Supplementary Online Content

Walther A, Breidenstein J, Miller R. Association of testosterone treatment with alleviation of depressive symptoms in men: a systematic review and meta-analysis. *JAMA Psychiatry*. Published online November 14, 2018. doi:10.1001/jamapsychiatry.2018.2734

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

eAppendix. Search Strategy (Extended)

The search was originally carried out on October, 9 2017 and updated on March, 5 2018.

Databases: PubMed/Medline, EMBASE, Scopus, Web of Science, PsycINFO, Cochrane Controlled Trials Register (database inception to 2018 March 5)

Exemplary for PubMed Database:

(((testosterone) AND (administration and dosage)) AND mood): 660 hits
(((testosterone) AND (adverse effects)) AND mood): 921 hits
(((testosterone) AND (deficiency)) AND mood): 398 hits
(((testosterone) AND (standards)) AND mood): 45 hits
(((testosterone) AND (therapeutic use)) AND mood): 1417 hits
(((testosterone) AND (therapy)) AND mood): 1487 hits
(((testosterone) AND (treatment)) AND mood): 2581 hits
(((testosterone) AND (supplementation)) AND mood): 181 hits
7690 TOTAL

after removal of duplicates: 3091
after formal assessment (of the excluded: 85 reviews, 2 meta-analyses, 6 case-studies, 2 meeting abstracts, 1 study protocol, 1 twin study, 3 practical guidelines, 2 books): 2989
human studies (without animals): 1392
without women: 874
without children: 837
without athletic studies: 758
without contraceptive studies: 728
without non-testosterone treatments: 548
without in vitro studies: 469
titles and abstracts finally (manually) screened on relevance including only RCTs: 54
without studies using psychometrically non-validated depression measures: 27

eTable 1. Characteristics of included RCTs

Author, year	Population	Duration	Groups (no. randomized)	Age, yr, mean (SD)	Baseline total T, mean (SD), nmol/L	Depression scale (baseline mean of TT and placebo group)
Grinspoon, 2000	AIDS wasting syndrome	24 wks	Placebo (26) IM TE, 300mg/3wk (26)	41.7 (1.5)	15.6 (1.9)	BDI-I 14.8 vs. 16.3
Rabkin, 2000	AIDS wasting syndrome	6 wks	Placebo (35) IM TC, 400mg/2wk (39)	39.0 (8.2)	13.1 (4.3)	BDI-I 14.2 vs. 13.9
Pope, 2000	Healthy men	6 wks	Placebo (56) IM TC, up to 600mg/wk (56)	27.8	16.9 (5.4)	HDRS 0.9 vs. 1.0
Seidman, 2001	Hypogonadal and MDD	6 wks	Placebo (17) IM TE, 200mg/wk (13)	52 (10)	9.2 (1.8)	BDI-I 23.5 vs. 19.3
Pope, 2003	Refractory Depression	8 wks	Placebo (10) 1% gel, 100mg/d (12)	49.2 (9.1)	9.8 (1.8)	BDI-II 23.1 vs. 23.6
Malkin, 2004	Hypogonadal and ischaemic heart disease	4 wks	Placebo (10) IM Sustanon*, 100mg/2wk (11)	60.8 (4.6)	4.2 (0.5)	BDI-II 9.0 vs. 7.0
Pugh, 2004	Congestive heart failure	12 wks	Placebo (10) IM Sustanon*, 100mg/2wk (10)	62 (9.3)	14.1 (6.3)	BDI-II 7.3 vs. 7.3
Kenny, 2004	Mild cognitive impairment	12 wks	Placebo (5) IM TE, 200mg/3wk (6)	80 (4.0)	14.4 (5.3)	GDS-15 2.7 vs. 4.6
Rabkin, 2004	AIDS wasting syndrome and MDD	8 wks	Placebo (39) IM TC, 400mg/2wk (38)	41 (7.7)	20.6 (9.6)	HDRS 17.8 vs. 16.8
Cavallini, 2004	Older men symptomatic for low T	24 wks	Placebo (45) Oral TU, 160mg/d (40)	63.5 (3.5)	10.2 (2.0)	BRMS 7.0 vs. 7.0
Haren, 2005	Older men	48 wks	Placebo (37) Oral TU, 160mg/d (39)	68.5 (6)	16.2 (4.6)	GDS-30 6.3 vs. 5.7
Seidman, 2005	Treatment-resistant depressed men	6 wks	Placebo (13) IM TE, up to 600mg/wk (13)	46.4 (10.8)	14.5 (7.4)	HDRS 22.8 vs. 22.6
Orengo, 2005	MDD	12 wks	Placebo (12) 1% gel, 50mg/d (12)	63 (8.5)	9.5 (2.1)	HDRS 15.7 vs. 15.7
Lu, 2006	Mild Alzheimer	24 wks	Placebo (24)	66.1 (7.7)	12.7 (4.0)	BDI-I 5.3 vs. 5.6

	Disease and healthy older men		1% gel, 75mg/d (23)			
Vaughan, 2007	T below the range of normal for young adult men	144 wks (36 mo)	Placebo (23) IM TE, 200mg/2wk (24)	710.8 (4.0)	10.1 (1.7)	BDI-I 3.3 vs. 5.1
Svartberg, 2008	Older men	52 wks	Placebo (18) IM TU, 1000mg/12wk (18)	69.0 (5.0)	8.3 (1.9)	BDI-II 5.1 vs. 4.8
Seidman, 2009	Dysthymia	6 wks	Placebo (10) IM TC, 200mg/10d (13)	50.6 (7.0)	11.8 (3.2)	HDRS 14.5 vs. 13.5
Shores, 2009	Dysthymia or minor depression	12 wks	Placebo (16) 1% gel, 75mg/d (17)	59.4 (6.4)	9.7 (3.9)	HDRS 12.7 vs. 13.8
Giltay, 2010	Hypogonadal and metabolic syndrome	30 wks	Placebo (71) IM TU, 1000mg/12wk (113)	52.1 (9.7)	8.0 (0.5)	BDI-I 9.5 vs. 9.3
Pope, 2010	Treatment-resistant men with MDD	6 wks	Placebo (49) 1% gel, 50mg/d (46)	50.3 (7.7)	11.6 (1.2)	HDRS 17.3 vs. 18.2
Stout, 2012	Chronic heart failure	12 wks	Placebo (20) IM Sustanon*, 100mg/2wk (20)	67.2 (7.1)	10.7 (2.6)	BDI-II 10.4 vs. 7.1
Zhang, 2012	Positive score on ADAM questionnaire	24 wks	Placebo (Vitamin E/C) (80) Oral TU, 120 or 160mg/d (depending on baseline T level) (80)	60.3 (6.7)	7.9 (0.8)	HADS-D 4.9 vs. 4.8
Hackett, 2013	Type 2 Diabetes and symptomatic for low T	30 wks	Placebo (102) IM TU, 1000mg/12wk (97)	61.6 (9.8)	9.1 (3.5)	HADS-D 7.9 vs. 7.3
Mirdamadi, 2014	Congestive heart failure	12	Placebo (25) IM TE, 250mg/4wk (25)	60.5 (5.0)	Not reported and not otherwise retrievable	BDI-I 4.6 vs. 4.6
Borst, 2014	Hypogonadal	52 wks	Placebo (16) IM TE, 125mg/wk (14)	70.0 (8.9)	8.8 (2.9)	GDS-15 2.4 vs. 2.1
Cherrier, 2015	Mild cognitive impairment	24 wks	Placebo (12) 1% gel, 50-100mg/d (10)	70.5 (8.2)	10.3 (3.0)	GDS-30 7.2 vs. 4.1
Snyder, 2016	Older men symptomatic for	52 wks	Placebo (234) 1% gel, 50mg/d (230)	72.2 (5.8)	8.2 (2.3)	PHQ-9 6.6 vs. 6.6

	low T					

T = testosterone, TT = total testosterone, AIDS = Acquired Immune Deficiency Syndrome, IM = Intramuscular, TC = testosterone cypionate, TE = testosterone enanthate, TRT = testosterone replacement therapy, TU = testosterone undecanoate, wk = week, wks = weeks, MDD = major depressive disorder, ADAM = Androgen Deficiency in Aging Men questionnaire.
*Blend of testosterone propionate, testosterone phenylpropionate, testosterone isocaproate, and testosterone decanoate.

eTable 2. Risk of Bias of Included Randomized Controlled Trials

Author, year	Adequate sequence generation	Allocation concealment	Selective Reporting	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data addressed (efficacy outcomes)	Incomplete outcome data addressed (harm outcomes)
Grinspoon, 2000	Low	Low	Unclear	Low	Low	High	Unclear
Rabkin, 2000	Low	Low	Low	Low	Low	Low	Low
Pope, 2000	Unclear	Unclear	Low	Low	Low	Low	Unclear
Seidman, 2001	Unclear	Unclear	Unclear	Low	Low	Low	Low
Pope, 2003	Low	Low	Low	Low	Low	Low	Low
Malkin, 2004	Low	Unclear	Low	High	Low	Low	Low
Pugh, 2004	Unclear	Unclear	Unclear	Low	Unclear	High	Low
Kenny, 2004	Unclear	Unclear	Unclear	Unclear	Low	Low	Low
Rabkin, 2004	Low	Unclear	Unclear	Low	Low	Low	Low
Cavallini, 2004	Unclear	Unclear	Unclear	Low	Unclear	Unclear	High
Haren, 2005	Low	Unclear	Unclear	Unclear	High	High	Low
Seidman, 2005	Low	Low	Unclear	Low	Low	Unclear	High
Orengo, 2005	Low	Unclear	Unclear	Low	Unclear	High	High
Lu, 2006	Unclear	Unclear	Unclear	Unclear	Low	Low	Low
Vaughan, 2007	Low	Low	Unclear	Low	Low	Unclear	Unclear
Svartberg, 2008	Unclear	Unclear	Unclear	Low	Low	Low	Unclear
Seidman, 2009	Low	Low	Unclear	Low	Low	Low	Unclear
Shores, 2009	Low	Low	Unclear	Low	Low	High	Unclear
Giltay, 2010	Unclear	Low	Unclear	Low	Low	Low	Low
Pope, 2010	Low	Unclear	Low	Low	Low	Low	Low
Stout, 2012	Unclear	Unclear	Unclear	Unclear	Low	High	High
Zhang, 2012	Unclear	Low	Unclear	Low	High	Low	Low
Hackett, 2013	Unclear	Low	Unclear	Low	Low	Low	Low
Mirdamadi, 2014	Unclear	Unclear	Unclear	Low	Unclear	Low	Unclear
Borst, 2014	Low	Unclear	Unclear	Low	High	High	High
Cherrier, 2015	Unclear	Unclear	Low	Low	Unclear	Low	Low
Snyder, 2016	Low	Low	Low	Low	Low	Low	Low

eTable 3. Jadad Scoring of Included Randomized Controlled Trials

Author, year	Study described as random	Randomization scheme described and appropriate	Study described as double-blind	Method of (double) blinding appropriate	Description of dropouts and withdrawals available	Jadad Score
Grinspoon, 2000	Yes / Yes	No / No	Yes / Yes	Yes / Yes	Yes / Yes	3 / 3
Rabkin, 2000	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	5 / 5
Pope, 2000	Yes / Yes	No / No	Yes / Yes	Yes / No	Yes / Yes	3 / 1
Seidman, 2001	Yes / Yes	No / No	Yes / Yes	Yes / Yes	Yes / Yes	3 / 3
Pope, 2003	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	5 / 5
Malkin, 2004	Yes / Yes	Yes / Yes	No / No	Yes / No	Yes / Yes	4 / 2
Pugh, 2004	Yes / Yes	No / No	Yes / Yes	Yes / No	No / No	2 / 0
Kenny, 2004	Yes / Yes	No / No	Yes / Yes	Yes / Yes	Yes / Yes	3 / 3
Rabkin, 2004	Yes / Yes	Yes / Yes	Yes / Yes	Yes / No	Yes / Yes	5 / 3
Cavallini, 2004	Yes / Yes	No / No	Yes / No	Yes / Yes	Yes / Yes	3 / 2
Haren, 2005	Yes / Yes	Yes / Yes	Yes / Yes	Yes / No	Yes / Yes	5 / 3
Seidman, 2005	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	5 / 5
Orengo, 2005	Yes / Yes	Yes / Yes	Yes / Yes	No / No	Yes / Yes	3 / 3
Lu, 2006	Yes / Yes	No / No	Yes / Yes	Yes / Yes	Yes / Yes	3 / 3
Vaughan, 2007	Yes / Yes	Yes / Yes	Yes / No	Yes / Yes	Yes / Yes	5 / 3
Svartberg, 2008	Yes / Yes	No / No	Yes / Yes	Yes / Yes	Yes / Yes	3 / 3
Seidman, 2009	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	5 / 5
Shores, 2009	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	5 / 5
Giltay, 2010	Yes / Yes	No / Yes	Yes / Yes	Yes / Yes	Yes / Yes	3 / 5
Pope, 2010	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	5 / 5
Stout, 2012	Yes / Yes	Yes / Yes	Yes / Yes	Yes / No	Yes / Yes	5 / 3
Zhang, 2012	Yes / Yes	No / No	No / No	Yes / Yes	No / no	1 / 1
Hackett, 2013	Yes / Yes	No / Yes	Yes / Yes	Yes / No	Yes / Yes	3 / 3
Mirdamadi, 2014	Yes / Yes	No / No	Yes / Yes	Yes / No	No / Yes	2 / 1
Borst, 2014	Yes / Yes	No / No	Yes / No	Yes / No	Yes / Yes	3 / 0
Cherrier, 2015	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	5 / 5
Snyder, 2016	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	5 / 5

eTable 4. Psychometric Instruments With Cut-off Levels According to Authors

	Cut-off	
Measure	**Mild**	**Moderate-Severe**
BDI-I	10-18	19-29
BDI-II	14-19	20-28
HDRS	8-16	17-23
MADRS	7-19	20-34
PHQ-9	5-9	10-14
GDS-15	5-9	10-15
GDS-30	10-19	20-30
BRMS	11-14	15-24
HADS-D	8-10	>11

Note: Cut-off levels are based on test instructions of the psychometric tests.[29-38]

eTable 5. Extraction and Derivation of Central Tendency, Dispersion Measures, and Hedges' *g*

Larry Hedges (1981) proposed the following standardized effect measure for continuous outcomes of two treatment groups A and B:

$g = (M_A - M_B) / SD^*$

where *M* denotes the mean outcome in the respective treatment group and *SD** denotes the pooled standard deviation that is weighted based on their sample sizes *N*:

$SD^* = [(N_A - 1) * SD^2_A + (N_B - 1) * SD^2_B / (N_A + N_B - 2)]^{0.5}$

To calculate *g*, the measures of central tendency M_A and M_B, and the dispersion measures SD_A and SD_B were either directly extracted from all included RCTs, or derived based on other reported data as described in the following table:

Author, year	Depression scale	Extracted central tendency	Derived central tendency	Extracted dispersion	Derived dispersion
Grinspoon, 2000	BDI-I	Last paragraph of results section ("Effects of testosterone administration"): *M* for each group. 9.2 for intervention group vs. 10.8 for control group.	No derivation necessary.	Last paragraph of results section ("Effects of testosterone administration"): Baseline and post-treatment *SE* for each group. 1.4 and 1.5 for intervention group vs. 1.6 and 1.6 for control group.	Baseline and post-treatment dispersion measures were pooled. Conversion by means of formula: $SD = SE * \sqrt{n}$ 6.03 for intervention group vs. 6.79 for control group.
Rabkin, 2000	BDI-I	Table 2. Measures "controlled for baseline values". *M* for each group. 7.2 vs. 10.8.	No derivation necessary.	Table 1 (baseline dispersion) and Table 2 (post-treatment dispersion). 8 *SD* vs. 1.1 *SE* (6.78 *SD*) and 9.6 *SD* vs. 1.1 *SE* (6.22 *SD*).	Baseline and post-treatment dispersion measures were pooled. Conversion by means of formula: $SD = SE * \sqrt{n}$ 7.39 vs. 7.91.
Rabkin, 2000	HDRS	Table 2. Measures "controlled for baseline values".	No derivation necessary.	Table 1 (baseline dispersion) and Table 2 (post-treatment	Baseline and post-treatment dispersion measures were pooled.

		M for each group. 3.3 vs. 6.4.		dispersion). 6.4 *SD* vs. 0.7 *SE* (4.93 *SD*) and 5.8 *SD* vs. 0.8 *SE* (4.53 *SD*).	Conversion by means of formula: $SD = SE * \sqrt{n}$ 5.35 vs. 5.17.
Pope, 2000	HDRS	Table 2. *M* for each group. 0.8 vs. 0.8.	No derivation necessary.	Table 2. Baseline and post-treatment dispersion measures. 1.6 *SD* vs. 1.4 *SD* and 1.6 *SD* vs. 1.2 *SD*.	Baseline and post-treatment dispersion measures were pooled. 1.5 *SD* vs. 1.4 *SD*.
Seidman, 2001	BDI-I	Table 1. Baseline scores for each group: 23.5 vs. 19.3 Results section, paragraph "Depression severity" change scores for each group: -8.8 vs. -7.2	Addition of baseline and change scores yields *M* for each group. 14.7 vs. 12.1	Table 1. 8.6 *SD* vs. 7 *SD*.	*SDs* represent baseline dispersion measures as post treatment dispersion measures were not reported.
Seidman, 2001	HDRS	Table 1. Baseline scores for each group: 22.23 vs. 20.1 Results section, paragraph "Depression severity" change scores for each group: -10.1 vs. -10.5	Addition of baseline and change scores yields *M* for each group. 12.13 vs. 9.6	Table 1. 5.1 *SD* vs. 4.7 *SD*	*SDs* represent baseline dispersion measures as post-treatment dispersion measures were not reported.
Pope, 2003	BDI-II	Table 1. Baseline scores for each group: 23.1 vs. 23.6 Table 3. Change scores for each group: -5.5 vs. -2	Addition of baseline and change scores yields *M* for each group. 17.6 vs. 21.6	Table 1. 4.3 *SD* vs. 7 *SD*	*SDs* represent baseline dispersion measures as post-treatment dispersion measures were not reported.
Pope, 2003	HDRS	Table 1. Baseline scores for each group: 21.8 vs. 21.3 Table 3. Change scores for each group: -7.4 vs. -0.3	Addition of baseline and change scores yields *M* for each group. 14.4 vs. 21	Table 1. 5.9 *SD* vs. 4.1 *SD*	*SDs* represent baseline dispersion measures as post-treatment dispersion measures were not reported.
Malkin, 2004	BDI-II	Table 3. *M* for each group. 4 vs. 7	No derivation necessary.	Table 3. 5.1 *SD* vs. 5.75 *SD*	No derivation necessary.

Pugh, 2004	BDI-II	Third paragraph in results section. Data only reported for intervention group.	Imputation of missing central tendency and dispersion using the following procedure:
			`#assume successful randomization so that baseline BDI(treatment) = BDI(control)`
			`7.3 - 1.6 #mean post-treatment BDI score (treatment group)`
			`7.3 - 1.5 #mean post-treatment BDI score (control group)`
			`6 #sd post treatment BDI score (treatment group)`
			`rho <- (7.3^2 + 6^2 - 0.7^2) / (2*7.3^2 + 6^2) #common recovered pre-post correlation`
			`7.3 #assumed sd baseline BDI scores (control group)`
			`6.1 #recovered sd post-treatment BDI scores (control group)`		

Kenny, 2004	GDS-15	Table 3. *M* for each group. 1.4 vs. 4	No derivation necessary.	Table 3. Baseline and post-treatment *SD* for each group. 1.8 and 1.1 vs. 3.6 and 3	Pooled *SD* for each group. 1.45 vs. 3.3

Rabkin, 2004	HDRS	Imputation of central tendencies and dispersions using the following procedure:
		`sims <- 1000000`
		`#simulate baseline HDRS scores based on reported moments`
		`treat <- rbinom(sims, size = round(17.8^2 / (17.8-4.2^2)), prob = 1 - (4.2^2 / 17.8))`
		`ctrl <- rbinom(sims, size = round(16.8^2 / (16.8-3.3^2)), prob = 1 - (3.3^2 / 16.8))`
		`par(mfrow=c(1,2)); hist(treat); hist(ctrl)`
		
		`#estimate SD of HDRS change assuming r(pre,post) = 0.75 and variance homogeneity`
		`streat <- 4.2*sqrt(2*(1-0.75))`
		`sctrl <- 3.3*sqrt(2*(1-0.75))`
		
		`#estimate mean HDRS changes to approximate the reported response rates (see Furukawa et al., 2005)`
		`library(GA)`
		`optfun <- function(p){`
		`  set.seed(1234)`
		`  ptreat <- treat - rnorm(sims, p[1], streat)`
		`  ptreat <- ifelse(ptreat < 0, 0, round(ptreat))`
		`  pctrl <- treat - rnorm(sims, p[2], sctrl)`
		`  pctrl <- ifelse(pctrl < 0, 0, round(pctrl))`
		`  -1*(((sum(ptreat < 0.5*treat) / sims) - 0.61)^2 + ((sum(pctrl < 0.5*ctrl) / sims) - 0.51)^2)`
		`}`
		`fit <- ga(type="real-valued", fitness = optfun, min=c(0,0), max=c(17.8,16.8), maxiter = 10)`
		`#suitable parameter set: p = c(10.13, 9.59)`
		
		`#simulate post treatment HRDS scores`
		`set.seed(1234)`
		`ptreat <- treat - rnorm(sims, fit@solution[1,1], streat); ptreat <- ifelse(ptreat < 0, 0, round(ptreat))`
		`pctrl <- treat - rnorm(sims, fit@solution[1,2], sctrl); pctrl <- ifelse(pctrl < 0, 0, round(pctrl))`
		`par(mfrow=c(1,2)); hist(ptreat); hist(pctrl)`
		
		`#calculate moments of post-treatment HRDS scores and response rates`
		`mean(ptreat); sd(ptreat); (sum(ptreat < 0.5*treat) / sims)`
		`mean(pctrl); sd(pctrl); (sum(pctrl < 0.5*ctrl) / sims)`			

Cavallini, 2004	BRMS	Table 1. *Mdn* for each group. 5 vs. 7	Imputation sensu Hozo et al. (2005); *M* = *Mdn*	Table 1. Baseline and post-treatment ranges for each group. 5-8 vs. 5-8 and 3-6 vs. 5-8	Imputation sensu Hozo et al. (2005); Pooled *SDs* = Ranges / 4 = 0.75
Haren, 2005	GDS-30	Table 4. Baseline scores for each group: 6.28 vs. 5.7	Addition of baseline and change scores yields *M* for	Table 1. 3.8 *SD* vs. 4.4 *SD*	*SDs* represent baseline dispersion measures as post-treatment

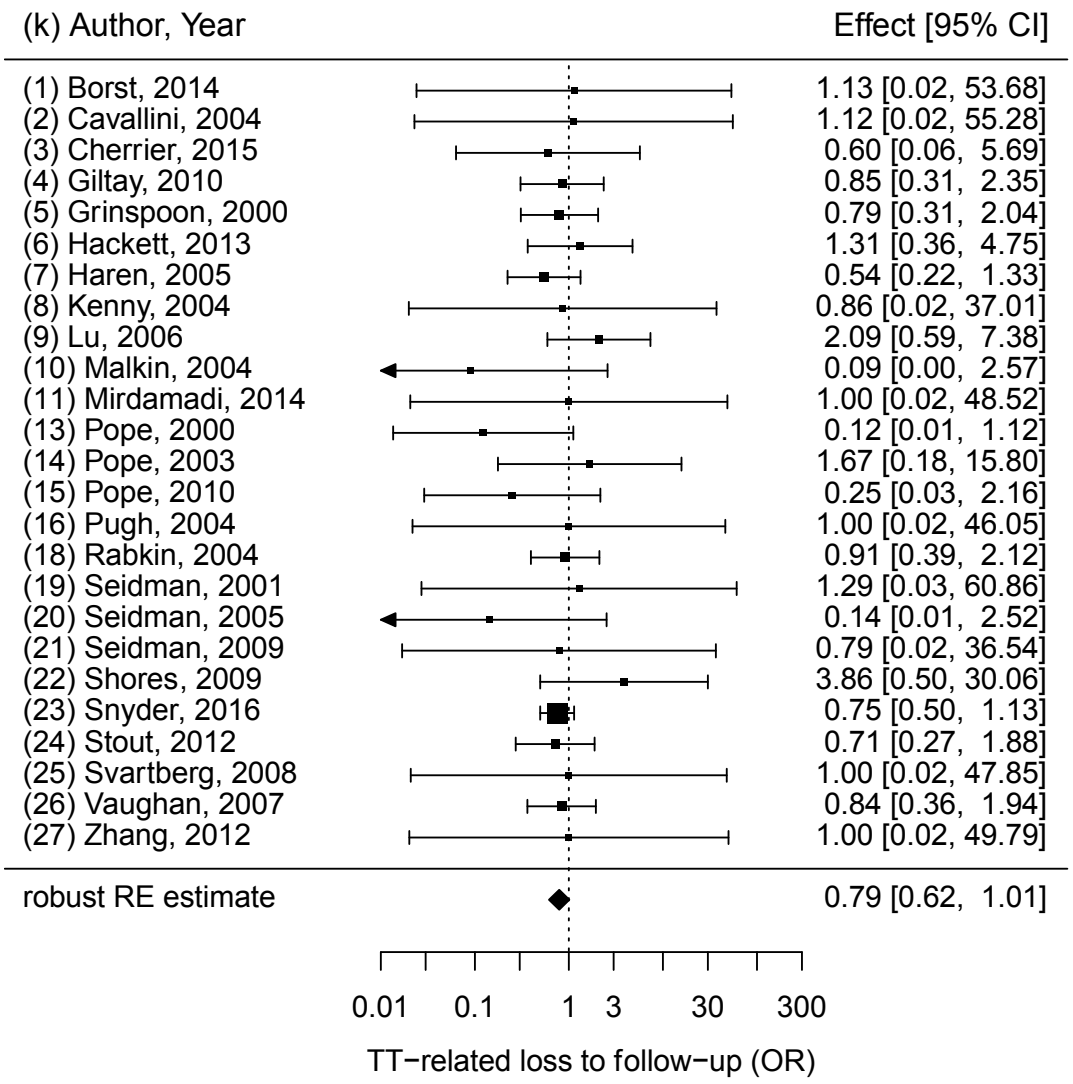
		Change scores for each group: -0.95 vs. -1.27	each group. 5.33 vs. 4.43		dispersion measures were not reported.
Seidman, 2005	HDRS	Results section, paragraph "Depression severity". *M* for each group. 14.4 vs. 15.2	No derivation necessary.	Results section, paragraph "Depression severity". 9.1 *SD* vs. 9.1 *SD*	*SDs* represent post-treatment dispersion measures as baseline dispersion measures were not reported.
Orengo, 2005	HDRS	Results section, second paragraph. *M* for each group. 9.2 vs. 10.4	No derivation necessary.	Results section, second paragraph. 4.1 *SD* vs. 5.4 *SD*	No derivation necessary.
Lu, 2006	BDI-I	Table 3. *M* for each group. 6.5 vs. 9.1	No derivation necessary.	Table 3. Pre and post *SD* for each group. 4.3 and 2.5 vs. 4.9 and 3.8	Pooled *SD* for each group. 3.4 vs. 4.35
Vaughan, 2007	BDI-I	Table 1. *M* for each group. 3.1 vs. 4.8	No derivation necessary.	Table 1. Baseline and post-treatment *SE* for each group. 0.6 and 0.6 vs. 1 and 1.2	Baseline and post-treatment dispersion *SE* were pooled. Conversion by means of formula: $SD = SE * \sqrt{n}$ 2.71 vs. 4.72
Svartberg, 2008	BDI-II	Table 4. *M* for each group. 3.8 vs. 4.3	No derivation necessary.	Table 4. Baseline and post-treatment *SD* for each group. 4.3 vs. 4.8 and 1.3 vs. 2.8	Pooled *SD* for each group. 4.55 vs. 2.05
Seidman, 2009	HDRS	Figure 1. Data extraction using WebPlot Digitizer.	*M* for each group. 6.9 vs. 11.7	Table 1. Single post-treatment scores of subjects.	Manual calculation of *SD* by taking single post-treatment scores of each participant using R statistical software. 4.11 vs. 6.14
Shores, 2009	HDRS	Table 2. *M* for each group. 8.4 vs. 11.4	No derivation necessary.	Table 2. Baseline and post-treatment *SD* for each group. 3.4 vs. 5 and 4.4 vs. 4.4	Pooled *SD* for each group. 4.2 vs. 4.4

Giltay, 2010	BDI-I	Table 2. *M* for each group adjusted for age, body mass index, smoking status, total testosterone level, and prevalent diabetes mellitus.	No derivation necessary.	Table 2. 95% CI. Calculation of *SDs* based on reported data:
				`(mean(c(abs(6.1 - 7.3), abs(6.1 - 5.1))) / qnorm(0.975))*sqrt(113-1)` `#SD of post-treatment BDI score (treatment group)`
				`(mean(c(abs(7.2 - 5.8), abs(7.2 - 9.0))) / qnorm(0.975))*sqrt(65-1)` `#SD of post-treatment BDI score (control group)`	

		6 vs. 7.7			
Pope, 2010	HDRS	Table 2. LOCF method for missing data on the participants with at least one post-baseline evaluation. *M* for each group. 13.4 vs. 15.2	No derivation necessary.	Table 1 (baseline *SD*) and Table 2 (post-treatment *SD*). 3.8 vs. 7.1 and 4.2 vs. 6.3	Pooled *SD* for each group. 5.45 vs. 5.25
Pope, 2010	MADRS	Table 2. LOCF method for missing data on the participants with at least one post-baseline evaluation. *M* for each group. 17.9 vs. 19.7	No derivation necessary.	Table 1 (baseline *SD*) and Table 2 (post-treatment *SD*). 6.3 vs. 9.1 and 5.9 vs. 8.5	Pooled *SD* for each group. 7.7 vs. 7.2
Stout, 2012	BDI-II	Table 5. *M* for each group. 6.6 vs. 7.1	No derivation necessary.	Table 5. Baseline and post-treatment *SD* for each group. 8.7 vs. 3.8 and 5.2 vs. 3.4	Pooled *SD* for each group. 6.25 vs. 4.3
Zhang, 2012	HADS-D	Table 2 and 3. *M* for each group. 2.39 vs. 4.29	No derivation necessary.	Table 2 and 3. Baseline and post-treatment *SE* for each group. 0.6 vs. 0.3 and 0.6 vs. 0.7	Baseline and post-treatment dispersion *SE* were pooled. Conversion by means of formula: $SD = SE * \sqrt{n}$ 4.02 vs. 5.81.
Hackett, 2013	HADS-D	Table 1. Baseline scores for each group: 7.9 vs. 7.26 Results section, "Depression and Anxiety Scores" change scores for each group: -1.05 vs. -0.41	Addition of baseline and change scores yields *M* for each group. 6.85 vs. 6.85	Table 1. Baseline *SD* for each group. 3.91 vs. 4.1	*SDs* represent baseline dispersion measures as post-treatment dispersion measures were not reported.
Mirdamadi, 2014	BDI-I	Table 4. *M* for each group. 5 vs. 5.55	No derivation necessary.	Table 4. Baseline and post-treatment *SD* for each group. 4.41 vs. 6.28 and 3.14 vs. 5.5	Pooled *SD* for each group. 5.35 vs. 4.32
Borst, 2014	GDS-15	Table 2. *M* for each group. 0.88 vs. 2.92	No derivation necessary.	Table 2. Baseline and post-treatment *SD* for each group. 1.76 vs. 0.64 and 1.93 vs. 3.26.	Pooled *SD* for each group. 1.2 vs. 2.6
Cherrier, 2015	GDS-30	Table 3. *M* for each group.	No derivation necessary	Table 3. Baseline *SE* for each	Only baseline *SEs* were reported.

		4.4 vs. 6.8		group. 1.3 vs. 1.2	Conversion by means of formula: *SD* = *SE* * $\sqrt{n}$ 4.11 vs. 4.16
Snyder, 2016	PHQ-9	Table 3. Baseline scores for each group: 6.6 vs. 6.6 Change scores for each group: -1.8 vs. -1.1	Addition of baseline and change scores yields *M* for each group. 4.8 vs. 5.5	Table 3. Baseline *SD* for each group. 4 vs. 4	*SDs* represent baseline dispersion measures as post-treatment dispersion measures were not reported.

eFigure 1. Forest plot of Treatment Acceptability



Acceptability of TT (odds ratio of loss to follow-up) in the respective study, and their meta-analytical estimate. Estimates below 1 represent less loss in response to TT as compared to placebo.

eTable 6. Robust Meta-regression of the Effectiveness of Testosterone Treatment (TT) on Various Study-Level Moderators After Removal of Influential Studies

	Prediction					NHST		
	Manifestation	Estimate	SE	$CI_{2.5\%}$	$CI_{97.5\%}$	*N*	χ^2 (*df*)	*p*
Baseline characteristics								
mean age	40 years	0.159	0.080	0.002	0.316	26	0.817 (1)	0.366
	60 years	0.227	0.057	0.115	0.340			
	80 years	0.296	0.108	0.084	0.508			
Testosterone status	eugonadal	0.122	0.096	-0.067	0.310	24	1.389 (1)	0.239
	hypogonadal	0.260	0.067	0.129	0.391			
HIV infection	yes	0.284	0.172	-0.052	0.620	26	0.200 (1)	0.655
	no	0.203	0.058	0.090	0.316			
symptomatology level	severe	0.460	0.138	0.191	0.730	19	3.926 (2)	0.140
	mild	0.198	0.057	0.086	0.309			
	subclinical	0.698	0.507	-0.297	1.692			
symptom variability (CV)	20 %	0.167	0.114	-0.057	0.391	25	0.358 (1)	0.550
	50 %	0.212	0.057	0.100	0.324			
	100 %	0.286	0.108	0.075	0.497			
Treatment characteristics								
treatment dose	0.1 g / week	0.123	0.087	-0.047	0.293	24	2.703 (1)	0.100
	0.3 g / week	0.244	0.062	0.122	0.365			
	1.0 g / week	0.667	0.276	0.127	1.207			
treatment duration	5 weeks	0.208	0.073	0.064	0.352	26	0.017 (1)	0.896
	20 weeks	0.213	0.058	0.100	0.326			
	100 weeks	0.238	0.184	-0.123	0.598			
administration	intramuscular	0.143	0.065	0.015	0.272	22	3.477 (2)	0.062
	oral	–	–	–	–			
	transdermal	0.439	0.144	0.156	0.722			

Note. CV = coefficient of variation, SE = standard error, CI = confidence interval, NHST = null-hypothesis significance test

eReferences

1. Grinspoon S, Corcoran C, Stanley T, et al. Effects of Hypogonadism and Testosterone Administration on Depression Indices in HIV-Infected Men. *J Clin Endocrinol Metab*. 2000;85(1):60-65.
2. Rabkin JG, Wagner GJ, Rabkin R. A double-blind, placebo-controlled trial of testosterone therapy for HIV-positive men with hypogonadal symptoms. *Arch Gen Psychiatry*. 2000;57(2):141-7; discussion 155-6. doi:10.1001/archpsyc.57.2.141.
3. Haren MT, Wittert GA, Chapman IM, Coates P, Morley JE. Effect of oral testosterone undecanoate on visuospatial cognition, mood and quality of life in elderly men with low-normal gonadal status. *Maturitas*. 2005;50:124-133. doi:10.1016/j.maturitas.2004.05.002.
4. Seidman SN, Miyazaki M, Roose SP. Intramuscular testosterone supplementation to selective serotonin reuptake inhibitor in treatment-resistant depressed men: Randomized placebo-controlled clinical trial. *J Clin Psychopharmacol*. 2005;25(6):584-588. doi:10.1097/01.jcp.0000185424.23515.e5.
5. Orengo CA, Fullerton L, Kunik ME. Safety and efficacy of testosterone gel 1% augmentation in depressed men with partial response to antidepressant therapy. *J Geriatr Psychiatry Neurol*. 2005;18(1):20-24. doi:10.1177/0891988704271767.
6. Lu H, Masterman A, Mulnard R, et al. Effects of Testosterone on Cognition and Mood in Male Patients With Mild Alzheimer Disease and Healthy Elderly Men. *Arch Neurol*. 2006;63(2):177.
7. Vaughan C, Goldstein FC, Tenover JL. Exogenous testosterone alone or with finasteride does not improve measurements of cognition in healthy older men with low serum testosterone. *J Androl*. 2007;28(6):875-882. doi:10.2164/jandrol.107.002931.
8. Svartberg J, Agledahl I, Figenschau Y, Sildnes T, Waterloo K, Jorde R. Testosterone treatment in elderly men with subnormal testosterone levels improves body composition and BMD in the hip. *Int J Impot Res*. 2008;20(4):378-387. doi:10.1038/ijir.2008.19.
9. Seidman SN, Orr G, Raviv G, et al. Effects of Testosterone Replacement in Middle-Aged Men With Dysthymia. *J Clin Psychopharmacol*. 2009;29(3):216-221. doi:10.1097/JCP.0b013e3181a39137.
10. Shores MM, Kivlahan DR, Sadak TI, Li EJ, Matsumoto AM. A randomized, double-blind, placebo-controlled study of testosterone treatment in hypogonadal older men with subthreshold depression (dysthymia or minor depression). *J Clin Psychiatry*. 2009;70(7):1009-1016. doi:10.4088/JCP.08m04478.
11. Giltay EJ, Tishova YA, Mskhalaya GJ, Gooren LJG, Saad F, Kalinchenko SY. Effects of Testosterone Supplementation on Depressive Symptoms and Sexual Dysfunction in Hypogonadal Men with the Metabolic Syndrome. *J Sex Med*. 2010;7(7):2572-2582. doi:10.1111/j.1743-6109.2010.01859.x.
12. Pope HG, Amiaz R, Brennan BP, et al. Parallel-group placebo-controlled trial of testosterone gel in men with major depressive disorder displaying an incomplete response to standard antidepressant treatment. *J Clin Psychopharmacol*. 2010;30(2):126-134. doi:10.1097/JCP.0b013e3181d207ca.
13. Pope HG, Kouri EM, Hudson JI. Effects of supraphysiologic doses of testosterone on mood and aggression in normal men: A randomized controlled trial. *Arch Gen Psychiatry*. 2000;57(2):133-140; discussion 155-156. doi:10.1001/archpsyc.57.2.133.
14. Stout M, Tew GA, Doll H, et al. Testosterone therapy during exercise rehabilitation in

male patients with chronic heart failure who have low testosterone status: A double-blind randomized controlled feasibility study. *Am Heart J*. 2012;164(6):893-901. doi:10.1016/j.ahj.2012.09.016.
15. Zhang X wei, Liu Z hua, Hu X wei, et al. Androgen replacement therapy improves psychological distress and health-related quality of life in late onset hypogonadism patients in Chinese population. *Chin Med J (Engl)*. 2012;125(21):3806-3810. doi:10.3760/cma.j.issn.0366-6999.2012.21.011.
16. Hackett G, Cole N, Bhartia M, Kennedy D, Raju J, Wilkinson P. Testosterone Replacement Therapy with Long-Acting Testosterone Undecanoate Improves Sexual Function and Quality-of-Life Parameters vs. Placebo in a Population of Men with Type 2 Diabetes. *J Sex Med*. 2013;10(6):1612-1627. doi:10.1111/jsm.12146.
17. Mirdamadi A, Garakyaraghi M, Pourmoghaddas A, Bahmani A, Mahmoudi H, Gharipour M. Beneficial effects of testosterone therapy on functional capacity, cardiovascular parameters, and quality of life in patients with congestive heart failure. *Biomed Res Int*. 2014;2014(392432). doi:http://dx.doi.org/10.1155/2014/392432.
18. Borst SE, Yarrow JF, Fernandez C, et al. Cognitive effects of testosterone and finasteride administration in older Hypogonadal men. *Clin Interv Aging*. 2014;9:1327-1333. doi:10.2147/CIA.S61760.
19. Cherrier MM, Anderson K, Shofer J, Millard S, Matsumoto AM. Testosterone treatment of men with mild cognitive impairment and low testosterone levels. *Am J Alzheimers Dis Other Demen*. 2015;30(4):421-430. doi:10.1177/1533317514556874.
20. Snyder PJ, Bhasin S, Cunningham GR, et al. Effects of testosterone treatment in older men. *N Engl J Med*. 2016;374(7):611-624. doi:10.1056/NEJMoa1506119.
21. Seidman S, Spatz E, Rizzo C, Roose S. Testosterone Replacement Therapy for Hypogonadal Men With Major Depressive Disorder: A Randomized, Placebo-Controlled Clinical Trial. *J Clin Psychiatry*. 2001;157(11):1884. doi:10.1176/appi.ajp.157.11.1884.
22. Pope HG, Cohane GH, Kanayama G, et al. Testosterone Gel Supplementation for Men With Refractory Depression : A Randomized, Placebo-controlled Trial. *Am J Psychiatry*. 2003;160(15):105-111.
23. Malkin CJ, Pugh PJ, Morris PD, et al. Testosterone replacement in hypogonadal men with angina improves ischaemic threshold and quality of life. *Heart*. 2004;90(8):871-876. doi:10.1136/hrt.2003.021121.
24. Pugh PJ, Jones R, West JN, Jones TH, Channer KS. Testosterone treatment for men with chronic heart failure. *Heart*. 2004;90(4):446-447. doi:10.1136/hrt.2003.014639.
25. Kenny AM, Fabregas G, Song C, Biskup B, Bellantonio S. Effects of testosterone on behavior, depression, and cognitive function in older men with mild cognitive loss. *J Gerontol A Biol Sci Med Sci*. 2004;59(1):75-78. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/005/CN-00460005/frame.html.
26. Rabkin JG, Wagner GJ, McElhiney MC, Rabkin R, Lin SH. Testosterone versus fluoxetine for depression and fatigue in HIV/AIDS: A placebo-controlled trial. *J Clin Psychopharmacol*. 2004;24(4):379-385. doi:10.1097/01.jcp.0000132442.35478.3c.
27. Cavallini G, Caracciolo S, Vitali G, Modenini F, Biagiotti G. Carnitine versus androgen administration in the treatment of sexual dysfunction, depressed mood, and fatigue associated with male aging. *Urology*. 2004;63(4):641-646. doi:10.1016/j.urology.2003.11.009.
28. Jadad AR, Moore RA, Carroll D, et al. Assessing the Quality of Reports of Randomized

Clinical Trials : Is Blinding Necessary ? *Control Clin Trials*. 1996;12(January 1995):1-12. doi:10.1016/0197-2456(95)00134-4.
29. Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clin Psychol Rev*. 1988;8:77-100. doi:10.1016/0272-7358(88)90050-5.
30. Bech P. *Clinical Psychometrics*. Chichester, West Sussex, UK: John Wiley & Sons; 2012.
31. Addington D, Addington J, Schissel B. A depression rating scale for schizophrenics. *Schizophr Res*. 1990;3(4):247-251.
32. Addington D, Addington J, Maticka-Tyndale E, Joyce J. Reliability and validity of a depression rating scale for schizophrenics. *Schizophr Res*. 1992;6(3):201-208.
33. Kroenke K, Spitzer RL, Williams JBW. The phq-9. *J Gen Intern Med*. 2001;16(9):606-613.
34. McDowell I. *Measuring Health: A Guide to Rating Scales and Questionnaires*. Oxford University Press, USA; 2006.
35. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res*. 1982;17(1):37-49.
36. Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS) Recent Evidence and Development of a Shorter Version. *Clin Gerontol*. 2008;7115(1986). doi:10.1300/J018v05n01.
37. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361-370.
38. Zimmerman M, Martinez JH, Young D, Chelminski I, Dalrymple K. Severity classification on the Hamilton depression rating scale. *J Affect Disord*. 2013;150(2):384-388.
39. Furukawa TA, Cipriani A, Barbui C, Brambilla P, Watanabe N. Imputing response rates from means and standard deviations in meta-analyses. *Int Clin Psychopharmacol*. 2005;20(1):49-52.
40. Hedges L V. Distribution theory for Glass's estimator of effect size and related estimators. *J Educ Stat*. 1981;6(2):107-128.
41. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*. 2005;5(1):13.