>
</table>

**DISCUSSION**

<table>
<thead>
<tr>
<th>Section/topic</th>
<th>#</th>
<th>Checklist Item</th>
<th>Reported on page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of evidence</td>
<td>24</td>
<td>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).</td>
<td>13ff</td>
</tr>
<tr>
<td>Section/topic</td>
<td>#</td>
<td>Checklist item</td>
<td>Reported on page #</td>
</tr>
<tr>
<td>---------------</td>
<td>----</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Limitations</td>
<td>25</td>
<td>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).</td>
<td>eAppendix 3</td>
</tr>
<tr>
<td>Conclusions</td>
<td>26</td>
<td>Provide a general interpretation of the results in the context of other evidence, and implications for future research.</td>
<td>14, 15, 16</td>
</tr>
<tr>
<td><strong>FUNDING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funding</td>
<td>27</td>
<td>Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.</td>
<td>See conflict of interest</td>
</tr>
</tbody>
</table>
**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.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Code</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>study</td>
<td>Study name comprising lead author and year of publication.</td>
<td>Free specification</td>
<td>carlbring2007 carlbring2007a</td>
</tr>
<tr>
<td>incl</td>
<td>Effect size cannot be included in statistical analysis</td>
<td>0 = exclude 1 = include</td>
<td>1</td>
</tr>
<tr>
<td>sid</td>
<td>Consecutive number for every publication</td>
<td>Range: [1, ∞]</td>
<td>1</td>
</tr>
<tr>
<td>sno</td>
<td>Consecutive number for every sample</td>
<td>Range: [1, ∞]</td>
<td>1</td>
</tr>
<tr>
<td>cntry</td>
<td>Country of conduct</td>
<td>Free specification</td>
<td>DE</td>
</tr>
<tr>
<td>pubtype</td>
<td>Publication type</td>
<td>1 = Peer-reviewed Journal 2 = Book 3 = Thesis (Master / PhD) 4 = Poster 5 = Other</td>
<td>1</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Code</td>
<td>Example</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------</td>
<td>-----------------------</td>
<td>---------</td>
</tr>
<tr>
<td>n1</td>
<td>Sample size $N$</td>
<td>Range: [2, ∞]</td>
<td>100</td>
</tr>
<tr>
<td>n2</td>
<td>Sample size of experimental group</td>
<td>Range: [2, ∞]</td>
<td>100</td>
</tr>
<tr>
<td>n3</td>
<td>Sample size of control group</td>
<td>Range: [2, ∞]</td>
<td>100</td>
</tr>
<tr>
<td>sample</td>
<td>Description of sample</td>
<td>Free specification</td>
<td>Adults</td>
</tr>
<tr>
<td>samtype</td>
<td>Description of sample (coded)</td>
<td>1 = students (undergraduates, college)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 = Adults, mixed sample</td>
<td></td>
</tr>
<tr>
<td>sexn</td>
<td>Number of women</td>
<td>Range: [0, ∞]</td>
<td>40</td>
</tr>
<tr>
<td>sexpr</td>
<td>Proportion of women (in %)</td>
<td>Range: [0, 100]</td>
<td>40</td>
</tr>
<tr>
<td>age</td>
<td>Mean age (in years)</td>
<td>Range: [13, ∞]</td>
<td>16.86</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Code</td>
<td>Example</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------</td>
<td>---------------</td>
<td>---------</td>
</tr>
<tr>
<td>m1_anx</td>
<td>Mean anxiety score in the experimental group</td>
<td>Range: [(-\infty, \infty)]</td>
<td>7.00</td>
</tr>
<tr>
<td>sd1_anx</td>
<td>Standard deviation of anxiety score in the experimental group</td>
<td>Range: [0, (\infty)]</td>
<td>1.00</td>
</tr>
<tr>
<td>m1_depr</td>
<td>Mean depression score in the experimental group</td>
<td>Range: [(-\infty, \infty)]</td>
<td>7.00</td>
</tr>
<tr>
<td>sd1_depr</td>
<td>Standard deviation of depression score in the experimental group</td>
<td>Range: [0, (\infty)]</td>
<td>1.00</td>
</tr>
<tr>
<td>m2_anx</td>
<td>Mean anxiety score in the control group</td>
<td>Range: [(-\infty, \infty)]</td>
<td>7.00</td>
</tr>
<tr>
<td>sd2_anx</td>
<td>Standard deviation of anxiety score in the control group</td>
<td>Range: [0, (\infty)]</td>
<td>1.00</td>
</tr>
<tr>
<td>m2_depr</td>
<td>Mean depression score in the control group</td>
<td>Range: [(-\infty, \infty)]</td>
<td>7.00</td>
</tr>
<tr>
<td>sd2_depr</td>
<td>Standard deviation of depression score in the control group</td>
<td>Range: [0, (\infty)]</td>
<td>1.00</td>
</tr>
<tr>
<td>tpos</td>
<td>Proportion of psychiatric diagnoses in the experimental group</td>
<td>Range: [0, 1]</td>
<td>0.50</td>
</tr>
<tr>
<td>tneg</td>
<td>Converse probability of tpos</td>
<td>Range: [0, 1]</td>
<td>0.50</td>
</tr>
<tr>
<td>cpos</td>
<td>Proportion of psychiatric diagnoses in the control group</td>
<td>Range: [0, 1]</td>
<td>0.50</td>
</tr>
<tr>
<td>cneg</td>
<td>Converse probability of cpos</td>
<td>Range: [0, 1]</td>
<td>0.50</td>
</tr>
<tr>
<td>d</td>
<td>Cohen's d</td>
<td>Range: [(-\infty, \infty)]</td>
<td>1.50</td>
</tr>
</tbody>
</table>

**Effect sizes**

(1 = experimental group; 2 = control group)
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Code</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>Odds ratio</td>
<td>Range: [1, ∞]</td>
<td>1.50</td>
</tr>
<tr>
<td>pval</td>
<td>p-value reported for the odds ratio</td>
<td>Range: [0, 1]</td>
<td>0.05</td>
</tr>
<tr>
<td>ciup</td>
<td>Upper limit of the confidence interval reported for the odds ratio</td>
<td>Range: [-∞, ∞]</td>
<td>1.50</td>
</tr>
<tr>
<td>cilow</td>
<td>Lower limit of the confidence interval reported for the odds ratio</td>
<td>Range: [-∞, ∞]</td>
<td>1.50</td>
</tr>
<tr>
<td>sei</td>
<td>Standard error (sometimes reported instead of confidence intervals)</td>
<td>Range: [-∞, ∞]</td>
<td>1.50</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Code</td>
<td>Example</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td><strong>Moderators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>construct</td>
<td>Psychiatric disorder: Anxiety disorder or depression</td>
<td>1 = Anxiety</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 = Depression</td>
<td></td>
</tr>
<tr>
<td>ak</td>
<td>Assessment of thyroid peroxidase antibodies</td>
<td>0 = No</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Yes</td>
<td></td>
</tr>
<tr>
<td>measure</td>
<td>Assessment of psychiatric disorder via categorical or dimensional measure.</td>
<td>0 = Categorical</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Dimensional</td>
<td></td>
</tr>
<tr>
<td>qual1</td>
<td>Category Selection</td>
<td>0 = 0 Stars</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>- Is the case definition adequate?</td>
<td>1 = 1 Star</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Representativeness of cases</td>
<td>2 = 2 Stars</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Selection of controls</td>
<td>3 = 3 Stars</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Definition of controls</td>
<td>4 = 4 Stars</td>
<td></td>
</tr>
<tr>
<td>qual2</td>
<td>Category Comparability: Comparability of cases and controls on the basis of</td>
<td>0 = 0 Stars</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>the design or analysis</td>
<td>1 = 1 Star</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Study controls for thyroid disease</td>
<td>2 = 2 Stars</td>
<td></td>
</tr>
<tr>
<td>qual3</td>
<td>Category Exposure</td>
<td>0 = 0 Stars</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>- Ascertainment of exposure</td>
<td>1 = 1 Star</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Same method of ascertainment for cases and controls</td>
<td>2 = 2 Stars</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Non-response rate</td>
<td>3 = 3 Stars</td>
<td></td>
</tr>
</tbody>
</table>
**eTable 4. The Newcastle-Ottawa Quality Assessment Scale**\(^\text{22}\) for Case-Control Studies

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

Annotation: Answers which are awarded with one star are printed in cursive characters.
<table>
<thead>
<tr>
<th>author(s), year</th>
<th>No. of patients</th>
<th>mean age (years)</th>
<th>sex (% of females)</th>
<th>instrument(s)</th>
<th>country</th>
<th>prevalence rates for depression (in %)</th>
<th>prevalence rates for anxiety disorders (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>van de Ven et al., 2016</td>
<td>25</td>
<td>906</td>
<td>56.8</td>
<td>NA</td>
<td>BDI</td>
<td>Netherlands</td>
<td>EG 19.1</td>
</tr>
<tr>
<td>Quinque, 2015</td>
<td>18</td>
<td>40.0</td>
<td>88.9</td>
<td>BDI, HAMD</td>
<td>Germany</td>
<td>NA</td>
<td>EG 15.3</td>
</tr>
<tr>
<td>Kirim et al., 2012</td>
<td>201</td>
<td>37.7</td>
<td>98.0</td>
<td>HAMD</td>
<td>Turkey</td>
<td>NA</td>
<td>EG 12.1</td>
</tr>
<tr>
<td>Gulseren et al., 2006</td>
<td>116</td>
<td>41.5</td>
<td>89.7</td>
<td>HARS, HAMD</td>
<td>Turkey</td>
<td>NA</td>
<td>EG 12.1</td>
</tr>
<tr>
<td>Demartini et al., 2014</td>
<td>246</td>
<td>NA</td>
<td>92.7</td>
<td>HAMD, MADRS</td>
<td>Italy</td>
<td>EG 63.8</td>
<td>EG 28.5</td>
</tr>
<tr>
<td>Ittermann et al., 2015</td>
<td>1714</td>
<td>51.2</td>
<td>NA</td>
<td>M-CIDI, BDI</td>
<td>Germany</td>
<td>EG 16.7</td>
<td>EG 13.4</td>
</tr>
<tr>
<td>Constant et al., 2006</td>
<td>49</td>
<td>49.5</td>
<td>57.1</td>
<td>STAI, BDI</td>
<td>Belgium</td>
<td>NA</td>
<td>EG 8.2</td>
</tr>
<tr>
<td>Yalcin et al., 2017</td>
<td>124</td>
<td>41.4</td>
<td>82.2</td>
<td>BAI, BDI</td>
<td>Turkey</td>
<td>EG 17.3</td>
<td>EG 3.3</td>
</tr>
<tr>
<td>Bunevicius et al., 2007</td>
<td>348</td>
<td>52.0</td>
<td>100</td>
<td>HADS</td>
<td>Lithuania</td>
<td>EG 20.5</td>
<td>EG 12.5</td>
</tr>
<tr>
<td>Krysiak et al., 2016</td>
<td>69</td>
<td>29.48</td>
<td>100</td>
<td>BDI</td>
<td>Poland</td>
<td>EG 32.0</td>
<td>EG 17.0</td>
</tr>
<tr>
<td>Carta et al., 2004</td>
<td>222</td>
<td>NA</td>
<td>57.2</td>
<td>CIDIS</td>
<td>Italy</td>
<td>EG 30.9</td>
<td>EG 18.9</td>
</tr>
<tr>
<td>Engum et al., 2005</td>
<td>30,175</td>
<td>NA</td>
<td>NA</td>
<td>HADS</td>
<td>Norway</td>
<td>EG 11.6</td>
<td>EG 13.2</td>
</tr>
<tr>
<td>Pop et al., 1998</td>
<td>583</td>
<td>49.9</td>
<td>100</td>
<td>EDS</td>
<td>Netherlands</td>
<td>EG NA</td>
<td>EG NA</td>
</tr>
<tr>
<td>Carta et al., 2005</td>
<td>95</td>
<td>NA</td>
<td>NA</td>
<td>CIDIS</td>
<td>Italy</td>
<td>EG NA</td>
<td>EG NA</td>
</tr>
<tr>
<td>Zettinig et al., 2003</td>
<td>76</td>
<td>45.8</td>
<td>80.3</td>
<td>SAS, SDS</td>
<td>Austria</td>
<td>EG NA</td>
<td>EG NA</td>
</tr>
<tr>
<td>Ayhan et al., 2014</td>
<td>119</td>
<td>34.4</td>
<td>95.0</td>
<td>BAI, BDI</td>
<td>Turkey</td>
<td>EG 33.3</td>
<td>EG 5.9</td>
</tr>
<tr>
<td>author(s), year</td>
<td>No. of patients</td>
<td>mean age (years)</td>
<td>sex (% of females)</td>
<td>instrument(s)</td>
<td>country</td>
<td>prevalence rates for depression (in %)</td>
<td>prevalence rates for anxiety disorders (in %)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Grabe et al., 2005</td>
<td>1006</td>
<td>47.0</td>
<td>100</td>
<td>ZCS</td>
<td>Germany</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Schinhammer, 2010</td>
<td>50</td>
<td>45.9</td>
<td>NA</td>
<td>STAI, BDI</td>
<td>Germany</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Franke, 2013</td>
<td>57</td>
<td>44.2</td>
<td>86.0</td>
<td>BDI</td>
<td>Germany</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Fjaellegaard et al., 2015</td>
<td>8214</td>
<td>53.2</td>
<td>51.9</td>
<td>MDI</td>
<td>Denmark</td>
<td>1.7</td>
<td>2.0</td>
</tr>
</tbody>
</table>

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; MDI, Major Depression Inventory; 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 (β), 95% CI, and P Values for Meta-Regression Models

<table>
<thead>
<tr>
<th>Psychiatric disorder</th>
<th>study quality</th>
<th></th>
<th>proportion of females</th>
<th></th>
<th>mean age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β (±CI)</td>
<td>p</td>
<td>β (±CI)</td>
<td>p</td>
<td>β (±CI)</td>
</tr>
<tr>
<td>depression</td>
<td>-0.051 (-0.525,0.422)</td>
<td>0.832</td>
<td>2.647 (-1.119,6.413)</td>
<td>0.168</td>
<td>-0.103 (-0.164,-0.042)</td>
</tr>
<tr>
<td>anxiety</td>
<td>0.226 (-0.240,0.693)</td>
<td>0.342</td>
<td>-2.209 (-6.568,2.151)</td>
<td>0.494</td>
<td>0.009 (-0.114,0.131)</td>
</tr>
</tbody>
</table>

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

**Abbreviations:** CI, confidence interval
eReferences


